

ABSTRACT

Title of Document: CONSCIOUSNESS, CONCEPTS AND
CONTENT

Bénédicte Veillet, Ph.D., 2008

Directed By: Professor Peter Carruthers, Department of
Philosophy

Concepts figure prominently in the defense and elaboration of representational accounts of phenomenal consciousness. Indeed, any adequate defense of (reductive) representationalism will require an appeal to so-called phenomenal concepts to deflect a group of related anti-physicalist (and hence anti-representationalist) arguments. What's more, an elaboration of representationalism requires a detailed account of the representational content of phenomenally conscious experience.

The goal of this dissertation is to contribute to the defense and elaboration of representationalism as it relates to concepts, first with a defense of demonstrative/recognition accounts of phenomenal concepts (and a defense of the more general physicalist strategy in which they figure); and second, with the development of a partially conceptual account of perceptual experience.

CONSCIOUSNESS, CONCEPTS AND CONTENT

By

Bénédicte Veillet

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Ph.D.
2008

Advisory Committee:
Professor Peter Carruthers, Chair
Professor Georges Rey
Professor Paul Pietroski
Associate Professor Allen Stairs
Professor Amanda Woodward

© Copyright by
Bénédicte Veillet
Ph.D.

Acknowledgements

Many thanks to Peter Carruthers for the great feedback and advice; and for being encouraging and supportive when I needed it most. Many thanks to Georges Rey for the pages (and pages) of comments and afternoon-long conversations. Many thanks to the rest of my committee.

Many thanks to my best “mates” (house/office) Ryan, Lizzie and Matt for being calm when I wasn’t.

Et finalement, merci à ACDEF, d’avoir toujours eu confiance en moi, et d’être maintenant si contents pour (avec) moi, malgré toutes nos transitions.

Table of Contents

Acknowledgements.....	ii
Table of Contents	iii
List of Figures	v
Introduction.....	1
1. What is phenomenal consciousness?	1
2. What is physicalism? What is anti-physicalism?.....	4
3. What is (reductive) representationalism?.....	6
3.1. Computational/Representational theory of mind (CRTM).....	7
3.2. Representationalism and phenomenal character.....	10
4. Concepts and reductive representationalism.....	12
4.1. Phenomenal Concepts.....	14
4.2. Concepts and Experience.....	15
Part I – Phenomenal Concepts	17
1. Introduction.....	17
1.1. Anti-physicalist arguments	19
1.2. Data and explanation.....	25
2. Phenomenal Concepts.....	30
2.1. Phenomenal concepts and other concepts.....	33
2.2. Phenomenal concepts and their referents.....	43
2.3. Accounts of phenomenal concepts.....	56
2.4. Phenomenal concepts and what it’s like	62
3. Putting phenomenal concepts to work (the phenomenal concept strategy)	64
3.1. Conceivability judgments and non-derivability.....	64
3.2. What Mary learns.....	65
3.3. The core contrast.....	77
3.4. Concluding remark.....	84
4. General anti-physicalist arguments.....	84
4.1. Property dualist arguments	85
4.2. Chalmers	122
Part II – Concepts and Experience.....	151
1. Introduction.....	151
1.1. Inverted Earth detour	152
1.2. Plan of Part II.....	156
2. The conceptual/nonconceptual debate	156
2.1. <i>Possession</i> conceptualism.....	157
2.2. <i>Constituent</i> conceptualism.....	159
2.3. Differences.....	162
2.4. Partial Conceptualism	168
3. The case for nonconceptualism.....	171
3.1. Continuity	171
3.2. Systematicity.....	174
3.3. Richness and fineness of grain.....	177

3.4. Last remarks.....	194
4. The case against conceptualism.....	195
4.1. Justifying beliefs.....	196
4.2. Seeing as.....	198
4.3. Concept acquisition.....	202
4.4. Concluding.....	208
5. Partial conceptualism.....	208
5.1. One more motivation.....	209
5.2. Some partial conceptualist models.....	210
5.3. Back to Block.....	215
Conclusion.....	217
Bibliography.....	219

List of Figures

Figure I.1. Filling in the blanks (Hesperus/Phosphorus)

Figure I.2. Filling in the blanks (White)

Figure I.3. The regress argument

Figure I.4. Filing in the blanks (White reminder)

Figure I.5. Filling in the blanks—according to the physicalist

Figure I.6. Filling in the blanks—a posteriori

Figure II.1. Possession nonconceptualism

Figure II.2. Possession conceptualism

Figure II.3. Possession *and* deployment conceptualism

Introduction

Concepts figure prominently in the defense and elaboration of representational accounts of phenomenal consciousness. Indeed, any adequate defense of (reductive) representationalism will require an appeal to so-called phenomenal *concepts* to deflect a group of related anti-physicalist (and hence anti-representationalist) arguments. What's more, an elaboration of representationalism requires a detailed account of the representational content of phenomenally conscious experience.

The goal of this dissertation is to contribute to the defense and elaboration of representationalism as it relates to concepts, first with a defense of demonstrative/recognition accounts of phenomenal concepts (and a defense of the more general physicalist strategy in which they figure); and second, with the development of a partially conceptual account of experience. The goal of this introduction is to set the stage and provide the background the reader needs to make sense of these issues and see why they matter.

1. What is phenomenal consciousness?

Perceptual experiences, bodily sensations, emotions, etc., are mental states that (often) feel a certain way to those who undergo them. There is something it's like to see the fresh green of a new leaf, to hear a bird singing, to touch grass with bare feet, to feel pain, to feel dizzy or terrified. Presumably these states feel different to their

subject—what it’s like for me to see fresh green is different from what it’s like for me to hear a bird.

The phenomenology, or phenomenal character, or phenomenal feel of an experience (sensation or emotion) is just what it feels like for a subject to have that experience (sensation or emotion). If what it’s like for you to taste English peas is different from what it’s like for you to taste string beans, then these two perceptual experiences do not have the same phenomenal character or feel. We’ll say that a *mental state* is phenomenally conscious if there is something it is like for the subject to undergo that state. And we can say, derivatively, that a creature is phenomenally conscious if and only if it has some phenomenally conscious states.

I take it to be rather obvious that there *are* phenomenal characters, as I’ve “defined” them.¹ A number of philosophers have denied that there are *qualia* (e.g. Dennett 1988)—and since the term ‘qualia’ is sometimes used to mean, simply, phenomenal characters, it may seem as though philosophers have (rather forcefully) denied that there are phenomenal characters. But the term ‘qualia’ is slippery; some claim, for instance, that “qualia are ineffable or non-physical or ‘given’ to their subjects incorrigibly (without the possibility of error)” (Tye 2008); some add that they are atomic, unanalyzable, simple, private and homogenous (Dennett 1988). Certainly, to say that there are qualia, in this *strong* or *bold* sense, is to say a lot more than that there are phenomenal characters. So if there is a debate about whether or not there are qualia, it is one about whether or not there are *strong qualia* (qualia in the

¹ Of course I haven’t “defined” phenomenal characters in any strict sense. As Block (1995) writes “I cannot define P[henomenal]-consciousness in any remotely noncircular way. ...The best one can do for P-consciousness ...is *point* to the phenomenon” (380—page numbers are to the 1997 reprint). Pointing is what I’ve done.

sense of ineffable, non-physical, incorrigible features of our experience) and not qualia in the “modest” sense of phenomenal characters. Still, there are a number of substantial disagreements about phenomenal characters and their nature. First, we may wonder exactly which states are phenomenally conscious. Many perceptual states have phenomenal characters, like the seeing of new green, or the hearing of birdsong, but that’s not to say that *all* perceptual states are phenomenally conscious. A great many such states might *not* be phenomenally conscious—as, for instance, the visual (and so in some sense *perceptual*) states that David Marr (1982) posits in early vision. Similarly, it is not obvious that every bodily sensation or every emotion is phenomenally conscious. And it is unclear what other mental states (beliefs, desires) have phenomenal characters. Second, we may wonder which creatures have phenomenally conscious states—do rats have them? Do bees? Do infants? (See for instance Tye 1997, 2000, Carruthers 2000, 2004). Finally, we might wonder about how phenomenal characters are related to the brain, its properties, and the natural world more generally. It is this last question—about the relationship between phenomenal characters and the physical world—that lies at the heart of this dissertation.

2. What is physicalism? What is anti-physicalism?

Physicalists believe that phenomenal characters can be reduced to non-phenomenal (and ultimately non-mental) things or properties.² Anti-physicalists (dualists), on the other hand, deny that such a reduction is possible. Phenomenal characters are irreducibly *phenomenal* and in some sense non-physical.

It may help to think of the disagreement about phenomenal characters as similar to the 19th century disagreement about *life*. “Physicalists” about *life*, for instance, were those who believed (as most everyone does today) that the phenomena we associate with being alive could be reduced to non-“living” phenomena, e.g. chemical and physical phenomena. They thought that there was nothing more to being alive than being chemically or physically made up a certain way. On the other hand, anti-physicalists about *life* (or rather *vitalists*) were those (rather prevalent at the time) who believed that life could not be reduced to non-“living”, chemical and physical phenomena and processes. Life, they thought, is irreducible. Being alive is a matter of an organism having a *life-force*, an *élan vital*, something over and beyond a particular biological and chemical makeup.

In much the same way, physicalists about phenomenal characters believe that there is nothing to being phenomenally conscious over and above having brain states of a certain sort (with certain functional or representational properties). Anti-physicalists, meanwhile, deny that this is the case. According to the latter, phenomenal characters are irreducible; being phenomenally conscious is a matter of

² There are a number of different kinds of physicalism. Type vs. token physicalism; supervenience physicalism; Stoljar’s t-physicalism and o-physicalism (2001). But the very general characterization above will suffice for our purposes.

an organism having some *phenomenal*, non-physical properties, something over and above a particular brain makeup. Because anti-physicalists believe that there must be two radically different kinds of properties in the world—the physical, natural ones (which physics, chemistry and biology tell us about) *and* phenomenal, non-physical ones—they are sometimes called *property dualists*.³

Of course, anti-physicalists about *life* are rare today. “The spectacular successes of molecular biology make it virtually certain that biological phenomena [like life] are just very special cases of physical phenomena” (Rey 1997, 22). Physicalists about phenomenal characters hope that phenomenal characters will ultimately prove to be biological, chemical, physical sorts of things—just as life ultimately proved to be a chemical, physical sort of thing. But anti-physicalists go on to make a rather strong claim: they argue that *regardless* of what progress and discoveries our sciences might make, phenomenal characters *will remain* irreducible. It is not simply that phenomenal characters cannot be reduced to biological or chemical phenomena *that we know of now*; but that phenomenal characters are not reducible to *any possible or imaginable* biological or chemical or physical phenomena.

³ A *substance* dualist believes that there is an immaterial, non-physical *substance*—non-physical *stuff* (like blobs of ectoplasm, say). Descartes famously thought that the mind itself was made up of an immaterial substance. The property dualist, however, doesn't think there is a non-physical *substance* but rather that physical things (like brains) have non-physical properties.

3. What is (reductive) representationalism?

Since physicalists believe that phenomenal characters can be reduced to physical sorts of things, it would make sense for them to pay special attention to the advances and successes in the sciences of the mind (including psychology, linguistics, neuroscience, or more generally cognitive science). After all, we think that the physicalist about *life* was right to take seriously the scientific developments in chemistry and biology—she claimed that life was reducible to chemical and biological phenomena. Similarly then, a physicalist about *phenomenal character* should take into account developments in the relevant sciences.

Over the past 50 years or so, one research program has proved to be especially successful in cognitive science: the computational/representational program. As a result, a physicalist theory of phenomenal characters incorporating some insights of the computational/representational framework has become increasingly prominent (see Dretske 1995, Tye 2000, Rosenthal 1995, Lycan 1996, Carruthers 2000, Rey 1998). I will call these types of theories *reductive representational* theories (or reductive representationalism). I will say a bit, first, about the computational/representational program (3.1), before coming back to reductive representationalism as a theory of phenomenal characters (3.2).

3.1. Computational/Representational theory of mind (CRTM⁴)

According to CRTM mental states involve relations to mental representations. It will help to think of mental representations as mental symbols which stand for other things—as symbols usually do. The symbol ‘=’ stands for ‘is the same as’ or ‘is equal to’ and words too are symbols. The English word ‘tree’ stands for actual trees. So we can say that the word ‘tree’ is a representation and that its *content* (what it stands for) is [tree]. (I put contents in brackets.) The French word ‘arbre’ is a symbol too, and like ‘tree’, its content is [tree].

According to CRTM, there are symbols in the head (mental representations) that stand for things outside the head. And to undergo certain mental states (like beliefs, desires, hopes) is to be related in some way to strings of mental symbols with particular contents. For instance, to believe that grass is green is to be related in some way (the belief way) to a mental representation (a string of symbols in the head) that stands for *grass is green*, or as we will say, whose *content* is [grass is green]. To *hope* that grass is green is to be related in a different way (the hope way) to the same string of symbols that stands for *grass is green*. To *perceive* a red tomato is to be related (presumably, again in a different way, say, the visual way) to a mental representation with a particular content. (How to cash out the content of perceptual representations is an important and challenging question that proponents of CRTM must answer. The question will be relevant to my project as well, see section 4 of this introduction).

Moreover, according to CRTM mental *processes* involve the transformation, manipulation and storage of these mental representations. Deductive reasoning, for instance, would involve the manipulation of mental symbols according to certain

⁴ See Rey 1997

rules (see Rey 1997, pgs 211 to 221 for a discussion of what manipulations might be involved in deductive reasoning, induction, abduction and decision making).

Thinking of the mind in this way (as transforming and storing strings of mental symbols) has provided a helpful theoretical framework in the cognitive sciences. Thagard, in his Stanford Encyclopedia entry on Cognitive Science, writes: “the central hypothesis of cognitive science is that thinking can best be understood in terms of representational structures in the mind and computational procedures that operate on those structures” (2007).^{5,6} Here are two examples of how CRTM has been insightful.

First is in the study of perception. Marr, in 1982, proposed to think about the *visual* system (and other information processing systems) as a computational system describable at three levels. At the first level, the visual system is describable in terms of *what it does*. It requires that we answer the following kind of questions: “what is the goal of the computation, why is it appropriate, and what is the logic of the strategy by which it can be carried out?” (1982, 25) So we may take the goal of the visual system to be the building of a three-dimensional, colored representation of the world from various inputs (light intensity, wavelength, etc.). And doing this may require the system to perform a number of intermediate tasks, which we can, in turn, characterize in terms of their intermediate goals, such as the building of the representation of an edge, of color, of a surface, of motion, etc. The second level of a computational theory is algorithmic: it is the level at which we attempt to describe the step by step

⁵ In the introduction to his 2005 *Mind: An introduction to cognitive science*, he writes again that “most cognitive scientists agree that knowledge in the mind consists of *mental representations*” (4).

⁶ Of course, this isn’t to say that there is no further disagreement, among cognitive scientists, about the nature of mental representations and of computations.

transitions between states of the system, which would take it from input representation (of light intensities, say) to output representation (of an edge, say). The second level, then, characterizes *how* the system might be doing what it does as described by the first computational level. The third level is the level of implementation. The question, here, is the following: “how can the [...] algorithm be implemented physically,” for instance, in the brain (1982, 25)? This kind of computational framework is the “framework within which most current theories of visual perceptions are cast” (Palmer 1999, 71). In other words, most current theories of visual perceptions are in the business of figuring out how representations are manipulated and transformed to yield other representations.

The computational model can also explain interesting facts about the way we *think*, e.g., the fact that our thought is productive and systematic (Fodor 1987). I will focus here on productivity. Consider the fact that we can understand sentences we’ve never heard before—sentences that combine words in ways we’ve never heard them combined. And consider the fact that we can produce such sentences too. To produce them, it must be that we can think thoughts we’ve never thought before. And if we had all the time in the world, it seems that there would be no end to the new thoughts we could think—that Rambo’s cat just had a bad hair day; that purple giraffes take their time when bowling, etc. This suggests that we could store an endless number of thoughts in our heads. Yet we are finite beings. So how can we produce endlessly many thoughts? Here is an answer within a CRTM: there are a finite number of stored and (relatively) simple mental representations, like a mental representation standing for cat, one for Rambo, one for hair. And what enables us to think an endless number

of new thoughts is the fact that these simple mental representations can be combined and recombined in many, *many* ways. Though I'd never thought that Rambo's cat had a bad hair day, I may have thought that Rambo has long hair, and that cats have bad days, etc. Thinking a new thought, then, is merely my combining mental representations in a new way. Such an explanation does require that we think of some mental representations as being *structured*, i.e., made up of more simple representations. So when I believe that grass is green, I am related to a *structured* representation that takes several simpler representations as constituents—one that stands for grass, another for green. In so far as some people embrace a computational theory of mind while denying that mental representations are structured in this way (e.g. Smolensky), not all computational theories of mind will be able to explain productivity. But it is nonetheless a positive feature of CRTMs that some of their instantiations can explain facts like productivity

3.2. Representationalism and phenomenal character

Now, if reductive representationalists can make the case that *phenomenal characters* can be reduced to the representational contents of certain states (like visual states), then they will have found a place for phenomenal characters within an empirically rather successful theory of the mind. And this, I think, makes reductive representationalism especially worth investigating.

It should be noted that most of the philosophers who are interested in phenomenal characters are representationalists in some sense. They usually think that perceptual systems work roughly as CRTM would say that they do. Still, they deny what the *reductive* representationalist claim is true, namely that phenomenal

characters are reducible to representational contents. Chalmers for instance does believe that perceptual experiences are representational states that do indeed involve relations to mental representations (1996); but he argues that phenomenal character is the one thing which can't be explained representationally. And ultimately he reaches an anti-physicalist conclusion.⁷ In the rest of my dissertation I will use the term 'representationalist' (tout *court*) to stand for "*reductive* representationalist" unless otherwise noted.

One last thing: representationalists are for the most part *functional* representationalists (save for Carruthers 2000). This means they believe that phenomenal characters can be reduced to representational contents *iff* their representations themselves *play the right functional role*. This matters quite a bit. After all, imagine that the content of a subject's red visual experience (she's focusing on a red wall) is the content [red]. Unfortunately many representations, besides her visual representation, will have that content: the word 'red' on this page for instance also has the content [red]. And if they claim that phenomenal characters are reducible to representational contents *alone*, then they will be committed to the claim that any two representations with those *same* contents will have the same phenomenal characters too. So the representationalist would have to say that if Sara's visual red experience has the content [red] and the word 'red' has the content [red], then both Sara's visual red experience and the word 'red' *have the same phenomenal character*. That is clearly counterintuitive. Sara's experience of red does feel like something

⁷ Block (2003), like Chalmers, thinks that perceptual states are representational, but, like Chalmers, he does not think phenomenal characters are reducible to experiential contents. Unlike Chalmers, he does not draw from this an anti-physicalist conclusion—phenomenal characters have more to do with intrinsic features of the *vehicles* of contents.

(there is something it's like to undergo it). But the word 'red' is not even the *kind* of thing that could have a phenomenal character (*seeing* the word 'red' would have a phenomenal character, but the representationalist I've described is committed to saying that the word 'red' *itself* has a particular phenomenal character).⁸ To avoid such conclusions, the representationalist might add a *functional* aspect and claim that phenomenal characters are reducible to representational contents, whose representations play the right role (whose representations are poised to impact beliefs in some way for instance, see Tye 1995, 2000).⁹

4. Concepts and reductive representationalism

Concepts figure prominently in the defense and elaboration of representational accounts of phenomenal consciousness. But I have yet to say what concepts are. Our starting point will be this: concepts are the constituents of beliefs. Here is one way to make sense of that claim within CRTM:¹⁰ when a thinker believes that grass is green, she is related (in some way, say, the belief way) to a mental representation, whose constituents are the concepts GRASS and GREEN (I use small caps for concepts).

⁸ Carruthers needn't worry about this particular problem. Even though he is not a functional representationalist, he thinks that phenomenal characters are reducible to "dual" contents – like [red/seems red], which can be acquired only when certain simpler contents (like [red]) can be targeted by a higher-order thought system. Clearly, the content of the word 'red' is not a content that can be targeted by a higher-order thought system. So there is no worry here for Carruthers' view.

⁹ Lycan (2006) points out that Dretske and Tye are both functional representationalists (Block 2003 calls them *quasi*-representationalists). Here is a quote from Lycan: "The representational theory of qualia cannot be purely representational, but must appeal to some further factor, to distinguish visual representations from other sorts of representations of redness. Dretske (1995) cites only the fact that visual representation is sensory and what he calls "systemic." Tye (1995) requires that the representation be nonconceptual and "poised"." (2006)

¹⁰ Which is not to say that if CRTM is roughly right, concepts must be mental representations. After all, some vocal defenders of CRTM do think that there is something wrong with the psychological view of concepts (see Rey)

Concepts on this view (which Laurence and Margolis 2007 call the *psychological view*) are mental representations which can be combined and recombined, and to which we can be, at the very least, belief-related. Though I will talk of concepts as mental representations throughout, I don't think that many of my conclusions require that this be *the* way to think about concepts.¹¹ (It could be that concepts are abstract constituents of propositions (what Laurence and Margolis call the *Semantic view* of concepts)¹².

Concepts, then, play an important role in the defense and elaboration of reductive representationalism. Indeed, the representationalist must deal with a group of related anti-physicalist (and hence anti-representationalist) arguments; and any adequate reply to this family of arguments requires an appeal to so-called *phenomenal concepts*—concepts deployed in thought to pick out, via introspection, our phenomenal feels. Moreover, representationalism claims that a red phenomenal character can be reduced to the content of the red perceptual experience. Any fully developed representationalist account will therefore require an account of the representational content of these phenomenally conscious experiences. And an adequate account of those experiential contents must explain the relation between them and *concepts*. I spell this out in more detail below.

¹¹ Though it might be incompatible with the view that concepts are abilities. See Laurence and Margolis (2006).

¹² I also don't think I will have to commit myself to a particular account of the *structure* of concepts (for instance prototypes vs. classical structures).

4.1. Phenomenal Concepts

Anti-physicalists like to underscore some well-known “data” regarding our phenomenally conscious experience, which, they argue, shows that physicalism is false. They point out, for instance, that we can conceive of creatures representationally identical to us but with inverted phenomenal characters, or with no phenomenal characters at all. They point out that Mary, the brilliant color scientist raised in a black-and-white room, seems to learn something new upon leaving the room and seeing red for the first time, even if we suppose she had learned everything scientific there was to know about color experience. They point out that there is, and must be, an explanatory gap in physicalist explanations of phenomenal feels, including representationalist ones. For explaining the feel of a red experience (explaining *this* feel, pointing to a red experience) in terms of experiential contents will never feel satisfactory the way that typical explanations are in science, e.g., the explanation of water boiling in terms of H₂O molecules and their properties.

The physicalist has had most success replying to this line of argument by taking the non-physicalist’s “data” as a given, but by insisting that it can be explained entirely by appealing to certain (physicalist) features of our *phenomenal concepts*. The anti-physicalist is sympathetic to the move. After all, she is ready to grant that an appeal to phenomenal concepts is needed to fully explain her data. However, she still claims that it is impossible to fully explain that data using the phenomenal concept strategy *and* remain a physicalist about these concepts and/or their referents.

The goal of Part I is to show that the anti-physicalist is wrong. Phenomenal concepts have a number of features (they are conceptually isolated, recognitional—in

a weak sense—and refer directly) which can both explain the relevant data and themselves be physically explicable. I begin Part I with a discussion of the anti-physicalist “data” (Part I, section 1) and an introduction of phenomenal concepts (section 2), spelling out the ways in which they relate to other concepts (2.1) and to their referents (2.2). I then show how phenomenal concepts can be used to explain the anti-physicalist data (section 3), before arguing that the two main anti-physicalist objections to the phenomenal concept “strategy” fail (section 4).

4.2. Concepts and Experience

Representationalists claim that phenomenal characters can be reduced to the representational contents of experience. They must go on to say something about the features of these contents, including how they relate to *concepts*. This is especially important because the representationalist must explain how experiences (which have phenomenal characters) differ from other mental states, most importantly propositional attitudes like beliefs (which most agree do not have phenomenal characters¹³).

Since concepts are, as I’ve said, the constituents of belief, spelling out the difference between experience and propositional attitudes will require spelling out the relation between *experience* and concepts—and contrasting that relation with the relation between *belief* and concepts.¹⁴

¹³ There are exceptions here of course. See Chalmers 2003.

¹⁴ At the beginning of Part II I also argue that it is more important for the representationalist to spell out the relation between experience and concepts than she realizes. Indeed, what she says about experience and concepts will affect the way the representationalist can deal with a famous argument against representationalism (Block’s Inverted Earth 1990).

The conceptual/nonconceptual debate is one that attempts to address this very question about the relation between experience and concepts. Part II begins with a number of distinctions, which allow me to identify the most interesting (and relevant) aspect of that rather messy debate (section 2). I take *conceptualists* to be those who maintain that the constituents of experience and the constituents of belief are of the same kind—such that, in principle, constituents of experience could be constituents of belief. Nonconceptualists, on the other hand, maintain that the constituents of experience and the constituents of belief are of a different kind. I argue, ultimately, that the best account of experiential content is a (more rarely defended) one according to which *some* of the constituents of experience are like those of belief and some are not. I do this by showing first that arguments *for* nonconceptualism don't quite succeed (section 3)—at most, we can conclude from these arguments that either nonconceptualism succeeds or partial conceptualism does. I then argue that nonconceptualism fails (section 4). I end Part II with a discussion of partial conceptualism (section 5).

Part I – Phenomenal Concepts

1. Introduction

The phenomenal concept strategy is arguably the most promising strategy available to the representationalist (and fellow physicalist) against an entire family of anti-physicalist arguments. These anti-physicalists like to underscore well-known “data” about our phenomenally conscious experience; this data, they argue, shows that physicalism is false. They point out, for instance, that we can conceive of inverters and zombies, or that there is an explanatory gap—and they maintain that these conceivability judgments, or the existence of the explanatory gap, cannot be explained by a(ny) physicalist account of phenomenal feels.

The physicalist’s most promising line of response has involved taking the non-physicalist’s data for granted while insisting that it can be explained entirely by appealing to certain (physicalist) features of our *thinking* about phenomenal feels, more specifically to features of the constituents of our phenomenal thoughts, i.e. *phenomenal concepts*. The anti-physicalist is not dismissive of the physicalist’s strategy entirely, for she is ready to grant—indeed she believes—that an appeal to phenomenal concepts is needed to fully explain her data. Still the physicalist is wrong, she claims, for it is impossible to 1) explain the data in question fully by appealing to phenomenal concepts *and* 2) remain a physicalist about these concepts and/or their referents. The anti-physicalist typically argues against the physicalist in

either of two ways. First, making a quite general point against the physicalist's strategy, the anti-physicalist will argue that the features of phenomenal concepts that do the crucial explanatory work, *whatever those might be*, cannot possibly fit within a physicalist framework. Chalmers (2007) makes just this kind of claim. Take F to be the features of phenomenal concepts the physicalist believes will explain the anti-physicalist data. Chalmers argues (recruiting zombies in the process) that a) either F do not in fact fully explain the data, or b) F are incompatible with a physicalist universe. The details about features F (how individual physicalists might spell out what these features are—and they seem to disagree about that) are irrelevant to Chalmers' argument. Property dualists who don't share Chalmers' fondness for zombies nonetheless argue along similar general lines—concluding that if phenomenal concepts do in fact have features F (which will successfully explain the relevant data), they (phenomenal concepts) must pick out non-physical properties.

Another kind of anti-physicalist's move against the physicalist's strategy involves singling out particular (detailed) physicalist accounts of phenomenal concepts, criticizing them one at a time. For instance, Levine (2007) argues specifically against demonstrative accounts of phenomenal concepts which, he claims, cannot explain *all* the relevant data—Levine more specifically contends that such accounts cannot explain the significance and substantiveness of what Mary learns upon leaving her room.

In any case, the goal of this half of the thesis is to show that it *is* indeed possible to explain the relevant data and remain a physicalist. Making this case will require, first, setting out in some detail the anti-physicalist's data (1.1. and 1.2.); a

discussion of what phenomenal concepts are (section 2); a discussion of the phenomenal concept strategy—namely of how physicalists use phenomenal concepts to explain the anti-physicalist’s data (section 3); and finally a discussion of general arguments against the phenomenal concept strategy (section 4).

1.1. Anti-physicalist arguments

The anti-physicalist offers three related arguments against physicalism—and it is in these arguments that we find the data the anti-physicalist urges us to explain. I start off, then, by reviewing the three arguments.

1.1.1. Conceivability Arguments

I begin with conceivability arguments strictly so-called, i.e. the inverted qualia argument and the absent qualia argument. There is a sense in which *all* the arguments discussed in this section may fall under the loose heading of ‘conceivability arguments’ (see, for instance, Levine 2001), but absent and inverted qualia arguments appeal to conceivability explicitly, making them especially deserving of the name.

The anti-physicalist points out that our phenomenal thinking is such that the folk find a number of scenarios conceivable. For instance, the folk find conceivable that two people could be physically (functionally, representationally) identical and yet have “inverted” feels: when one of them, call her Adi, looks at a yellow flower, her experience has the same phenomenal feel as her twin’s experience (call her I-Adi) when the latter looks at another flower, identical to the first in every way except for its color, which is blue. A roughly identical conceivability judgment is said to be arrived at spontaneously by children as they start thinking about the experiences of

others. In the philosophical literature, we find mention of such conceivability judgments as far back as John Locke's *Essay*. He claims it isn't obviously false that

the same Object should produce in several Men's Minds different Ideas at the same time; e.g. if the Idea, that a Violet produced in one Man's Mind by his Eyes, were the same that a Marigold produces in another Man's, and vice versa. (1689/1975, II, xxxii, 15)

That the folk can conceive of inverts has some degree of plausibility, then. Some anti-physicalists go on to make another (less plausible) claim about what the folk find conceivable, namely that they find Adi's zombie twin (Z-Adi) conceivable. Z-Adi, like I-Adi, is a creature physically (functionally, representationally) identical to Adi, though Z-Adi's experiences, unlike Adi's, are "absent" such that if both twins looked at an identically colored flower, Adi's flower experience would feel one way to her while Z-Adi's experiences would feel *like nothing at all*. (Let me note that I have found it quite hard to convince some folk that they *can* conceive of zombies).

The fact that the folk can conceive of inverts and zombies *alone* does not ground any anti-physicalist conclusions, as the anti-physicalist herself is aware. Rather she reaches her conclusion by an appeal to the relation between what the folk find conceivable and what is in fact possible (in some relevant sense), as follows:

- (1) I-Adi (Z-Adi) is conceivable.
- (2) Whatever is conceivable is possible.
- (3) I-Adi (Z-Adi) is possible.
- (4) If I-Adi (Z-Adi) is possible, then physicalism is false.
- (5) Physicalism is false.

Physicalism does seem committed to the claim that any two physically identical individuals must be phenomenally identical too—physicalist functionalists and representationalists will be committed, more precisely, to the claim that any two

physically/functionally identical or representationally identical individuals must be phenomenally identical. Premise (4) is true then—there is some relevant sense of ‘possible’, such that if it is possible (in that sense) for Adi’s twin to have inverted feels, then two physically identical individuals fail to be phenomenally identical and physicalism is false.

What is less clear, however, is that premises (1) and (2) are true; more specifically, it is not clear that there is some kind of conceivability that entails the right kind of possibility (the kind that entails the falsity of physicalism in premise (3)), and that zombies are indeed conceivable in *that* way. In fact, some critics (see Kirk 2006) argue that this gives rise to an interesting tension in the anti-physicalist’s argument: the broader the sense of conceivability, the easier it is to make the case that zombies and inverts are conceivable (i.e. that premise (1) is true), but the harder it is to make the case that conceivability entails the right kind of possibility (i.e. that premise (2) is true). To see this, take ‘conceivable’ to mean *prima facie* conceivable, where something S is *prima facie* conceivable if “S is conceivable for that subject on first appearances” (Chalmers 1999, 8). Unpacked, assume that this means that, on first appearances, the subject cannot “detect any contradiction in the hypothesis expressed by S” (ibid). I- and Z-Adi are very likely to be conceivable in *that sense*: on first appearance, subjects may not detect any contradiction in the hypothesis that Adi may have an inverted or zombie twin. But unfortunately, it seems quite obvious that something being *prima facie* conceivable does not entail that it is possible in the relevant sense. After all, something may be conceivable on first appearances but not on further reflection—if *prima facie* conceivability is not even a reliable guide to

conceivability on further reflection, how can it be a guide to *possibility*? Restricting the sense of conceivability would indeed make it much more likely that conceivability might actually entail possibility, while making it much *less* plausible that inverts and zombies are indeed conceivable. Take, for instance, ‘conceivable’ to mean *ideally* conceivable, where something *S* is *ideally* conceivable “if an ideal reasoner could not rule out the hypothesis expressed by *S* a priori” (ibid). It is at least somewhat plausible that something which is *ideally* conceivable may indeed be possible in the relevant sense; but now it is far from obvious that I- and Z-Adi are so conceivable. Would an ideal reasoner find the thought of I- or Z-Adi to involve a contradiction? It isn’t obvious that she would—what’s more, it isn’t clear that, as *non*-ideal reasoners, we may ever be in a position to know whether or not these thoughts involve a contradiction.

The anti-physicalist can bypass *some* of these worries by construing conceivability arguments as arguments to the best explanation along the following lines:

- (1) I-Adi and Z-Adi are conceivable—that is, we can conceive of two physically (functionally, representationally) identical twins who aren’t phenomenally identical.
- (2) What best explains (1) is that there in fact *can* be two physically (functionally, representationally) identical twins who aren’t phenomenally identical—i.e. what best explains (1) is the falsity of physicalism
- (3) Therefore, physicalism is false.

In the rest of the discussion, I will think of conceivability arguments as arguments to the best explanation using a rather broad sense of ‘conceivability’. The folk do indeed, for the most part, judge zombies and inverts to be conceivable in some weak

sense of conceivable (say in the sense that there is no *a priori* contradiction in our *description* of zombies and inverts the way there would be in our description of, say, apples-that-aren't-apples). And one may have to explain why the folk should make such conceivability judgments.

1.1.2. *The Knowledge Argument*

The most well-known version of the knowledge argument is due to Frank Jackson (1986). He imagines Mary, a woman born and raised exclusively in a black-and-white room, her environment controlled in such a way that she never experiences any other colors. Mary becomes, as an adult, the world's leading color scientist—color science being as advanced as it can be, Mary actually comes to know *every scientific fact* (including every functional and representational fact) there is to know about color vision and color experience. Yet, when Mary is finally allowed to leave her black-and-white room, she learns something: she learns what it's like to actually experience color. What kind of fact is that, asks Jackson? Since we made Mary such that she knew *every scientific*—physical, functional, representational—fact about color prior to leaving her room, this new fact she learns upon leaving it cannot be one of *those*. It follows, then, that there are non-scientific, non-physical facts. Since physicalism can be construed as the thesis that every fact is a physical fact, Mary's story shows physicalism to be false. The argument is schematized by Jackson as follows:

- (1) Mary (before her release) knows everything *physical* there is to know about other people.
- (2) Mary (before her release) does not know *everything* there is to know about other people because she *learns* something about them on her release.
- (3) Therefore, there are truths about other people (and Mary herself) that escape the physicalist's story.

1.1.3. *The Explanatory Gap*

Reductive explanations, when they are successful, are satisfying. Why is it that water expands when it freezes? The answer goes something like (but is much more complicated than) this: water is made up of molecules of H₂O, and its constituent hydrogen and oxygen atoms have certain properties which allow them to bond with each other in various ways. Water expands because of the way hydrogen atoms bond when H₂O molecules have low energy. The explanation, if actually spelt out in all its lovely detail, would be satisfying, one might claim, because being merely told facts about molecules, their constituent elements and their properties would allow one to deduce the behavior of water at the macro-level. Knowing, that is, how atoms of hydrogen and oxygen behave would enable one to know what happens to water when it freezes.

Reductive explanations of phenomenal feels, unlike reductive explanations of the behavior of water at different temperatures, don't feel satisfying and never will, the argument goes. Why is it that seeing red *feels* this particular way? Why is it that experience, more generally, feels like anything at all? There seems to be no satisfying answer. Consider, for instance, a physicalist who claims that experiences feel like something because of the role (physical) experiential states play—a red experience feels the way it does because of its particular functional role. We can tell that such an explanation, even spelt out in detail, wouldn't quite do the trick. The important point may be this: that knowing how brain cells behave, or how functional states of the brain work, or how representational states interact still won't enable one to deduce what seeing red feels like. Why would it be the case that reductive explanations of

phenomenal feels—unlike reductive explanations of other physical phenomena—fail to feel satisfying? The best explanation may be that phenomenal feels—unlike other physical phenomena—are *not* physical phenomena.

Levine (2001) adds to this that there is a “core contrast” between usual scientific identities, like water is H₂O, and psychophysical identities such as *this_R feel* is representational property R. Though scientific identities might start off seeming arbitrary—why is water H₂O?—learning the relevant facts will dissipate the feeling. In fact, it would not make much sense for someone who does possess all the relevant facts to keep thinking that the proposed identity ‘water is H₂O’ was still arbitrary (see Levine 2001, 83 and 2007, 147). Psychophysical identities too might start off feeling arbitrary, however no amount of learning will make that feeling subside. Psychophysical identities *remain* arbitrary—no matter how much is learned. It always makes sense for someone to wonder whether *this_R feel* is indeed representational property R and not some other physical property.

1.2. Data and explanation

1.2.1. *The data*

The preceding three arguments are best construed, I think, as bringing to our attention four related observations which the anti-physicalist insists anyone—and that includes the physicalist—must explain. These observations include: i) that we make certain conceivability judgments, ii) non-derivability/non-deducibility; iii) the fact that Mary learns something substantial when she leaves her room; iv) a core contrast between typical identity claims in science and psychophysical identities.

(i) Quite obviously, the fact that the folk make certain conceivability judgments is at the core of conceivability arguments—and a physicalist who accepts this fact will need to explain why it is that the folk make such judgments.

(ii) Non-derivability and non-deducibility are related features of the Knowledge Argument and the Explanatory Gap. In keeping with what seems to be the standard notation, let 'P' stand for all the scientific/(micro)physical facts, and let 'Q' stand for phenomenal facts, like the fact that "*this* is what it feels like to see red".

At the core of both non-derivability and non-deducibility is the fact that 'P→Q' is not knowable a priori. Jackson's thought experiment makes the case that Mary cannot deduce phenomenal facts she does not know from the scientific/physical facts she learned about in her science books—facts about brains, functional/representational states and color. Merely knowing P (the scientific/physical facts) does not enable her to deduce Q (that red feels like *this*). In other words, 'P → Q' is not knowable a priori. This is indeed how Chalmers (2004) thinks of it: "the initial moral of the knowledge argument is that Q cannot be deduced from P by a priori reasoning. That is, the material conditional 'P → Q' is not knowable a priori."

(8) The Explanatory Gap is also in large part concerned with the fact that 'P → Q' is not knowable a priori. But it is arrived at by noticing that someone who knows P *and* Q (unlike Mary who knows P but not Q) will not be able to *derive* Q from P.

(iii) The anti-physicalist next presses us to explain why the knowledge that Mary learns when she leaves her room seems so substantive and significant. A satisfactory account of Mary's new knowledge requires more than an appeal to non-deducibility. After all, some facts could be such that they aren't deducible from other

facts, without the learning of these former facts being anything significant or substantive. Anna the astronomer may know *all* the facts about Hesperus (e.g., that it is 40 million kilometers away), but if she does not possess the concept PHOSPHORUS, she won't be able to deduce from the facts she knows that Phosphorus is 40 million kilometers away. Phosphorus-facts cannot be deduced a priori from Hesperus-facts. But imagine that Anna now acquires the concept PHOSPHORUS and is told that Hesperus is Phosphorus. If she really knew all the facts about Hesperus would acquiring that new concept be cognitively significant? Would Anna think that she just learned something striking, or surprising? Presumably she would not. At most she would learn that some think Hesperus is sometimes thought of as Phosphorus. But that hardly seems as though it would be very significant at all for Anna. And certainly it seems much less significant than what Mary learns when she learns that *this_R* is what it feels like to see red. It seems possible, then, for someone to account for non-derivability without accounting for the substantiveness of the knowledge acquired. But it is quite important, the anti-physicalist insists, that one explain not only why phenomenal facts cannot be deduced from physical facts, but also why Mary's learning of phenomenal facts is as significant and substantive as it seems to be.

(iv) Finally, the anti-physicalist demands that one explain the remaining core contrast between (ordinary) a posteriori scientific identity claims (like Hesperus is Phosphorus or water is H₂O) and what are, according to the physicalist, similar a posteriori identity claims involving phenomenal concepts (like *this_R* is representational property # 50). The latter (psychophysical) identity claims don't ever feel *satisfying*, whereas the former do. Why should that be?

Again, merely accounting for non-derivability is not enough, the anti-physicalist argues. Imagine now that Anna the astronomer knows a number of Hesperus facts and a number of Phosphorus facts without knowing that Hesperus is Phosphorus. Anna won't be able to derive Phosphorus-facts from Hesperus-facts. Yet the relevant identity claim (that Hesperus is Phosphorus) would not feel arbitrary for her for very long. Presumably, convincing Anna of the truth of the identity claim would involve showing her how Venus (the purported single referent of both her concepts), because of its trajectory say, would come to look to someone like Anna as though it has the properties she associates with the concept HESPERUS and the properties associated with the concept PHOSPHORUS. And if once we showed Anna this, she still thought the identity claim was arbitrary, "I believe we wouldn't understand what [she] was talking about" (Levine 2007, 147). Yet, if Mary "were to follow her exclamation [so *that's* what it's like to see red] with the question, "But why should it be like *that*?" we'd know what she means" (ibid). It makes sense to feel as though the psychophysical identity claim remains arbitrary, even after learning all there is to know about the relevant facts. And that needs to be explained.

Presented with this set of observations, which the anti-physicalist insists must be explained, the physicalist has (roughly) two options: she can either maintain that the observations in question are false/misguided—and explain why that is—or she can take them for granted, and show how they can be explained within a physicalist framework. The focus of Part I is on the merits of taking that second route.

1.2.2. *The role of phenomenal concepts*

How, then, can one account for this set of observations? Anti-physicalists and physicalists agree that an appeal to our phenomenal thoughts and their constituents (phenomenal concepts) is needed—even the anti-physicalists agree that merely positing anti-physical properties would do little to explain the observations in question. After all, the observations to be explained are *epistemic*. They concern what we can *conceive* of (zombies and inverts); what Mary can or cannot *know*, what she can or cannot *think*; what we can or cannot *deduce* or *derive*; which questions it makes sense for us to *consider*, and which would be we wouldn't *understand*. A full explanation of these observations will require, then, an account of phenomenal *belief*—and of those constituents of phenomenal beliefs seem to pick out phenomenal feels, phenomenal *concepts*.

The disagreement between physicalists and anti-physicalists is therefore a disagreement about whether an appeal to phenomenal concepts can successfully explain the set of observations within a physicalist framework. Physicalists believe they can; anti-physicalists believe they cannot. To put it another way, the core disagreement is one about whether it is possible 1) to explain the relevant observations fully, by appealing to phenomenal concepts, and 2) to remain a physicalist about phenomenal concepts and/or their referents. The term 'phenomenal concept strategy' (a term coined by Stoljar (2005)) refers to the *physicalist's* attempt at showing that it is possible to do both 1) *and* 2)—and that is how I will use the expression. Of course, anti-physicalists too must provide an account of how

phenomenal concepts—which on their view pick out non-physical properties—manage to fully explain the data.

It is crucial to keep in mind that the phenomenal concept strategy is a *defensive* physicalist strategy. Some critics (see Stoljar 2005) seem to take advocates of the strategy to be arguing that a successful appeal to phenomenal concepts *entails* the truth of physicalism, but that is not the case. A successful appeal to phenomenal concepts shows that physicalism is *consistent with* the existence of the Explanatory Gap, not that physicalism must be true.

2. Phenomenal Concepts

Phenomenal *thoughts* are thoughts about phenomenal feels, like the thought that green feels are annoying, or that *this_R* is what it feels like to see red. Some of the constituents of these phenomenal thoughts will presumably play an especially important role in explaining the anti-physicalist's data, if only because they are those constituents of phenomenal thoughts that actually refer to phenomenal feels (as for instance GREEN FEELS and THIS above).

We can use the expression 'phenomenal concepts' broadly to include *any* such constituent of a phenomenal thought that refers to phenomenal feels. Of course, to token phenomenal concepts (broadly construed)—i.e. to have a thought in which such a concept figures as a constituent—someone must be able to think about her own experience and its properties. It is important to realize, then, that creatures that are incapable of higher-order thought are, *ipso facto*, incapable of possessing and

deploying phenomenal concepts. Now, the fact that phenomenal concepts are higher-order (and require that someone introspect her own phenomenal feels) does *not* entail that higher-order thought is required in order for creatures to *have phenomenal feels at all*. The existence of higher-order concepts of experience is compatible with first-order theories of phenomenal feels. Tye, for instance, has forcefully defended (1995, 2000) a first-order representationalist theory of phenomenal feel: he claims that an experience is phenomenally conscious if and only if it has poised, abstract, nonconceptual content (PANIC). A creature can be in a PANIC state (hence be phenomenally conscious) according to Tye without being capable of higher-order thought. Yet, Tye himself acknowledges that to token phenomenal concepts—concepts that pick out these phenomenal feels—requires being able to *think* about PANIC states, something which naturally requires the capacity to introspect.¹⁵ Phenomenal concepts, as those constituents of thought that pick out our phenomenal feels, are necessarily higher-order concepts.

Fully explaining the data—conceivability judgments, what Mary learns, etc.—will require that we say more about phenomenal concepts, especially a *subset* of these concepts (broadly construed) whose members seem more intimately connected to phenomenal feels themselves. Some phenomenal concepts (broadly construed) seem *relational*, picking out phenomenal feels by their causes: like the concept THE FEEL

¹⁵ I take it that this is uncontroversial. However, some use the phrase ‘phenomenal concept’ to refer not to concepts *of* experience, but rather to those (sentential) representations to which phenomenal feels themselves are reduced. We see this in Rey (2007), who talks of the “gap between physical and phenomenal concepts (and/or non-conceptual contents; the distinction won’t be significant for purposes here.)” By phenomenal concepts, then, he means the nonconceptual contents that phenomenal feels are reducible to on his computational/representational theory of qualia. The distinction is certainly important to us here: by ‘phenomenal concepts’ we do mean concepts *of* experience—not nonconceptual contents (see Part II of the dissertation for a discussion of conceptual and nonconceptual contents.)

TYPICALLY CAUSED IN NORMAL MEMBERS OF MY COMMUNITY BY PARADIGMATIC RED THINGS or a similar (individual) concept THE FEEL TYPICALLY CAUSED IN ME BY PARADIGMATIC RED THINGS (see Chalmers 2003, 224). But other phenomenal concepts seem to refer more directly to phenomenal feels, as, for instance, the demonstrative concept THIS_R (pointing to a red phenomenal feel as I introspect). In the rest of Part I, I will use the expression ‘phenomenal concepts’ *narrowly* to refer to members of the relevant subset of these concepts (like THIS_R), unless otherwise noted. I do not mean to thereby suggest that phenomenal concepts (narrowly construed) are necessarily demonstrative concepts (as some indeed believe—see below)¹⁶. However, I find using the demonstrative THIS_R to be helpful in avoiding various kinds of confusions and will use it for this reason.¹⁷

The goal of this section is to make sense of (narrow) phenomenal concepts by spelling out the features physicalists widely assume that they have. I should note that physicalist accounts of phenomenal concepts (narrowly construed) seem to fall into three or four categories (demonstrative accounts, recognitional accounts and quotational accounts, along with a hybrid of the first two) and are presented as such. This is *not* how I will introduce phenomenal concepts in this section. What I will do,

¹⁶ Chalmers distinguishes between demonstrative phenomenal concepts and *pure* phenomenal concepts—the latter being the ones that play the crucial explanatory role. Demonstratives, like THIS QUALITY, pick out any phenomenal feel a thinker might be demonstrating. Pure phenomenal concepts, like *R* pick out *red phenomenal feels*. The demonstrative concept I use throughout (THIS_R) is, in some ways, more of a *pure* phenomenal concept (given Chalmers’ definition)—since it is “coupled” with a demonstration and thus anchored (it’s the concept THIS_R (pointing to a red phenomenal feel via introspection)). Chalmers himself acknowledges that “if someone wants to count pure phenomenal concepts...as ‘demonstrative’ in a broad sense..., there is no great harm in doing so, as long as the relevant distinctions are kept clear” (227). Again, my motivation for using the demonstrative THIS_R throughout the discussion *is* clarity.

¹⁷ Tye for instance uses the expression ‘RED_p’ for the phenomenal concept which picks out *red* phenomenal feels via introspection (2000—in his 2003 he uses RED*) and the expression RED_{NP} for the *nonphenomenal* concept red which picks out redness. This suggests that the two concepts (RED* and RED) are connected a priori which they most likely aren’t (see 2.1).

rather, is discuss first the (purported) relation physicalists believe exists between phenomenal concepts and other concepts (2.1); second, I will discuss their (purported) relation to their referent, i.e. to phenomenal feels (2.2). I will come back to general accounts of phenomenal concepts in subsection (2.3), arguing that we can best understand these accounts in light of the discussion in 2.1 and 2.2. In 2.4, I explain how I think we should understand the connection between the possession of phenomenal concepts (narrowly construed) and “knowing what it’s like”.

2.1. Phenomenal concepts and other concepts

Phenomenal concepts are related to other concepts in some interesting ways. Spelling out these relations, the physicalist believes, will help explain the data we started off with. Of course the anti-physicalist, too, will have to say something about the relations between phenomenal concepts and other concepts. To reiterate the (rather obvious) point made earlier, positing non-physical properties cannot help the anti-physicalist explain why phenomenal facts cannot be deduced from physical facts even by someone who knows all the physical facts. After all, what explains why Anna the astronomer cannot deduce Phosphorus-facts from Hesperus-facts has nothing to do with how many planets there are; but rather with the concepts she possesses—the concept HESPERUS and the concept PHOSPHORUS—and the relation between them. Similarly, that Mary cannot derive phenomenal facts from physical facts must be explained by an account of the concepts she possesses (phenomenal and physical concepts) and the relation between them. (Though nothing prevents the anti-physicalist from claiming, then, that to explain why phenomenal and physical

concepts are related as they are, we must posit non-physical properties, which is in fact what they do, see section 4.1.)

It is widely believed that phenomenal concepts are, in some way or other, conceptually isolated from physical concepts. Loar (1997) characterizes this conceptual isolation (alternatively called “cognitive irreducibility” or “conceptual independence”) as follows: phenomenal concepts “neither a priori imply, nor are implied by, physical-functional concepts” (295¹⁸). Tye (1995, 2000) writes that “no amount of a priori reflection on phenomenal concepts alone will reveal phenomenal-physical or phenomenal-functional connections” (2000, 30). Carruthers (2004) says that phenomenal concepts have “no conceptual ties with physical concepts, or with concepts of causal role and/or concepts of intentional content” (PG).¹⁹ My goal here is to spell out in more detail what it means to say that phenomenal concepts are conceptually isolated.

2.1.1. Clarifications

It may help to think of conceptually *linked* concepts as standing in a privileged relation to other concepts “generally by way of some type of inferential disposition” (Laurence and Margolis 1999, 5). It seems plausible to suppose that “one would have certain [a priori] dispositions linking *RED* and *COLOR*—for example, the disposition to infer *X IS COLORED* from *X IS RED*” (ibid) a priori. Or that one would have certain a priori dispositions linking *BACHELOR* and *UNAMARRIED MAN*.

¹⁸ Page numbers refer to the reprinted version of the paper in Chalmers’ 2002 anthology *Philosophy of Mind*.

¹⁹ See also Sturgeon (2000), Papineau (2002), Lycan (2002).

To say that phenomenal concepts are conceptually isolated, then, is to say that a thinker could possess the phenomenal concept $THIS_R$ (picking out a red phenomenal feel) without possessing any of the concepts of cognitive science—concepts like NEURON, S CONES, V1, BRAIN, REPRESENTATION, CONTENT, FUNCTIONAL ROLE, etc. That is, the thinker could think of some X that “X is $this_R$ feel” without being a priori disposed to think that “X is representational property #50”.

Interestingly, there seem to be no a priori connections between the phenomenal concept $THIS_R$ and the more general concept PHYSICAL STUFF. A thinker may think of some X that “X is $this_R$ feel” without thinking that “X is physical stuff”. In this way we might think that phenomenal concepts are unlike *many* other concepts (like WATER), which seem *a priori* connected with the concept PHYSICAL STUFF. Anyone who thinks of some X that “X is water” might be disposed to infer that “X is physical stuff” (as opposed to an abstract object).²⁰

Logical Connections

To claim that thinkers have no a priori disposition to link phenomenal concepts and physical concepts is not to say that phenomenal and physical concepts are not connected a priori in some *general* way, say logically. So those who defend (even a strong) conceptual isolation of phenomenal concepts need not (and do not) deny that a number of a priori logical connections will hold between these concepts and physical concepts. If Adi knows that some X is “either $this_R$ or representational property # 50 (rp50)”, then she will know *a priori* of X that “if it isn’t $this_R$, then it is

²⁰ This is not to say that phenomenal concepts are the *only* concepts that are *not* a priori connected to the concept PHYSICAL STUFF. We’ll return to this in the discussion of the core contrast (3.3).

rp50". Such an a priori logical connection might seem to hold between *any* concepts—and proponents of phenomenal concepts can grant that it holds between phenomenal and physical concepts as well. Still, these proponents will insist that phenomenal and physical concepts are conceptually isolated, meaning (more precisely) that there are no a priori connections between phenomenal concepts and *atomic* physical concepts.²¹

The Atomist

The claim about conceptual isolation appeals to a priori connections between concepts, and as such it may seem incompatible with the atomist account of concepts. After all, according to the atomist, there are no *a priori* connections between concepts. Take the concept BACHELOR. The atomist will deny that it is constitutive of possessing the concept BACHELOR that there be a connection between that concept and the concepts UNAMMARRIED and MALE. Hence there are no *a priori* connections between the concepts BACHELOR, UNAMMARRIED and MALE.

If the atomist denies that there is an *a priori* connection between BACHELOR and UNAMARRIED MALE, then she won't be impressed with our claim about phenomenal concepts. After all, to say that the concept THIS_R and the concept REPRESENTATIONAL PROPERTY # 50 (RP#50) aren't connected a priori, is to say nothing more than that phenomenal and physical concept pairs are like every other pair of concepts.

There is no way I can fully address this worry here. I do think there are plausible ways of dealing with it however, along the lines suggested by Laurence and

²¹ Thanks to Georges Rey for pointing this out.

Margolis (1999). They point out that though the atomist denies that there are *a priori* connections between concepts (connections that are constitutive of concept possession), she doesn't claim that there are no connections *at all* between concepts.

Like any other theorists, the atomist holds that people associate a considerable amount of information with any concept they possess. The only difference is that whereas other theorists say that much of the information is collateral (and that only a small part is constitutive of the concept itself), atomists say that *all* of it is collateral (Laurence and Margolis 1999, 65).

In other words, even on an atomist view, concepts will be connected to each other. It's just that none of these connections are *a priori*. Still, the atomist can claim that subjects will often *believe* (falsely) that some connections between their concepts are *a priori* while others are not.²² And so while subjects believe (falsely) that there are *a priori* connections between the concepts BACHELOR, and MALE, they won't similarly believe that there are *a priori* connections between phenomenal concepts and physical concepts. This will be enough for us to draw a contrast between BACHELOR and the phenomenal concept THIS_R even if atomism is true.²³

2.1.2. *How isolated are phenomenal concepts?*

Phenomenal concepts are thought to be isolated from physical concepts in such a way that there are *no a priori* connections between them. How plausible is that claim?

Not isolated at all?

Levine (2007) argues that phenomenal concepts aren't isolated at all. "We have a rich body of beliefs concerning the causes and effects of phenomenal states—composed of

²² I take it these beliefs about a priority needn't be ones that thinkers are consciously aware of.

²³ This is by no means to suggest that there are no objections to this move on behalf of the atomist. See Laurence and Margolis 1999 for more discussion.

both phenomenal and nonphenomenal concepts in the very same cognitive states” (152). Adi will be disposed to connect her belief that she is undergoing *this_R* feel to beliefs like the following: that this is the feel caused by seeing red; that this is the feel people in her community typically undergo when seeing red; etc. Levine concludes that “phenomenal concepts maintain [...] links to nonphenomenal concepts [...] so it seems as if cognitive isolation isn’t really the issue” (153).

Levine is right, of course. Phenomenal concepts will often be conceptually connected to nonphenomenal concepts—as those who believe phenomenal concepts to be conceptually isolated themselves will grant. Indeed, to say that phenomenal concepts are conceptually isolated is not to say that there are no connections between them and physical concepts; it is to say that there are no *a priori* connections between them and physical concepts. And it isn’t obvious that the connections Levine has in mind are *a priori* connections. Certainly the connection between *THIS_R* and the description “the feel people in my community ...” is *not* *a priori*.

Semi-isolated?

Some might insist, in the spirit of Levine, that there are a number of other physical concepts which are linked *a priori* to our phenomenal concepts—physical concepts which, though they are used in the cognitive sciences are also part of our ordinary everyday stock of concepts, like RED. So the phenomenal concept *THIS_R* (pointing to a red phenomenal feel) might seem connected *a priori* to the physical color concept RED.²⁴ However, I don’t think that it is. Imagine that Mary is finally allowed out of

²⁴ Rey in (written) conversation has suggested this. The plausibility of the connection between *THIS_R* and RED might seem more obvious to those (like Rey and Tye (2000, 2003)) who use expressions like

her black and white room but instead of being let out into the open, she is ushered into another room, one full of variously colored wallpaper (a variation on Nida-Rümelin's Marianna thought experiment (1996, 1998)). The wallpaper does not come labeled, and Mary is not allowed any equipment that would enable her to find out which panel of wallpaper is green, which is red, which is orange. In this odd room, as Mary stands looking at the panel of red wallpaper, she can token the phenomenal concept $THIS_R$. Yet she won't be disposed to think that she is seeing *red*—after all she doesn't know whether she's seeing red, or green, or orange. So the phenomenal concept $THIS_R$ is not a priori connected with the concept RED.

Now what about the concept COLOR? Certainly, Mary would be disposed to think that she is seeing *some color or other*. But could someone possess the phenomenal concept $THIS_R$ without connecting it with the concept COLOR? That is, could someone think of some X that "X is *this_R* feel (pointing to a red phenomenal feel)" and not be a priori disposed to infer that "X is related to color in some way or other"...say, as opposed to shape? We may wonder also about the connection between phenomenal concepts and concepts of the sense modality involved (like SEEING or HEARING). Could someone think of some X that "X is *this_R* feel" and not be a priori disposed to infer that "X is related in some way to seeing" as opposed to hearing? This is where things become increasingly unclear.²⁵ We could try to imagine a thinker who has always been blind. We could imagine a neuroscientist causing that

'RED*' or 'RED_p' to stand for the phenomenal concept that picks out red phenomenal feels. Which is one reason that I think this notation is misleading.

²⁵ Not that this should be surprising. Very obvious cases of *a priori* connections (as in the case of BACHELOR and UNMARRIED MALE) or lack thereof (as in the case of $THIS_R$ and BRAIN or RED) might be the exception rather than the rule. As it turns out, what some believe to be obvious *a priori* connections are controversially *a priori* (for instance Chalmers's (200X) claim that the concept WATER is a priori connected to the concepts CLEAR DRINKABLE LIQUID is far from obviously true. I come back to this point later in this section and in my discussion of the property dualist argument.

thinker's brain to token a red experience—and hence we could imagine that the thinker now possesses the phenomenal concept THIS_R. But could the thinker fail to know, then, that the feel she was introspecting was a *color* feel? It isn't obvious. Or we could try to imagine a thinker who never had any visual *or* auditory experiences. Again, a neuroscientist causes the thinker's brain to token a red experience—and the thinker as a result of thinking about that red experience tokens the concept THIS_R. Could she fail to know that the phenomenal feel was a *seeing* kind of phenomenal feel as opposed to an auditory one? Again, the answer is not immediately clear.

The lack of clear intuitions in these cases should not worry the proponent of phenomenal concepts, however. Ultimately, those who care about conceptual isolation need not claim that there are *no* a priori connections between phenomenal concepts and *any other concepts*. That sort of claim is too strong and is not needed. What will matter, ultimately, is not that phenomenal concepts are *completely* conceptually isolated, but rather that they are quite significantly conceptually isolated from the concepts of our physical, brain, and cognitive sciences, including the general concept PHYSICAL STUFF.

Strict a priori connections vs. pre-theoretical connections

So far, our test for a priori connections has been rather strict. To find out whether the concept THIS_R is conceptually linked with the concept RP#50, we should ask whether it is possible for someone to possess the concept THIS_R and not possess the concept RP#50. If the answer is “yes,” then the concepts are conceptually isolated; if the answer is “no,” then the concepts are conceptually linked.

But maybe this is *too* strict a test. After all, it seems possible (despite what I might have claimed) for someone to possess the concept THIS_R and not possess the concept PHYSICAL—which means that the concepts in question would be conceptually isolated. Moreover, it often seems that proponents of conceptual isolation do have in mind a slightly looser connection than the one I’ve been talking about. What matters might be what a thinker would come to believe pre-theoretically about the referent of her concept “armed only with her understanding of [the concept] and a bit of a priori reflection” (Byrne and Pryor 2004).²⁶ And now it becomes plausible to say that, armed with her grasp of the concept THIS_R and a bit of a priori reflection, a thinker might fail to draw the conclusion that *this_R* is physical. Or that she might even draw the conclusion that *this_R* is nonphysical. (Note that these are two different claims: the first says that there are no a priori connections between THIS_R and the concept PHYSICAL. The second – not entailed by the first – says that there is an a priori connection between THIS_R and the concept NONPHYSICAL.)

Indeed, some recent data suggests that a belief in a “nonphysical” mind may be part of our innate theory of mind. Bloom (2004) and others (see Kuhlmeier et. al 2004, Bering and Bjorklund 2004) have recently argued that human beings innately believe that the mind and the body are two different kinds of things. In one experiment, very young children were told the story of an alligator killing a mouse. Asked whether the mouse’s brain worked, or whether it would need food, the children overwhelmingly answered negatively. However, despite thinking that the mouse was

²⁶ Tye, for instance, says that “no a priori analysis can be given of [phenomenal concepts] in nonphenomenal terms” (2000, 26). The notion of analysis indeed fits in better with the “looser” notion of pre-theoretical connection (also see Jackson and Chalmers on conceptual analysis and consciousness).

dead, the children seemed to think that it still had a mental life (that it still loved its mother and liked cheese). This suggests that children do not seem to connect physical, bodily lives and mental lives (see Bering and Bjorklund 2004).

We can add that phenomenal concepts seem pre-theoretically connected to a number of other interesting concepts, at the very least concepts like INTROSPECTION or THINKING. Anyone who thinks of some X that “X is *this_R* feel” is disposed to think that “X is the thing that I’m introspecting”, or “X is the thing I’m thinking about” if nothing else. Again, we should be hesitant to accept any strong intuitions about strict a priori connections, so we needn’t claim that these concepts are connected a priori in a stricter sense—such that it would be impossible for someone to possess the concept THIS_R and not possess the concept INTROSPECTION or THINKING. But they seem quite obviously connected pre-theoretically.

Other pre-theoretical connections might hold between phenomenal concepts and the concepts that people have traditionally associated with qualia: concepts like PRIVATE, INEFFABLE etc. For before they learn much of anything about phenomenal consciousness, thinkers might indeed be disposed to infer that their feels are private, ineffable, etc. Again these connections are not *strictly* a priori. One can possess the phenomenal concept THIS_R without possessing the concept PRIVATE or INEFFABLE. They are a priori in a looser sense.

2.1.3 Summary

Phenomenal concepts are widely believed to be a priori conceptually isolated from physical concepts. There are two ways to think about this lack of a priori connection.

First, in a strict sense: phenomenal concepts aren't a priori connected to physical concepts iff it is possible for a thinker to possess a phenomenal concept and not possess any physical concept. Phenomenal concepts might indeed lack any such a priori connections to physical concepts (though there are unclear cases). Second, we can think of these connections as pre-theoretical connections. Phenomenal concepts might be a priori connected, in this looser sense, to a number of concepts. Most importantly phenomenal concepts seem a priori connected in this way to the concepts INTROSPECTION and THINKING, NONPHYSICAL and possibly also to the concepts PRIVATE, INEFFABLE, etc.

That phenomenal concepts *are* connected with some concepts a priori won't be a problem for the physicalist at all—as will become obvious when we put phenomenal concepts to work in section 3. And naturally it does not entail that phenomenal concepts are not, a posteriori, connected with physical concepts in a number of intricate ways.

2.2. Phenomenal concepts and their referents

Narrowly construed phenomenal concepts are thought to be related to their referents in two interesting ways. First, it seems that (narrow) phenomenal concepts can be tokened in someone's thought only if that person has actually undergone an experience with the referent phenomenal feel. Someone can think that *this_R* is what it feels like to see red, only if they've actually had a red experience. Second, it seems that phenomenal concepts refer to phenomenal feels *directly*. (So far as I can tell, these are two distinct,

separable features of phenomenal concepts and their referents, as will become clear in my discussion of each.)

2.2.1. *Undergoing the feel*

If phenomenal concepts can be tokened only once a thinker has actually undergone an experience with the referent feel, then phenomenal concepts should remind us of another kind of concept, namely, *recognitional* concepts. A concept is recognitional, it is often suggested, if possessing it (being able to think a thought involving it) requires the ability to recognize or re-identify things that fall under the concept. The concept RED is widely considered to be *the* paradigmatic recognitional concept:²⁷ possessing it (being able to think thoughts with the concept RED as a constituent), seems to require at the very least the ability to re-identify red objects as red (to recognize that they fall under the concept RED) relatively reliably at different points in time (see Fodor 1998, who goes on to “sort of prove” that there are no recognitional concepts). From which it follows, that a thinker can token the *recognitional* concept RED only if she has come into contact with (i.e., perceived) red objects before, just as a thinker can token the concept THIS (in the thought so *this* is what it’s like to see red) only if she has “come into contact” with referent red *phenomenal feels* before.

So it seems as though phenomenal concepts are a species of recognitional concepts. Note, however, that phenomenal concepts also seem *unlike* other recognitional concepts (such as RED). Though it is sometimes the case that someone will have the ability to recognize or re-identify a feel as falling under a phenomenal

²⁷ When Fodor (1998) argues, in Chapter 4 of *In Critical Condition*, that there are no recognitional concepts, he titles the chapter “There are no recognitional concepts; not even RED.”

concept she deployed earlier, it will be more often the case that we have experiences to which we can refer to using what *seem* to be phenomenal concepts, even while we lack the ability to recognize or re-identify the experience as having *that* feel when we undergo it later on. Adi's poi tasting experience at dinner one day and at lunch the next may trigger the same phenomenal concept, e.g. it's *this* kind of taste again. But Adi, at the paint store, thinks "so *this* is what it feels like to see firefly green" even though she lacks the ability to recognize this exact feel when she undergoes it again.²⁸ It would seem that her concept THIS is a phenomenal concept (in our narrow sense)—though not a *recognitional* concept in the traditional sense.

As it turns out, the re-identification requirement on recognitional concepts may be weakened. So we may claim that the possession of a (weak) recognitional concept does not require that the thinker have an ability (or disposition) to recognize that two things are of the same kind *at different points in time*—call such an ability a *diachronic* ability. Rather, all that may be required to token a weak recognitional concept may be the ability (or disposition) to recognize that two things are of the same type *at one given point in time*—call such an ability a *synchronic* ability (see Chuard 2006 for a nice discussion of re-identification conditions, though in another context). And indeed, Adi has dispositions to identify her feel as belonging to the same type as other feels when they are presented to her at the same time. If presented with a patch of lime green (another shade of light green) and a patch of firefly green at the same time, Adi would *not* be disposed to identify the respective feels caused in her by those patches as of the same kind. And another patch of firefly green,

²⁸ This is often the case in the case of very fine-grained phenomenal feels, though not for all (Raffman points out that we are very good at re-identifying the fine-grained shades of unique shades of red, green, yellow and blue (1995, see Levin (2007) for discussion)).

presented to her at the same time, would indeed cause Adi to think that there is another one of *these* feels.

We can conclude that some recognition/re-identification condition is required on phenomenal concepts (as for other (weak) recognitional concepts), though we can deny that what is required is *diachronic* recognition/re-identification. (Of course, it might be odd to use the term ‘recognition’ here. When we talk of recognition we do usually mean *over time*. So it may be less misleading to use the term ‘*re-identification*’).

Of course, some have argued that phenomenal concepts must be more robustly recognitional—namely, that possessing a phenomenal concept (and being able to deploy that concept in thought) requires diachronic recognition (i.e. the re-identification of an object as falling under the concept after a temporal gap). For if we deny that there is a diachronic recognition condition on phenomenal concepts, then it seems possible for someone to token a phenomenal concept in thought without that concept being stored in memory. But, one may claim, to even *be* a concept, a mental representation must be stored in memory. If phenomenal concepts aren’t stored in memory, then they aren’t even *concepts* at all.²⁹

I don’t think this need cause us worry. The term ‘concept’ may be used more or less technically to pick out different kinds of mental representations.³⁰ It may indeed be used by most cognitive scientists and “concept researchers” to refer to

²⁹ Prinz (2007) expresses something like this worry, though he argues for the strong claim that the things we want to call phenomenal concepts *cannot ever be stored in memory*. He concludes that there are no such things as phenomenal concepts.

³⁰ Or of abstract objects for the Fregeans. Macherie (2005 and forthcoming) argues that even in psychology the term ‘concept’ picks out three different kinds of things (exemplars, prototypes and theories).

those mental representations that are stored in memory (as Prinz (2007) claims). However, those interested in phenomenal concepts (physicalists and anti-physicalists alike) are interested first and foremost in phenomenal *thought*. For they claim that the set of observations put forth by the anti-physicalist, being epistemic, will have to be explained by our way of *thinking* about phenomenal feels, that is, by our phenomenal thoughts. Particular *constituents* of phenomenal thought will play an especially important role in explaining these observations, and those have been called phenomenal *concepts*. The term ‘concept’, in this debate, then, given our particular endeavor, is used to pick out those mental representations which are constituents of *thoughts*—regardless of what else may be true of them (whether or not, for instance, they are stored in memory)^{31,32}.

Where does this leave us? For Adi to token a phenomenal concept in her firefly green thought, must she have the ability to recognize (or re-identify) firefly green feels? That depends on whether this ability is synchronic or diachronic. It would seem as though Adi needn’t have the ability to recognize firefly green feels after a temporal gap. The recognition condition on phenomenal concepts could then turn out to be relatively weak: a phenomenal concept may be *recognitional* in the sense that possessing such a concept (i.e. being able to think with such a concept)

³¹ Prinz (2007) makes the following claim: that for mental representations to be *concepts*, they need to be “capable of being stored” (201). And he goes on to argue that no so-called phenomenal concept is capable of being stored. He concludes that there are no phenomenal concepts. Even if he is right and no so-called phenomenal concepts are even capable of being stored, we (those interested in the so-called phenomenal concept strategy) needn’t worry about his claim at all (as he has acknowledged in conversation). For, again, the point is that there are constituents of our phenomenal thoughts (whether or not they are concepts in some more technical sense) which can help explain the relevant anti-physicalist data.

³² In the second half of the thesis, the issue about the relation between concepts and memory will come up again.

requires synchronic identification.³³ Recognitional concepts, even those with weaker recognition/re-identification conditions, do have the feature we were emphasizing at the beginning of this section: their deployment in one's thoughts requires that she have actually experienced the property that she has the ability to recognize/re-identify. Possessing the recognitional concept RED is not merely a question of having the ability to discriminate between red objects and other objects, for presumably anyone who isn't colorblind would have the *ability* to make the relevant discriminations, even if she had never actually seen anything red. To token the recognitional concept RED, one must have the ability to re-identify red things *and* have come into perceptual contact with red things.

Similarly, for Adi to token the phenomenal concept THIS_G (referring to (general) green feels) she must have undergone the kind of feel in question. Though Mary, in her black-and-white room, may have the *ability* to discriminate between colored feels (an ability as of yet untapped), she cannot deploy the phenomenal concept THIS_G. Phenomenal concepts, that is, are *recognitional*.

³³ As a matter of fact, those who have claimed that phenomenal concepts are *recognitional* seem, for the most part, to have had a rather weak condition in mind. Phenomenal concepts are not recognitional concepts in quite the way that paradigmatic non-phenomenal concepts (like RED) are recognitional (and which seems to involve diachronic recognitional dispositions). Loar (1997) says that recognitional concepts are "grounded in dispositions to classify, by way of perceptual discriminations, certain objects, events, situations" (298). However, having the ability to *discriminate* between two experiential feels and to classify feels need not require that one be able to re-identify two feels as of the same kind *after some time has passed*. Carruthers (2004) and Tye (2000), like Loar, shy away from the rather strong characterization requiring recognition over time. Carruthers writes that "a concept is recognitional when it can be applied on the basis of perceptual or quasi-perceptual acquaintance with its instances" (pg ##). Adi's phenomenal concept THIS in her thought "*this* is what it feels like to see firefly green" is recognitional, then, simply because it can be deployed on the basis of acquaintance with her experiential feel. There is no requirement, here, that the feel be *recognized* later on in time.

2.2.2. *Phenomenal concepts refer directly*

It is said that phenomenal concepts are concepts that refer to phenomenal feels *directly* (see White 2007, Levine 2007, Levin 2007, Tye 2000 among others), a claim that is often spelt out in terms of reference-fixing. Tye writes that phenomenal concepts enable us to recognize phenomenal feels “via introspection without the use of any associated reference-fixing intermediaries” (28). White (2007) also claims that unlike other concepts such as HESPERUS and PHOSPHORUS, phenomenal concepts do not come to pick out their referents via a mental description associated with the concept by the thinker—a description which would *uniquely* determine the referent. HESPERUS comes to pick out Venus because the thinker associates with HESPERUS a description like “star that rises here in the evening”, which uniquely picks out Venus. The phenomenal concept THIS_R, on the other hand, comes to pick out, say, representational property # 50 (or @ for short), without the mediation of any such mental description of it. This claim is sometimes expressed in terms of ‘modes of presentation’ as follows: the relation between phenomenal concepts and their referents is not mediated by a *mode of presentation*. According to White:

the relation of “pain” to pain [...] is not mediated by a mode of presentation of pain. In this it differs from the referential relation commonly thought to hold between “Hesperus” or “Phosphorus” and Venus. The ordinary assumption is that the reference, for example, of “Hesperus” to Venus is mediated by a description such as “the first heavenly body visible in the evening” (211).

Unfortunately, the claim that phenomenal concepts refer directly is sometimes ambiguous. Some take it to mean *not merely* that the reference of phenomenal concepts is determined directly (i.e. not via a mental description), but also that phenomenal concepts are simply such that there are *no* mental descriptions associated

with them. It is important to note that the two claims are very different. To say that the reference of a concept is determined directly (however one is to cash out what the reference-fixing mechanism is) is not to say that there are no mental descriptions associated with that concept at all. After all, one might want to claim (as many in fact *have* claimed since Kripke) that what determines the reference of HESPERUS is not any description thinkers might associate with it but rather some causal/historical link between the thinker and Venus. This fact does not entail that thinkers do not associate *any* descriptions with their concept HESPERUS, but only that these descriptions are not reference-fixing.

Of course, the two claims might be linked in this way: if a concept is such that there are *no* descriptions associated with it, then descriptions associated with it could not possibly fix its reference—and something other than descriptions would have to play the reference-fixing role. Interestingly enough, many of the writers who claim that what fixes the reference of phenomenal concepts is not a set of descriptions seem to do so by arguing, first, that there are no descriptions associated with these concepts at all—hence, that something else must be fixing their reference (see Ismael 1999, Carruthers 2000, 2004). But here is another quite important thing to note: claiming that phenomenal concepts are not associated with any mental description is to make a claim about the relation between phenomenal concepts *and other concepts*, not between concepts and their referents. After all, mental descriptions are merely complex mental representations with concepts as constituents. If we have reason to think (as argued in section 2.1. on conceptual isolation) that phenomenal concepts are not likely to be *wholly* isolated from other concepts, then we have reason to think that

phenomenal concepts are associated (even a priori) with some mental descriptions, even if those descriptions are not reference-fixing.

In any case, let us assume, as many seem to do in the phenomenal concept literature, that what determines the reference of phenomenal concepts are not descriptions. What else might determine the reference of these concepts? There are currently three proposals on the table: demonstrations, recognitional abilities, and quotation. I take a look at each in turn.

Demonstrations

Even if we assume that what determines the reference of HESPERUS is a description that the thinker associates with the concept, we might think that (at least some) demonstratives, work rather differently. It might seem as though what determines their reference is an act of demonstration—maybe a pointing gesture (Kaplan 1989a) or a “directing intention” (Kaplan 1989b). Again, this is not to say, of course, that thinkers don’t associate various descriptions with their demonstrative concepts. Anna could associate with THIS STAR (pointing at Hesperus) a number of descriptions of the referent: it’s a heavenly body, it rises here, etc. But the reference of her demonstrative is not *fixed* by these descriptions. It is fixed rather by her act of pointing—or her “intention to point at a perceived individual on whom [s]he has focused” (Kaplan 1989b, 582).

Some of the writers who think that phenomenal concepts *are* demonstratives make claims of this sort—though they seem to want to say first that phenomenal concepts are not associated with *any* descriptions. Ismael (1999) claims that

phenomenal concepts *ostend* phenomenal feels, where ‘ostension’ is the “name for identification without the employment of [associated] representations” (356). He goes on to say, a few pages later, that

to ostend a part of the world one need not know anything about it; as a matter of fact, one must only be appropriately related to it. It is an *actual, external* relation between oneself and what one points at—regardless of what one knows, or thinks one knows, regardless, that is, of anything ‘in one’s head’—that makes *it*, rather than any other thing, the object of one’s ostension (359).

Ismael’s point is, I think, that phenomenal concepts, like some demonstratives, do not have any descriptions associated with them. From which it follows, of course, that no description can fix the reference for these concepts. Reference must be fixed, then, by some act of demonstration.

Recognitional Abilities

I have claimed above that (narrow) phenomenal concepts are like recognitional concepts in *one* respect: they can be tokened only by subjects who have “come into contact” with the referent kind, and have the ability to re-identify (whether diachronically or synchronically) other instances belonging to that referent kind. I made no claim, then, about what might fix the reference for these concepts. However, it seems plausible to assume that what fixes the reference of a recognitional concept is *not* a set of descriptions thinkers associate with that concept, but rather the very recognitional (re-identifying) abilities and dispositions of the thinker.

Let us go back to an example we used in our earlier discussion of recognitional concepts: Adi, in the paint store, thinks “so *this* is what it’s like to see firefly green”. What determines the referent of her recognitional concept *this*?

Presumably her recognitional abilities. Janet Levin (2007) claims that nothing other than this ability *could* determine reference here:

[T]he ability to recognize or reidentify is required to underwrite *determinate reference* to a particular property. The best way—perhaps the only physicalistically acceptable way—to determine whether someone’s current “pointing in” denotes what it’s like to see some particular shade of red, or a more coarse-grained phenomenal property (e.g., red in general, or color) [...] is to see what she is disposed to identify as *other* instances of that property (89).

So, if it turns out that Adi cannot re-identify (even at the same point in time) firefly green with other shades of light green, then we have reason to think that her recognitional concept THIS does not actually refer to firefly green (despite what Adi herself might think).

Loar (1990), too, endorses such a view of the relation between phenomenal concepts and their referents: phenomenal concepts are *direct* recognitional concepts, he claims, where direct recognitional concepts are recognitional concepts that refer “unmediated by a higher-order reference-fixer” (87). Loar adds that even “lacking a name [for an object] I may still come to recognize instances [of it] here and there” (1990, 88). Consider, for instance, subjects in Livingston, et. al’s experiment (1999), who acquired two recognitional concepts, GEX and ZOF, without being told explicitly, *and without explicitly learning*, what made one of the “creatures” on the screen a Gex as opposed to a Zof. What fixes the reference of their recognitional concepts GEX and ZOF is not a description which they associate with GEX or ZOF and which uniquely determine the reference of these concepts. After all, subjects *do not know* what differentiates Gexes and Zofes—so they do not even have any associated mental description which would uniquely pick out the referent. What determines the reference of these subjects’ concept GEX is simply their recognitional abilities. (In the

revised version of “Phenomenal States” (1997) Loar defines *direct* recognitional concepts rather differently, as those recognitional concepts that pick out their object or property via a *necessary mode of presentation*. I ignore this complication now, until section 4).

Carruthers (2000, 2004) argues that phenomenal concepts are *purely* recognitional:

A concept is purely recognitional when nothing in the grasp of that concept, as such, requires its user to apply or appeal to any other concept or belief (2004, 4).

Carruthers uses the example of “chicken sexers”, who are

people [who] can be trained to [identify the sex of] very young chicks entirely intuitively by handling them, without having any idea of what they are doing, or of the basis on which they effect their classifications (2000, 56)

Chicken sexers might have no name for the properties they’re picking out: we can imagine them thinking simply, “it’s one of these”; “now it’s one of those”. A chicken sexer, then, would have *almost* no beliefs about the nature of the property she is picking out and her recognitional concepts would be, correspondingly, *almost* purely recognitional—though not quite. Carruthers grants that it is unlikely that chicken sexers would have *no* associated descriptions of what it is they are picking out. Phenomenal concepts, then, are simply more “extreme” versions of recognitional concepts like those used by chicken sexers—recognitional concepts which have *no descriptions* associated with them whatsoever. It will therefore follow that what fixes the reference of these concepts cannot be associated descriptions—but rather something like recognitional abilities.

Quotation

Quotational concepts have an odd relation to their referents: they are concepts that “contain” or “embed” or “quote” their referents. What determines the reference of these concepts is simply their quoting it. Papineau (2002), Block (2002, 2007) and Balog (draft) offer quotational accounts of phenomenal concepts. On Papineau’s 2002 account, phenomenal concepts are “compound terms” of the form ‘the experience:----’ where the “---” stands for the experience itself (along with its phenomenal feel). The phenomenal concept contains the very experience it picks out. The quotation involved in such accounts should be understood by analogy to linguistic quotation. The referent of ‘I walk the line’ is the very sentence which is located between the quotation marks. Quotation can be iterated an unlimited number of times: ‘I walk the line’ refers to the *quoted* sentence which is located between the quotation marks, namely the quoted sentence ‘I walk the line’ and so on. Importantly, in ordinary text we can quote things other than sentences. Drawings or symbols, for instance, can be quoted. Following Balog, I will now use ‘*’ to signal mental quotation. The quotation of drawings is actually not unusual; it is sometimes used in novels in ways that feel natural enough.³⁴ So, Balog claims, just as we can quote in natural language (in English) we can quote in thought. If I work on a theory of concepts, I may have thoughts like this: *I wonder what the structure of a concept like *BACHELOR* is.* When I think about my own concept BACHELOR I am, plausibly, quoting it: I’m using a device that enables me to refer to, or rather to mention it (the concept) without *using* it. The idea now is this: that just as I can quote symbols (or

³⁴ See, for example, Dan Brown’s *Angels and Demons* (Illuminati anagrams) and possibly some Harry Potter books (tear-stained “handwritten” letters).

drawings) in a text, I can quote things other than concepts in thought: more specifically I can quote experiences. When I think *I love these (red-feels)* my concept THESE in fact quotes those very experiences: it really should look like this, *I love *®**. (I discuss the strengths and weaknesses of these three accounts later on in Part I (see section 3))

2.3. Accounts of phenomenal concepts

In this section, I have discussed various features that play important roles in physicalist accounts of phenomenal concepts. Phenomenal concepts are thought to be related to other concepts in a particular way: thinkers will have no a priori dispositions to link thoughts like “X is *this* feel” to thoughts about the nature of X (like “X is physical state such-and-such”). Phenomenal concepts are also thought to be related to their referents in two interesting ways: first, they are concepts that cannot be tokened unless the thinker has “come into contact” with the referent (has undergone the feel). Second, they are concepts whose reference is fixed not by descriptions but directly. There are three physicalist proposals concerning what that reference-fixing mechanism might be: acts of demonstration, recognitional abilities, or quotation.

As promised at the beginning of section 2, I will now consider the mainstream physicalist accounts of phenomenal concepts (demonstrative, recognitional, hybrids, and quotational accounts) and show how they are related to the features discussed in 2.1. and 2.2. and argue that we should be wary of quotational accounts (2.3.2)

2.3.1. Mapping the accounts

Various accounts of phenomenal concepts are often presented, in the literature, as being significantly unlike each other. I don't think they are. And that is what I show below.

According to *recognitional* accounts, phenomenal concepts are recognitional concepts of some kind. Why make this claim? Because some recognitional concepts have *three* important features: first, they too seem *conceptually isolated* (in some instances). Chicken sexers (discussed by Carruthers) will have dispositions to infer very little about the nature of what it is they are picking up on. Phenomenal concepts are simply a bit *more* isolated than the chicken sexer's (almost) conceptually isolated recognitional concepts. Second, it is plausible to claim that the reference of recognitional concepts is fixed not by associated descriptions but by recognitional abilities. Third, recognitional concepts cannot be tokened before the thinker has come into contact with the referent.

Demonstrative accounts of phenomenal concepts (defended by Perry (2001), Ismael (1999), and O'Dea (1999)) make the claim, rather unsurprisingly, that phenomenal concepts are demonstratives of some kind. Why make such a claim? For two main reasons: first, because (some) demonstratives are, in some circumstances, *conceptually isolated* from other concepts in ways that give rise to non-deducibility. Imagine, then, that Adi is kidnapped and wakes up locked in a strange room. In the room she finds a very detailed map of the state of Maryland, which she assumes—rightly—is the state she is in. She knows, of course, that she is *here* in some sense—though she doesn't know what coordinates on the map *here* points to. She knows that

this (pointing at the space around her) is her location—that she is in *this* corner of the room, and that *this* is where she sat an hour ago. But she has no way of linking her demonstratives with the map she studies. In fact, she can study it all she wants and come to know all there is to know about Maryland geography. Nothing will help, for what is missing is a connection between the spatial demonstratives she deploys and the geographical map. Second, it is plausible to think that the reference of demonstratives is not determined by an associated description but by an act of demonstration. To claim that phenomenal concepts are demonstrative concepts, then, is to claim that they have two of the features we discussed above: 1) they are conceptually isolated; 2) they refer directly. Defenders of demonstrative accounts usually come to have such a hybrid account. O’Dea (2002) writes that phenomenal concepts are “partly recognitional” (178); Perry (2001) argues that our phenomenal concepts have “a demonstrative/recognitional core” (141). What obviously motivates these writers is the desire to capture the recognitional feature of phenomenal concepts—a feature that mere demonstratives don’t necessarily share. Proponents of demonstrative accounts, like proponents of recognitional accounts, ultimately hope to claim, then, that phenomenal concepts have the *three* features I discussed in 2.1 and 2.2.

Quotational accounts also attempt to capture these three features. They argue that phenomenal concepts quote or embed their referents, phenomenal feels. Such embedding, they think, will give rise to conceptual isolation. After all, it seems plausible enough to say that the odd concept THIS:® would not be obviously connected to other concepts that, unlike it, do not quote referent feels. Also, it seems

plausible to claim that what determines the reference of a concept like THIS:® is the quotation within it. Finally, the concept THIS:® cannot be tokened without the referent phenomenal feel (®) itself being tokened.

2.3.2. Demonstrative/recognitionnal vs. quotational

As I just pointed out, these accounts have very much in common. Still, we might wonder which we should prefer—if we should be defend a demonstrative/recognitionnal account or a quotational one (since these two accounts are widely considered to be the two main rivals, I focus my discussion on them). Of course, a full answer to that question would require taking a look at how these different accounts handle the anti-physicalist data. Levine, for instance, claims that demonstrative/recognitionnal accounts can't explain what Mary learns when she leaves the room. And it might seem as though quotational accounts will do a better job than demonstrative/recognitionnal accounts at explaining the core contrast. I leave these issues for the next section (I argue there that Levine is wrong (3.3.3) and that quotational accounts can't explain the core contrast any better than any other physicalist account of phenomenal concepts.) What I do consider now are some independent reasons for thinking that demonstrative/recognitionnal, or quotational accounts, cannot be the right accounts of phenomenal concepts.

Quotational accounts are, in some ways, the most puzzling of the three accounts. After all, we have demonstrative thoughts, and we possess and deploy recognitionnal concepts; phenomenal concepts, on demonstrative and recognitionnal accounts, are just quirky versions of concepts we are used to. Quotational concepts,

however, seem to be an entirely new species of concepts. The first strike against quotational concepts concerns the nature of quotation itself: some consider quotation to be simply a kind of demonstrative pointing (Levine (2007) makes something like this point). When Adi quotes a sentence like ‘I walk the line’, some argue that she simply points demonstratively to the sentence within the quotation marks. If quotation can be reduced to demonstrative pointing, then quotational accounts of phenomenal concepts may be reducible to demonstrative accounts of phenomenal concepts. Now as it turns out, what motivates defenders of the quotational view is a dissatisfaction with demonstrative and/or recognitional accounts: Levine (online) and Balog (online) both argue along these lines. Balog, for instance, suggests that on recognitional/demonstrative accounts, a “phenomenal concept and its referent [are] distinct existences related by causation” (17/18). She continues:

But it seems that this leaves too much of a distance between, e.g., a phenomenal concept one applied to a particular pain as it occurs (let’s call the concept P) and the particular pain itself, as on this view their occurrence is independent.

The recognitional/demonstrative view leaves open the possibility of someone tokening the concept P while not being in pain. But that is unfortunate, she thinks: “anybody who tokens a first personish phenomenal concept of pain purporting to refer to a current state is really in pain” (ibid 18). The quotational account circumvents this problem by eliminating the “distance” between phenomenal concept and referent: anyone who has a thought with a phenomenal concept as a constituent will have the very feel the phenomenal concept “points to” as a constituent as well. Of course, it is still possible for a subject to be mistaken about feeling pain, as when Adi thinks “*this (heat feel)* is a pain feel”. Any categorization of

any feel as *pain, heat, fear*, etc. may be wrong. What the quotational account may avoid is the deployment of a phenomenal concept *in the absence of any feel whatever*, for on the quotational account, a deployed phenomenal concept will necessarily quote a feel: it will be impossible, then, for someone to think “I am having *this* feel” while she’s having no feels at all. Balog is eager to deny, then, the possibility of a “conceptual” zombie as she calls it—an individual identical to our subject Adi *conceptually* but who doesn’t have feels. Z-Adi, Balog wants to claim, couldn’t possess phenomenal concepts.

As it turns out, defenders of demonstrative/recognition accounts can, like Balog, deny that zombies could possess phenomenal concepts—if they construe phenomenal concepts *in terms of phenomenal feels*.³⁵ When I first introduced phenomenal concepts in this chapter, I introduced them by making reference to a subject’s phenomenal feels: a phenomenal concept is a concept that picks out phenomenal feels. It is clear that zombies cannot have phenomenal concepts so construed simply because, by supposition, they lack feels altogether. But I went on to characterize phenomenal concepts as somewhat conceptually isolated, (weakly) recognition concepts that pick out their referents directly, and *so* construed phenomenal concepts are such that zombies *can* possess them. But of course, it isn’t clear why we want to avoid claiming that zombies possess “phenomenal concepts” when construed in this way. So there is no reason to prefer quotational accounts to demonstrative/recognition ones so far.

Quotational accounts also come with a higher price tag, for it is hard to see how a phenomenal *feel* could be a component of thought as the quotationalist would

³⁵ Or “*first-personally*” (see Carruthers and Veillet 2007). This same distinction comes up in Part I 4.2.

have it. It is important to notice that quotational accounts are in fact incompatible with many physicalist accounts of phenomenal feels, more specifically with any, even partly, *functional* account of feels. For if, as the functional representationalist believes, what makes something a feel is in part its functional role R, then it is hard to see how something *with functional role R* could be a constituent of thought as well. After all, it seems that constituents of thought will have *their* functional roles, say role Q, and it is hard to see how the same *token* representation could have *both* functional roles so as to be both a phenomenal feel and a constituent of thought. If anything, then, we should be wary of quotational accounts.

2.4. Phenomenal concepts and what it's like

In this section (section 2), I have introduced phenomenal concepts and discussed what I take to be their interesting features. Before we can move on and put phenomenal concepts to work, there is one last thing it will be helpful to settle: the relationship between possessing (narrow) phenomenal concepts and knowing what it's like. This is especially important because some seem to think that the relationship in question is extremely tight, such that all there is to coming to know what it's like to see red is coming to possess the (narrow) phenomenal concept $THIS_R$.

The worry then is this: as mentioned in 2.2.1, narrow phenomenal concepts may be like recognitional concepts in that their possession might require some weak, *synchronic* recognitional abilities, i.e. abilities to re-identify the feel when presented with others simultaneously. But knowing what it's like (unlike the tokening of (weak) phenomenal concepts) would seem to require recognition of a feel *at different points*

in time. It seems to require other abilities too, like the ability to visualize the particular shade of color when it's not being experienced (and possibly the other abilities Lewis (1990) associated with what Mary learns when she learns what it's like). On the other hand, insisting that knowing what it's like requires strong diachronic recognition has somewhat counterintuitive implications. For it would follow that when we're looking at very fine-grained color—as when Adi was looking at firefly green in the paint store—most of us don't know what it's like to see these shades (see Raffman and Levin 2007). After all, most of us are unable to diachronically recognize fine-grained shades like firefly green, to remember and visualize them. But certainly it seems that we do know what it's like to see firefly green *while we're looking at it*—even if not at any later time.

The best way to think of this is the following (see Levin's 2007): there may be more or less robust ways of 'knowing what it's like'. On the one hand, Adi knows, as she looks at the firefly green patch, what it's like to see firefly green, even while she doesn't have robust (diachronic) recognitional abilities. In this (weak) sense, knowing what it's like *just is* deploying (weak) phenomenal concepts. But we can take knowing what it's like to require more; and in *that* sense of the expression, knowing what it's like requires having diachronic recognitional abilities. Adi knows, while looking at the firefly green patch, what it's like to see firefly green in some (weak) sense; she doesn't know what it's like in some more robust sense.

This wraps up our section 2. I introduced phenomenal concepts and their three features; I highlighted the connection between these features and traditional accounts of phenomenal concepts and finally the connection between phenomenal concepts

and “knowing what it’s like”. We are now ready to use these concepts to explain the anti-physicalist data.

3. Putting phenomenal concepts to work (the phenomenal concept strategy)

Now that we have a better idea of what the physicalist takes phenomenal concepts to be—what features she takes phenomenal concepts to have—we can be more precise about how they can be put to work. The goal of this section is to show how the physicalist thinks an appeal to phenomenal concepts (concepts with the features we discussed earlier) can explain all of the data.

3.1. Conceivability judgments and non-derivability

Conceptual isolation explains our conceivability judgments rather straightforwardly, the physicalist will claim. The folk are asked to think of a physical, or functional, or representational twin of Adi who, like Adi, is in functional state such and such; in physical state such and such, etc. ... However, since the folk lack any a priori dispositions to infer from “X is in functional state such and such” that “X has *this_g* feel”, the folk will lack any a priori dispositions to infer from “Adi is in functional state such and such” that “Adi has *this_g* feel”. The folk can therefore wonder whether Adi would have *this_g* feel, they can therefore believe that Adi might have *this_r* feel, or believe that Adi might have no feels at all. Anna our amateur astronomer can wonder whether Hesperus is Phosphorus, and can even believe that Hesperus is *not* Phosphorus, presumably because her concepts HESPERUS and PHOSPHORUS are

conceptually isolated, i.e. not a priori linked. And so it goes with the folk and phenomenal feels.

Non-deducibility and non-derivability similarly look like they can be straightforwardly accounted for by an appeal to conceptual isolation. After all, at the core of both non-deducibility and non-derivability is the idea that phenomenal facts and physical facts are not linked a priori. Knowledge of all the physical, functional, representational facts (the P facts) will not enable Mary to *deduce* that seeing green feels like *this_g* (the Q facts). But knowing phenomenal facts (Q) requires the deployment of phenomenal concepts. Since phenomenal concepts are not a priori linked to physical concepts, it follows that ‘P → Q’ is not a priori.

3.2. What Mary learns

As mentioned earlier, the mere fact that phenomenal concepts are conceptually isolated from physical concepts will not quite account for what seems substantive about what Mary learns upon leaving her room. HESPERUS and PHOSPHORUS are conceptually isolated, yet if Anna knows all there is to know about Hesperus, learning the concept PHOSPHORUS will be quite boring to her. It will be like learning that there is another name for an object one knew everything about in the first place—no big deal, really. But clearly, what goes on when Mary leaves the room is very much different. What she learns is cognitively significant. Can a physicalist account for that?

3.2.1. Perry's attempt: the Dretske analogy

Perry (2001) argues that beliefs can be detached from perception (and the resulting demonstrative beliefs), as his own belief that Fred Dretske wrote *Knowledge and the Flow of Information* turned out to be disconnected from his perceptual experiences of Dretske at a party—and the resulting demonstrative beliefs, such as “this man has white hair and is interesting to chat with”. When Perry finally learns that the man in question is actually Fred Dretske, the knowledge he gains is the “sort of knowledge that occurs when one attaches percept and notion” (122).

Many of us, Perry argues, are in a very similar position. We can think of our phenomenal feels demonstratively when we introspect—*this_R* feel. We can also think about the physical or functional or representational properties of our brains. But the two kinds of information are detached. Mary, despite knowing all there is to know about experiences of red, is in this situation also. She is unable, from within her room, to attach her expert, all-encompassing knowledge to the phenomenal demonstratives deployed as she introspects (after all, she is unable to token such phenomenal demonstratives). What happens to her when she leaves the room is cognitively significant, then. It is just as cognitively significant for Mary as learning that *this* man is Fred Dretske was for Perry at the party.

Though Perry believes that phenomenal concepts are essentially demonstrative, what plays a primary role in his explanation of what Mary learns when she leaves the room (or what he, Perry, learned at the party) is not *reference-fixing* by demonstration, but rather the conceptual isolation between demonstrative concepts and other concepts. What explains the significance of what Perry learns at the party seems to be the building of a connection between two mental representations (the

demonstrative concept THIS MAN and a perceptual representation of Dretske) where there was no such connection before. If Perry's analogy works, what explains the significance of what Mary learns when she leaves the room should be the building of a connection between two of her mental representations where there was no such connection before—and where the two mental representations in question are something like THIS_R FEEL and REPRESENTATIONAL PROPERTY 50, say, or RP50.

As stated, Perry's analogy is not quite right. After all, Mary, unlike Perry, does not possess, before she leaves her room, two unconnected concepts. Mary, rather, possesses only *one* of the two concepts; upon leaving her room she acquires another (the demonstrative THIS_R) *and* connects it with the first. To help bring out the fact that two things happen to Mary upon leaving her room, Nida-Rümelin (1996, 1998) asks us to consider another woman, Marianna. (This thought-experiment was mentioned in section 2.1.2.) Marianna is raised just like Mary and comes to know just as much as Mary; but when she is let out of her room, she is not allowed to step outside, rather she is ushered, first, to a room full of unlabeled colored wallpaper. In that room, Marianna can token phenomenal concepts—as when she thinks “so *this_R* is what it's like to see...some color”. But here is the point of bringing up the thought-experiment in this context. Marianna, at this point in time, is much more like Perry before he knows that this man is Fred Dretske than Mary is. Marianna possesses the concept RED; but she also possesses the concept THIS_R. But her concepts are not “attached”. When Marianna is *finally* let outside is when she finally gets to attach her phenomenal concepts THIS_R and her other “physical” concepts (like RED).

Perry is wrong, then, to claim that Mary is in just the position *he* is in at the party—and that what she learns is merely what *he* learns when he learns that this man is Fred Dretske. For a lot more is happening to Mary than is happening to him.

3.2.2. The three things that happen to Marianna

To be more precise, a total of three things happen to Marianna, because as she takes her first look at the colored room, *two* things happen to her. First, she gets to *experience color*. That is, Marianna’s brain gets to token a new property—not “new” in the sense that it is a property Marianna could not think about and know about while in her room, but new because her brain has never tokened it before. According to the physicalist, of course, that property is a *physical* property, say property ®: so Marianna’s brain gets to token a *physical* property it never tokened before (we’ll come back to this). Second, Marianna introspects that newly instantiated property and forms, as a result, the new higher-order phenomenal thought: “so *this_R* is what it feels like to see...some color or other” or “I am having *this_R feel*”. That is, she now introspects directly the functional property she could only think about theoretically before.

So here are the three things that happen to Marianna.

- (a) Her brain instantiates ®

Then she gets to think about ® demonstratively, or recognitionally by introspecting as follows:

- (b) I am having *this_R feel*

And finally she gets to connect her new phenomenal concept with other concepts, like so:

(c) Normal perceivers, when they see a ripe tomato, experience *this_Rfeel*.

The physicalist can clearly account for Marianna's learning of (c)—along just the lines that Perry sketches.³⁶ But what about Marianna's coming to think (b)? And what about (a)?

Marianna introspects

As Marianna steps out of her room and introspects, she is finally able to think that she is having *this_Rfeel*. Some physicalists (Tye 1995 and Perry 2001) have suggested that the knowledge she acquires is new *demonstrative/recognitionnal* knowledge. Inside her room, she could refer to ® using theoretical concepts such as RP50. When she thinks—after having left—that she is having *this_Rfeel*, she gets to pick out ® demonstratively or recognitionnally using the concept *this_R*. Marianna acquires a new way of thinking about ®: a demonstrative/recognitionnal way of thinking.

The anti-physicalist, however, points out that merely being able to think about someone (or something) demonstratively (or recognitionnally) does not seem to constitute a (robust) new way of thinking of that person (or that thing). Imagine that I know all there is to know about my neighbor before having ever met him. Upon first meeting him—hence, upon first being able to refer to him demonstratively or

³⁶ Stalnaker (2008), in his Locke lectures, expresses some worries about Marianna's learning of (b). After all, he claims, once Marianna has left her black-and-white room and come to see red, she has learned what seeing red feels like. She knows that (knows what seeing red feels like) before she learns (b). So what is it, Stalnaker asks, that she *doesn't* know? The answer seems rather straightforward: Marianna doesn't yet know *that she knows what seeing red feels like*. She knows that she knows what seeing some color feels like: she has no way, however, to connect that knowledge with her concept RED.

recognitionally—I get to think of him as *this man* (by pointing my finger at him), something I couldn't do before. But finally being able to think of my neighbor demonstratively (or recognitionally) does not seem to constitute a robustly different way of thinking about him. After all, by hypothesis I know *everything* about him. My new way of thinking about him can't be the result of my finding out that he has a property I didn't know he had—say that he's a serial killer. So when I finally meet him, I do not learn he has a new property of which I was ignorant. All I now get to do (which I couldn't do before) is to think “*this man is my neighbor*” or “*this man has brown hair*” instead of thinking only “John Connor is my neighbor” and “John Connor has brown hair”. But merely getting to think demonstrative (or recognitional) thoughts that I could not think before isn't cognitively very significant—it isn't really a new way of thinking about the referent of my thoughts. Levine (2007) makes just this point. On a scenario like the one just described, Levine writes, “no real new information is introduced via the demonstrative presentation” (157). He continues:

Mary doesn't seem to learn just that the state she can describe in such rich theoretical vocabulary is happening here and now; that it's *this* one. She forms a new conception, one with substantive and determinate content, of this state (157/158).

Since Marianna's coming to think of \textcircled{R} differently (when she finally gets to think (b), that she's having *this_R* feel) *is* cognitively significant for her, then it can't be merely because she started thinking of \textcircled{R} demonstratively. Something else must be going on with Marianna. But something else is indeed going on with her: her brain is finally instantiating \textcircled{R} .

Marianna's brain instantiates ®

Consider what a functional representationalist would claim happens to Marianna when she finally instantiates the representational property ®. She finally becomes experientially related to a mental representation (call it M_R) with a particular content (C_R), and (since we're considering functional representationalists) M_R plays a particular role in her psychology (say, it is poised to impact beliefs as Tye (2000) suggests). Marianna has never been experientially related to that mental representation type before. And M_R has never played its role in her psychology before.

Of course, Marianna knows all there is to know about M_R , about C_R and about the experience relation. But her knowing all this means (according to the functional representationalist) that she is belief-related to a number of mental representations *about* the experience-relation, M_R and C_R —representations whose content might be [M_R plays such and such a role and has content C_R]. Certainly Marianna can know all that without herself being experience-related to M_R and C_R .³⁷ Moreover, knowing a number of things *about* M_R and C_R isn't to be belief-related to a mental representation that is anything at all like M_R —or whose content is at all like C_R . What it does mean is that the content of many of Marianna's beliefs *involve* (are about) M_R and C_R ; but obviously the content of M_R itself isn't anything like [M_R plays this role]. The content of Marianna's experience of red presumably involves something about red. How we should spell out the content of Marianna's experience is by no means obvious (see Part II), but it is certainly not [M_R plays such and such a role].

³⁷ She supposedly does know enough to be able to put herself in that state, and if she had the equipment needed to manipulate her brain into that state, she could do it. But she's not allowed any such equipment in her room.

It follows that Marianna's instantiating \textcircled{R} should be significant for her. Not because she *learns* something new—i.e., not because she becomes *belief-related* to a substantively new, non-demonstrative representation (or concept) of \textcircled{R} . But because she is now experience-related to a new representation—new in the sense that she had never tokened it before.

3.3.3. *Two ways to account for what Mary learns*

The physicalist now has two options to account for what Mary learns when leaves her room.

Option 1: biting a bullet

The first physicalist option involves biting a bullet about Marianna's thought that (b): she is having *this_R* feel. Levine argues, remember, that Marianna cannot merely have come to think of \textcircled{R} demonstratively when she finally gets to think that (b) because merely being able to demonstrate something is not cognitively significant. But the physicalist can insist nonetheless that Marianna does merely come to think of \textcircled{R} demonstratively and accept Levine's claim that merely coming to think of something demonstratively is not cognitively significant. The physicalist, that is, can accept that Marianna's getting to think that (b) is *not* cognitively significant.

How is this a live option for the physicalist? Isn't this simply giving up on explaining what Marianna learns when she leaves the room? I think not. For the physicalist should point out what we've pointed out above, that there *is* something significant that happens to Marianna when she leaves her room: she finally gets to be

experientially related to a new representation. This allows the physicalist to claim that what is significant for Marianna is not her acquiring of new *knowledge* (i.e., it isn't her coming to be belief-related to anything substantially new). Rather, it is her standing in that experiential relation. The physicalist is not thereby denying that Marianna acquires new knowledge. After all, the physicalist thinks that she *does* come to think something she hadn't thought before when she thinks that (b)—it's just that *that* isn't the significant bit. The physicalist can go on: anti-physicalists mistakenly assume that it must be Mary's new *knowledge* that is significant—that it is something that she comes to *believe* that is significant. This is an understandable mistake, since Marianna comes to instantiate © and think about it at the same time, which makes her having the new thought *itself* seem like the significant event, even though it isn't. Still, the physicalist can claim, the instantiating of © *and* the thinking of a new thought *together* are significant—not thinking the thought alone.³⁸

Option 2: seeing-as

The second option for the anti-physicalist is to argue that Marianna can come to see © in a new way (not merely demonstratively) without our needing to posit a new property of ©. Let us consider an ordinary case of coming to think of someone in a new way: Amy goes from thinking about Art as just a friend to thinking of him as more than that. Does that entail that Amy ascribes to Art a new property, one that she didn't know Art had before? In some ways it does—maybe Amy now thinks that Art

³⁸ It may be that *neither* conjunct *alone* is significant. After all, Marianna could instantiate © without paying attention, or without being able to think about it. Certainly if that were the case the mere instantiation of © might not be very significant. Still the physicalist can deny that the thought considered alone is significant too. It really seems to be the conjunction of the two that has significance.

is attractive, or that he would make a good life partner. But is this to say that there are some things *about Art* that Amy did not know before? That there is a property of Art's (the property of *being attractive to Amy*) that Amy did not know he had? Not necessarily. Amy might come to see Art in a new way—yet it needn't follow that there must be a property *of Art* that Amy didn't know about before. Presumably, Amy's coming to think of Art differently has much to do with the way Amy herself perceives Art.

Now consider Annie who is looking at an ambiguous figure—say the famous duck-rabbit—but can only ever see it one way—say, *as a rabbit*. Assume, however, that she knows *everything* there is to know about the figure: she knows *that* it looks like a duck to some people under some circumstances even though it has never looked that way to her. It seems that when Annie finally gets to see the figure *as a duck*, she gets to think of it in a new way—*as* a figure that looks like a duck. She sees the figure *as* a duck. Yet it isn't clear that when she comes to see the picture in this new way that she learns about a new property of the figure, a property she didn't know it had. For what could this property be? Not the property of *looking like a duck*. Again, we can assume (very plausibly, I think) that Annie knows the figure has the property of *looking like a duck*. She may know everything there is to know about that property too—that it is, say, a dispositional property of the figure that can be reduced to the shape itself and its causal powers. It's just that she unfortunately has never managed to see it *as* a duck.

There would seem to be something cognitively significant about coming to see a friend differently (as in the case of Amy) or coming to see a figure differently

(as in the case of Annie and the duck-rabbit). Yet these new ways of thinking do not involve ascribing to the object of thought (or sight) a new property—a property that thinkers didn't know the objects possessed. Rather, these new ways of thinking involve new ways that thinkers *relate* to the objects in question. Of course, we can agree with the anti-physicalist that not *every* new way to relate to an object will lead to a significantly new way of thinking about that object. So we agree that being able (finally) to point at my neighbor may involve a new way of relating to him (via pointing), though not a cognitively significant new way of relating to him. However, being able (finally) to see something *as* a duck, or someone *as* a potential lover are cognitively significant new ways of relating to a figure, or a person. Yet these cognitively significant new ways of thinking needn't entail the existence of correspondingly new properties.

A physicalist could therefore insist that Mary's new way of thinking about ® upon leaving her room is very much like Amy's new way of thinking about Art, or Annie's new way of seeing the ambiguous figure. We can maintain that Mary knew, in her room, everything about ®. That includes, presumably, Mary's knowing that ® is a particular kind of phenomenal feel (which it is according to the physicalist); that people who see red undergo that particular feel (whatever it is), just as Mary herself, when she sees black, undergoes a particular feel. But still, when Mary finally leaves her room, she gets to "see" or "perceive" ® differently. She gets to introspect ® *as a feel*—she gets to think of it *as* an introspected feel. The fact that Mary has a new way of thinking of ® does not entail that she finds out that ® has a *new* property, one Mary didn't know it had before. She knew ® was a feel—but she had never

introspected it as a feel. Spelling out this type of physicalist reply would require the physicalist to say a lot more about what counts as a property and what does not. After all, an anti-physicalist might say that when Annie finally comes to see the duck-rabbit *as a rabbit*, there *is* a new property that she learns about, namely that it has the property of *looking like a rabbit to her*. We'll come back to these very issues in detail in 4.1. (The property dualist argument discussed there is in many ways connected to the anti-physicalist's point about Mary.)

3.2.4. *Where we sum up*

Here, then, are the things that happen to Marianna when she leaves her room.

- (a) Her brain instantiates ®.

She gets to think:

- (b) I am having *this_R* feel.

And she gets to think:

- (c) Normal perceivers, when they see a ripe tomato, experience *this_R* feel.

The physicalist can easily account for the significance of coming to think that (c): in coming to think (c), Marianna comes to link previously unconnected concepts. I have suggested two physicalist strategies for explaining what else goes on with Marianna. First, the physicalist can claim Marianna's coming to think that (b) alone is *not* significant—what is significant is her brain instantiating ® *together with* her coming to think that (b). It is a mistake, then, to attempt to capture what is significant about coming to think that (b) alone. Second, the physicalist can argue that what goes on when Marianna gets to think that (b) is akin to what goes on when someone finally

sees an ambiguous figure *as a duck*—which does not seem to require the positing of substantive new properties of the figure. Similarly, Marianna gets to see ⑧ differently when she thinks that (b)—and this need not require the positing of substantive new properties of ⑧.

3.3. The core contrast

Finally, there is the core contrast. Remember that what needs to be explained is the fact that psychophysical identities remain arbitrary, regardless of how much one is told in detail about phenomenal feels or about the physical properties of brains. Other ordinary, a posteriori identity claims (‘Hesperus is Phosphorus’ or ‘water is H₂O’), though they might feel arbitrary at first, do not go on to feel that way after one is told in detail about Venus or H₂O. What might explain this contrast between psychophysical identities and other kinds of identities?

Presumably, identity claims like ‘water is H₂O’ do not remain arbitrary for very long because one can learn “how the molecular structure H₂O is responsible for all the superficial properties by which we identify water” (Levine 2007, 127). Take the set of mental descriptions we associate with our concept WATER: a chemist will be able to show us how molecules of H₂O would come to act in the way that fit these descriptions (or else explain them away). Similarly, Anna becomes convinced that Hesperus is Phosphorus when she is shown how *one single* heavenly body can have (roughly) all the characteristics her mental descriptions associate with HESPERUS and with PHOSPHORUS. A physicalist might therefore attempt to explain the core contrast by contrasting the descriptions associated with ordinary concepts (like WATER and

H₂O) and the descriptions associated with phenomenal concepts (call these phenomenal descriptions).

3.3.1. Two remarks before we start

A priori vs. a posteriori descriptions

In our discussion of conceptual isolation, we pointed out that phenomenal concepts are connected with very few (if any) concepts a priori (in the strict sense). A physicalist could conclude, then, that *there are no descriptions* associated with phenomenal concepts such as THIS_R. After all, mental descriptions are made up of concepts—if no concepts are a priori connected with phenomenal concepts, then no mental description will be a priori connected with them either.

Making such a claim would presumably help explain why there is a core contrast. After all, if there are no descriptions associated with phenomenal concepts, one cannot become convinced that *this_R* feel is representational property #50 by being shown that representational property #50 is responsible for the characteristics that these (non-existent) mental descriptions associate with THIS_R.

Unfortunately, this is a non-starter. First, phenomenal concepts are not entirely conceptually isolated (especially if we focus on the looser pre-theoretical connections). Second, the descriptions to which we would appeal to in ordinary cases are far from being a priori. What makes the claim that water = H₂O lose its arbitrariness? The fact that someone can explain to a thinker how molecules of H₂O can come to act in the ways that she thinks water acts. Part of this explanation will involve explaining why it is that water boils at high temperatures—and certainly the

belief that water boils at high temperatures is not a priori. So if a physicalist wants to contrast the descriptions associated with phenomenal concepts and the descriptions associated with ordinary concepts, she can't forget that thinkers will associate many a posteriori descriptions with their phenomenal concepts (of the kind Levine was alluding to, see 2.1). For instance, thinkers who believe of some X that X is *this_R* feel will be disposed to infer that X is the feel I experience when I see red objects. Clearly, the description "feel I experience when I see red objects" is not connected with the phenomenal concept *THIS_R* a priori. Still, dealing with the core contrast adequately requires acknowledging that phenomenal concepts will have many such descriptions associated with them.

The folk and the "philosophically minded"

Levine claims that psychophysical identity claims will feel arbitrary no matter how much neuroscientists and cognitive scientists might tell us about the mind. But feel arbitrary to *whom*?

Many of the folk seem to see nothing wrong at all with psychophysical identity claims. On their first day, my philosophy of mind students seemed happy to think that the mind was the brain and that feeling in love just was having high levels of oxytocin. Not only did they not seem to think these were arbitrary identity claims, they seemed puzzled as to why someone might doubt them. So it isn't the folk's reaction to these identity claims that leads Levine to think that there is a core contrast.

Of course, most people don't think about these things very carefully (not that there is any reason why they should). What is interesting to notice is that they *can*

come to see why the identities in question might seem arbitrary; they can come to appreciate the core contrast. And that is all that matters. Why is it that when the folk give it some thought, they start thinking that these identity claims feel (and always will feel) arbitrary?

3.3.2. A *NONPHYSICAL* description

In our discussion of conceptual isolation, we pointed out that phenomenal concepts seem pre-theoretically connected to a number of concepts, including the concept *NONPHYSICAL*. In that way they might be unlike many ordinary concepts—the concept *WATER* and *HESPERUS* are not pre-theoretically connected to the concept *NONPHYSICAL*. Rather, thinkers seem to have pre-theoretical dispositions to infer from “X is water” that “X is physical stuff”—and from “X is Paris” that “X is a place”. These pre-theoretical inferential dispositions reveal that our concepts *PARIS* and *WATER* are associated with descriptions concerning the nature of Paris and water—what kind of things those are.

If this is the case, then we might think that it is no wonder that thinkers resist linking them to concepts connected with *PHYSICAL*. I also suggested that there might be pre-theoretical connections between phenomenal concepts and concepts like *PRIVATE* and *INEFFABLE*. These inferential connections—from “X is *this* feel” to “X is private” or “x is ineffable”—may very well make it even harder for the subject to build inferential connections between phenomenal concepts and the physical concepts of neurology, cognitive science, etc.

The anti-physicalist may want to ask why it is that phenomenal concepts like *THIS_F FEEL* turn out to be connected with concepts like NONPHYSICAL. Isn't the best explanation for *that* fact simply that phenomenal feels *are indeed* nonphysical? The answer here is simply no. If phenomenal concepts are linked to the concept NONPHYSICAL, this will be a fact about our naïve psychology (as Bloom claims it is), not about the ontology of the world. After all, it is widely believed that we are born with stores of "specialized knowledge"—what we might call naive theories (whether or not they truly deserve to be called theories is irrelevant here). The contents of our naïve theories is determined experimentally; it turns out, for instance, that our naïve physics seems to link CARRIED OBJECT with concepts of falling straight down—subjects assume, that is, that carried objects, when dropped, fall straight down. But no one takes the connections between those concepts to hold true *because* that is actually the way it is in the world. Classical mechanics—to say nothing of relativity theory or quantum mechanics—contradicts naïve physics many times over; naïve physics, in the end, reflects merely the way we *think* about the world. Those who press on and ask why we think of the world this way will get, at the most, an evolutionary story about how a certain set of beliefs might be useful to organisms like us.

What is true for our folk physics concepts may very well be true for the concepts involved in our naïve theory of mind, or naïve psychology. Naïve psychology, like naïve physics, reflects first and foremost the way we *think* about things, in this case about our own minds and phenomenal states. Why we think of it this way may, too, be given some evolutionary explanation. But the fact remains that it does not require an anti-physicalist ontology.

3.3.3. Worries

I now consider some worries one might have about the physicalist's account of the core contrast I just described.

Other NONPHYSICAL descriptions

The existence of a core contrast has been explained by an appeal to a connection between phenomenal concepts and the concept NONPHYSICAL. But here is what this should lead us to expect: *if* there are any nonphenomenal concepts which are connected to the concept NONPHYSICAL, then we should see similar core contrasts. Rey (1992) suggests that concepts like NATION, CLUB, ASSOCIATION, SYMPHONY, etc. might be linked with the concept NONPHYSICAL. Yet we do not have similar dualist intuitions about clubs and associations. Hence, a mere connection to the concept NONPHYSICAL can't explain why there is a core contrast.

Though it does turn out to be surprisingly hard to satisfactorily identify clubs or nations with *physical* descriptions, I doubt that the concepts CLUB and NATION are connected with the concept NONPHYSICAL in the way that phenomenal concepts are. Perhaps the fact that it is so hard to make such identifications will drive someone to connect the concept CLUB with the concept NONPHYSICAL. But physicalists are claiming that phenomenal concepts are *pre-theoretically* connected to the concept NONPHYSICAL. And it is *that fact* which is supposed to make the proposed identity claim that much harder to accept.

There do indeed seem to be nonphenomenal concepts that are connected pre-theoretically to the concept NONPHYSICAL, the concepts GHOST, SPIRIT, GOD and SOUL. And of course, if we tried to identify ghosts, god or the soul with something physical we would see the same kind of resistance that we see in the phenomenal case.³⁹

What about the folk?

The second worry is this: assume that what explains the core contrast is the fact that we pre-theoretically connect phenomenal concepts and the concept NONPHYSICAL. The folk, like the “philosophically-minded”, will pre-theoretically connect phenomenal concepts and the concept NONPHYSICAL. Why is it, then, that the folk (my students on their first day) seem perfectly happy with psychophysical identity claims?

Here is one story we might tell. Pre-theoretical connections can become less prominent and less salient. My students had read articles and studies that suggested to them that love was just high levels of oxytocin. And as a result the pre-theoretical connections between their phenomenal concepts and the concept NONPHYSICAL may have faded into the background. But this isn't to say that the connections in question are broken. That they are still there may explain why the same students who claim that the mind just is the brain don't bat an eyelash when faced with a mind-switching movie plot. Yet, since the pre-theoretical connections haven't disappeared, it is possible to carefully draw the focus back to them. And it's when these pre-theoretical connections become salient again that the core contrast can be appreciated.

³⁹ The concept NUMBER too might be connected pre-theoretically with the concept NONPHYSICAL, though it's less obvious. Of course trying to identify numbers with physical things does seem problematic too.

3.4. Concluding remark

I have, in this section, sketched how a physicalist might go about explaining the anti-physicalist data. But some anti-physicalists believe that the phenomenal concept strategy is doomed for more general reasons. It is to these anti-physicalist arguments that I now turn.

4. General anti-physicalist arguments

General anti-physicalist arguments are meant to undermine *any* physicalist account of phenomenal concepts, but not by focusing on which bits of the data phenomenal concepts cannot explain. Rather, the focus is on the features of phenomenal concepts (conceptual isolation in particular) which the anti-physicalist argues phenomenal concepts could not have within a physicalist framework. Here is the argument in broad strokes:

- 1) If (at least part of) the relevant data is to be explained, phenomenal concepts and physical concepts must be conceptually isolated.
- 2) Phenomenal concepts and physical concepts can be conceptually isolated only if phenomenal concepts refer to non-physical properties.
- 3) Therefore, if (at least part of) the relevant data is to be explained, phenomenal concepts must refer to non-physical properties.

In this section, I consider two general arguments against the physicalist phenomenal concept strategy, which build on the general schema just presented. The first I will call the *property dualist argument* (though it's really a family of arguments, see 4.1.); the second is an argument from David Chalmers (2007) (4.2.).

4.1. Property dualist arguments

Property dualist arguments attempt to make a case for premise 2)—that phenomenal concepts and physical concepts can be conceptually isolated only if phenomenal concepts refer to non-physical properties. The argument for 2) rests on two further premises. A very basic (and truncated) version of the first—call it, as has been done, the Semantic Premise—goes something like this:

In order for two concepts to pick out the same object (or property) *and* to be conceptually isolated, the object (or property) in question must have at least two properties. (see Loar 1990/97 White (1999, 2003, 2007) and what Block calls the “Max Black objection” (2007))

It follows that for the phenomenal concept THIS_R and the physical concept REPRESENTATIONAL PROPERTY 50 (RP#50) to be conceptually isolated and pick out the same physical property ®, ® in turn must have two properties. Of course, the Semantic Premise alone does *not* entail that 2) is true. The two properties of ® may very well turn out to be two *physical* properties of ® and 2) claims that phenomenal and physical concepts can be conceptually isolated only if phenomenal concepts pick out *non-physical* properties. The anti-physicalist’s case for 2) therefore requires a second premise to the effect that one of the properties of ® (in the psychophysical case) must be non-physical.

The goal of this subsection (4.1) is to argue, naturally, that the anti-physicalist argument for 2) fails—it is not true that phenomenal concepts and physical concepts can be conceptually isolated only if phenomenal concepts refer to non-physical properties.

4.1.1. The Semantic Premise

The Semantic Premise, in its most basic form, claims that in order for two concepts to co-refer and be conceptually isolated, the target object (or property) must have two properties.

Some Clarifications

Discussions of the Semantic Premise usually involve talk of modes of presentation. The Premise itself is sometimes expressed as follows: that in order for two concepts to co-refer and be conceptually isolated, they must pick out the target object (or property) via two different modes of presentation. Unfortunately, the expression ‘modes of presentation’ is often used ambiguously either to refer to a complex mental representation (or mental description) associated with a given concept, or to refer to properties of the concept’s referent. Block (2007) points out this ambiguous usage in writings by Smart (1959) and Perry (2001) and, to avoid equivocation, distinguishes between “conceptual modes of presentation” (CMoPs) and “metaphysical modes of presentation” (MMoPs). (White (2007) draws out the same distinction using different vocabulary: he talks of “representational modes of presentation” (RMPs) and “nonrepresentational modes of presentation” (NMPs)). Representational modes of presentation (or conceptual modes of presentation) are “aspects of the way we represent the world and not the world itself” (White 2007, 210)—they are associated mental descriptions of the kind we’ve already talked about earlier in Part I. Nonrepresentational modes of presentation (or metaphysical modes of presentation), on the other hand, are “features or properties of items in the world” (ibid). The

Semantic Premise, then, can be expressed more precisely as follows: in order for two concepts to co-refer and be conceptually isolated, they must be associated with different conceptual or representational modes of presentation (different complex mental representations); and, in turn, each of these different representational or conceptual modes of presentation must be associated with a metaphysical mode of presentation (a property) of the target object or property.

To further complicate matters, conceptual modes of presentation are usually thought to play a number of (different) roles—even by those who use the Semantic Premise in arguing for anti-physicalism. Block (2007, 264) notes that representational or conceptual modes of presentation are assumed to be those mental descriptions associated with concepts that 1) fix reference, 2) account for cognitive significance and 3) make metaphysical (nonrepresentational) modes of presentation accessible a priori. This explains the claim by some writers, such as White, that phenomenal concepts do not have conceptual modes of presentation. After all, phenomenal concepts are supposed to refer directly. So even if it turns out that there are some mental descriptions associated with these concepts, those descriptions, since they are not reference-fixing, do not count as conceptual modes of presentation.

Like Block, I find reason to doubt that the same mental description would have to play all three roles: fix reference, account for cognitive significance, and make MMoPs accessible a priori (see also Burge 1977 and Byrne and Pryor 2006). Here I will assume that concepts can refer directly even while we can appeal to (non-reference-fixing) descriptions to account for their cognitive significance. After all, the anti-physicalist's interest in these conceptual modes of presentation (in this discussion

of the Semantic Premise) is an interest in cognitive significance. Her goal is to spell out what is required for a thinker to have two concepts that, though they co-refer, are conceptually isolated; that is, learning that her concepts co-refer would be cognitively significant for her. This does not require that the conceptual modes of presentation (associated mental descriptions) determine reference or make metaphysical modes of presentation accessible a priori. Therefore, when I talk of conceptual modes of presentation here (more likely I'll talk of associated mental descriptions), I do not mean to claim that these descriptions determine reference. The Semantic Premise, recast now to avoid talk of modes of presentation, goes something like this:

In order for two concepts to co-refer and be conceptually isolated, the concepts must be associated with different mental descriptions; and, in turn, they must be associated with a property of the target object (or property).

Figure 1 below depicts how this is meant to work for the conceptually isolated, co-referring concepts HESPERUS and PHOSPHORUS

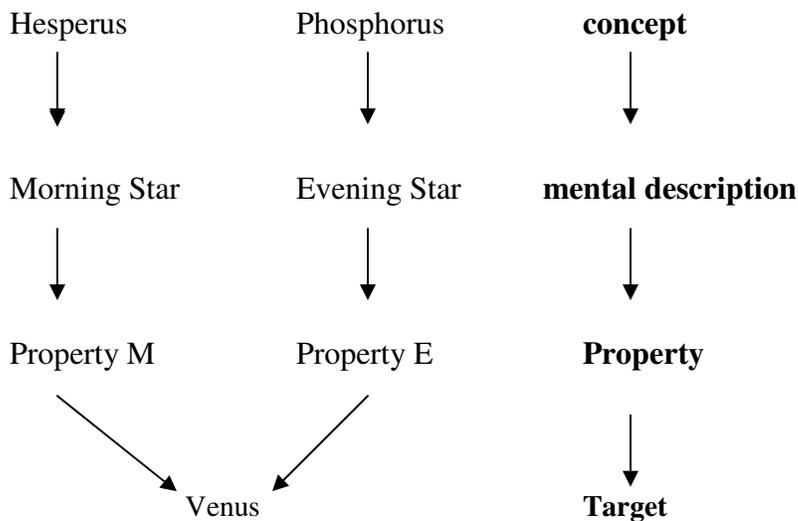


Figure I.1. Filling in the blanks (Hesperus/Phosphorus)

Why think the Semantic Premise is true?

White (2007) provides a “four-stage” argument for the Semantic Premise. The argument consists in a list of four requirements, the first two of which are the most crucial to our current inquiry. First, White claims, we must be able to

say how the world presents itself to the subject who believes incompatible or contradictory things about the same object by providing a set of possible worlds that are the way that subjects takes the actual world to be. (222)

A thinker who has two concepts for the same object (or property) may come to believe contradictory things about that object (or property) as when Anna the astronomer comes to believe that Hesperus is bright and Phosphorus is not bright—or when she thinks that Hesperus is not Phosphorus. Clearly Anna is not, in these cases, irrational, as she would be were she thinking that Hesperus is bright and Hesperus is not bright—or that Hesperus is not Hesperus. Let’s call this rational Anna, Rational Anna. When her good friend *Irrational* Anna makes an irrational claim, there is no coherent way that she takes the world to be—there are no possible worlds that are the way Irrational Anna takes the world to be. But there is a coherent way that Rational Anna takes the world to be and we need to be able to say what that world is like. So when Rational Anna thinks that Hesperus is not Phosphorus, what are the possible worlds that fit her beliefs? The answer isn’t obvious. The identity claim “Hesperus = Phosphorus”, though a posteriori, is considered to be necessarily true. There are no worlds, then, in which Hesperus is not Phosphorus. According to White, if we maintain that there are no possible worlds corresponding to Anna’s thought that

Hesperus is not Phosphorus, then Rational Anna is really no different from Irrational Anna.

White goes on to claim that the only way to distinguish Rational Anna from Irrational Anna is to posit two different mental descriptions of the referent by Rational Anna and two corresponding properties of Venus. The possible worlds corresponding to Rational Anna's thoughts [beliefs?] are those worlds in which there are two stars; one which has whatever property Anna associates with Hesperus and another which has whatever property Anna associates with Phosphorus. This is, in effect, White's second "requirement":

We must satisfy the first requirement by providing two distinct properties of the object in question [Venus] which correspond to the subject's [concepts] and which are such that there is a possible world at which they are instantiated by different objects. (222)

Note that the possible world corresponding to Anna's thought that Hesperus is not Phosphorus is not a world in which Anna's thought is *true*. Since we assume that it is *necessarily* true that Hesperus is Phosphorus, the world corresponding to Anna's thought cannot be a world in which *Hesperus is not Phosphorus* (there is no such world)—i.e. it cannot be a world in which her belief that Hesperus is not Phosphorus is true. Rather, the world corresponding to her belief is a world in which there are two heavenly bodies, one that is the brightest in the evening, one that is the brightest in the morning (and neither of which need be Hesperus/Phosphorus/Venus). Finally, notice that the mental descriptions associated with a thinker's concepts are assumed to be accessible to the subject—even if they needn't be in the forefront of her mind. Asked what she thinks of Hesperus, Anna should be able to say something about the descriptions she associates with it.

The Semantic Premise—in full

The Semantic Premise I have been discussing so far is actually a truncated version of the ones Loar and White target. Here is something like Loar's version:

In order for two concepts to pick out the same object (or property) and to be conceptually isolated, the object (or property) in question must have at least two properties; *and at least one of these properties must be a contingent property of the object (or property) picked out.*

What motivates this last clause is presumably the same thing that motivates the more basic version of the Premise—a need to distinguish Rational Anna from Irrational Anna. If there must be possible worlds corresponding to Rational Anna's thought that Hesperus is not Phosphorus, those worlds must be worlds in which there are two stars, one of which has the property of being the brightest heavenly body in the morning (property M) but not the property of being the brightest heavenly body in the evening (property E), and one of which has property E but not property M. What allows for such a possible world—one in which one star has property M but not E, and another star has property E but not M—is the fact that at least one of these properties is a contingent property of Venus (in this case both). If instead we consider Simple Anna who associates with HESPERUS the description “thing that's Hesperus” and with PHOSPHORUS the description “thing that's Phosphorus”, then we might have a case in which the two corresponding properties (*being Hesperus* and *being Phosphorus*) actually are necessary properties of Venus (remember that Hesperus = Phosphorus = Venus *necessarily*). And there would be no world corresponding to Simple Anna's thought that Hesperus is not Phosphorus: there would be no world in which there are *two* heavenly bodies, one of which has the property of *being*

Hesperus but not the property of *being Phosphorus*, and one of which has the property *being Phosphorus* but not the property of *being Hesperus*. Indeed, since *Hesperus is Phosphorus* the property *being Hesperus* just is the property *being Phosphorus*. And therefore, no heavenly body can have the property of *being Hesperus* without also having the property of *being Phosphorus*.

Again, take chemically naïve Abbie who believes that water is not H_2O . What world corresponds to her thought? Well, she must have two mental descriptions corresponding to each of her concepts, and there must be two properties of the target “stuff” corresponding to each of these descriptions. Take these to be the property of filling lakes and oceans (property L for lakes) and the property of being made up of two atoms of hydrogen and one of oxygen (property of *being H_2O*). Now it may be that the property of being H_2O is a necessary, essential property of the target stuff. But as long as property L is a contingent property of that stuff—as it surely seems to be—there will be a set of worlds corresponding to Rational Abbie’s thought. That will be the set of worlds in which some stuff is H_2O and some *other* stuff has property L. Indeed, if property L were a necessary property of the stuff—like the property of being H_2O —there would be no worlds corresponding to Rational Abbie’s thought, for there would be no world in which some stuff is H_2O and some other stuff has property L. Whatever stuff was H_2O , would necessarily have property L as well.

As (roughly) stated by Loar, the Semantic Premise is subject to counterexamples (see Block (2007)). For instance, pretend for the sake of example that Abel is an only child. Give a thinker two concepts of Abel ($ABEL_1$ and $ABEL_2$), and two different mental descriptions of him each corresponding to the following two

properties of Abel: first, that of *being the person who originated from a particular sperm cell of Adam*, and second, that of *being the person who originated from a particular egg cell of Eve* (see Block 2007 and White 2007 for discussion). Now, both properties would seem to be necessary properties of Abel. Abel is *necessarily* the person who originated from that sperm cell of Adam and that egg cell of Eve. If the Semantic Premise as stated is true, then we should expect there to be no possible world corresponding to the subject's belief that Abel₁ is not Abel₂. (The justification for accepting the Semantic Premise, remember, is to ensure that there is a set of worlds that corresponds to every rational thought.) However, as Block points out, there *is* a possible world corresponding to the thought that Abel₁ is not Abel₂. It is a world in which there are two people, one of which originated from a particular sperm cell of Adam *but not from the particular egg cell of Eve*, and the other of which originated from a particular egg cell of Eve *but not from the particular sperm cell of Adam*. Since there is a possible world corresponding to the rational thought in this case, there is no reason to claim—as the Semantic Premise does—that at least one property picked out by one concept must be contingent in order for us to distinguish between rational thinkers and irrational ones.

As it turns out, there is a legitimate worry about necessary properties in the vicinity (see White 2007). There can be a world in which there are two people, one of whom originated from a particular sperm cell of Adam (but not from the particular egg cell of Eve), and another which originated the other way around, only because the two properties in question are only *contingently co-instantiated*. Indeed the mere fact that the two properties in question are necessary properties of Abel should not worry

us; what matters is that they aren't necessarily co-occurring. For if they were, then, *necessarily*, anyone who originated from that sperm cell of Adam had to originate from that sperm cell of Eve, and there would then be no world corresponding to someone's thought that $Abel_1$ is not $Abel_2$. Thus, there would be no way to distinguish that thinker from the irrational one who thinks that $Abel_1$ is not $Abel_1$. So, what is crucial here is the fact that the two properties of the target object are only contingently co-instantiated. We can coherently describe the world Abbie thinks about when she denies that water is H_2O because the property of being in lakes and oceans (property L) is only contingently co-instantiated with the property of *being H_2O* . As a result, there are worlds in which these two properties can come apart and each be instantiated by a different object—worlds which actually are the way Abbie thinks the actual world is. Similarly for Anna the astronomer in our standard example. She can think that Hesperus is not Phosphorus only because the property of rising here in the evening (property E) and setting here in the morning (property M) are contingently co-instantiated in Venus; there are worlds where they can come apart and two different objects can instantiate one of these properties (but not the other). Of course, the complete version of the Semantic Premise did sound plausible as first stated (in the Simple-Minded Anna). But note that in that example, the two corresponding properties (*being Hesperus* and *being Phosphorus*) are not merely necessary properties of Venus—they are, indeed, necessarily co-instantiated as well.

The complete version of the Semantic Premise, then, looks like this:

in order for two concepts to pick out the same object (or property) *and* to be conceptually isolated, the object (or property) in question must have at least two properties *and those must be only contingently co-instantiated*.

4.1.2. From the Semantic Premise to anti-physicalism

Assuming, as the anti-physicalist does, that the Semantic Premise is true, we can make the following claim: the phenomenal concept THIS_R and the physical concept REPRESENTATIONAL PROPERTY 50 (RP#50) co-refer (they both pick out \textcircled{R}) and are conceptually isolated (in some weak sense). Therefore, thinkers must have a) two different descriptions of the referent, and b) the referent \textcircled{R} in turn must have two corresponding properties which are only contingently co-instantiated.

The first step of the anti-physicalist argument involves filling in the blanks, i.e., characterizing the two relevant descriptions and corresponding properties in the case of the phenomenal concept-physical concept pair. The description associated with RP#50 is rather easy to come by; presumably, it's something like "property of having content such-and-such" ("property of having C" for short), its corresponding property being *being the property of having content C*. But what description ought we associate with the phenomenal concept THIS_R ? Since it is assumed that the description in question (the conceptual mode of presentation) should fix the reference [referent?] of the phenomenal concept, we might conclude (see White) that there is no description associated with phenomenal concepts—or, at the very least, no "non-phenomenal" description, i.e., no description that doesn't take as a constituent the very phenomenal concept with which it's associated. White writes in his 1986: "there is no physicalistic description that one could plausibly suppose is coreferential a priori with an expression like "Smith's pain at t" (353). As he discusses the phenomenal concept picking out pain feels, White, in his 2007, claims that "the only [associated] description could be something like 'the state of mine that is hurtful'"

(226). From which it seems to follow that the only description a thinker could associate with $THIS_R$ in our example would have to be something like “the state of mine that feels like $this_R$ ”. And if that’s right, the corresponding property of \textcircled{R} should be something like the property of *being $this_R$ feel* (see figure 2 on the next page).

As I’ve already mentioned, this instantiation of the Semantic Premise *alone* won’t yield the conclusion that a property of \textcircled{R} is *nonphysical*—one could insist, after all, that the property of *being $this_R$ feel* is a physical property. The anti-physicalist seems to have at least two arguments (a regress argument and one based on contingent co-instantiation) for thinking that that’s not a possibility—that *being $this_R$ feel* cannot be a physical property. I present each in turn.

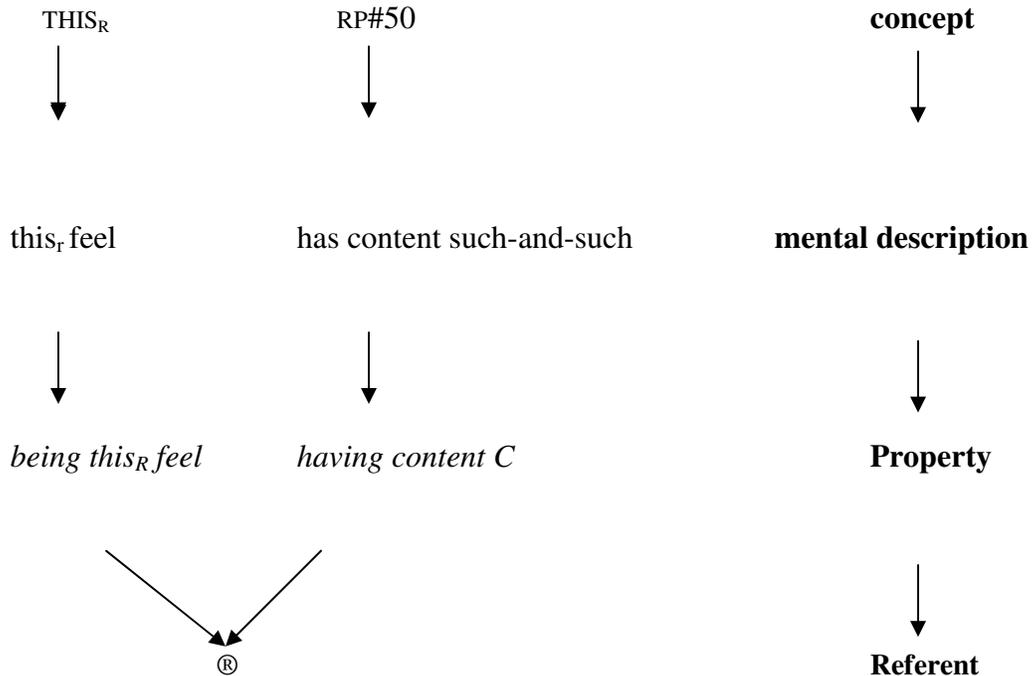


Figure I.2. Filling in the Blanks (White)

The Regress Argument

All that is strictly entailed by the Semantic Premise (it seems) is that *being this_R feel* is not identical to the physical property *having content C*, since the Premise claims that where there are two concepts, there must be two *distinct* mental descriptions and two *distinct* properties. So a physicalist might want to claim that *being this_R feel* is identical to some physical property other than *having C*. But, that won't do, the anti-physicalist argues. Assume that *being this_R feel* is identical to *some (any) physical property P*—*being this_R feel = being P*. Then it should be possible for us to pick out that property in two ways, using two distinct concepts. First by using the concept PROPERTY OF BEING P. Second by using a concept like PROPERTY OF BEING THIS_R FEEL. Since *being this_R feel = being property P*, the two concepts above co-refer. But, if that's right and the Semantic Premise is true, then the two *distinct* concepts must be associated with two *distinct* descriptions and the descriptions must be associated with two *distinct* properties of the referent. Following White's example in associating a description with the concept PROPERTY OF BEING THIS_R FEEL, we get the following: "the property of being *this_R feel*". The corresponding property is, rather straightforwardly, the property of *being the property of being this_R feel* (starred in figure 3 on the next page).

What are we to say, now, of the status of this new property *being the property of being this_R feel*? The physicalist might insist that it is identical to a physical property P', but the same reasoning will then apply again. The anti-physicalist will insist that there could be two distinct concepts picking out the property, and hence

Contingent co-instantiation

The regress argument turns merely on the short version of the Semantic Premise, making no reference to the additional clause according to which the two properties of the referent must be only contingently co-instantiated. The argument which turns on this clause takes us back to White's first requirement—that we provide a possible world corresponding to rational thoughts. Adi can rationally think that *this_R feel* is not representational property #50. The world corresponding to her thought would be a world in which there are two things, one of which has the property of *being this_R feel* (but not of *having C*) and the other one of which has the property of *having C* (but not of *being this_R feel*). Yet, if physicalism is true, the properties corresponding to Adi's mental descriptions (*being this_R feel* and *having C*) are necessarily co-instantiated. After all, what physicalism claims is that *this_R feel* just *is* having content C. In whatever world we find one of these properties instantiated (in whatever world some state has the property of *being this_R feel*), we will find the other instantiated as well (the state will have the property of *having content C*). Just as the properties of *being Hesperus* and *being Phosphorus* are necessarily co-instantiated, so are the properties of *being this_R feel* and *having C*. It follows that if physicalism is true, no possible world will correspond to Adi's thought that *this_R feel* is not representational property #50. Being able to describe the world corresponding to Adi's thought would require the property of *being this_R feel* to be a property of ® that is only contingently co-instantiated with *having C*—or only contingently co-instantiated with any other physical property. But according to physicalism, the property of *being this_R feel* is necessarily co-instantiated with some physical property. It follows that being able to

describe the world corresponding to Adi's rational thought requires that the property *being this_Rfeel* not be a physical property.

I will now argue that the anti-physicalist arguments fail. The arguments fail, I think, for at least two independent reasons: first, because we have good reasons for thinking that the Semantic Premise is false. Second, because even if we grant that the Semantic Premise is true, the anti-physicalist conclusion does not follow. In 4.1.3 I focus on the Semantic Premise—why we might think it is false, and how its falsity might affect the anti-physicalist arguments. In the next section (4.1.4), I argue that even if we assume that the Semantic Premise is true, we still have reason to think that the anti-physicalist arguments do not go through.

4.1.3. Back to the Semantic Premise

According to the Semantic Premise, there are two important links between the thinker and what she picks out in thought. First, there are links between her concepts and mental descriptions (or conceptual modes of representation); and second, there are links between mental descriptions and properties of the referent. Moreover, the Premise is motivated in large part by the need to distinguish rational and irrational beliefs, which (White claims) requires that there be a possible world corresponding to the world as Anna sees it. There are, therefore, (at least) three ways to attack the Semantic Premise. First we could deny that distinguishing rational and irrational beliefs requires that there be possible worlds corresponding to rational beliefs. Second, we could deny that there must be two different mental descriptions

associated with two distinct concepts. Third, we could deny that there must be two properties of the referent corresponding to each mental description.

Block argues that there is no need to appeal to *possible* worlds to explain rational thought. We can get by with an appeal to *epistemically possible situations*—where a situation is “something that may or may not be possible” (2007, 269). Though I agree with Block, I will grant, for the sake of argument, White’s first requirement. I start off below with a few comments about the second claim—that where there are two concepts there must be two different mental descriptions—though I ultimately grant the existence of such a link as well. I then move on to show that we have serious reason to doubt the existence of the second link (between mental descriptions and properties).

Concepts and mental descriptions

It may seem as though there are cases in which a thinker possesses two distinct concepts even while she associates with each concept the same mental description. (Associating one mental description with one concept and none with the other does entail associating *different* mental descriptions with each.) Block (2007) mentions two such examples, one involving a proper name (Paderewski) and one, from Loar (1988), involving a general term (the French ‘chat’ for ‘cat’). Take a subject, then, who comes to falsely believe that there were two Paderewskis living at the beginning of the 20th century; one a Polish composer, one a Polish politician. The subject goes on, over the years, to forget what she used to think distinguished the two Paderewskis and “remembers only that both were famous Polish figures at the turn of the twentieth

century” (266). It may seem, in such a case, that though our thinker has (plausibly enough) two concepts (PADEREWSKI₁ and PADEREWSKI₂), she does not have two different mental descriptions associated with each. Now consider an English speaker, Allie, who learns from a monolingual French teacher the term ‘chat’—“and then is taught the term ‘chat’ again by the same forgetful teacher exhibiting the same cats” the next day (ibid). Allie goes on to assume that there are two different kinds of creatures (chat₁ and chat₂) which are different in a way *she* can’t discern. Over the years, she forgets the context of acquisition of her concepts CHAT₁ and CHAT₂ (forgetting, for instance, that one was taught to her on Monday, the other on Tuesday). Allie is then such that, though she plausibly possesses two concepts (CHAT₁ and CHAT₂), she associates the same description with each concept.

Now, anyone who wants to conclude, on the basis of these examples, that we here have two concepts but only *one* associated mental description must say something about what, if not these descriptions, makes the two concepts in question *distinct*. Block claims that thinkers, in these cases, have two different “mental files”, however that is to be cashed out—each file specifying that there are two things involved (two Paderewskis, or kinds of creatures). Of course, it is not obvious (nor does Block think it is) what makes these files distinct. We can explain why a thinker came to create two distinct files in these cases, but presumably *that* explanation will appeal to the fact that thinkers had a different description associated with each at the time of acquisition. Of course, the different descriptions can fade with time (presumably that is what happens when Allie forgets), but once they’ve faded

completely, we may want to know why we should think that there are *still* two distinct files—and not that the two files themselves have faded into one.

There may be very good ways of individuating mental files without appealing to mental descriptions. Block suggests that the difference may be a “semantic difference”, or that we could suppose that “there is a need for something more than semantics—something cognitive but non-semantic” in individuating the files (2007, 266). I do not mean to suggest, here, that there is no plausible way of spelling out the difference between files. What I will do, however, is grant the proponent of the Semantic Premise a need for different mental descriptions. The Semantic Premise fails for more important reasons, which we’ll take a look at next.

From descriptions to properties

Granting that there is some kind of difference in the mental descriptions associated with the relevant concepts, must each of these descriptions be associated with a property of the referent? The claim is plausible enough in familiar cases. Anna the astronomer possesses two concepts—HESPERUS and PHOSPHORUS—and associates a different description with each (“the star that sets here in the morning”, “the star that rises there in the evening”). And Venus indeed seems to have two properties—corresponding to each one of Anna’s mental descriptions—i.e., the property of *being brightest heavenly body in the morning* (property M) and that of *being brightest heavenly body in the evening* (property E).

Though plausible in this case (and some others like water/H₂O), the claim seems very hard to accept in other instances. Take the ‘chat’ example just discussed,

and consider Remembering Allie who, unlike Allie, remembers the context of acquisition of her concepts. She associates with each of her CHAT concepts the following mental descriptions: “the kind of creature I learned about on Monday from Mme Pignon” and “the kind of creature I learned about on Tuesday from Mme Pignon”. Are we to conclude that cats (the creatures themselves) actually have two corresponding properties—the property of *being the kind of creature that Remembering Allie learned about on Monday from Mme Pignon* and that of *being the kind of creature that Remembering learned about on Monday from Mme Pignon*? Or consider Romantic Anna the astronomer, who learns about Hesperus in a poem, according to which it is the star that Hera gave to Zeus as a wedding present—and who comes to believe that this is true. The description she associates with Hesperus is just that: “the star that Hera gave to Zeus as a wedding present”. Does it follow that Venus itself (the planet) must have two properties including that of *being the star that Hera gave to Zeus as a wedding present*? Or, in a similar vein, imagine that, interested in other celestial bodies, Anna comes to possess two concepts for picking out the sun—one of which (SUN) she associates with the description “star that revolves around the Earth”. Should we conclude that, despite what our scientists tell us, the sun actually has the property of *being the star that revolves around the Earth*? The case (suggested to me in conversation by Peter Carruthers) can be modified to draw from the Semantic Premise an even more implausible conclusion, namely that we should sometimes claim that one object can have *incompatible* properties. Imagine that Anna’s second concept for picking out the sun (SOL) is “the motionless star at the center of the solar system”. (Anna the astronomer is really not a very good

astronomer—she has two false beliefs about the sun). According to the Semantic Premise, to explain the rationality of Anna’s belief that Sol is not the sun, the sun itself must have at least two properties, that of *being the star that revolves around the Earth* and that of *being the motionless star at the center of the solar system*. But these properties are incompatible: one thing cannot both revolve around the Earth and not move.

The Semantic Premise, then, has plenty of counterexamples. Holding onto it (as White does) requires biting a substantial bullet, and this is in fact what White seems to want to do. He considers a case in which advanced astrophysicists determine that the two properties of Venus, which (Regular, not Romantic) Anna thought to be properties of Venus (the property of *being the brightest heavenly body in the morning* (property M) and that of *brightest heavenly body in the evening* (property E))

are actually both explained by a single underlying property of Venus’s trajectory—say the property of being *T*. And imagine that being *T* has far greater explanatory power than any of the commonsense properties of Venus to which we currently appeal. Suppose finally, that on the grounds that properties must pull their weight in a causal-explanatory scheme, it is concluded that there is only one property of Venus—the property of being *T*—by virtue of which each of the two [mental descriptions] pick[...] it out. (219)

Again, appealing to his first requirement—according to which there must be a possible world corresponding to Anna’s thought that Hesperus is not Phosphorus—White concludes that “we are committed to the existence of properties that do not pull their weight in a causal-explanatory scheme” (ibid). Presumably, these are properties like M and E in the example he considers, and the property of *being the star that Hera gave to Zeus as a wedding present* or of *being the star that revolves around the Earth* in the examples I presented. The properties in question turn out to be rather

weak properties; properties that do not earn their explanatory keep (to use one of Georges Rey's lovely phrases). They play no explanatory role, that is, in the best theory of the referents involved (say of Venus or of cats or of the sun) despite the fact they are supposedly properties *of these referents* (of Venus, or cats, or the sun). They seem to be "fine-grained quasi-linguistic-cognitive" properties (Block 2007, 266).

It is true, of course, that according to White these properties play *some* explanatory role—they explain how subjects can be rational when they think certain things, by providing possible worlds corresponding to rational thoughts. I now argue that White's fine-grained quasi-linguistic-cognitive properties do not even play the role of providing a possible world corresponding to rational beliefs. For we need not claim that Venus, in the actual world, has properties M and E in order to meet White's first requirement (namely provide a possible world corresponding to Rational Anna's thought). The world corresponding to her thought is indeed a possible world in which there are two objects (neither of which need be Venus)—one of which has property M, the other of which has property E. But I do not see why it should follow that Venus, *in the actual world*, must have these properties. Similarly, there is a possible world corresponding to Romantic Anna's thought that Hesperus is not Phosphorus: it's a world in which there are two objects (neither of which need be Venus), one of which has the property of being a star that Hera gave to Zeus as a wedding present, the other of which has property E. But I do not see why it should follow that Venus, *in the actual world*, must have these properties. And there is a possible world corresponding to Remembering Allie's thought that chats₁ are not chats₂: it is the world in which there are two different kinds of creatures and Mme

Pignon taught Allie about the first kind of creature on Monday and about the second kind of creature on Tuesday. I do not see why it follows that cats, in the actual world, must have the properties in question. We can say the same of Anna's sun thought: the world corresponding to her thought that the Sun is not Sol is a world in which there are two stars—one of which revolves around the Earth, the other of which is the motionless star at the center of the Solar System. But the sun, *in the actual world*, need not have these properties.

Ultimately, what enables us to provide a possible world corresponding to the rational thoughts in question is not the fact that *actual* Venus has *actual* properties in the *actual* world. What enables us to provide this possible world is rather an appeal to properties that the relevant thinkers *believe* are properties of Venus in the actual world—properties which thinkers believe are properties of Venus given their mental descriptions of Venus, whether or not they actually are. The possible worlds corresponding to their thoughts are then worlds in which there are two objects, each having the properties *thinkers associated (given their mental descriptions)* with Venus, or of cats, or of the sun. But I see no reason to conclude that Venus, or cats, or the sun, *in the actual world*, must have the properties thinkers believe them to have. Even if we agree with White that there must be a possible world corresponding to rational thoughts, then, there is no reason to posit properties of actual world referents corresponding to each mental description a thinker might have of them.

To conclude: the Semantic Premise is not in good shape. Even if we grant (as I have, for the sake of argument) that two concepts must be associated with two different mental descriptions, we have reason to deny that two properties of the actual

referent must correspond to each of these descriptions. There are a number of cases in which this claim leads to highly implausible conclusions. Biting the bullet requires positing fine-grained, quasi-linguistic-cognitive properties which do not earn their explanatory keep.

Back to the anti-physicalist arguments

If the Semantic Premise is false and we need not posit two properties of the referent \textcircled{R} in the psychophysical case, then presumably we need not worry about the possible regress which would arise if the Premise were true. If we have no reason to believe that \textcircled{R} actually *has* two properties (one of them *being this_R feel*), then the anti-physicalist can't go on to argue that the (non-existent) property in question cannot be physical. The first argument for anti-physicalism therefore fails.

The argument from contingent co-instantiation, however, may still have bite even if the Semantic Premise is false. I claimed that we need not posit actual properties of actual referents—rather, all that matters is that thinkers *believe* the referent to have some property (corresponding to their mental description) not that the referent actually have it. But providing a possible world corresponding to a rational thought seems to require that these associated properties (the properties thinkers *believe* are properties of the referents whether or not they actually are) be only contingently co-instantiated. The world corresponding to Romantic Anna's rational thought is a world in which the properties she associates with Venus given her mental description (that of *being the star that Hera gave to Zeus as a wedding present* and of *being the brightest heavenly body in the evening*) can be instantiated by different

objects. But how can we provide a world corresponding to Adi's thought that *this_R* feel is not representational property #50 if the properties Adi associates with ® (in her mental descriptions) are necessarily co-instantiated, as they seem to be? This brings us to our second argument against the anti-physicalist.

4.1.4. Where the anti-physicalist goes wrong: the *a priori* requirement

I have just argued that the Semantic Premise might be false. Though its falsity would disarm the Regress Argument, it would leave a version of the second argument (from contingent co-instantiation) untouched. I now argue that even if we assume that the Semantic Premise is true, *both* anti-physicalist arguments fail.

As I noted earlier, the first step in the anti-physicalist's argument requires "filling in the blanks", namely providing a mental description to associate with the phenomenal concept *THIS_R* (and providing thereby a corresponding properties). In most of our toy examples (including all the ones provided in the section just above), the associated descriptions are just stipulated. But it doesn't seem as though we can simply stipulate what mental description thinkers associate with their phenomenal concepts. At least White doesn't think we can: thinkers, he claims, associate only one description with any one of their phenomenal concepts—and that description is itself phenomenal, i.e. it includes the very phenomenal concept it is associated with (like "the state of mine that is *painful*" or "the state of mine that feels like *this_R*"). But why think that this is right? Remember that White reaches this conclusion because he believes that: "there is no physicalistic description that one could plausibly suppose is coreferential *a priori* with an expression like "Smith's pain at t" (1986, 353 italics

mine). White seems to assume here the relevant mental description can only be one that thinkers associate with their concepts *a priori*. We'll call this the *a priori requirement*.

Now White might seem to have a point: phenomenal concepts are not connected a priori (in the strict sense we discussed in 2.1) with many other concepts. Assuming that *a priori* descriptions are those descriptions whose constituents are concepts a priori connected to phenomenal concepts, it will follow that phenomenal concepts will be connected a priori with rather few descriptions. More specifically, since physicalists themselves grant that phenomenal concepts will *not* be a priori connected to “physicalistic” concepts—like BRAIN, REPRESENTATION, or even PHYSICAL—White concludes that no description involving these concepts can be associated with phenomenal concepts *a priori*, leaving only *phenomenal* descriptions to do the job. Since we're now assuming the Semantic Premise is true, there will be an odd (phenomenal looking) property corresponding to the phenomenal description. Below is a reminder of how an anti-physicalist like White would characterize the relevant description and property, which he claims ultimately leads to the conclusion that physicalism is false, either via a regress, or because, were it true, we wouldn't be able to explain why Adi is rational to think that *this_R* is not rp#50.

There are (at least) two ways to spell out White's *a priori requirement*. After all, in our discussion of a priori connections (2.1), we distinguished between *strict* a priori connections and pre-theoretical a priori connections. White can either mean then that 1) descriptions associated with *THIS_R* when filling in the blanks can only be composed of concepts *strictly* a priori connected with it. Or he can mean that 2)

descriptions associated with $THIS_R$ when filling in the blanks can only be composed of concepts *pre-theoretically* connected to $THIS_R$. Neither option will get White what he wants. Choosing option 1) might indeed yield the conclusion that only phenomenal descriptions are associated with phenomenal concepts, but choosing it will leave White unable to explain the difference between Rational and Irrational Anna. Choosing option 2) simply won't yield the conclusion that only phenomenal descriptions are associated with phenomenal concepts.

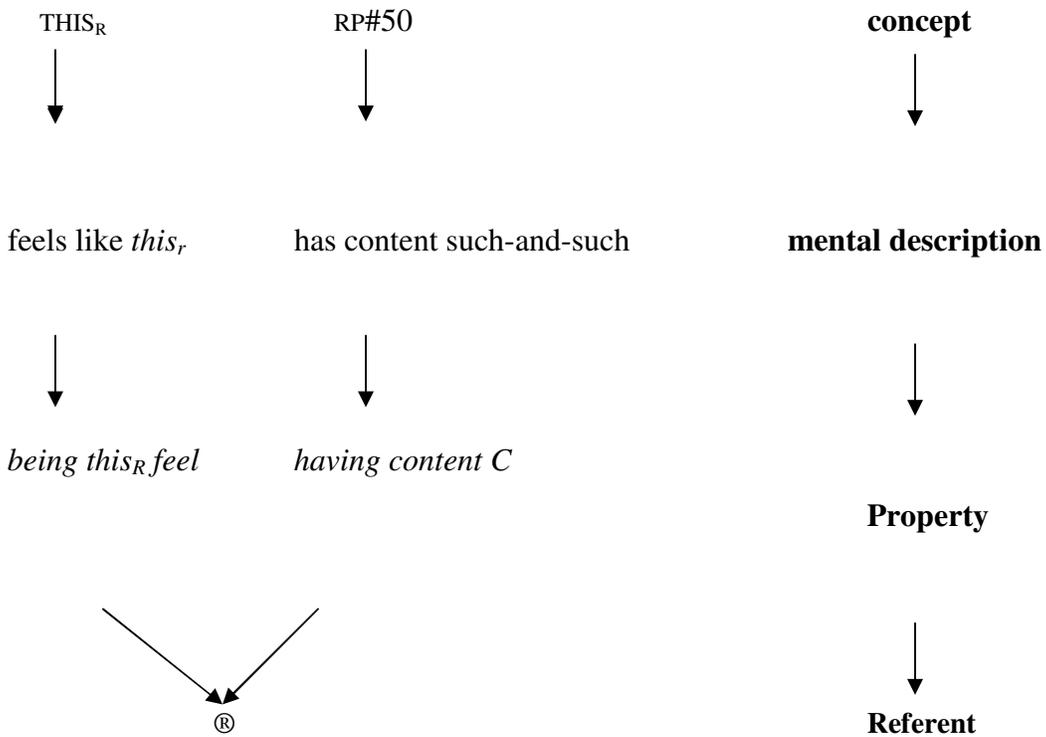


Figure I.4. Filling in the blanks (White reminder)

Strictly a priori

Let us assume, then, that the only descriptions we can associate with the phenomenal concept $THIS_R$ are descriptions composed of concepts strictly a priori connected with

it. A concept C is *not* strictly a priori connected with THIS_R, if it is possible for a thinker to possess the concept THIS_R and not possess C—in other words a thinker can believe of some X that “X is *this_R* feel” without being disposed to infer that “X is ...C...” And indeed most concepts will *not* be strictly a priori connected with THIS_R, not the concept BRAIN, or even the concept RED. Let us grant White, that if *this* is how we should spell out the a priori requirement then only phenomenal descriptions will be associated with phenomenal concepts.

Of course this a priori requirement is a general requirement. It demands that the descriptions we associate with HESPERUS or PHOSPHORUS or WATER also be descriptions that are connected with them *strictly a priori*. But it is far from obvious that the descriptions usually associated (by White himself) with these concepts are connected with them *strictly a priori*. Does the description “brightest heavenly body visible in the evening” seem associated with HESPERUS a priori in this strict sense? Couldn’t someone think of some X that is Hesperus without being disposed to infer that X is the brightest heavenly body visible in the evening? It certainly seems so. Someone might associate with it only the description “Venus in the evening according to the Greeks”; or, like Romantic Anna, someone might associate with it the description “star that Hera gave to Zeus as a wedding present.” So: does it seem impossible for a thinker to possess the concept HESPERUS without possessing the concept BRIGHTEST? It does not. Does it seem impossible for a thinker to possess the concept HESPERUS without possessing the concept EVENING? It does not. The one concept (in that “standard” description) that is most plausibly connected with HESPERUS a priori is the concept HEAVENLY BODY.

We can make the same point about PHOSPHORUS—with a similar conclusion. Which leaves the two a priori descriptions for HESPERUS and PHOSPHORUS looking something like: “heavenly body that’s Hesperus” and “heavenly body that’s Phosphorus”. (These are basically the description that Simple Anna with these two concepts). The point generalizes to other cases. Should we agree with White when he says:

Consider...the property of being the natural kind that falls as rain, fills the lakes and oceans, and glows from faucets *here* (or at the actual world). It seems clear that the connection between this property and “water” is a priori for normal subjects.

White can’t mean, here, that the connection between these concepts is a priori in the strict sense. Do we really want to say that a thinker cannot possess the concept WATER without possessing the concepts NATURAL KIND, RAIN, LAKES, OCEANS, FAUCETS, FALLING, FILLING, FLOWING? I would think not. It might turn out then that the only strictly a priori description associated with water a priori is something like “the stuff that’s water”.⁴¹

Why should this worry White? Because he believes that it is by appealing to these a priori descriptions that we get to explain the difference between rational and irrational thinkers. What explains why it is rational for Anna to think that Hesperus is not Phosphorus is the fact that, corresponding to her two associated descriptions, are two contingently co-instantiated properties. In the standard case we say that what explains why she is rational is the fact that there can be (in a possible world) a heavenly body that is the brightest in the evening (but not in the morning) and another heavenly body that is the brightest in the morning (but not in the evening). But as I

⁴¹ See Byrne and Pryor, p1.

just pointed out these descriptions do not seem to be *a priori* connected with the concepts HESPERUS and PHOSPHORUS—at least in the strict sense. And unfortunately, the descriptions that *are* a priori connected with HESPERUS and PHOSPHORUS won't be able to explain why Anna is ratiōn.

Indeed, the two plausible a priori descriptions are the following: “heavenly body that's Hesperus” and “heavenly body that's Phosphorus”. And they correspond to two *necessarily co-instantiated* properties. Since Hesperus necessarily is Phosphorus, the heavenly body that's Hesperus is necessarily the heavenly body that's Phosphorus. From which it follows that there is no world in which there are *two* things—one of which is the heavenly body that's Hesperus *but not* the heavenly body that's Phosphorus, the other one of which is the heavenly body that's Phosphorus *but not* the heavenly body that's Hesperus. So there is no world corresponding to Rational Anna thought that Hesperus is not Phosphorus after all. And if *that's* right, then the a priori requirement doesn't merely entail that i) were physicalism true, it would be irrational for Adi to think that *this_R* is not *rp#50* (from which White concludes that physicalism is false). The a priori requirement also entails that that ii) if HESPERUS and PHOSPHORUS co-refer, it would be irrational for Anna to think that Hesperus is not Phosphorus

The point, again, is this: White should not spell out his a priori requirement so strictly, because *even in regular, toy examples* strictly a priori descriptions will not be able to explain rational thinking (by providing a possible world corresponding to the rational thought). The *a priori* description associated with phenomenal concepts

can't explain why Adi's thought is rational; but neither can the a priori description associated with HESPERUS explain why Anna's thought is rational.

Pre-theoretical connections

White might think it is plausible that the *only* a priori description associated with the concept THIS_R is the description “state of mine that feels like *this_R*”. But what about *pre-theoretical* connections like the following descriptions: “the state of mine I'm thinking about now₁”, or “the state of mine I'm introspecting now₂”. Plausibly enough, a thinker who thinks of some X that “X is *this_R* feel” will be disposed to infer that “X is the state of hers that she's thinking about now”. And these descriptions will allow the physicalist to provide a world corresponding to Adi's thought that *this_R* is not rp#50—and to deal with the regress argument (see figure 5, next page).

How does this help us deal with Adi's thought that *this_R* is not representational property #50? Well, it now seems that the two relevant properties—*being the kind of thing introspected by Adi now* and *having content C*—are only contingently co-instantiated. The world corresponding to Adi when she thinks that *this_R* feel is not representational property #50 is a world in which there are two properties—one of which is the kind of thing introspected by Adi now (but which does not have content C), the other one of which has content C (but isn't the kind of thing introspected by Adi now). Even assuming that the Semantic Premise is true, this way of filling in the blanks enables us to provide a possible world corresponding to Anna's rational thought without giving up on physicalism

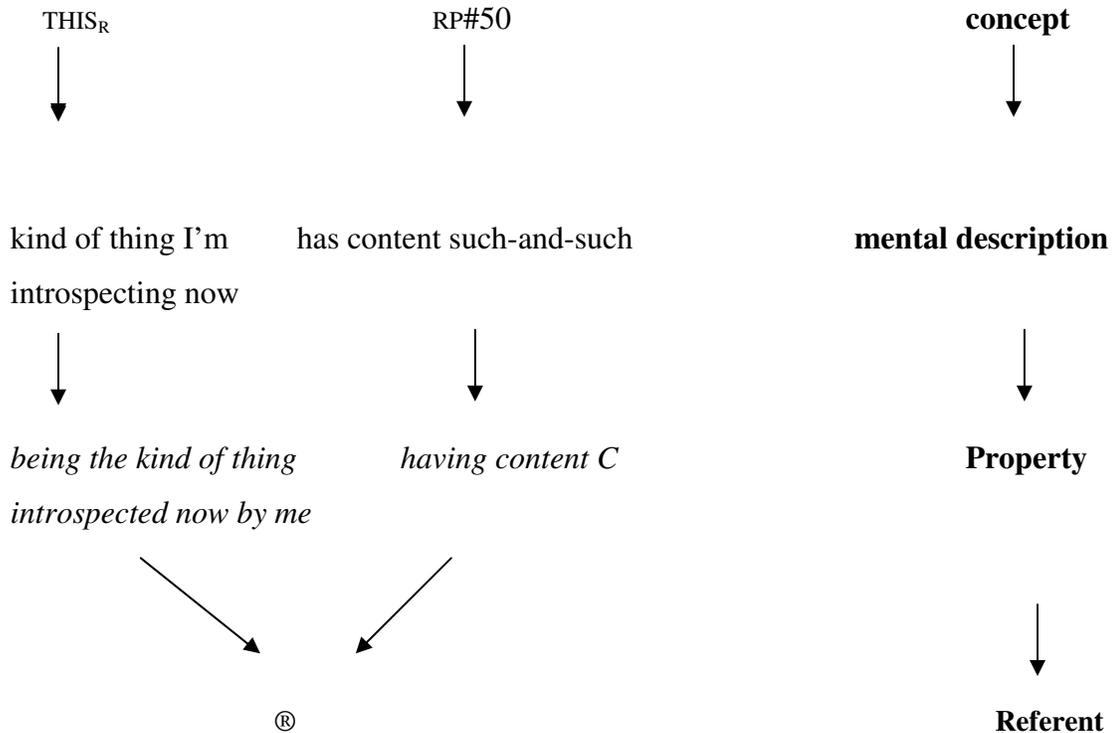


Figure I.5. Filling the blanks—according to the physicalist.

And we can also deal with the regress argument (even while assuming the Semantic Premise is true). Remember that the physicalist was committed to a regress because she claimed that kinds of role that the descriptions thinkers associate with words might fill (I generalize their claims here to concepts). First, associated mental descriptions might fill what Byrne and Pryor call the *a priori role*. Some descriptions might play the reference-fixing role. Finally, some descriptions might play *the Frege role*. The latter are those we use to explain the cognitive significance of the thought that Bob Dylan is Robert Zimmerman “many philosophers appeal to differences in the properties that [thinkers] associate with the name *Bob Dylan* and *Robert Zimmerman*.” (Presumably, the descriptions that explain cognitive significance will

also be able to explain why a subject who thinks that Bob Dylan is not Robert Zimmerman would be rational).

Byrne and Pryor go on to argue that the descriptions that fill the a priori role need not fill the Frege role. After all, they write, we seem to associate few significant descriptions with our concepts *a priori*—even our *nonphenomenal* concepts. Thinkers will associate *a priori* with BOB DYLAN descriptions like “is sentient” and “is Bob Dylan”—and with ROBERT ZIMMERMAN descriptions like “is sentient” and “is Robert Zimmerman”. And they add: “since these properties are associated with both names, they cannot help explain the difference in cognitive significance (3).” They argue that the converse seems true too: descriptions that fill the Frege role need not fill the a priori role. “*Being the author of Mr. Tambourine Man* for example, might fill the Frege role for *Bob Dylan* simply because it is a very well-known a posteriori fact that Dylan wrote *Mr Tambourine Man*.”

If White should get rid of the a priori requirement, what should he put in its place? One answer is simply: nothing. Or, he could add another requirement, for instance a *thinness*. I argue that none of this will help.

No a priori requirement

If Byrne and Pryor are right, then it’s not obvious why we should claim (as White does) that the associated mental descriptions—which will explain the rationality of thinking certain things—should be a priori. The relevant mental descriptions should, first and foremost, fill the Frege Role—and they can fill that role without filling the a priori role. We should move away from a priori descriptions like “heavenly body

that's Hesperus"—which though a priori cannot play the Frege role and explain the rationality of Anna's beliefs. Similarly, the physicalist should move away from a priori descriptions like "state of mine that feels like *this_R*"—which though a priori cannot fill the Frege role and explain the rationality of Adi's beliefs.

And now the physicalist can help herself to a whole slew of descriptions which she grants (agreeing with Levine see 2.1) are *a posteriori* connected with phenomenal concepts. Adi might associate with THIS_R the description "thing I get when I look at ripe tomatoes".

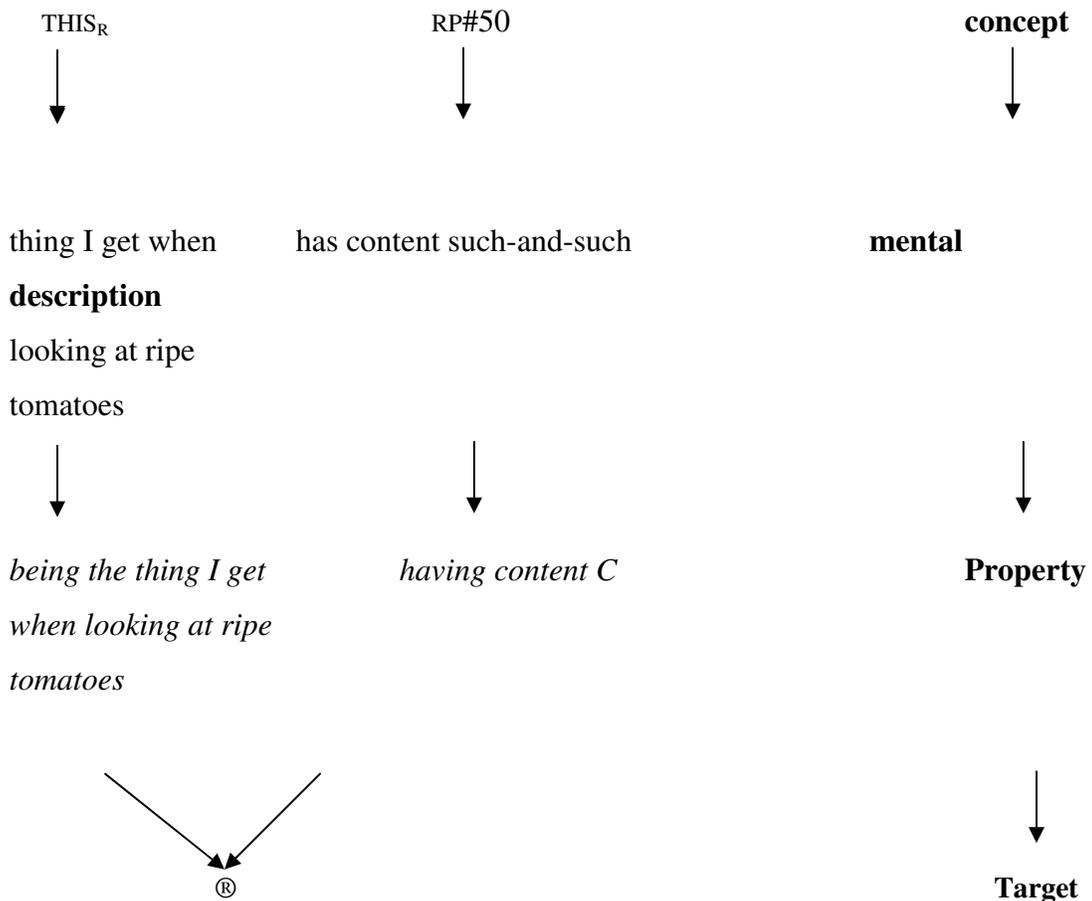


Figure I.5. Filling in the blanks—a posteriori.

When she thinks that *this_R* feel is not *rp#50*, here is the world corresponding to her thought then: a world in which there are two things, one of which she gets when looking at ripe tomatoes but which doesn't have content C (say because of facts about ripe tomatoes in that environment); and one of which has content C but, in that environment, isn't the feel she gets when she looks at ripe tomatoes. If White gives up on the a priori requirement altogether, then, the physicalist has really nothing to worry about.

The thinness requirement

White seems to think that there are *two* requirements on the relevant descriptions: they should be a priori *and thin*. Could White then replace the a priori requirement with the thinness requirement?

Thin descriptions are descriptions that ascribe to the referent *thin* properties. And thin properties are properties such that “there is nothing to [them] over and above what is understood by the subject who understands the predicate[s] that express[...them]” (223). He puts it differently too, claiming that thin properties are “properties that confer no empirically discoverable essence on the things in which it is instantiated” (2007, 233). The description he believes is associated with WATER a priori (natural kind in lakes etc.) is a *thick* one he says, because it ascribes to the referent a *thick* property—a property which could turn out, unbeknownst to the speaker, to have a hidden essence (to be identical for instance with another property). The description “state of mine that I’m introspecting now” is a *thick* description as well, because its referent could turn out to have some hidden essence (it could turn

out to be identical to some other property unbeknownst to the thinker). So it can't be the relevant description. The description "state of mine that's *this_R* feel", on the other hand, is thin. It does not have a hidden essence; it's just *this_R* feel, period, nothing more. Therefore only *that* description can be *the* relevant when we fill in the blanks.

This requirement, again, won't do—for it, too, has counterintuitive implications in the standard cases. White claims that the description associated with water is thick. It would follow that a thinker who associates that description with WATER would not be rational to believe that water is not H₂O. Similarly--the description "thing that's Hesperus" is not thin—since it ascribes to the referent (Venus) a property that *does* confer on it an empirically discoverable essence. It will be obvious to Simple Anna that there is something more to *being Hesperus* than merely what she knows about it. And there is, since Hesperus turns out to be Phosphorus (to be Venus). The thinness requirement won't do.

A recap

Even assuming that the Semantic Premise is true, property dualism doesn't follow. White's property dualist argument relies, crucially, on his a priori requirement, which claims that the associated descriptions relevant to rationality must be a priori. We saw that there were two ways to interpret this a priori requirement. Interpreted strictly, it leaves White unable to account for rational thoughts in most standard cases. Interpreted more loosely, it leaves White unable to conclude that physicalism is false. It leaves White with a two options: get rid of the a priori requirement altogether—which unsurprisingly leaves White unable to conclude that physicalism is false—or

replace it by the thinness requirement—which again leaves White unable to account for rational thoughts in most standard cases. Even if the Semantic Premise is true, it does not follow that physicalism is false.

4.1.5. The big picture

In this section (4.1.), I have argued that the property dualist argument against physicalism fails. The argument rests in large part on an instantiation of the Semantic Premise according to which two distinct concepts can co-refer only if there are distinct mental descriptions associated with each concept, and two distinct properties of the referent corresponding to the two mental descriptions. The anti-physicalist argues that in the psychophysical case, one of the two distinct properties of the referent must be non-physical. If both properties are physical, we end up either with an infinite regress or with the inability to provide a possible world that corresponds to rational thoughts like “*this_R* feel is not representational property #50”.

I first argued that the Semantic Premise has counterintuitive implications and should therefore be rejected. Then I argued that the regress and the worry about rational thoughts arise only if the mental descriptions associated with phenomenal concepts are characterized “phenomenally” (i.e., using phenomenal concepts). White believes this because he relies on an priori requirement which, I showed, puts in him in a bind.

4.2. Chalmers⁴²

In this last section, I consider an argument against the phenomenal concept strategy due to Chalmers (2006). Like the property dualist arguments discussed in the last section (4.1.), Chalmers argument is one which concludes that no appeal to phenomenal concepts—regardless of how they are characterized—can constitute an adequate defense of physicalism. More specifically, Chalmers argues that phenomenal concepts can't *both* effectively defuse the anti-physicalist arguments *and* be physically explicable themselves.

4.2.1. A dilemma for the physicalist

Here is the argument as he sees it:

- (1) Either we can conceive that Chalmers' zombie duplicate (call him "Zombie Chalmers") lacks phenomenal concepts, or we can't conceive that he lacks such concepts.
- (2) If we *can* conceive of Zombie Chalmers lacking phenomenal concepts, then a new explanatory gap is formed and phenomenal concepts turn out to be physically inexplicable.
- (3) If we *can't* conceive of Zombie Chalmers lacking phenomenal concepts, then phenomenal concepts can't explain the explanatory gap.
- (4) It follows that either phenomenal concepts aren't physically explicable or they don't explain the explanatory gap.

The argument seems powerful. Premise (1) looks like a necessary truth. Premise (2) looks to be true. For anything that Chalmers has that Zombie Chalmers can be imagined to lack (given that the latter is physically, functionally, and intentionally identical to Chalmers) will be physically inexplicable. Premise (3) also seems true. For if Zombie Chalmers can't be conceived to lack phenomenal concepts, then that must mean that those concepts are physically or functionally explicable; but Zombie

⁴² This section is a very slightly modified version of Carruthers and Veillet (2007).

Chalmers lacks phenomenal consciousness—so we have agreed that physical and functional facts can't explain phenomenal consciousness; in which case phenomenal concepts won't be able to do the work required of them, either. Moreover, the argument as a whole appears valid.

On further reflection, however, the argument as it stands can be seen to be problematic. For in order for (1) to be a necessary truth, the phrase “phenomenal concepts” will have to be taken univocally. But then when we see that term at work in the two premises that follow, it seems that it must be taken in a different way in each. In premise (2) we are to assume that we can conceive of *Zombie Chalmers* lacking phenomenal concepts, which seems to require characterizing phenomenal concepts in terms of phenomenal feels. By definition zombies lack phenomenal feels. Therefore, if we take phenomenal concepts to be those concepts that pick out phenomenal feels, then surely zombies must lack them too.

The usage in Premise (3), in contrast, seems to require a different characterization of phenomenal concepts—as conceptually isolated concepts which are partly recognitional and refer to their *physical* referents directly. Such a characterization, note, doesn't require bringing up phenomenal feels themselves (under that description at least). So we can conceive of *Zombie Chalmers* having these phenomenal concepts—after all, since he shares all of Chalmers' physical, functional, and intentional properties, *Zombie Chalmers* must also possess conceptually isolated partly recognitional concepts that refer directly and pick out a property of his brain.

If 'phenomenal concepts' can't be interpreted univocally throughout the

argument then, as it stands, it commits a fallacy of equivocation. Naturally, it would make life easy for physicalists if Chalmers' argument could be defeated so easily! But in fact it can be reformulated to avoid the difficulty, by framing a version of Premise (1) that no longer purports to be a necessary truth. Thus:

- (1*) Phenomenal concepts can either be characterized in terms of phenomenal feels, or they can be characterized wholly physically.
- (2a) If phenomenal concepts are characterized in terms of phenomenal feels, then we can conceive of Zombie Chalmers lacking such concepts.
- (2b) If we can conceive of Zombie Chalmers lacking phenomenal concepts, then a new explanatory gap is formed and phenomenal concepts turn out to be physically inexplicable.
- (3a) If phenomenal concepts are characterized purely physically, then we can't conceive of Zombie Chalmers lacking such concepts.
- (3b) If we can't conceive of Zombie Chalmers lacking phenomenal concepts, then phenomenal concepts can't explain the explanatory gap.
- (4) It follows that neither way of characterizing phenomenal concepts can help with the problem of phenomenal consciousness – either they introduce a new explanatory problem, or they can't do the explanatory work required.

This argument commits no fallacy that we can see, and all of its premises present at least the appearance of truth. So is the phenomenal concept strategy defeated? We believe not. For we think that there are sufficient grounds for denying the truth of Premises (2b) and Premise (3b). We consider those in turn.

4.2.2. *A new Explanatory Gap*⁴³

Defenders of the phenomenal concept strategy should concede that if (2a) is true and we can conceive of Zombies Chalmers lacking phenomenal concepts, then phenomenal concepts won't be physically explicable. This is not quite to say that the physicalist—or anyone else for that matter—should agree that if (2a) is true then (2b) must be true, since (2b) doesn't merely state that phenomenal concepts aren't

⁴³ The final version of Carruthers and Veillet's "Phenomenal Concept Strategy" didn't include this section.

physically explicable. It also states that this is the case *because* a new explanatory gap is formed. In this section we will challenge the claim that the gap is a new one. Of course, challenging the move from (2a) to (2b) on these grounds won't help the physicalist much if (2a) still turns out to entail, as we believe that it does, that phenomenal concepts aren't physically explicable for some other reason. But this deserves to be cleared up regardless of how helpful it is to our purposes.

Chalmers believes that a new explanatory gap is formed: why? Remember that on this characterization of phenomenal concepts, it is possible to imagine that Chalmers' zombie twin should lack them. The argument for a new gap in explanation therefore parallels very closely Chalmers' original zombie argument. The original argument maintained that if we can conceive of two physical duplicates, one of whom is phenomenally conscious and one of whom isn't, then *phenomenal feels* cannot be physically explained. So if we can conceive of two physical duplicates, one of whom possesses phenomenal concepts and one of whom doesn't, then we should conclude that *phenomenal concepts* (characterized in the first-person way) aren't physically explicable, either. The first gap in explanation, which the physicalist agrees is real, is a gap between physical explanations and phenomenal feels. What we have here is a second gap, between physical explanations and phenomenal *concepts*.

The explanatory gap argument when applied to *concepts*—phenomenal or otherwise—seems to us to be invalid, however. Putnam (1975) in his famous Twin Earth thought-experiment introduced us to Twin Earth, a planet just like Earth except that the identical-looking stuff in the lakes, rivers, and so on, isn't H₂O but XYZ. The latter is a substance that can only be distinguished from H₂O by means of

sophisticated laboratory tests. And on Twin Earth lives Twin Oscar, who is a microphysical duplicate of Earthling Oscar (abstracting from the fact that his body contains XYZ whereas Oscar's contains H₂O – by hypothesis this makes not the smallest difference to their cellular, neurological, or cognitive processes). According to Putnam and those with externalist leanings when it comes to intentional content (like Chalmers in the paper that we are discussing – see below), it turns out that Twin Oscar possesses the concept TWATER and not his twin's corresponding concept WATER. We now have two physical duplicates (Oscar and Twin Oscar) one of whom (Oscar) possesses the concept WATER while the other one (Twin Oscar) doesn't.

This case is very much like that Chalmers and his zombie twin. Chalmers possesses phenomenal concepts; his twin doesn't. And there exists a physicalist story as to why that might be the case: Chalmers is causally related, in the right sort of way, to phenomenal states; his twin isn't.⁴⁴ Of course this explanation won't ultimately be very satisfying, since the physicalist's explanation of what it is to possess a phenomenal concept (characterized in this first way) appeals to a relation to *phenomenal feels*. And he (the physicalist) has agreed that there is a gap between phenomenal feels and the physical. But notice that what the physicalist is facing, here, is the *old* gap over again. Phenomenal concepts turn out to be physically inexplicable simply because they are characterized in terms of phenomenal feels which are themselves physically inexplicable. Given this first account of phenomenal concepts, the physicalist's strategy simply fails to deal adequately with the original explanatory

⁴⁴ Note that the causal relationship here needn't be direct, in order for the point to go through. One could hold, with Chalmers (1996), that phenomenal properties are tied to physical ones by basic laws without themselves having physical effects, while still making the point that we make in the text.

gap:⁴⁵ there is no *new* gap.

Even if that's right, premise (2b) can be weakened by striking out all references to a *new* explanatory gap. And in that case defenders of the phenomenal concept strategy should concede that it is true. For those who adopt this strategy say that it is because we make use of phenomenal concepts that zombies are conceivable, there is an explanatory gap, and so on. No matter how detailed a description I am given in physical, functional, or intentional terms, it will always be possible for me to think, "Still, all that might be true while *this* state was absent or different." And defenders of the phenomenal concept strategy claim that it is the conceptual isolation of the phenomenal concept THIS that makes such thoughts thinkable, and thus that gives rise to the conceivability of zombies and the explanatory gap.

By the same token, then, if phenomenal concepts are characterized in terms of phenomenal feels, I shall be able to conceive of zombies lacking such concepts. If I characterize a given phenomenal concept as, "The concept that I hereby deploy when thinking about *this* state", then no matter how detailed a description of someone I am given in physical, functional, or intentional terms, it will always be possible for me to think, "Still, all of that might be true while the concept that I hereby deploy when thinking about *this* state was absent or different." And since this thought is thinkable, the explanatory gap remains. Given that we can conceive of a physical, and functional, and intentional duplicate who would nevertheless lack phenomenal

⁴⁵ Indeed, under this characterization of phenomenal concepts, a physicalist who attempts to embrace the first horn of Chalmers' dilemma would be doing the following: he would be conceding that there is an explanatory gap between phenomenal states and the physical while claiming that phenomenal concepts can explain the existence of this gap. But then he describes phenomenal concepts in terms of phenomenal states. It shouldn't be surprising that the phenomenal concept strategy, thus construed, should fail.

concepts (characterized in first-person terms), there will be an explanatory gap between all physical, functional, and intentional facts and the existence of such concepts (so characterized).

Suitably weakened, the first horn of Chalmers' dilemma should be embraced, then: if phenomenal concepts are characterized in terms of phenomenal feels, then the explanatory gap remains, and phenomenal concepts themselves turn out to be physically inexplicable. But this only presents a difficulty for the phenomenal concept strategy if the premises on the other horn of the dilemma are also true. Otherwise we can claim that our physicalist account of phenomenal concepts can *explain why* there is an explanatory gap between all physical, functional, and intentional facts and the existence of such concepts, just as it can explain why there is an explanatory gap between all physical, functional, and intentional facts and the existence of phenomenally conscious mental states themselves. In both cases the explanation will turn on the conceptual isolation attributed to phenomenal concepts.

4.2.3. Explaining our epistemic situation

Chalmers' defense of the claim made in Premise (3b) is quite complex, turning crucially on his discussion of what he calls "epistemic situations". Throughout the discussion of this conditional, however, it should be borne in mind that phenomenal concepts are to be understood in purely physical terms, as conceptually isolated partly recognitional concepts deployed in the presence of certain physical states.

Let us recall the original explanatory gap problem, the conceivability of zombies, and the argument from Mary's new knowledge. And let us, in addition,

consider claims like, “I am phenomenally conscious.” These problems (and others like them) and this claim (and others like it) form what Chalmers’ calls our *epistemic situation* when it comes to phenomenal consciousness. Proponents of the phenomenal concept strategy believe that our possession of phenomenal concepts can explain our epistemic situation. We have already seen how the physicalist will argue that phenomenal concepts explain why there is a gap in explanation, why zombies and inverters are conceivable, and what Mary learns. And when I say, “I am phenomenally conscious”, it may be that I am in fact saying something like: “I have experiences like *these*”, where THESE is a phenomenal concept. We are now in a position to schematize Chalmers’ argument for Premise (3b) as follows:

- (i) If zombies do indeed possess phenomenal concepts (which must be the case if Zombie Chalmers can’t conceivably lack phenomenal concepts, characterized in the physicalist way), but don’t share our epistemic situation, then *our* having phenomenal concepts can hardly explain *our* epistemic situation.
- (ii) Zombies don’t share our epistemic situation.
- (iii) It follows that the possession of phenomenal concepts can’t explain our epistemic situation (given a third-person characterization of phenomenal concepts).

According to the first premise of this argument, if Chalmers and Zombie Chalmers don’t share the same epistemic situation, then phenomenal concepts can’t explain our epistemic situation. Chalmers provides an argument for this claim which parallels, again, the original arguments from zombies and the explanatory gap. The original arguments can be summarized like this: if you can imagine two physical duplicates, one phenomenally conscious and the other not, then phenomenal consciousness can’t be explained in physical terms. Now we can say this: if we can imagine two duplicates both possessing phenomenal concepts, one in our epistemic situation and

the other not, then our epistemic situation isn't explicable in terms of phenomenal concepts. We will grant Chalmers the truth of this premise.

Premise (ii) asserts that Chalmers and his zombie twin don't share the same epistemic situation. This is more questionable. According to Chalmers (2007, 176), for two duplicates to share the same epistemic situation is for their corresponding beliefs to have the same truth-values and the same epistemic status "as justified or unjustified, and as cognitively significant or insignificant". Corresponding beliefs, Chalmers goes on to say, need not have the same contents. Oscar and Twin Oscar, he argues, share the same epistemic situation.⁴⁶ Oscar's belief that water [H₂O] is refreshing and Twin Oscar's corresponding belief that twater [XYZ] is refreshing will both be true, even if the two beliefs don't have the same content. Chalmers argues that he and his zombie twin, unlike Oscar and Twin Oscar, do *not* share the same epistemic situation. Chalmers' belief that he is phenomenally conscious is true, whereas Zombie Chalmers' belief that he is phenomenally conscious is false. Or think back to Mary, and imagine her possessing a zombie twin. Mary gains new introspectible knowledge when she is finally freed from her room, whereas Twin Mary doesn't gain all of the same knowledge. So they don't seem to share the same epistemic situation. Chalmers concludes that our zombie twins cannot share our epistemic situation.

We now propose to argue that Premise (ii) is false, however, and that zombies

⁴⁶ For those unfamiliar with the famous Twin Earth thought-experiment (Putnam, 1975), Twin Oscar is a microphysical duplicate of Earthling Oscar who lives on Twin Earth, a planet just like Earth except that the identical-looking stuff in the lakes, rivers, and so on, isn't H₂O but XYZ. The latter is a substance that can only be distinguished from H₂O by means of sophisticated laboratory tests. (Of course Twin Oscar cannot be a *complete* duplicate of Oscar, since his body contains XYZ whereas Oscar's contains H₂O. But by hypothesis this is supposed to make not the smallest difference to their cellular, neurological, or cognitive processes.)

do share our epistemic situation (in one good sense of the notion of “epistemic situation” – we will return to this point in section 4.2.4.).

Chalmers compares zombie duplicates to Oscar and Twin Oscar. Oscar, on Earth, is entertaining a thought that he would express with the words, “Water is refreshing.” Our intuition is that Oscar is referring to H₂O. When Twin Oscar thinks a thought that he, too, would express with the words, “Water is refreshing”, our intuition is that he is referring to XYZ, and not H₂O. Oscar and Twin Oscar both possess concepts that they deploy under the same circumstances (when they are thirsty), which are associated with certain kinds of perceptual states (seeing a colorless liquid), and so forth. But, according to the externalist, those corresponding concepts will have different contents. The content of Oscar’s concept is tied to H₂O, whereas the content of Twin Oscar’s concept is tied to XYZ. Chalmers seems ready to accept the externalist conclusion. He argues that Oscar and Twin Oscar have corresponding beliefs with the same truth-values but different *contents*. When they say, “This is water”, both are right, although they are talking about different things: Oscar is talking about water (H₂O), his twin is talking about twater (XYZ) (Chalmers, 2006, 11). And yet despite this, they share the same epistemic situation.

What, then, prevents us from saying the same thing about Chalmers and his zombie twin? Chalmers and Zombie Chalmers both have concepts that they deploy in similar circumstances in the presence of certain perceptual states, that are conceptually isolated, and so on. An externalist (of the sort that Chalmers seems to be throughout his paper) could very well say that the contents of Chalmers’ phenomenal concepts differ from the contents of his zombie twin’s phenomenal concepts. The

content of one of Chalmers' phenomenal concepts will turn out to involve a phenomenal feel, whereas the content of his twin's corresponding phenomenal concept can't possibly involve such a state.⁴⁷ According to Chalmers it seems plausible that the content of a zombie's phenomenal concepts would be *schmenomenal* feels. (These would be states that have the same physical, functional, and intentional properties as Chalmers' states, but that aren't phenomenally conscious; see 2006, 19.) The physicalist would then argue that Chalmers' and Zombie Chalmers' corresponding beliefs have the same truth-values and are justified in similar ways, but they are quite importantly *about* different things. So Chalmers and Zombie Chalmers can share the same epistemic situation after all, just as do Oscar and his twin.

Chalmers argues that defending this kind of reply, "requires either deflating the phenomenal knowledge of conscious beings, or [...] inflating the corresponding knowledge of zombies" (2007, 185). He goes on to argue that either strategy has counterintuitive consequences. No one thinks that Zombie Mary learns just *as much as* Mary (an implication of the inflationary move). No one thinks that Mary learns just *as little as* Zombie Mary does (an implication of the deflationary move). When we think of zombies, we aren't conceiving of creatures possessing something epistemically just as good as consciousness. We are conceiving of deprived creatures with impoverished knowledge of themselves.

But Chalmers is surely confused here. Arguing that zombies' phenomenal

⁴⁷ This isn't to say that phenomenal concepts are *characterized* in terms of those phenomenal states, as they are on a first-person interpretation. Again, throughout this section of the paper phenomenal concepts are characterized as conceptually isolated concepts deployed in the right sorts of circumstances. But (and this is our point) there is no reason to think that our phenomenal concepts have the same content as our zombie twins' corresponding concepts.

concepts have different contents enables us to say the following about Mary and her zombie twin: they both gain the same *amount* of knowledge, but (and this is crucial) it is the same amount of knowledge *about different things*. Mary's knowledge is knowledge *of phenomenal feels*, Zombie Mary's knowledge is knowledge *of schmenomenal feels*, just like Oscar's knowledge is *of water* (H₂O) and his twin's is knowledge *of twater* (XYZ). Physicalists needn't deflate the knowledge gained by Mary or inflate the knowledge gained by Zombie Mary in order for the phenomenal concept strategy to work. All we need to point out is that the *objects* of their knowledge are very different.

Physicalists can now deal with a variety of claims quite effectively. Consider, for instance, the discussion that Chalmers imagines between a zombie eliminativist and a zombie realist. The eliminativist argues that there is no such thing as phenomenal consciousness and the realist maintains that there is such a thing. Here is what Chalmers says about them:

When such a debate is held in the actual world, the [...] materialist and the property dualist agree that the zombie realist is right, and the zombie eliminativist is wrong. But it is plausible that in a zombie scenario, the zombie realist would be wrong, and the zombie eliminativist would be right. (2006, 12.)

But in the zombie scenario, it is just as plausible that the zombies would simply not be talking about *phenomenal* consciousness. Their debate is about the existence of *schmenomenal* consciousness. And the zombie realist, like his twin in the actual world, may very well be right; his beliefs, like those of his twin, may very well be true.

We can say the same type of thing when it comes to the explanatory gap, or

the conceivability of zombies. Zombies are thinking about schmenomenal consciousness using their phenomenal concepts, which are conceptually isolated from their other concepts and partly recognitional. They will conclude from their reflections that there is a gap in explanation between *schmenomenal* consciousness and their physical world. They will also conclude that it is conceivable for someone to be physically, functionally, and intentionally identical to them and yet lack *this* (where the concept THIS that they deploy picks out a schmenomenal state). And so forth.

This difference-in-content move now allows us to deal with a variety of first-person claims as well. Zombie Mary, after she leaves her room, may well come to believe something that she would express by saying, “*This* is an experience of blue.” What will make this belief true isn’t her actually having a *phenomenal* experience of blue, but rather her having a *schmenomenal* experience – whatever that turns out to be. And so both her beliefs and Mary’s beliefs could plausibly have the same truth-values. Similarly when Chalmers says, “I am phenomenally conscious”, and his zombie twin utters the same string of words, both are in fact saying something different. To assume that they are saying the same thing (that they are both talking about *phenomenal* consciousness) is to assume that the contents of their states and concepts will be the same. But if there is no reason to assume this about Oscar and Twin Oscar, then there is no reason to assume this about Chalmers and his zombie twin. Zombie Chalmers is really saying that he is *schmenomenally* conscious, and we have every reason to think that he is right in thinking that, just as Chalmers is right in thinking he (Chalmers) is *phenomenally* conscious.

Chalmers, to block this line of reply, may now resort to our intuitions about zombies. We have claimed that they will turn out to have something epistemically just as good as phenomenal consciousness, namely schmenomenal consciousness. But doesn't that seem wrong? When we are conceiving of zombies, aren't we conceiving of beings with nothing at all that is epistemically like consciousness?

Well, on our view zombies are still zombies in that they are *not* phenomenally conscious. Their perceptual states don't have phenomenal feels. In this respect it *is* all dark inside. Yet they have something playing a certain role in their psychology—a role analogous to the role that phenomenal consciousness plays in ours. They have something *epistemically* just as good as consciousness, but they don't have anything that is *phenomenally* as good. And it seems that this is what matters here. The schmenomenal states they undergo do not *feel* like anything. Even though their schmenomenal beliefs are *true* when our corresponding phenomenal beliefs are, their beliefs are, sadly enough, not *about* the same good stuff as our corresponding beliefs—they are not about the *feel* of experiences. Zombies are still, it seems, in quite a dreadful situation. So our intuitions about zombies are preserved.

4.2.4. *Of Zombies and Zombie Zombies.*

We have shown that there are good reasons to resist Chalmers' claim that zombies fail to share our epistemic situation. If he can't make this case, then he can't argue successfully for Premise (3b). And so it *isn't* true that if zombies conceivably possess phenomenal concepts, then phenomenal concepts can't do the work that physicalists want them to do. Or at least, we have been given no reason to believe that this is so.

There is, however, a further line of reply open to Chalmers, which we consider in the present section.

About Epistemic Situations.

We think that Chalmers will object that in conceiving of an “epistemic situation” in such a way that both Chalmers and Zombie Chalmers share the same epistemic situation, the facts crucial to our actual epistemic situation have been omitted. For when I make the judgment that I might express by saying, “*This* is a blue-cup experience”, I don’t *just* deploy a conceptually isolated concept in the presence of an intentional state representing the presence of a blue cup. In addition, I deploy such a concept on the basis of my awareness of *this* type of mental state (a phenomenal feel). And by hypothesis, Zombie Chalmers doesn’t have awareness of any such state. While Chalmers and Zombie Chalmers have much in common – in particular, they make similar judgments in similar circumstances (all of which can be true) and the epistemic liaisons of those judgments (when characterized in third-person terms) are all precisely parallel to one another – there are also crucial differences. For Chalmers’ judgments are grounded in the presence of mental states like *these*, and *those*, and *this*, and *that* (where the indexicals here express phenomenal concepts), whereas Zombie Chalmers’ judgments are not. This seems like it might be an important – indeed, vital – part of Chalmers’ epistemic situation. In which case the crucial premise in the argument outlined in 4.2.2. is true: zombies don’t share our epistemic situation.

Another way of expressing the point just made would be this: Chalmers may

deny that the distinction between a concept and its conceptual mode of presentation (or associated mental description) finds any application in connection with phenomenal consciousness. Since H₂O and XYZ are presented to Oscar and Twin Oscar in the same way, we can say of them that (1) they possess concepts that play similar roles in their mental lives, and (2) they apply those concepts on the basis of the same associated mental descriptions. Only when these two conditions are met can we say that the twins share the same epistemic situation. Phenomenal feels, in contrast, provide their *phenomenal* associated descriptions, which are essential to them (Kripke, 1972). It follows that a phenomenal feel and another distinct (schmenomenal) property *cannot* be presented to Chalmers and his zombie twin in the same way. So the pair of them possess, at most, (1): concepts that play similar roles in their mental lives. Since they can't possibly apply those concepts on the basis of the mental descriptions, they cannot share the same epistemic situation, just as Chalmers maintains. Seen in this light, Chalmers ought to concede that it was a tactical error (or at best misleading) for him to have introduced Oscar and Twin Oscar into the discussion.

Recall, however, the distinction drawn at the very beginning of this section (4.2.) between two characterizations of phenomenal concepts – a distinction similar to one Chalmers himself makes between phenomenal and schmenomenal concepts. We claimed there (again roughly as Chalmers himself does) that we could think of phenomenal concepts as applied *either* in response to phenomenal feels *or* in response physical (functional/representational) properties (for example). According to the second horn of Chalmers' argument that we have been considering since the outset of

4.2.2, moreover, phenomenal concepts are to be characterized in *physical* terms. So both Chalmers and Zombie Chalmers should be said to employ concepts whose applications are prompted by the presence of certain distinctive sorts of intentional / functional state, where those concepts are conceptually isolated from others. In which case, to introduce the feel of the state into our description of the conceptual mode of presentation of Chalmers' concepts is to switch illegitimately to the "phenomenal" characterization of those concepts. And if we do restrict ourselves to a physical account of the concepts involved, in contrast, then the comparison with Oscar and Twin Oscar is entirely appropriate: in both cases we have pairs of people whose concepts have similar associated mental descriptions and play the same conceptual roles, but where those concepts happen to pick out different things.

We have alleged that the response that we made on Chalmers' behalf would re-introduce (illegitimately) the first (phenomenal) characterization of phenomenal concepts into the defense of Premise (3). Chalmers might reply, however, that this allegation is unfounded. For it isn't the characterization of phenomenal concepts that is in question, here. What is at issue isn't what we mean by "phenomenal concept". Rather, what is in question is the presence, or absence, of the states picked out by such concepts, when those concepts are used by their possessors. It is the presence of *this state* (the state, not the concept of the state here deployed) that is partly distinctive of Chalmers' epistemic situation, and which marks its difference from Zombie Chalmers' epistemic situation.

But now a problem of a different sort emerges. If Chalmers' epistemic situation is partly characterized in terms of the presence of *this state* (a phenomenal

feel), which we can imagine Zombie Chalmers to lack, then this amounts to saying that it is an important part of Chalmers' epistemic situation that he has phenomenally conscious mental states, whereas Zombie Chalmers doesn't. And doesn't that now beg the question? For this is something that is supposed to be granted on all hands. Defenders of the phenomenal concept strategy, too, allow that we can conceive of someone who is physically, functionally, and intentionally identical to Chalmers (that is, Zombie Chalmers), but who lacks any of the phenomenally conscious mental states that Chalmers enjoys. And we claim to be capable of *explaining* how such a thing can be conceivable in a way that doesn't presuppose the existence of anything beyond the physical, the functional, and/or the intentional.

Asserting that this strategy cannot work because phenomenal states themselves are part of what is distinctive of Chalmers' epistemic situation, and pointing out that the strategy can't explain *them*, is to insist that the phenomenal concept strategy should explain phenomenal consciousness. But that was never at issue. The phenomenal concept strategy is a strategy for explaining the conceivability of zombies, the explanatory gap, and so forth, not for explaining phenomenal consciousness *per se*. To put the point somewhat differently, the phrase "our epistemic situation" is supposed to be a handy label for the various phenomena that the phenomenal concept strategy is intended to explain (the conceivability of zombies etc.). But since that strategy was never intended as a reductive explanation of phenomenal consciousness as such, "our epistemic situation" should *not* be understood in such a way as to encompass phenomenal feelings.⁴⁸

⁴⁸ It is important to note, too, that a physicalist who deploys the phenomenal concept strategy is *not* here arguing *for* physicalism. Stoljar (2005) goes wrong on just this point. For he claims that the

The true dialectical situation is as follows, we believe. Insofar as they argue legitimately, Chalmers and other anti-physicalists are asserting that the *best explanation* of the conceivability of zombies, the conceivability of experiential inversions, the explanatory gap, and so on is that our experiences possess distinctive properties (call them “qualia”) that cannot be reductively explained in physical, functional, or intentional terms. Chalmers might concede that we do possess phenomenal concepts, characterized in something like the way that the proponent of the phenomenal concept strategy characterizes them (conceptual isolation and so forth). But he denies that an appeal to these concepts alone can explain what needs to be explained (the possibility of zombies, the explanatory gap, and so forth). His opponent, in contrast, asserts that we don’t need to appeal to any special properties of phenomenally conscious experience to do the work: the entire explanatory burden can be taken up by appeal to the phenomenal concepts in terms of which we think about those experiences.

4.2.5. *Zombie-Zombie Chalmers*

In order to move this debate forwards, we need to introduce a further character into the story: *Zombie-Zombie Chalmers*. Recall that *Zombie Chalmers* has been allowed

physicalist’s reply to conceivability arguments comes in two stages, the first of which is that the conceptual isolation of phenomenal concepts/truths entails that the conditional, (1) *If P, then P**, is a posteriori necessary (where *P* is a summary of all physical truths, and *P** is a summary of all phenomenal truths). But physicalists who adopt the phenomenal concept strategy aren’t attempting to show the truth of this entailment. Making the case that (1) is a necessary truth would, it is true, be making the case for physicalism. But the phenomenal concept strategy is only intended to be defensive. The physicalist is only arguing that the conceivability arguments don’t show that physicalism is *false*, despite what their proponents claim: there is another explanation for why we can conceive of these things, an explanation that appeals to phenomenal concepts. So Stoljar misses the fact that the phenomenal concept strategy is essentially a defensive strategy. It is a strategy that physicalists employ to show that the key anti-physicalist arguments fail. It isn’t meant to make a positive case for the truth of physicalism, or for the necessary a posteriori truth of (1).

to possess phenomenal concepts, characterized in a third-person way. For example, he has concepts that are applied recognitionally on the basis of his perceptual and imagistic states, and which are conceptually isolated from all of his other concepts (whether physical, functional, or intentional). Possessing such concepts, Zombie Chalmers will be able to conceive of a zombie version of himself (Zombie-Zombie Chalmers). If on a given occasion he uses the word “this” to express one of his phenomenal concepts, then he will be able to entertain thoughts that he might articulate by saying, “There might exist someone who is physically, functionally, and intentionally identical with myself, but who nevertheless lacks anything resembling *this* type of state.” Since his phenomenal concept is conceptually isolated, there will be no hidden contradiction in this thought that he would be capable of detecting a priori.

Likewise if Zombie Chalmers uses the word “this” to express a phenomenal concept that applies to one of his percepts of color. (For these purposes, Zombie Chalmers’ perceptions of color need to be characterized purely functionally and intentionally, of course. They are perceptual states with a fine-grained intentional content representing properties of surfaces that impact the latter’s reflection of light, perhaps.) Then he, too, will fall subject to the Mary thought-experiment. He will be inclined to think, “Mary brought up in her black and white room couldn’t know what it is like to undergo *this* type of perceptual state, no matter how much she knows about the physical, functional, and intentional properties of color vision.” And he will be inclined to think this precisely because the concept that he expresses by “this” is a conceptually isolated one.

By the same token, Zombie Chalmers will think that there is an explanatory gap between all physical, functional, and intentional facts, on the one hand, and his own mental states (characterized using phenomenal concepts), on the other. Because those concepts are conceptually isolated ones, he will be able to think, “No matter how much you tell me about the physical, functional, and intentional facts involved in perception, it will still be possible that all of what you tell me should be true, while states of *this* sort are absent or inverted.” So he, too, will be inclined to think that there is something mysterious about his perceptual (and imagistic, and emotional) states, which puts them outside the reach of physicalist explanation.

It is plain that it is Zombie Chalmers’ possession of phenomenal concepts that explains why he should find the existence of Zombie-Zombie Chalmers conceivable. And likewise it is his possession of such concepts that explains the conceivability to him of perceptual inversions, that explains why he thinks Mary would learn something new, and that explains why he would think that there is an explanatory gap between the character of his own mental states and all physical, functional, and intentional facts. Plainly, since Zombie Chalmers is being conceived to lack any phenomenally conscious states, it cannot be the presence of such states in him that explains the conceivability of Zombie-Zombie Chalmers, and the rest.

Zombie Chalmers, when presented with the phenomenal concept strategy for explaining the conceivability of Zombie-Zombie Chalmers and so forth, might even be inclined to insist that this strategy can’t explain what is distinctive of his own epistemic situation. He will allow that Zombie-Zombie Chalmers would make parallel judgments to himself, of course, and would act in exactly similar ways, and

on similar grounds. But he will be inclined to insist that something crucial is left out by the phenomenal concept strategy. What is left out is that he (Zombie Chalmers) bases his judgments on the presence of states like *this*, and *this*, and *that*, whereas, by hypothesis, Zombie-Zombie Chalmers is being conceived to lack such states.

Now we can bring it all back home. For in connection with everything that Chalmers thinks, and for every possibility that Chalmers can conceive, and for every argument that Chalmers can offer, Zombie Chalmers can offer a parallel one. Of course, from our perspective, conceiving all of this along with Chalmers, we are conceiving that they are thinking about different things: Chalmers is thinking about phenomenal states, whereas Zombie Chalmers is thinking about schmenomenal states. But this difference plays no role in explaining what each is capable of thinking. On the contrary, it is their mutual possession of phenomenal concepts (characterized in the third-person way) that does that. Since it can't be the fact that Zombie Chalmers possesses phenomenal states that explains his capacity to conceive of Zombie-Zombie Chalmers and the rest (for by hypothesis, he possesses no such states), we shouldn't allow that Chalmers' possession of phenomenal states plays any role in explaining how he can conceive of Zombie Chalmers, either.

This “zombie-zombie argument”, as one might call it, seems to us to decisively shift the burden of proof in this area onto the anti-physicalist.⁴⁹ Since an appeal to phenomenal concepts (characterized in a third-person way as conceptually isolated and so on) can explain everything that Zombie Chalmers is inclined to think and say (and in particular, since it can explain the conceivability to Zombie Chalmers

⁴⁹ Remember, though, that the argument isn't supposed to be an argument in support of physicalism. It is rather a defensive argument intended to undermine a set of arguments *against* physicalism (the arguments from zombies, explanatory gaps, and so forth).

of Zombie-Zombie Chalmers), and since everything that Zombie Chalmers is inclined to think and say, Chalmers is also inclined to think and say and vice versa (controlling for what will seem from Chalmers' perspective to be differences of content), the most reasonable conclusion to draw is that it is Chalmers' possession of phenomenal concepts, too, that explains the conceivability of zombies, the explanatory gap, and so forth.

4.2.6. Replies to Objections.

Chalmers will surely reply as follows: the zombie-zombie argument presupposes that when Zombie Chalmers claims, "I am phenomenally conscious", he says something *true*, and yet (Chalmers will insist) it is much more plausible that this claim is false. Surely, in the zombie world, there is no phenomenal consciousness, and so Zombie Chalmers' claim, in that world, that he is phenomenally conscious must be false.⁵⁰

This can't possibly be a good reply to the argument of the present paper, however. Certainly it can't be if it assumes that Zombie Chalmers' concept PHENOMENAL CONSCIOUSNESS refers to phenomenal consciousness. For as we have shown at the beginning of 4.2., Zombie's Chalmers' phenomenal concepts plausibly refer to his perceptual states (characterized purely functionally and intentionally). Actually, it isn't in the least plausible that a zombie's phenomenal concepts (characterized physically) should be referring to the zombie's (non-existent) phenomenal states (which would make what he says wrong). This would be like saying that Twin Oscar's twater concept actually refers to H₂O, in which case he is

⁵⁰ Chalmers (2006) makes a very similar reply to an argument by Balog (1999) that parallels ours (but deployed in the service of a different conclusion: Balog is interested in denying that there is a link between conceivability and possibility).

wrong every time he says, “This water tastes good.” But clearly that is just absurd. No theory of concepts does (or should) yield such a counterintuitive claim. Zombie Chalmers is correct when he says that he is conscious, because he isn’t saying that he has phenomenal states as *we* understand them. He is correct because he means that he has schmenomenal states, and he has them.

As we have argued, all of Zombie Chalmers’ beliefs turn out to have the same truth-values as Chalmers’ corresponding ones. As a realist about phenomenal consciousness, Chalmers here on Earth will say, “There are phenomenal states”, and he will be right. His zombie twin will utter the same words but will mean that there are schmenomenal (i.e. physical, functional, and/or intentional) states, and he, too, will be right. Likewise if someone here on Earth denies that there are phenomenal states and turns out to be wrong, his zombie twin will likewise turn out to be wrong in the zombie world, since he will be denying, there, that there are schmenomenal (e.g. functional and/or intentional) states.

In fact, it seems that such pairs of corresponding beliefs will turn out *not* to have the same truth value only if dualism is true. If dualism is true and Chalmers says, “Phenomenal states aren’t physical”, then he will be right; but his zombie twin uttering the same words will mean that schmenomenal (e.g. functional and/or intentional) states aren’t physical, and he will be wrong; for by hypothesis his schmenomenal states *are* physical. Since Chalmers’ overall goal is to argue for dualism and against physicalism, he begs the question when he assumes that his zombie twin’s corresponding beliefs don’t have the same truth-values as his own.⁵¹

⁵¹ Our own argument, in contrast, isn’t question-begging. For as we pointed out in Section 5.1, the phenomenal concept strategy is only intended as a defense of physicalism against anti-physicalist

Chalmers is very likely to adopt a rather different tactic, however: he will argue that the zombie's phenomenal statements are false, not because they refer to phenomenal states that he doesn't have, but because they fail to refer altogether. The right analogy isn't between Earth and Twin Earth but rather between Earth and Dry Earth. Dry Oscar's claims about water (e.g., that it is refreshing) are false because he is subject to some sort of grand illusion: there is no such thing as water in his environment. If this is the right analogy then we would have to grant Chalmers that the epistemic situation of zombies isn't, as a matter of fact, the same as ours. But we have two responses to make to this argument. One is to deny that this *is* the right analogy. The other is to say that even if it is, we can still run a version of the zombie-zombie argument. Let us elaborate.

How *could* Zombie Chalmers' phenomenal concepts fail to refer? For these are concepts that, on their third-person characterization, are applied in a recognitional way in the presence of content-bearing mental states of a distinctive sort (perceptual and imagistic states). How could these concepts fail to refer to the very states that prompt their application? One option would be to claim that there is something *else* built into their content. For example, as Chalmers once suggested (1996, p. 204), they might include the commitment that they should *not* refer to any physical or functional property. But this would be inconsistent with the claim that phenomenal concepts are conceptually isolated. Concepts that are so isolated must lack any commitments of this sort.

Another option would be to claim that the presence of phenomenal

arguments, not as an independent argument in support of physicalism, nor as a purported reductive explanation of phenomenal consciousness itself.

consciousness is a *constitutive* aspect of the *content* of a phenomenal concept. In which case Zombie Chalmers' "thoughts" involving phenomenal concepts will be either false or truth-valueless because employing a contentless concept. (Chalmers develops such a position at length in his 2003.) But this option is entirely question-begging in the present context. Chalmers (2003) develops his account of the content of phenomenal concepts within the framework of his own anti-physicalist position, assuming that there are irreducible qualia and such like. But that position is supposed to be established on the basis of arguments from the conceivability of zombies and so forth, and hence cannot be taken for granted in the evaluation of those arguments. Moreover the horn of Chalmers' dilemma we have been addressing for most of the paper (sections 4 and on) presupposes the third-person characterization of phenomenal concepts. And given such a characterization, there is no reason whatever to think that the thoughts of Zombie Chalmers, employing such a concept, should be empty.

Even if we allow that Zombie Chalmers' phenomenal concepts might fail to refer, however, we can still run a version of the zombie-zombie argument. For we surely need to explain the inferences that the zombie makes, and the reasons why he thinks (granted, mistakenly) that he can conceive of a zombie version of himself. The fact that the zombie's beliefs are false (because containing an empty term) doesn't mean we are under no obligation to explain his reasoning and his behavior. We can explain why it is that little John wants to be nice by appealing, in part, to his (false) belief that Santa will only give him presents if he is nice. Although his concept SANTA fails to refer, it still plays a role in his reasoning and behavior. What, then, explains

the zombie's reasoning and behavior? Clearly, the presence of phenomenal feels can't explain that reasoning. Just as in the case in which we assume that the zombie's phenomenal concepts refer to physical states, so in the case in which his concepts are empty, his reasoning can't be explained by an appeal to phenomenal *states*. The only thing that can truly explain the relevant bits of reasoning is the fact that Zombie Chalmers has a concept (in the original case, referring to a physical property, now being allowed to be empty) which is conceptually isolated from all physical, functional, and intentional concepts.

What emerges, then, is that the zombie-zombie argument can still work even if we allow that Chalmers and Zombie Chalmers don't share the same epistemic situation (because all of the latter's beliefs involving phenomenal concepts are false by virtue of failing to refer). Since it is the conceptual isolation of Zombie Chalmers' (empty) phenomenal concepts that explains the conceivability to him of Zombie-Zombie Chalmers and so forth, parity of reasoning suggests that in Chalmers' case, too, it is the conceptual isolation of his phenomenal concepts and *not* the presence of phenomenal consciousness itself which explains the various problematic thought experiments. We want to emphasize, however, that we are actually very unwilling to allow that the corresponding beliefs of Chalmers and Zombie Chalmers should differ in truth value. We think that it is *much* more plausible that Zombie Chalmers' phenomenal concepts should refer successfully to his schmenomenal states.

4.2.7. Concluding

It is worth noting in closing that there is both a weaker and a stronger conclusion that might be drawn from our defense of the phenomenal concept strategy. The weaker conclusion is that the arguments from zombies, from the explanatory gap, and so forth, to the mysterious and/or non-physical nature of phenomenal consciousness is decisively blocked. For everyone can agree that our phenomenal concepts fit some or other variant of the third-person descriptions canvassed in Section 2. Everyone can agree that it is possible for us to form concepts of experience that are purely recognitional, or that “quote” percepts or images, or whatever. What they will disagree about is whether our phenomenal concepts are *exhausted by* such factors. Anti-physicalists will insist that something has been left out, namely that those concepts pick out non-relational, non-intentional properties of experience like *these*. So if the zombie and explanatory gap thought experiments can be fully explained in terms of our possession of phenomenal concepts, then there is no longer any *argument* from those thought experiments to the existence of qualia, the mysteriousness of consciousness, property dualism, and so forth. Such claims might still be correct, but the arguments for them have collapsed.

The stronger conclusion that might be drawn from our discussion is this. Once we see that all the puzzling factors can be explained in terms of our deployment of phenomenal concepts; and perhaps especially once we see in those terms that even the conceived-of zombies will be able to conceive of zombie versions of themselves, then the most plausible conclusion to draw overall is that there is nothing *more* to our

phenomenal concepts than is described in the third-person description. (Remember, however, that the third-person description is *not* supposed to be any sort of analysis or partial definition of our phenomenal concepts.) So the most reasonable conclusion is that a phenomenal state just *is* a perceptual state with a certain distinctive sort of intentional content (non-conceptual, perhaps) that occurs in such a way as to ground the application of phenomenal concepts. Hence we can conclude that phenomenal consciousness can be fully reductively explained (somehow – of course there are a number of mutually inconsistent competing accounts, here)⁵² in physical, functional, and/or intentional terms.

We have provided a number of reasons for thinking that Chalmers' argument against the phenomenal concept strategy is unsuccessful. On the contrary, that strategy still stands as providing a powerful response to a wide range of anti-physicalist thought-experiments, enabling us to draw the anti-physicalist sting from the latter.

⁵² One of the issues outstanding will concern the selection of the *best* third-person description of the nature and role of phenomenal concepts. About this matter we have said nothing.

Part II – Concepts and Experience

1. Introduction

According to representationalists, phenomenal characters can be reduced to representational contents. It is no wonder, then, that spelling out a representationalist account involves saying quite a bit about these representational contents, their roles, as well as their relations to other contents and to the external world. After all, if the mind itself is representational, virtually every mental state will involve a relation to a representation and its content; but presumably not every mental state will be phenomenally conscious. The states that *are* phenomenally conscious, like perceptual experiences, must therefore either have contents *unlike* the contents of other states (including subpersonal states and propositional attitude states); or the role that perceptual representations play in cognition must be different from the role played by the representations involved in propositional attitudes (remember that representationalists are for the most part *functionalist*-representationalists); or a bit of both.

Concepts play a central role in the representationalist's attempt to cash out the difference between the representational contents to which phenomenal characters are reducible (i.e., experiential contents) and the representational contents of propositional attitudes (most importantly, belief). Again this should come as no surprise. As mentioned in the introduction, concepts are (pre-theoretically) the

constituents of beliefs. Spelling out the difference between experiential contents and belief contents should therefore require saying something about the relation between experiential contents and concepts. Very generally, conceptualists about experiential contents will claim that the contents of experience are related to concepts in the way that belief is related to concepts. That is, experiential content, for the conceptualist, is importantly similar to belief content. Nonconceptualists, on the other hand, deny that experiential contents are related to concepts in the way that belief contents are: experiential contents, for the nonconceptualist, are importantly different from belief contents.

Ultimately, it also looks as though the relation between phenomenal contents and concepts will have an impact on what the representationalist can and should say about the relationship between experiential contents and the external world. Seeing why that is requires a detour through Inverted Earth.

1.1. Inverted Earth detour

Block (1990) presents a well-known argument against representationalist accounts of phenomenal characters. The argument requires that we imagine a duplicate of our world in which each thing is identical to each thing here on Earth except for its color, which is inverted. Every object on this Inverted Earth is the complementary color of its counterpart object here on Earth, such that ripe bananas on Inverted Earth are blue (not yellow), grass is red (not green), etc. We are asked to imagine an Earthling subject, let it be Sara, who is whisked away, unbeknownst to her, to Inverted Earth and equipped, while she travels, with a pair of inverting contact lenses. Here on Earth these lenses would make each thing look its complementary color: bananas blue,

grass red etc. On Inverted Earth too they have this effect; however, since every thing there is the complementary of every thing here, the lenses, when worn by Sara on Inverted Earth, make every thing there look just as it would here: bananas yellow, grass green, etc. As a result, Sara, upon waking, notices nothing and lives out her life on Inverted Earth.

Block makes the following two claims: 1) the phenomenal character or feel of Sara's Inverted visual experience of a blue VW bug will stay the same as the phenomenal character of her visual experience of the blue bug on Earth; but 2) the *phenomenal content* of her Inverted blue bug experience will eventually switch from representing blueness (as it did on Earth) to representing yellowness (which is the *real* color of the bug on Inverted Earth). At some time *t* in her Inverted life, then, the phenomenal character of her bug experience will have the same phenomenal character as her Earth bug experience, though the experiences will *not* both have the same phenomenal content. Hence, we have sameness in phenomenal character without sameness in phenomenal content, and Block concludes that reductive representationalism must be false.

- 1) The phenomenal character of Sara's VW bug experience on Inverted Earth at *t* is the same as the phenomenal character of the experience she had of the counterpart bug on Earth.
- 2) The phenomenal *content* of Sara's VW bug experience on Inverted Earth at *t* is *not* the same as the phenomenal content of her Earth (counterpart) bug experience.
- 3) Reductive representationalist must be false.

A representationalist may attempt to resist Block's conclusion by denying either of the argument's two premises. The first—according to which the phenomenal character of Sara's VW bug experience on Earth and on Inverted Earth always

remains the same—seems strongly intuitive. There is little reason to think that our subject’s visual system would undergo any major changes during her lifetime. If the inverted lenses continue to function throughout that lifetime, they will keep on transforming visual information in such a way that objects will keep looking to her as though they are the complementary of their “actual” colors. And, assuming that on Inverted Earth the bug remains, throughout her lifetime, the same color—i.e., the complementary color of its Earth counterpart—it seems as though the bug will keep looking to her like its counterpart on Earth would. However, since our intuitions sometimes get it wrong, some representationalists have toyed with the idea of arguing that we get it wrong here (see Tye 2000, 6.2 for an attempt at this sort of reply).

It has seemed much easier, however, for representationalists to deny premise 2), i.e., to deny that the content of Sara’s Inverted bug experience will eventually change. Block himself is quite willing to admit that premise 2) is true only if externalism about experiential content is true—more especially a particular kind of “causal” externalism, according to which the content of an experience depends on what *causes* it. If what *causes* Sara’s bug experience on Inverted Earth is a *yellow* bug, then a causal externalist might claim that her experience will represent yellowness. But, and this is the worrisome bit, Inverted Earth remains a problem for the representationalist even if we deny that (causal) externalism about phenomenal content is true.

Block makes the claim that the contents of our subject’s color *concepts* would, like the content of her experience, shift after a while spent in her new environment—the content of her concept BLUE, for instance, would shift from [blue] to [yellow] and

so the content of any belief taking the concept BLUE as a constituent would shift along with it. Representationalists usually grant this, though they may not think it affects them much. After all, a shift in the content of a subject's belief needn't entail a shift in her experiential content, and it is the latter that the representationist is most concerned with. But clearly, if experiential contents are (even partially) conceptual—if they take, say, some concepts as constituents—then a shift in the contents of Sara's *concepts* could result in a shift in her experiential contents. In fact, if phenomenal contents are conceptual, resisting the conclusion of Block's argument will require that the representationalist argue not just that externalism is false *of experiential content* but that it is false of our *concepts* as well. Here are the live options for the representationalist:

- (1) She may argue that the content of experience is narrow *and* experience is wholly nonconceptual.
- (2) She may argue that the content of experience is narrow, that experience is wholly (or partly) conceptual, *and* that the content of concepts is (at least partly) narrow.

The first strategy is, in some ways easier: it requires arguing for two claims only. It is the purpose of this chapter to determine not which looks easier (i.e., requires arguing for fewer claims), but which is most plausible, by spelling out what the relation between concepts and experiential contents might be.

In any case, it should be clear that spelling out the relation between experiential contents and concepts is very important for the representationalist. Not merely because the representationalist needs to say something about the difference between experiential contents and other contents (like belief-contents), but also because it may impact the way in which she deals with Block's Inverted Earth.

1.2. Plan of Part II

We'll begin the next section with a discussion of the debate between conceptualists and nonconceptualists—separating out *possession* conceptualists/nonconceptualists and *constituent* conceptualists/nonconceptualists. The distinction is crucial to the debate (see Byrne 2004, Bermudez and Cahen 2008, Speaks 2005). We'll spend the rest of Part II showing that partial conceptualism is the best alternative by first considering the best arguments for nonconceptualism (section 3). I argue that they fail (though maybe not entirely). Though they make the case that conceptualism is false, they do not rule out the third alternative: partial conceptualism. In section 4, we'll consider the best arguments against nonconceptualist accounts and present an argument from concept acquisition which I think succeeds. We'll conclude with a discussion of partial conceptualism (section 5).

2. The conceptual/nonconceptual debate

The conceptual/nonconceptual debate is complex and multifaceted, yet this much seems relatively clear: the central disagreement between conceptualists and nonconceptualists is one about how similar experience and belief are. All proponents in the debate agree that experience, like belief, is a representational state involving a relation to a contentful mental representation. But there is wide disagreement about

what, beyond this basic structure, experience and propositional attitudes have in common, especially when it comes to their relation to concepts.⁵³

As mentioned earlier, conceptualists are those who think that experiences and beliefs are importantly similar; nonconceptualists deny that claim. It follows that taking a stand in this debate—as either a conceptualist or a nonconceptualist about experience—requires that one have, operative in the background, an account of *belief* and *its* relation to concepts. After all, what account one has of experience is defined in contrast to what one might say about belief: experience is either *similar to* or *different from* belief. Insofar as there is disagreement about *belief* and its relation to concepts, there is room for disagreement about what shape the conceptual/nonconceptual debate should take. Since at least *two* different relations between *beliefs* and concepts have figured prominently in the literature, there are at least two conceptual/nonconceptual debates. First, the debate between *possession* conceptualists and possession nonconceptualists; second, the debate between *constituent* conceptualists and constituent nonconceptualists.

2.1 *Possession* conceptualism

Here is the first proposal about the relation between beliefs and concepts: for a thinker to be able to have a particular belief (say, the belief that there is a banana on the table), she must *possess* the relevant concepts (at the very least, the concepts

⁵³ I take it that one can argue for conceptualism (or nonconceptualism) about states other than perceptual experiences—most especially about representational states in subpersonal systems (see Bermudez 2003). I want to make it clear, then, that when I talk about conceptualists (or nonconceptualists) I mean only conceptualists (or nonconceptualists) *about perceptual experience*.

BANANA and TABLE). This proposal about belief focuses on the *conditions* on belief: for a state to be a *belief* that there is a banana on the table, the believer herself must *possess* the relevant concepts. With this in mind, here is what we can say about *experience*.

Possession conceptualists will say that the same is true of experience. In other words, the conceptualist claims that the conditions on experience are the same as the conditions on belief, such that for a perceiver to have a particular experience—to see that there is banana on the table—she must possess the relevant concepts BANANA and TABLE. If it turns out that Sara does not possess the concept BANANA, then her experience would be more accurately described as her seeing that there is a yellow object on the table. And if she doesn't possess the concept YELLOW, the conceptualist would want to say that Sara sees that there is some colored object on the table. In characterizing what Sara perceives, we should be sensitive to how she apprehends the world. And the way she apprehends the world, the conceptualist adds, is a function of the concepts she possesses (see Bermudez and Cahen 2008).

Possession nonconceptualists, however, maintain that Sara can be adequately described as seeing that there is a banana on the table, even if she does *not* possess the concept BANANA. Though possession nonconceptualists, like possession conceptualists, believe that in characterizing what Sara perceives we should be sensitive to the way in which she apprehends the world, they believe that the way she apprehends the world is *not* a function of the concepts she possesses. Tye and Crane are paradigm possession *nonconceptualists*. For Tye, to say that “a mental content is nonconceptual is to say that its subject need not possess any of the concepts that we,

as theorists, exercise when we state the correctness conditions for that content” (Tye 2000, 62). And Crane argues that a person “X is in a state with nonconceptual content iff X does not have to possess the concepts that characterize its content in order to be in that state“ (Crane 1992, 149).

2.2. *Constituent* conceptualism

The second proposal about the relation between concepts and beliefs is more directly one about what the *constituents* of the belief states are. This proposal can be spelt out at two different levels depending on yet another variable; namely, on one’s account of what concepts are. There are those (they might constitute a majority, as Byrne (2004) claims) who take concepts to be abstract constituents of propositions, and those (still prominent enough, see for instance Tye (1995, 2000) and Dretske (1981, 1995)) who take concepts to be mental representations of some particular sort. These views were mentioned briefly in the introduction—as the *semantic* view of concepts and the *psychological* view of concepts (Laurence and Margolis 2007). We’ll now say a little more about them.

2.2.1. *The psychological and the semantic view of concepts*

The psychological view

We mentioned the psychological view of concepts briefly in the introduction. According to it (Laurence and Margolis 2007), concepts are mental representations, which can be combined in various ways to form more complex representations. The concept GREEN and the concept GRASS can be combined to form the more complex

representation GRASS IS GREEN—a representation I am belief-related to if I believe that grass is green. This view of concepts fits in nicely within the computational/representational theory of mind and this, combined with the fact that it can be used to account for the productivity of thought, “provides considerable motivation for adopting the psychological view of concepts” (Laurence and Margolis 2007, 5).

Laurence and Margolis go on to defend this psychological view, but surprisingly, they say little about *what* makes a mental representation a concept. Presumably not all mental representations are concepts, even if all concepts are mental representations. At the very least, we will say, concepts are those mental representations that are the constituents of our propositional attitudes (like beliefs). Many claim concepts are mental representations that are *stored in (long-term) memory* (see also Tye 1995, 2000; Carruthers 2000, Machery 2005, Prinz 2007). Or that they are the mental representations used in the “higher cognitive processes (categorization, inductive and deductive reasoning..., etc.)” (Machery 2005, 444).

To say that concepts are the constituents of belief, on the psychological view, then, is to say that the mental representations we are belief-related to are the right kind of mental representations (the ones that are stored in memory, or used in higher cognitive processes).

The semantic view

According to the semantic view, concepts are abstract constituents of abstract objects, i.e., propositions. To believe that grass is green is to be related to a mental

representation with the content [grass is green]. And one way to think of these contents is in terms of abstract objects (propositions), which are bearers of truth-value (they are either true or false) and mind-independent. Some go on to say (following in the footsteps of Frege; see Peacocke 1992, Zalta 2001) that these propositions are structured and, like sentences, are composed of more “basic” constituents. These more basic constituents are concepts. In line with the literature, we will call these abstract constituents *Fregean senses*.

To say that beliefs take concepts as constituents, on the semantic view, means that the content of belief-representation will be a Fregean proposition—one that takes Fregean senses as constituents. With this in mind, some will say the following about *experience*.

2.2.2. *Constituent conceptualists vs. constituent nonconceptualists*

Constituent conceptualists will claim that the content of experience, like that of belief, is a Fregean proposition; or she may think that the mental representations a subject is experientially related to are composed of stored representations. Constituent *nonconceptualists*, on the other hand, will argue that experience is such that it takes radically different kinds of constituents. Experience may take as constituents actual objects or properties (as on the Russellian view) or possible worlds instead Fregean senses; or experience may take as constituent mental representations that do not play a role in reasoning instead of mental representation that play such a role.

The debate between constituent nonconceptualists and constituent conceptualists then is a debate about the *kind* of content (or the kind of mental

representation) involved in believing and experiencing. Constituent conceptualists argue that the contents of experience and belief are of the same kind such that the content of an experience could, in principle, be the content of a thought. Brewer, for instance, says: “a mental state with conceptual content [...] is one whose content is the content of a possible judgment by the subject” (Brewer 2005, 217). Or to put it in terms of representations: constituent conceptualists argue that the representations subjects are related to when they experience (experience-representations for short) could be belief-representations. Constituent *nonconceptualists*, by contrast, argue that the content, or mental representations, involved in thinking and experiencing are of a different kind; the content of an experience could *not* be the content of a thought; an experience representation could *not* be a belief-representation.

2.3. Differences

It is important to note that these two debates are different debates, and the positions they carve out are different positions.

2.3.1. *The possession nonconceptualist*

For instance, it is possible for one to be a *possession nonconceptualist* while being either an *constituent nonconceptualist* or an *constituent conceptualist*. To see this, let us assume, along with the *possession nonconceptualist*, that our subject Sara can see that there is a banana on the table, even when she doesn't possess the concept BANANA, or the concept TABLE, or any other relevant concept like the concept OBJECT, YELLOW, etc. To say that Sara does not possess these concepts, on the consensus

view, is to say that Sara cannot have *beliefs* in which these concepts figure as constituents. This view (possession nonconceptualism) is, naturally, compatible with constituent nonconceptualism: the fact that Sara cannot have thoughts in which the concept BANANA figures is perfectly compatible with Sara's *experience* being such that concepts cannot figure in it at all. But, more importantly, possession nonconceptualism is compatible with constituent *conceptualism*. The fact that Sara cannot have *beliefs* in which the concept BANANA figures is perfectly compatible with Sara's having *experiences* in which some concepts—maybe even the concept BANANA—figure. Concerns about which concepts a thinker possesses are concerns about which particular concepts can figure into that thinker's *beliefs*. And the fact that *particular* concepts (i.e., BANANA or TABLE) can or cannot figure in a thinker's *beliefs* does not entail anything about which particular concepts can or cannot figure in the thinker's *experience*. Nor does it entail that no concepts *whatever* can figure in a thinker's experience. Of course, this position is a strange one to occupy. Again, on this view, Sara might not be able to think *that there is a banana on the table*—the concept BANANA might not be able to figure in her beliefs. Yet the concept BANANA could figure in her experience. However odd the position, it is important to notice that it *is* an available, consistent option.

Now it may be that possession nonconceptualists take themselves to be constituent nonconceptualists as well. For instance, they may assume that

(A) if a particular concept cannot figure into a subject's beliefs, then that concept cannot figure into the subject's experience either.

So, possession nonconceptualists believe that a subject can see that there is a banana on the table without possessing (i.e., being able to think with) *any* of the relevant

concepts (BANANA, TABLE, YELLOW, FRUIT, OBJECT, etc.). But if (A) is true then, none of these relevant concepts would be able to figure in the subject’s experience either. If, however, no relevant concept can figure in a subject’s experience of a banana, it seems crazy to insist that experience still takes concepts as constituents—after all, what concepts would those be? Concepts that are *irrelevant* to the current experience, say the concept DEMORACY or ROBOT? Experience, we would have to say, is such that, whatever its constituents are, they aren’t concepts. In other words, it is *absolutely* nonconceptual.

Possession nonconceptualism *and* assumption (A) might entail constituent nonconceptualism, but—and this is what matters here— possession nonconceptualism *alone* does not entail constituent nonconceptualism. And some writers (Byrne 2004, for instance) simply deny assumption (A). He concludes that even if a particular concept cannot figure into Sara’s thought, it is possible for that very concept to figure in her experience. It is possible, then, to be a possession nonconceptualist *and* a constituent conceptualist.

	Constituent Conceptualism	Constituent Nonconceptualism
Possession Nonconceptualism	√	√

Figure II.1 Possession nonconceptualism

2.3.2. *The possession conceptualist*

Possession conceptualism vs. constituent (non)conceptualism

Possession conceptualism might be compatible with either constituent conceptualism or constituent nonconceptualism. To see *this*, assume along with the possession

conceptualist that subjects must possess (i.e., be able to think with) the concept BANANA to be able to see that there is a banana on the table. And notice that, by itself, the possession conceptualist thesis tells us nothing about whether or not it can take concepts as constituents. This is because the thesis doesn't make any claims about *why* it is that a subject cannot see that there is a banana unless she possesses the concept BANANA.

We may believe, for instance, that the possession conceptualist thesis holds because, unless a concept *C* that stand for some thing *T* can figure in thought, one cannot bear the *experience relation* to another (even nonconceptual) representation of *T*. In other words, we might think that, if Sara cannot think with the concept BANANA (if she cannot be “belief-related” to a complex mental representation of which BANANA is a constituent), then she cannot be *experience-related* to any complex representation which has as a constituent some representation of bananas, even if that representation is not the concept BANANA. I'm not sure why anyone would think that this is true. Still, the fact that it is available makes it such that possession conceptualism is *compatible with* constituent *nonconceptualism*.

Of course, we could believe, rather more plausibly, that the *possession* conceptualist thesis holds because, unless a concept (like BANANA) can figure in belief, it cannot figure in experience. And if a concept for some thing cannot figure in experience, then a subject cannot see *that* there is that thing. In other words, we may believe that if Sara cannot think with the concept BANANA, then that concept cannot figure in her experience (notice that this is assumption (A) all over again). And if the concept BANANA cannot figure into Sara's experience, then (since we can reason this

way about the other concepts relevant to Sara’s thought) it follows that no relevant concept can be a constituent of Sara’s thought. This particular elaboration of the possession conceptualist view (as conjoined with (A)) entails constituent conceptualism: seeing that there is an X right there requires that a concept for that X figure into the perceiver’s *experience* of it. And this entails that concepts *can* be constituents of experience, which the constituent conceptualist believes is true.

	Constituent Conceptualism	Constituent Nonconceptualism
Possession Nonconceptualism	√	√
Possession Conceptualism	√	√

Figure II.2. Possession conceptualism

Concept deployment: a complication

We have assumed, so far, that possession conceptualism says the following: a subject can *see* that there is a banana on the table, only if she possesses (i.e., can think with) *the concept* BANANA. And this claim is made within the background of a claim about *belief*; namely, that a subject can *believe* that there is a banana on the table only if she possesses the concept BANANA. However, some possession conceptualists claim somewhat more than that *about belief*—as a result, what they say about experience is different too.

Here is what some possession conceptualists think is true of belief: a subject can *believe* that there is a banana on the table, only if she 1) possess the concepts BANANA and TABLE and she 2) *deploys* those concepts in thought. The possession conceptualist concludes that experience is like belief in the following way: Sara can

see that there is a banana on the table, only if Sara 1) possess the concepts BANANA and TABLE and she 2) deploys those concepts *in experience* (McDowell 1994, Brewer 1999, and Sedivy 1996 are thought to be such possession conceptualists; see Siegel 2005).

What is it to *deploy* a concept, whether it be in thought or in experience? I take it that it requires that the concept in question actually *figure in* a belief or experience. Possessing a concept requires only that the concept *be able to* figure in someone's thought—to say that Sara possesses the concept BANANA is to say that the concept BANANA *can* figure in her thought, that it has the capacity to figure in her thought. Deploying a concept requires more than this ability, it requires that the concept *actually* figure in a given thought. To say that Sara is deploying the concept BANANA in a thought is to say that the concept BANANA *does actually* figure in that thought.

Understood this way, possession conceptualism is incompatible with constituent *nonconceptualism*. After all, possession conceptualism now claims that for Sara to see that there is a banana on the table, the concept BANANA 1) must be capable of figuring in Sara's *thought* and 2) must *actually* figure in her *experience*. Sara's experience, then, has to be such that it *can* take concepts as constituents. But that is what the constituent *nonconceptualist* denies.

	Constituent Conceptualism	Constituent Nonconceptualism
Possession Nonconceptualism	√	√
Possession Conceptualism (possession only)	√	√
Possession Conceptualism (deployment)	√	X

Figure II.3. Possession *and* deployment conceptualism

2.4. Partial Conceptualism

The rest of Part II will focus, for the most part, on the debate between *constituent* conceptualists and *constituent* nonconceptualists—in what follows, then, ‘conceptualist’ and ‘nonconceptualist’ will mean *constituent* conceptualist and *constituent* nonconceptualist. Most discussions of the relation between concepts and experience have been between *possession* conceptualists and *possession* nonconceptualists—something which even those leading the discussions failed to realize until the last few years. Tye, for instance, wants to argue that experience has a different *kind* of content from belief—something *constituent* nonconceptualist believe—but his arguments for the conclusion have been, as he himself acknowledged (see Tye 2005), arguments for *possession* nonconceptualism.

It will be more important for the purposes of the representationalist to say something about whether or not concepts *can be* the constituents of experience. As mentioned in the introduction, the Inverted Earth thought-experiment will affect the representationalist differently depending on what she says about concepts as constituents of experience (i.e., depending on the upshot of the *constituent* debate).

What has been said so far may have given the reader the impression that the conceptual/nonconceptual debate is a debate between *two* rival views of experience. This impression is misleading: there is a third option available, and it is the option I favor and will defend in Part II. We'll call this third alternative *partial conceptualism* though it could just as well be called partial nonconceptualism. Like the other two accounts, partial conceptualism about experience can be construed either at the level of representations or at the level of content. We can say, then, that according to partial conceptualism, the content of experience is a hybrid content, one that can take two different forms. In the first form, the content of experience includes at least two *kinds* of propositions: a Fregean proposition and a proposition that isn't Fregean, in which case we can think of the content of experience as layered, with a Fregean layer and another—say Russellian—layer. Or, if one wants to deny that there are independent “layers” of experience in this way, we can think of the content of an experience as one proposition with two kinds of constituents: some Fregean constituents and some other kinds of constituents.⁵⁴ Alternatively, we can say that, according to partial conceptualism, the mental representations subjects are experientially related to take two kinds of constituents: first, those “basic” mental representations that are stored and processed in the right way (i.e., concepts); second, other mental representations that can't be constituents of thoughts.

Interestingly enough, the literature itself often makes the conceptual/nonconceptual debate look as though there are only two live options. As a result, arguments *for* nonconceptualism are often merely arguments *against*

⁵⁴ Jeff Speaks suggests a view like the latter—though focused around internalism and externalism. He denies that we need to think of a “level” of narrow content, imagining instead a content with internalist *and* externalist constituents (Speaks, forthcoming).

conceptual accounts, and they fail to rule out partial conceptualism.⁵⁵ Similarly, arguments for conceptualism are often arguments against nonconceptualist accounts, and they too fail to rule out partial conceptualism.⁵⁶ Byrne (2005) adds that conceptualism should be the default position in this debate. It is not obvious why that would be: most may agree that thought is conceptual, but it is hard to see, without argument, why the reasons for believing thought to be conceptual necessarily apply to experience. Byrne is right, of course, when he says that “all parties agree, in effect, that perceiving is very much like a traditional propositional attitude, such as believing or intending” (23). We have acknowledged the fact that proponents in the debate all agree that thinking *and* experiencing involve 1) a relation to 2) a mental representation 3) with a certain content. The meaty question, however, is what, beyond this structure, experience and thought share. And I fail to see how the mere fact that they share this representational structure gives us a reason to think that the mental representations involved, or the contents involved, are of the same *kind*.

In any case, now that we have a better grip on the relevant distinctions, let’s take a look at the arguments *for* nonconceptualism.

⁵⁵ This is true of the most famous nonconceptualist argument—the argument from richness and fineness of grain (see Evans 1982, Dretske 1981, Tye 1995, 2000, Heck 2000, Kelly 2001)—but also of the argument from situation-dependence (Kelly), and the argument from continuity (Peacocke 2001).

⁵⁶ This is true of epistemological arguments against nonconceptualism (McDowell 1995, Brewer 1999, 2003).

3. The case for nonconceptualism

As mentioned earlier, arguments presented in support of nonconceptualism are, for the most part, simply arguments *against* conceptualism. In this section, we'll look at three strands of argument for nonconceptualism (or against conceptualism): first, the least consequential of these arguments, i.e., the argument from continuity (3.1); then an argument from systematicity (3.2.); and finally the argument from richness and fineness of grain (3.3).

This last argument (section 3.3) is complex. An elaboration of it requires discussing some of the important moves made in response to the initial arguments by conceptualists, and the replies offered, in turn, by nonconceptualists. Those nonconceptualist replies amount to, naturally, more arguments against conceptualism. And though they could be discussed separately (as in Speaks 2005), they are best understood when located within the larger discussion of richness and fineness of grain. In any case, the upshot is that section 3.3, though it presents one *type* of argument against conceptualism, really sets out several arguments against it.

3.1. Continuity

The argument from continuity is perhaps the least convincing argument against conceptualism, but it is worth rehearsing. It starts with a claim that is intuitive enough; namely, that 1) non-human animals (at the very least primates, but including possibly other mammals) have perceptual systems not altogether unlike ours. It follows that 2) their perceptual states are not altogether unlike ours either—we might

say it is plausible that some of the representations (or contents) that make up *our* perceptual states are shared by non-human animals. However, the argument continues, 3) non-human animals do not possess any concepts. This leads to 4) the content of their perceptual states must be nonconceptual, then, since they are contents that can be had even when one does not possess any corresponding concept. It follows that 5) some of the representations (contents) that make up *our* perceptual states (those we have in common with non-human animals) are nonconceptual.

There are several things to note here. First, the argument's conclusion is explicitly *not* that human perceptual states are wholly nonconceptual. At most, the argument, if it succeeds, shows that human perceptual states are partially nonconceptual. Peacocke himself writes: "It follows that *some* perceptual representational content is nonconceptual" (2001, 613-4). Second, the argument as presented here (and in the literature) is an argument against *possession* conceptualism: its conclusion is that it is possible to have perceptual representations (contents) without possessing any of the relevant concepts, as is the case (according to premise 3)) with non-human animals and infants. But the mere fact that possession conceptualism is false (in the case of animal's perceptual states) does not entail that *constituent* conceptualism is also false. (In section 2.3.1. we saw that possession *nonconceptualism* is compatible with constituent conceptualism.) In fact, it is part of Byrne's *constituent* conceptualist strategy, in this case, to grant that possession nonconceptualism is true for animals; to grant, that is, that animals *can* see that there is a banana on the table without possessing the concepts BANANA and TABLE; and to

argue that the perceptual states of animals could be made up of concepts nonetheless (as the constituent conceptualism claims they are).

One might reasonably hold [that animals don't possess concepts] together with the view that perceptual content, in humans and lower animals, is the same kind of content that can be believed—hence denying [the argument's conclusion] (Byrne 2005, 10).

Unfortunately, in some respects, Byrne's position is slightly odder than he himself has realized. For the argument from continuity takes as a premise the claim that animals do not possess any concepts. Which presumably means that animals cannot *think* with any concepts, and therefore that their "belief" states (whatever those are) cannot take concepts as constituents. It is rather strange to maintain, as Byrne would have to if he grants the premises of the continuity argument (as he seems to), that the perceptual states of animals take concepts as constituents while their belief states do not.

In any case, the claim gets a bit stranger, I think, if we remember that to claim that a perceptual state is conceptual or nonconceptual is to contrast it with *belief*. But presumably we shouldn't contrast perceptual states with *anyone's* belief. Human perceptual states are nonconceptual if they turn out to be unlike *human* belief—conceptual if they are *like* human belief. So we might think that the perceptual states of non-human animals are *nonconceptual* if they turn out to be unlike *animal* belief—and conceptual if they turn out to be *like* animal belief. And according to premise 3), animal *belief* doesn't take concepts as constituents. So if animal perceptual states *do* take concepts as constituents, then their perceptual states are, it turns out, *unlike* their beliefs (they take different kinds of constituents), and hence their perceptual states are, technically, nonconceptual (even if their constituents are all concepts!).

Of course, this is merely a technicality, and it shouldn't worry Byrne at all. For as long as perceptual content in animals is the kind of content that can be believed *by a human being*, the conclusion of the argument from continuity can be blocked. For it can then be true that humans and animals share part of their perceptual contents, and that they don't possess any concepts—even while it is false to conclude that the content that we share is *nonconceptual*.

In any case, we now come to another worry about this argument: why assume that non-human animal thought is so drastically unlike human thought, as premise 3) would have us do? Though the first and second premise of the argument (that non-human animals have perceptual systems and states at least in part like ours) are widely believed to be true, the same cannot be said of the third premise. The claim that animal thought is drastically unlike human thought—if it should count as thought at all—is controversial enough to raise doubts about the success of the argument. Moreover, if we have reason to believe that animal thought is so unlike human thought that it doesn't deserve to be called thought—maybe we should call it *proto-thought* (see Byrne)—, why think that animal *perception* is so much like human perception that they share some content? “If lower animals merely proto-think”, Byrne asks, “why don't they merely proto-perceive?” (10)

3.2. Systematicity

Fodor (1975, 1988, 1990) famously noted that human thought is productive and systematic. This gives us reason, he argued, to make a claim about the kinds of representations involved in thinking. A nonconceptualist may want to deny, then, that

experience is systematic—denying thereby that we have reason to think that the kinds of representations involved in experiencing are the same kind of representations involved in thinking. This is, in essence, one of Tye’s arguments for nonconceptualism (see especially his 1995).

Productivity was mentioned in the introduction. Thought is productive because we seem able to entertain an infinite number of new thoughts, like the thought that purple giraffes take their time while bowling. But given that we are finite beings, there is no way that we could store an infinite number of representations. We can, however, explain productivity by positing a finite set of “simpler” representations and a system to combine and recombine them. Thought is also systematic in the following way: the ability to think certain thoughts will be systematically connected with the ability to think certain other thoughts. Consider the thought that Mary loves John. No native speaker of English will have the ability to think that Mary loves John without also having the ability to think that *John loves Mary*. We can generalize and say that having the ability to think any thought with content p entails having the ability to think thoughts with contents $L(p)$ where $L(p)$ are the logical permutations of p (Rey 1997). This ability is best explained, we might think, by an underlying sentential representational structure: thinking that Mary loves John involves the tokening of a complex mental representation with a simpler component “corresponding to” John, one “corresponding to” Mary, and one “corresponding to” the two-place predicate love. Thinking that John loves Mary involves the tokening of the same three representations, combined differently.⁵⁷

⁵⁷ Thoughts are also systematic in a somewhat different way, i.e., inferentially. Not only is the ability to have certain thoughts systematically connected with the ability to think other thoughts (as with

Tye argues that there are two reasons for thinking that the constituents of experience are unlike those of thought (and propositional attitudes more generally). First “there are no general systematic connections between pains of the sort found in thought” (1995, 120); second, unlike thought, “pain is not productive; we cannot generate endlessly many new kinds of pain in the way that we can generate endlessly many new thoughts” (ibid). This is not Tye’s only argument. For, he adds, scientific evidence supports an alternative view about the representational structure of experience. So, he says, “we know that in visual perception, the retinal image is reconstructed in the visual cortex, so that in quite a literal sense adjacent parts of the cortex represent adjacent parts of the retinal image.” What’s more, “topographic organization of this sort is also found in the somatosensory cortex. There is, for example, an orderly topographic representation of the surface of the human body that is dedicated to touch” (1995, 120). He concludes that perceptual representations are “patterns of active (or filled) cells occurring in topographically structured three-dimensional arrays or matrices” (1995, 121). (See also his 1991.) We can think of the activity in a given cell as “representing (in the manner of a simple symbol) that there is tissue damage at the body region to which the cell is dedicated” (ibid). The case can be generalized for color perception and, ultimately, to any experience. Tye therefore concludes that experiential representations are unlike belief representations: their *constituents* are quite different. The simple symbols that represent damage at a certain body region are *not* concepts—they aren’t *stored memory* representations.

“John loves Mary”), but the ability to make certain inferences is systematically connected to the ability to make certain others. No one can infer P from $P \& Q \& R$ without also being able to infer P from $P \& Q$ (see Fodor and McLaughlin 1990, Aydede 2004).

Tye's argument for nonconceptualism isn't very convincing. Though Tye claims that experience, unlike belief, isn't systematic, it is far from obvious that he's right. After all, there are different kinds of pain—diffuse pains, stabbing pains, pressing pains—and it may well be the case that the having the ability to experience a stabbing headache requires the ability to experience stabbing toothaches or other sorts of pains (assuming one can experience toothaches at all). This may amount to some sort of systematicity. And even if the case for systematicity is hard to make for pains, it seems quite easy to make for visual experiences (I take it the same applies to auditory and tactile experiences): a creature who can see a red square and a blue circle will in all likelihood also be able to see a blue square and a red circle. Some even try to make the case that bee representations are systematic (Tetzlaff and Rey, forthcoming). The experimental data seems to suggest, Tetzlaff and Rey argue, that bees have a number of systematic states, such that if they represent the hive as a certain angle x from the feeder, then they can represent the feeder as being at that very angle x from the hive. Now, these representations may not be “perceptual” representations of the bees; rather, they may be *thoughts*—this is how Tetzlaff and Rey think of them at least. Still, if it turns out that bees have systematic (thought-) representations despite the relative simplicity of their brains, then it seems plausible that our extremely more complex visual system would have such representations too.

3.3. Richness and fineness of grain

The argument from richness and fineness of grain is perhaps the best-known argument for nonconceptualism. Moreover, it is an argument whose central premise is

granted by both nonconceptualists *and* conceptualists. Experience is rich and fine-grained; that much isn't really contentious. The disagreement is rather a disagreement about whether conceptualists have the resources to account for the richness and fineness of grain of experience. What is it, then to say that experience is rich and fine-grained?

3.3.1. *Rich and fine-grained*

The claim that experience is rich isn't the claim that experience is as rich as we sometimes think it is. The latter claim is controversial; the former much less so. So, it may seem to us as though our experience is like a high resolution digital picture, as though, *every* thing before us—the banana, the table, etc.—is represented exactly, in sharp detail. Much data suggests that experience isn't quite *that* rich and *that* detailed: drastic changes made to scenes we are perceiving can go unnoticed if attention isn't directed the right way (see for instance Simons and Chabris 1999). If our experience really were so rich, we would, it seems, notice a difference immediately. But being skeptical about what some call the “picture-view” of experience (see Dennett 1991, O'Regan 1992, Noe and O'Regan (2002)) isn't quite being skeptical that experience is rich in the sense that matters for the conceptual/nonconceptual debate. For the claim here is simply that experience is richer *than thought*, not that it is as rich as a picture. The richness claim, then, is this: Sara's experience of a banana on a table is richer than her thought that there is a banana on the table. Her thought contains information only about the banana and its general location relative to the table. Her experience of a banana on the table, however, necessarily contains information about

more than just that: it necessarily contains information as to whether the banana is on the right side of the table, or on the left, or in the center; whether the banana is yellow, or spotted, or black or green, etc.⁵⁸ In that sense, then, experience is rich.

Experience is also considered to be *fine-grained*; that is, it represents precise and determinate properties, relations, etc., in a way that thought does not. Sara's experience represents the banana as being at an *exact* place on the table, not just generally on the left or on the right, but exactly *this* far from a particular edge. Her experience represents the color of the banana as being a determinate shade of yellow—say, yellow₁₇—and not just any yellow.

Richness and fineness of grain are quite different features of experience—though they are often (and for good reason, as we'll see) discussed together. It is worth noticing, for instance, that they could come apart: experience could be fine-grained without being rich—and vice versa. An experience of a uniformly yellow wall may be not be rich—at least not compared to most experiences we undergo. (Though it might be richer than the corresponding thought that the room is yellow, since it would represent the location of one's body with respect to the walls in the room, etc.) Still, the experience can be fine-grained, for it is a particular *shade* of yellow that the room is represented as having. Similarly, experience could be rich without being fine-grained, as when Sara's experience of the banana represents it as being on the left side of the table and green. Neither of those properties are specified

⁵⁸ This tracks Dretske's (1981) and Kelly's (2001) distinction between analog and digital representations: analog representations carry a lot more information than digital ones.

in the corresponding (simple) thought that the banana is on the table (so the experience is rich), but neither *need be*, in principle, fine-grained.⁵⁹

3.3.2. *The arguments: a first pass*

To say that experience is rich, then, is to say that it is richer than *thought*. To say that experience is fine-grained, is to say that it is fine-grained in a way that thought is not. Making both of these claims, then, involves drawing a *contrast* between experience and thought as *nonconceptualists* want to do. Here is one way to get from richness to nonconceptualism:

- (1) Experience is richer than thought.
- (2) Since experience is richer than thought, there must be times when experience represents a property for which a thinker possesses no concept.
- (3) Therefore, there must be times when experience is nonconceptual.

And here is how we get from fineness of grain to nonconceptualism:

- (1) Experience is more fine-grained than thought, i.e., it represents properties like yellow₁₇.
- (2) Often, subjects do not possess concepts that are fine-grained enough, concepts like YELLOW₁₇.
- (3) Therefore, experience sometimes represents properties for which a subject does not possess a concept, i.e., experience is sometimes nonconceptual.

First, we should notice that these arguments are not arguments for the conclusion that experience is *always* nonconceptual. At the most, these show that experience is *sometimes* nonconceptual. Second, like the continuity argument, these arguments are arguments against *possession* conceptualism, and explicitly so. The point is that

⁵⁹ It is not clear, however, that our experience can fail to be fine-grained at least when it comes to certain properties: representing something to be yellow, it would seem, requires that I represent it as being a certain *particular shade* of yellow. It may be that some experiences of shape can fail to be fine-grained, as might be the case when we're inebriated.

experience represents more properties than one may possess concepts for, or properties that are so fine-grained that one obviously doesn't possess concepts for them. If the arguments succeed, they show, at most, that *possession* conceptualism is false. This last point has driven those interested in the debate between *constituent* nonconceptualists and *constituent* nonconceptualists to quickly dismiss the richness and fineness of grain arguments. After all, it is possible, on the *constituent* conceptualist view, for a concept to figure in experience even if it cannot figure in thought (i.e., even if the subject in question does not *possess* it, that is, cannot think with it). So even if experience is so rich that it represents a property (say the property *being an antelope*) for which Sara doesn't possess a concept (she can't think with the concept ANTELOPE), nothing prevents that very concept from figuring in her experience, if not in her thought. Likewise, her experience might be so fine-grained that it represents yellow₁₇ even while Sara does not possess the concept YELLOW₁₇ (i.e., cannot think with the concept YELLOW₁₇). However, nothing prevents YELLOW₁₇ from figuring in Sara's *experience*, if not in her thought. For these reasons, Byrne and Speaks conclude that "the richness [/fineness of grain] of experience is not relevant to the question of whether the contents of perception are absolutely nonconceptual" (7). This dismissal, however, is too quick; fineness of grain and richness do pose a challenge for the *constituent* conceptualist.

3.3.3. *Against constituent conceptualism: a second pass*

This second version of the argument from fineness of grain is easiest to see on the psychological construal of the view. The constituent conceptualist holds, remember,

that the complex mental representations a subject is experientially related to are made up exclusively of concepts. But concepts—on the psychological view—are mental representations *of the right kind*. And there seems to be general agreement among representationalists that the “right kind” of mental representation will be, at the very least, a “*stored memory* representation, which one brings to bear in an appropriate manner (by, for example, activating the representation and applying it to the sensory input)” (Tye 1995, 139). Carruthers, in that spirit, writes that, so far as he is concerned, concepts are “discrete, memorable, recombinable components of thoughts and judgments, whose tokenings in thought will play an important role in inference” (2000, 135). After all, the constituents of thought must be able to support inferences, and this can be done only if two token mental representations can be recognized (re-identified) as being two tokens of the same type. This is true of very simple computing machines. Imagine, then, a machine that can compute modus ponens:

For example, if it encountered “Fa” and “Fa→Gb” on the input portion, it would print out “Gb” on the output portion; and it would do so for any such physical patterns that entokened well-formed sentences in the language (Rey, 213).

Clearly, such a machine can compute modus ponens only if it is able to identify the two tokens of ‘Fa’ as tokens of the same type. The machine must “remember”. If concepts are mental representations, then, they better be *stored* mental representations. And the constituent conceptualist is now in trouble, for she finds herself making the highly implausible claim that every property represented in experience—for instance yellow₁₇—is represented *conceptually*, i.e., by a *stored* mental representation. The claim simply seems to go against the empirical data, which suggests that many of the properties represented in experience cannot in fact be

re-identified by subjects even seconds after the initial experience (see Hurvich (1981) and Raffman (1996)). We discussed this somewhat in Part I: Sara might single out, on a chart, a paint chip of just the color she would like for her living room. But she is clumsy, drops the chart, and, upon staring at it again, finds herself unable to re-identify the shade she had selected just a few seconds earlier. The empirical facts suggest, as Tye puts it, that “normal perceivers typically have no schema in *memory* for red₂₉ or red₃₂” (104), though these properties are nonetheless represented in their experience.⁶⁰ We may reformulate the argument from fineness of grain as follows:

- (1) A concept is a stored mental representation
- (2) Empirical data suggests that fine-grained properties represented in experience aren't represented by stored mental representations.
- (3) Empirical data suggests that fine-grained properties represented in experience aren't represented by concepts.
- (4) Therefore, experience isn't wholly conceptual.

There may be an argument from richness here as well. After all, memory, like all cognitive resources, is limited. If every property that is represented in experience *did* get stored, as the conceptualist would claim it must, it would quickly create an “information overload” (Tye 2005, 520). Experience is rich enough that it seems impossible, computationally, for *all* the information represented to be stored in memory without severely hindering the functioning of the system as a whole. We may conclude that if some properties represented in experience aren't represented by

⁶⁰ This is true even of those who claim to have a *photographic* memory. The technical term used in the scientific literature is *eidetic* memory (see Palmer 1999, 593). Subjects with eidetic memory can scan a picture for 30 seconds or so, but after it has been removed, they claim to see it, still, in all its vividness and detail. Eidetic memories seem to be unlike other mental images. They are unlike afterimages; they don't move around as subjects move their eyes, and subjects “typically experience them as ‘outside the head’” (Palmer 1999, 593). There is, however, a fair amount of skepticism about the accuracy of eidetic memory. Though it is good, it is far from perfect. Subjects will not remember every detail of the scene; in fact, they sometimes “make up” details. Moreover, the data does not suggest that subjects with eidetic memory can recognize very fine-grained shades.

stored mental representations, they cannot be wholly represented by *concepts*, and conceptualism must be false. The argument can be outlined as follows:

- (1) A concept is a stored mental representation
- (2) Experience is so rich that, at some point, it is bound to represent a property for which there is no stored mental representation.
- (3) At some point, some representations in experience won't be concepts.
- (4) Therefore, experience isn't wholly conceptual.

This kind of argument can be leveled at constituent conceptualism even if the constituent conceptualist in question takes concepts to be, not mental representations, but abstract constituents of propositions. For even though concepts are, on this view, abstract entities, we may believe—as even McDowell himself does—that for these Fregean senses to truly be concepts, they need to be associated with an ability, on the part of the thinker, to re-identify the property later on in time.

What ensures that [something] is a concept—what ensures that thoughts that exploit it have the necessary distance from what would determine them to be true—is that the associated capacity can persist into the future, if only for a short time. (1995, 57)

Since it seems empirically false that the abstract constituents of experiential content are associated with a capacity to re-identify the property yellow₁₇ or yellow₁₈ even a short time into the future, then it would seem that constituent conceptualism is false.

Though these arguments do spell trouble for the constituent conceptualist, there is a straight-forward reply available to her. Remember that constituent conceptualism is the view according to which whatever the constituents of experience turn out to be, they are possible constituents of belief. The fineness of grain and richness arguments show that all the constituents of experience cannot be concepts, but it isn't obvious that all the constituents of *belief* are concepts either. If it turns out that there are constituents of thought that *aren't* concepts—that aren't stored memory

representations—then experience can take these latter representations as constituents as well and be wholly conceptual. Demonstratives seem like the best candidates for these constituents of belief.

3.3.4. Demonstratives

Demonstratives usually play the following role in the conceptual/nonconceptual debate: they are used by *possession* conceptualists as a reply to the argument from richness leveled against *them*. According to the *possession* conceptualist, one cannot experience some property *p* unless one possesses a concept for *p*. But if experience is rich, then it represents properties (yellow₁₇) for which a subject will possess no concept. The demonstrative *possession* conceptualist reply is this: subjects may not possess *general* concepts for every property represented by their rich experience (like yellow₁₇), but they could possess a *demonstrative* for each property. Experience, it turns out, could represent the very many properties it represents *demonstratively*.

The role that I want demonstratives to play here is not quite the same. The goal is to see whether we can use demonstratives to rescue (temporarily, at least) the *constituent* conceptualist by claiming that demonstratives might be those constituents of thought that do not require being stored in memory. This may strike those familiar with the debate between *possession* conceptualist and *possession* nonconceptualist as odd. After all, one important objection to the *possession* conceptualist's demonstrative move is to insist that demonstratives, like general concepts, require memory storage. McDowell makes this a requirement: demonstratives are *demonstrative concepts* and, as such, they too must come associated with a capacity

by the thinker to re-identify the object that falls under the demonstrative (see also Kelly 2001). But if it turns out that demonstratives, like concepts, require memory storage, then they can be of no use to the *constituent* conceptualist. The constituent conceptualist, that is, must deny that there is a re-identification condition on demonstrative concepts. In fact, she may have to deny that demonstratives are, strictly speaking, concepts. This, as already mentioned, does not take away from the conceptualist essence of the view: for as long as demonstratives—whether or not they are concepts—are constituents of belief, the constituent conceptualist can allow for them to be constituents of experience. (Remember that the constituent conceptualist's claim is that experience takes the same constituents as *belief*, no matter what the constituents of the latter are.)

So, the conceptualist does not need to hold on to the re-identification condition on demonstratives. What's more, the re-identification condition seems much too strong for demonstratives. Kelly's (2001) defense of the condition is not very convincing. Imagine, he says, that Sara is presented with triangle-square pairs, the triangle always being presented on her left and the square on her right. When asked whether these are the same shapes, Sara consistently answers no. After this task is completed, she is shown ten triangles in a row. Each time, she is asked the same question: is this the same shape you saw earlier on your left? Half the time Sara answers that it is, half the time that it isn't. Though she can clearly discriminate between a triangle and a square when they are presented simultaneously, Sara cannot re-identify a triangle when presented with one. The right conclusion, Kelly argues, is that Sara has no idea what the shape she saw on her left is. Though we can attempt to

explain her behavior in various ways, it is “impossible for us to allow that such a person possesses the *concept* expressed by the phrase ‘that shape’ (said while pointing to what is in fact a triangle)” (Kelly 2001, 13).

Kelly’s conclusion doesn’t quite follow. Sara can *think*, while sitting at the circus watching a tiger jump through a fiery hoop, that *this tiger is amazing*. And of course, there may be no way she could reliably re-identify said tiger, no matter how quickly after entertaining her thought she was asked to do it. After all, all tigers look pretty much the same to her, and as she watches *this one* perform she is not paying attention to those features of it which would help her tell it apart from other tigers. We can extend this case, it seems, from *token* demonstratives like *THIS TIGER* to type-demonstratives. The demonstrative expression Sara utters at the circus refers to *the token tiger* she is pointing to, but demonstrative expressions like *THIS SHADE* or *THIS SHAPE* demonstrate a *type*. When Sara thinks that this shade would be pretty on her living room walls, she is not thinking about *the particular token patch* of color she is currently looking at. She means to refer to *its shade*, and she is thinking that she wishes that *it* (the shade) could be the shade of her living room walls too. Similarly, when we have thoughts about this shape, we aren’t usually thinking about *the particular token triangle*, say, but about its shape, which can be the shape of many other things. Thinking about types (using type-demonstratives in thought) doesn’t seem to require that we be able to re-identify the type demonstrated any more than thinking about tokens does. I can think, while at the hardware store, that I want *this shade* for my living room walls even if there is no way I could ever re-identify the particular shade again. This is just what seems to be going on with Sara in the triangle

experiment. She may think that *this shape is magical* (pointing to a triangle), though she can't in fact re-identify that shape again. What makes Sara's triangle case so puzzling, unlike the hardware store one, is that the difference between a triangle and a square is so salient to us it is hard to imagine what must be going on in the mind of someone who can't tell the difference between them. But again, we seem to have no problem at all grasping this sort of thing in the case of color—or even of very intricate shapes.

For the purposes of Part II, we will take demonstratives to be constituents of thought that need not be stored in memory. Determinate color shades and all the many properties represented in experience, the conceptualist will claim, are represented *demonstratively*.

Two problems for the demonstrative view

Demonstrative accounts have faced a number of criticisms; we'll review two of them now. First, Sean Kelly (2001) argues that demonstratives are too “coarse-grained” to do the work that conceptualists want them to do. Kelly argues that demonstrative accounts cannot account for the fact that experience is situation-dependent: a uniformly colored white wall will, depending on lighting for instance, look slightly different shades. At sunset, it might in fact look orange; at any given time, that wall will look off-white, gray, light yellow, say.⁶¹ How can this experience be captured *demonstratively*? After all, any demonstrative pointing to the *color* of the wall will have the *same* content: the wall is uniformly colored. If there is one property (here a

⁶¹ It is interesting to notice, actually, that the wall does, in some sense, look *white*. It sometimes takes work to come to see that, though the wall *is* uniformly white, it actually does look yellow or shades of gray. We'll come back to this point at the end of the section.

color property), then all the demonstratives in experience will be pointing to *it* and constituent conceptualism cannot capture the fineness of grain of experience after all.⁶²

Misrepresentation and hallucinations also seem problematic for demonstrative accounts. We might wonder, the argument goes, what happens when a subject hallucinates a banana. What is the content of the demonstrative in the subject's experience? After all there is no yellow there to be "picked up" by the demonstrative. It would therefore seem to be empty. Moreover, we might wonder how to explain an experience that *misrepresents* a yellow banana as green, say. If a demonstrative points directly to features of the world, it would seem that the content of the misperceiver's color demonstration and the content of a *normal* subject's demonstration would be the same. After all, the '*this*' points to objective yellow in both cases.

Why pointers can't do the job

The pointer demonstrative account might appear promising, but I'll now argue that it fails to provide a real option for the conceptualist. Here is why: though demonstratives in experience and demonstratives in thought are pointers, and hence, in that respect mental constituents of the same kind, they aren't *enough* of the same kind. There are, after all, many ways of typing any two things. A banana and a loaf of bread are of the same kind in that they are both edible foodstuffs, but they are of different kinds in that a banana is a fruit while a loaf of bread is not. And so it goes for any two pointers. A pointer in experience and a pointer in belief are of the same

⁶² Though I talk of "objective color properties" I do not mean to be committed to a realist account of color. The example could make use of other properties; color properties are just easiest to talk about.

kind in that they are both pointers. In fact, some might want to claim that there are pointers within the encapsulated visual system. Pylyshyn (2003) posits visual tags, which he calls FINSTs and which he compares to demonstrative pointers. FINSTs are deployed by the early visual system (pre-attentively) to track proto-objects. In some respects, then, a FINST, a pointer in experience, and a pointer in belief are all of the same kind. But the conceptualist doesn't care about just *any* which way of typing the constituents of experience: for instance, it isn't enough that the constituents of experience, like the constituents of thought, are mental representations and hence, in that respect, the same kind of things. The constituents of experience have to be enough like the constituents of thought that they (the constituents of *experience*) could, in principle, be constituents of thought. It isn't enough to say, then, that demonstratives in experience are pointers and hence, that they are, in some way, like demonstratives in thought. The conceptualist must further argue that the pointers in experience are so similar to those in thought that pointers that are now in experience could, in principle, become constituents of thought. FINSTs, though they are pointers, aren't the right kind of pointers—for surely *they* could not, even in principle, be constituents of thought. This isn't to say that the deployment of FINSTs within early vision isn't necessary for the deployment of demonstratives in thought. Still, the two pointers are of a different kind, such that a FINST couldn't figure in thought, for FINSTs are characterized (at least partly) functionally: they are the immediate (and mostly automatic) outputs of a mechanism that “is related to focal attention but is more primitive and operates earlier in the information-processing stream.” They allow the visual “system to pick out a small number of [...] *primitive visual objects* or

proto-objects, in order subsequently to determine certain of their properties” (Pylyshyn 2003, 201). Demonstratives in thought may be pointers, but they are *not* the immediate, mostly automatic, outputs of the FINST mechanism. Their role is not to allow the visual system to pick out some number of proto-objects so that their properties can be determined. So FINSTs are pointers of the wrong kind, in that they could not be, even in principle, constituents of thought.

Pointers in experience, like FINSTs, are the wrong kind of pointers. First, because it seems plausible that pointers in experience, like FINSTs, should be characterized at least partly functionally. Rey (1997) suggests that representations in experience, unlike those in thought, enter into “characteristic processing” as a result of being “parameterized in specific ways.” In experience a pointer to a shade will be parameterized for “hue, lightness, saturation and relative position”, whereas the concept YELLOW (or the concept YELLOW₁₇) figuring in Sara’s belief isn’t so parameterized and, as result, doesn’t enter into characteristic processing.

Second, pointers in experience do not point to the same sort of mental representation as do demonstratives in thought. Demonstratives in *thought* point to experiential representations; demonstratives in *experience* point to some prior representation—say, a feature map. One could not “pluck”, as it were, a pointer from one program and insert it into another—the same goes here. One cannot pluck a pointer in experience and insert it, even in principle, into thought. And it won’t help the conceptualist to argue that pointers in experience and pointers in thought point to the *same* representation. Imagine for a moment that pointers in experience, like pointers in thought, point to experiential representations. Then presumably the

pointers in experience can't point to more pointers in experience. They must, that is, point to a "part" of experience that isn't itself demonstrative. But what kinds of constituents will that part of experience take, if not demonstratives? The conceptualist's only possible answer is: more concepts. But what concepts? Not demonstratives; but certainly not general concepts either. (After all, if pointers in experience "pointed to" general concepts, then it's not clear that an appeal to pointers could capture fineness of grain.) If that is right, though, the relevant part of experience would have to be *nonconceptual*—something the conceptualist isn't willing to accept.

Finally, the kind of mental pointing that goes on in belief seems importantly connected with *attention*.⁶³ Levine writes: "the point is, when I demonstrate that fly on the wall, visually *attending* to it seems to be an essential component of the process" (online, 13, emphasis mine). Chuard claims that there is an attention constraint on demonstratives: "if a subject *S* forms a *demonstrative* concept *C* for a property *f*, *S* is able to focus her attention on an instance of *f* in her perceptual field" (2006, ... see also Evans 1982). However, the mental pointers that the conceptualist claims figure in experience do not seem related to attention in the same way. In a number of cases, it seems as though some properties are represented in experience without a subject's attention being directed at them. Here are two examples (the first is from Chuard, under review).

First, consider Sara, who is back at the paint store, looking at a color chart. She has focused her attention on the color chip whose color is the one she thinks she

⁶³ There are many kinds of attentional mechanisms (see Palmer 1999, Ch. 11 for a discussion of visual attention, which we'll discuss in the next section).

wants for her living room. Yet, as she looks intently at that color chip, her experience does not stop representing the shade of the chips closest to her focus of attention. That is, these shades remain represented in her experience, despite the fact that Sara is not focusing her attention on them. Or consider the case of situation-dependency we discussed earlier. A uniformly white wall will actually look (at any one time) light gray here, and then light yellow there, and then dark gray. But this is something which subjects might not notice until their attention is drawn to it (when they first take a painting class, say). Presumably, their experience represents these various shades even before they take a painting class. This is why, I take it, novice painters sometimes try to render the wall in question by painting a uniformly white surface, only to find that their painting does not “look” right (and sometimes they can’t quite say why it doesn’t look right). If the experience of naïve subjects did not represent these shades of gray and yellow (but represented only the one shade of white) before their attention was drawn to it, then naïve painters would think that a uniformly white surface does look just right (if it’s the right shade)—that it matches their experience. But they don’t—and it is not merely a problem of finding the right shade. This, I think, is evidence that the subjects’ experiences represent these shades prior to their focusing their attention on them.

So again, it seems that the mental pointers that the conceptualist wants to posit as constituents of experience are not the right kinds of pointers. Whereas attention seems to play an important role in determining what the mental pointer in thought actually points to, attention will play little or no role in the *experiential* mental pointing.

3.3.5. *Where we conclude*

The arguments from richness and from fineness of grain, though they initially seem to be arguments against *possession* conceptualism (and hence for possession nonconceptualism) can be recast as arguments against *constituent* conceptualism. And these arguments are *somewhat* successful. After all, the demonstrative move can't quite rescue the conceptualist. As such, I do think that these arguments are successful arguments *against* conceptualism. However, the arguments fail as arguments *for* constituent nonconceptualism. After all, as mentioned in section 2, there *is* a middle position—and, at most, the arguments we discussed here (in 3.3) are arguments for this middle position. The arguments from richness and fineness of grain show, at most, that experience must be *partly* nonconceptual.

3.4. Last remarks

My goal in section 3 was to present some important arguments for *nonconceptualism*—arguments whose conclusions are usually that experience is *entirely* nonconceptual (entirely made up of nonconceptual elements). The first two kinds of arguments were very unconvincing. The fact that the experiences of infants and animals seem to be enough like ours that they should share (some) constituents does not entail 1) that the constituents of their experiences cannot be entirely made of concepts that they cannot use to think and 2) that the constituents of their experience cannot be *partly* made up of concepts (which they cannot use to think), as the *partial* conceptualist would claim.

Also, experience, like belief, seems systematic, and systematicity therefore gives us no reason to think that experience cannot be conceptual (or partially conceptual). The arguments from richness and fineness of grain come the closest to making the case that nonconceptualism is true. I do think these arguments show that experience cannot be *entirely* made up of concepts (as the conceptualist would like to claim). However, the arguments do not show that experience is thereby *entirely* made up of nonconceptual constituents. At the most, they show that experience is partially conceptual. To make the case that experience is indeed partially conceptual, I need to argue that experience cannot be *entirely* made up of nonconceptual elements. We'll consider arguments *against* nonconceptualism.

4. The case against conceptualism

Arguments against nonconceptualism are usually arguments *for* conceptualism. If the arguments from fineness of grain and richness go through, however, they won't be arguments for *conceptualism* tout court, but rather for partial conceptualism. We'll discuss, in this section, three important arguments against nonconceptualism: first, an argument (perhaps the best-known) defended by McDowell (1995) and Brewer (1999, 2003), according to which our perceptual beliefs cannot be justified if experience is nonconceptual (4.1). Second, Noë's (1999) argument (4.2). And finally, what I take to be the most compelling argument, an argument from concept-acquisition (4.3).

4.1. Justifying beliefs

Experience can justify beliefs, “that much seems obvious,” writes Byrne (2005). Sara’s *seeing* that there is a banana on the table provides her with a justification, it seems, for *believing* that there is a banana on the table. Her seeing that there is a dragonfruit on the table would justify her belief that there is a dragonfruit on the table, but it would not justify her belief that there is a banana on the table. How does this fact, which “seems obvious”, support conceptualism?

By itself, of course, it doesn’t. But McDowell (1995) and Brewer (1999) make an additional claim: experience can justify belief *only if* experiences and beliefs take the same kind of constituents. (Or, put in terms of content, as McDowell and Brewer do, only if experiential contents and belief contents are of the same kind.) McDowell is driven to this conclusion by a further assumption that experience can be only one of two things: 1) a representational state with concepts as constituents, or 2) a mental state *that isn’t representational at all*—a state rather more like a sense datum. Now, given the limited options here—and the serious shortcomings of the second one—we may well conclude, as McDowell does, that experience can justify belief only if it takes concepts as constituents.

But surely, if this is McDowell and Brewer’s reasoning, we should point out that experience can be neither of the two states presented earlier. After all, it seems that mental states can be representational even if their constituents couldn’t be constituents of thought. That is, presumably, the case for “subpersonal” perceptual states, i.e., states within encapsulated systems of the kind that Marr (1982) posits in early vision (see also Pylyshyn 2003). Let’s leave aside McDowell’s false dichotomy

then: I'll assume, as I have all along, that states (like subpersonal states) might be representational without being *conceptual*. Now, what follows? Let us assume that experiences do justify belief. Can we explain how that happens *only* if we claim that experience has the same kinds of constituent as thought?

Nonconceptualists, naturally, think that *they* too can explain how experience justifies belief, even if turns out that the constituents of experience aren't concepts. They point out that experience has accuracy conditions, if not truth conditions. So we can talk of experiences being inaccurate, as Sara's would certainly be if she hallucinated a banana on the table. Most of the time, Sara's experience is indeed accurate and represents a banana on the table only when there is a banana on the table. Nonconceptualists have argued (see Peacocke (2001) and Heck (2000)) that representational states with accuracy conditions *can* justify belief. It is enough, they argue, that Sara's seeing of a banana has conditions under which it is accurate; here, conditions under which there is, in fact, a banana on the table. An experience with such accuracy conditions would justify, on this view, the belief which is *true* in the same conditions. This is how Heck (2000) puts it:

If, for example, the information carried by a given perceptual state is a scenario, a set of ways in which the space around the observer might be arranged, as on Peacocke's view, there will be no bar whatsoever to perceptions' standing in semantic relations with beliefs: some beliefs about how space is arranged will be inconsistent with its being arranged in one of the ways the scenario includes; others, required by it; others, made probable by it; others, in context, could be reliably inferred from it; and so on. (505)

In conclusion, the epistemological argument just presented doesn't quite succeed. But there are others to consider.

4.2. Seeing as

Everyone in the debate, it seems, acknowledges that there is a difference between *seeing* and *seeing as*. Sara can see that there is a banana on the table without seeing it *as* a banana. Even a possession conceptualist may agree, then, that for Sara to see a banana *as a banana*, she has to possess the concept BANANA, even while insisting that Sara's mere seeing that there is a banana on the table does *not* require concept possession

Two things are worth noticing immediately. First, granting that *seeing as* requires concept possession does not commit one to possession conceptualism. After all, possession conceptualism is the view that to have *any* experience, one must possess the relevant concepts. But someone might grant that only *some* experiences involve *seeing as* and hence, that only *some* experiences require that we possess the relevant concepts, not *all*.

Second, though one may grant that *seeing as* requires concept possession, nothing is thereby entailed about the kind of constituents experience may take. This is the point, made early on in section 1, that the possession-only version of possession conceptualism is compatible with both constituent conceptualism and constituent nonconceptualism.

One may attempt, however, to supplement the claim that *seeing as* requires concept possession so as to yield a more interesting conclusion. There are at least two ways to go about accomplishing this goal: one may argue that *all* seeing is *seeing as*. If that is right, then we have an argument for *possession* conceptualism. (This appears to be, roughly, Noë's argument in his (1999).)

- (1) *Seeing as* requires that subjects possess the relevant concepts.
- (2) All seeing is *seeing as*.
- (3) Therefore, all seeing requires that subjects possess the relevant concepts (possession conceptualism about (visual) experience as a whole).

There is an alternative, however. Note that we have only mentioned, so far, the first version of possession conceptualism—the possession-only version, as we have called it. But we did discuss, in section 2, a second version of possession conceptualism—one that appeals not just to concept possession but to concept deployment as well. That second version of possession conceptualism makes the claim that Sara cannot see that there is a banana on the table unless she 1) possesses the concept BANANA and 2) deploys it in experience. In section 2, I went on to explain that this second version of possession conceptualism, if true, entails that constituent *nonconceptualism* must be false. One may then argue as follows:

- (1) *Seeing as* requires that subjects possess the relevant concepts *and deploy these concepts in experience*.
- (2) Experiences of *seeing as* can't be absolutely nonconceptual.
- (3) Experience as a whole isn't absolutely nonconceptual.

The conclusions of both arguments are interesting because they either implicate experience as a whole—the first argument—or make a claim about the *kind* of constituents experience can take—the second argument. They each have a premise in need of defense. A defense of the first argument requires a defense of the claim that *all* seeing is *seeing as*. A defense of the second argument requires a defense of its first premise—that *seeing as* requires not only possessing the relevant concept but *deploying* that concept in experience. Though it seems intuitively true that *seeing as*

requires possessing the relevant concept, it is far less obvious that it requires *deploying* that concept in experience.

One may wish to combine the two arguments just presented to make a *stronger* argument against constituent nonconceptualism. The combined argument would go like this:

- (1) All *seeing as* requires *both* concept possession and concept deployment.
- (2) All seeing is *seeing as*.
- (3) All seeing requires concept possession and concept deployment.
- (4) All seeing is incompatible with constituent nonconceptualism.

Let us now turn to a discussion of the two first premises in turn.

Noë argues (1999) that all seeing is *seeing as*. Experience, he writes, necessarily presents things to us as being a certain way. The claim is certainly intuitive: when Sara first sees the dragonfruit, her experience does not present the thing in front of her as being a dragonfruit, but it does present the thing in front of her as being some way or other—as some kind of thing or other. Sara, though she does not see the fruit *as a dragonfruit* sees it *as some thing*—as an object, as a fruit, as big, etc. So, we may conclude, seeing some thing always involves seeing it, at the very least, *as an object*. “To have visual experiences is not to judge that things are some way or other, but it is to represent things as being some way or other” (7).

Granting, then, that all seeing is *seeing as*, do we have any reason to believe that all *seeing as* requires concept *possession* and *deployment* as premise (1) would have it? I have said, already, that it seems intuitive that *seeing as* would require concept possession, but less so that it would require concept *deployment*. I’ll now argue that 1) *seeing as* doesn’t *always* require concept possession, and that 2) in some

instances it requires concept possession *and* deployment. The first premise, I'll conclude, doesn't quite hold up.

Grant that all seeing is *seeing as*, such that any seeing is necessarily a seeing of some thing *as an object, as colored*, presumably *as three-dimensional*, etc. Does such *seeing as* require that a subject *possess* the concepts OBJECT, COLORED and THREE-DIMENSIONAL—i.e., that the subject be able to think with these concepts? Some evidence suggests that seeing something *as* an object, or *as* colored doesn't require possessing these concepts. Pylyshyn (2003) claims that “individuation of what, for now, I will call *visual objects* is a primitive operation” of the *early* visual system (173).⁶⁴ The footnote here is telling: “in every case considered here, “object” is understood in terms of whether something is perceived *as* an individual” (italics mine). But the primitives in early perceptual representations aren't of the right sort to be constituents of thought. Some ways of seeing, then, may not depend on the concepts we possess—i.e., can think with—as much as they depend on the adequate functioning of our visual system. Though concept possession may not be required to *see* things *as* objects, colored etc... we may insist that concept possession is indeed required to see things as anything beyond these visual primitives. Seeing some thing as a dragonfruit, then, may indeed require possessing the concept DRAGONFRUIT.

What, now, of concept deployment? Does seeing something *as* a dragonfruit require possessing the concept *and deploying it* in experience? Here and there Noë hints that he does believe concept deployment (he calls it ‘exercise’) is required to see some thing *as* what it is. For instance, “when we have perceptual experiences, we *exercise* our grasp of concepts” (1) and “experiences require the mastery and exercise

⁶⁴ Also see Biederman (1993).

of concepts” (7). But there is no argument offered for this conclusion. In fact, one may want to *deny* that *seeing as* requires concept deployment by insisting that concept possession may be enough to causally influence the inputs of visual processing (which will come up in the next section 4.3.) So, when Sara sees something, not merely *as a thing* anymore *but as a dragonfruit*, the pattern of saccadic movement over the thing may change. Seeing the duck-rabbit figure *as* a duck, or *as* a rabbit doesn’t require that the concepts DUCK and RABBIT be deployed in each experience. It may simply involve different saccadic patterns over the figure, which yield different inputs for the visual system and hence, different experiential contents. Now, I will argue, in 4.3., that causal explanations of that sort aren’t always available; they aren’t available, for instance, in the case of color.

We are left with a (limited) argument against nonconceptualism:

- (1) All seeing is *seeing as*.
- (2) Most *seeing as* requires either concept possession or concept possession *and* deployment.
- (3) Therefore, most experiences aren’t wholly nonconceptual.

4.3. Concept acquisition

Acquiring concepts can change the way things look, feel, sound or taste to us. Training to distinguish between very complex but subtly different visual patterns—‘gexes’ and ‘zofs’—makes such patterns *look* importantly different to experimental subjects (Livingston et. al 1999); intensive bird watching can turn what used to look like countless gray birds on a beach into what now look to the expert like a family of knots and three groups of plovers (Carruthers 2000, see also Pylyshyn 2003, Tanaka and Taylor 1991); becoming a wine connoisseur can modify one’s phenomenal

experience of Pinot Noir (Melcher and Schooler 1996); studying interior decorating or going to art school can change one's phenomenal experience of color (Burns and Shepp 1988); and so on and so forth.

Assuming that concept acquisition *does change* the content of a subject's experience, how might we account for that change?⁶⁵ There are two possible explanations. First, one may claim that one's newly acquired concepts *enter in* the content of one's experience we'll call such an explanation a *constituent* explanation. Second, one may claim that one's newly acquired concepts *causally modify* the perceptual processing of one's experiential content—we'll call such an explanation a *causal* explanation. To illustrate, consider our subject Sara: at t_1 she is a naïve birdwatcher, but at t_2 she has become an expert. Her experience at t_1 and her experience at t_2 are different: at t_1 a bird-filled beach looks to Sara like just that, a beach filled with birds. At t_2 , however, the beach looks to Sara as though it is filled with knots and plovers (the example is from Carruthers 2000). On the constituent account, the explanation for the change is the following: the concepts KNOT and PLOVER figure in Sara's experience at t_2 though they did not at t_1 .⁶⁶ According to the

⁶⁵ One might maintain that what seems like a change in experiential content is actually not a change *in the content* itself but rather a change in what we can infer from that content. The data presented in the following section on "causal mechanisms" strongly suggests that this is not a live option.

⁶⁶ Note that for this particular explanation to be adequate, one must believe something like assumption (A): that if a particular concept cannot figure in someone's thought, then it cannot figure into someone's experience, such that if the concept *KNOT* cannot figure into the subject's thought (as when she's naïve), that concept cannot figure into her experience either. To see this, assume that (A) is false: assume that the concept *KNOT* cannot figure into the subject's thought (she's naïve) but is nonetheless a constituent of her experience. Now, if that's the case, the constituent story will lack the resources to explain the subject's changed experience once she's become an expert. Because according to the constituent account, what makes the expert's experience different from the naïve subject's is the fact that the concept *KNOT* is a constituent of the expert's experience but *not of the naïve subject's*. If one wants to maintain that (A) is false (as Byrne does for instance), one cannot give the constituent explanation. Of course, one may endorse some more restricted version of (A), such as: if a *high-level property concept* cannot be a constituent of thought, then it cannot be a constituent of experience either.

causal account, acquiring the concepts KNOT and PLOVER changes something in the visual processing of the birds on the beach, giving rise to a different experience. The argument against nonconceptualism goes like this:

- (1) Constituent explanations are unavailable to the nonconceptualist.
- (2) Causal explanations, however, fail to account for *all* the changes in experiential content caused by concept acquisition.
- (3) Therefore, nonconceptualists are unable to fully account for the concept acquisition data.

It should be rather obvious that the first premise is true. Constituent explanations require that a subject's newly acquired concepts—KNOT and PLOVER—become part of her experiential content at t_2 . But the (constituent) nonconceptualist denies that concepts can be constituents of experiential content—*ever*. A nonconceptualist hoping to explain the concept acquisition data, then, has no choice but to give a causal explanation. However, as we'll now see, causal explanations fail to account for all the changes caused by concept acquisition.

4.3.1. Causal mechanisms

Cognition can impact perception at the very least at two loci: very early on in perceptual processing and rather late in processing (see Pylyshyn 2003).⁶⁷ The mechanism in play is *selective attention*, and there are indeed two kinds of attention—early and late (Palmer 1999). A fair amount of data suggests that experts have learned where to look—where to direct attention. Biederman and Shiffrar (1987) show that “what distinguishes good chicken sexers from poor ones is, roughly, where they look and what distinguished feature they look for” (Pylyshyn 2003, 86). Good

⁶⁷ I assume that Pylyshyn is right and that the processing that goes on “in between” is encapsulated and therefore impenetrable to cognition.

chicken sexers usually train through repeated trials, i.e., without being told what to look for explicitly. Unsurprisingly, then, the data shows that explicitly telling naïve chicken sexers what to look for and where speeds up the learning process considerably (Shiffrar 1978). Birdwatchers, subjects in Livingston, et. al's experiments, wine tasters, etc., seem to learn, at the very least, to focus their attention the right way. Now, where one looks presumably has a tremendous impact on what becomes the input of the visual system: "Overt eye movements determine what optical information is available to the visual system; covert selective attention determines what subset of this information gets full processing" (Palmer 1999, 532). Experts, then, by directing their attention differently from naïve subjects, change the input to their visual systems and thereby modify the output experience.

Selective attention later on in processing involves "one perceptual feature being weighted more or less than another and/or combined in different...ways in the post perceptual categorization process" (87). Chess masters seem to be experts whose knowledge impacts perception in the later stages of processing. Chess masters often seem to display a "rapid visual processing and better visual memory" of chess board configurations (84) but, interestingly enough, not of just *any* such configuration. When the pieces on the board are arranged in a random manner, chess experts behave pretty much like beginners—they are no better at processing the particular arrangements or at remembering them. It is only when the pieces on the board are arranged in a way that is consistent with the rules of chess that the experts are better at processing and remembering the boards. Some interpret "the data as showing that ...chess masters have developed a very large repertoire...of patterns that they use to

classify or encode a large number of relevant patterns” (Pylyshyn, 85). At this late stage, then, cognition seems to impact perception by accessing data stored in long-term memory (again, see Pylyshyn, 85) and by allowing what is being processed to be compared to stored representations. The functioning of late *auditory* attention also suggests that stored representations are activated. The fact that subjects will hear their own names even when it is broadcast to an unattended channel suggests that the information in unattended channels is processed and, when some of it is recognized as one of a handful of “dictionary units with permanently lowered thresholds,” attention is directed to it. It looks, then, as though stored representations must be activated for late attention to impact perception. But these stored representations seem to be very much like concepts. And if concepts must be deployed for late attention to affect perception, an explanation of changes in experiential content using late attention looks less like a purely *causal* explanation (in the sense I’ve meant here) and more like a constituent explanation.

4.3.2. *Problems for nonconceptualism*

Nonconceptualism faces *two* problems here. First, there seems to be a consensus that cognition does in fact impact perception *both* via early attention and via late attention. If accounts of experience must allow for both kinds of influence, *and* if, as seems to be the case, explanations appealing to late attention really amount to constituent explanations, then the nonconceptualist is in trouble. For the nonconceptualist will not be able to allow for the part played by late attention. But even if it turns out that late attention plays virtually no role in perception, the nonconceptualist still has a

second problem: a causal story which appeals only to early selective attention cannot fully explain the effects of concept acquisition.

Imagine, then, a subject S and a uniformly red postcard, which looks, at t_1 , red to her. At t_2 , S has graduated from art school, and her postcard, as a result, looks scarlet to her. Can we really explain the change between t_1 and t_2 simply by appealing to a change in early selective attention? It seems not. A difference in saccadic eye movements (or even covert attention) can make quite a difference in the processing of shape (and of a complex scene) because the features of the shape or scene a subject's eyes saccade to the *most* dictate, to a large extent, what information is processed by her visual system. The naïve subject looking at a bird-filled beach will display one kind of pattern of saccadic exploration, and, as a result, the information processed will be of a certain sort and the resulting experience will have a certain content. Having just become an expert birdwatcher, the subject will display a new pattern of saccadic exploration, with saccades to features of the scene that were previously ignored. The new pattern of exploration is responsible for the new information processed and hence, in the end, for the subject's new experiential content. Can we account for the change in our art school graduate's experience this way? She may move her eyes over the picture differently, but there is no reason for this new saccadic pattern to cause information to be processed that wasn't processed before. After all, no matter where the subject's eyes saccade to, they encounter the *same* color information that they encountered when she was naïve. The expert birdwatcher, by *looking* where the naïve subject does not, provides her visual system with different shape information. But the art school graduate cannot, simply by *looking* where the

naïve subject does not, provide *her* visual system with different *color* information; the same color information is available everywhere. And since we cannot make the case that the *input* to the subject's visual system would be different at t_2 , we cannot make the case that her experiential content would be different at t_2 either. Nonconceptualist accounts of experience, it would seem, cannot fully explain the changes in experiential content due to concept acquisition.

4.4. Concluding

At the end of section 3, we concluded that experience could not be *entirely* conceptual. However, that conclusion left us with two alternatives: either experience could be *entirely* nonconceptual, or partially conceptual. The argument from concept-acquisition, however, makes the case that experience cannot be *entirely* nonconceptual. To account for the changes in Sara's experience of color, we need to say that her newly acquired concept gets to figure into her experience. And so this brings us to the following conclusion: experience is partially conceptual.

5. Partial conceptualism

The goal of this section is to spell out in a bit more detail what a partial conceptualist account might look like. First, I discuss one more motivation for partial conceptualism—the view captures very nicely a phenomenon we've talked about

before, i.e. situation-dependence. I then present some possible partial conceptualist models of experience (5.2). I end section 5 spelling out where this leaves us concerning Block's Inverted Earth.

5.1. One more motivation

The arguments reviewed here may all draw our attention to important features of experience; however, as I hope I have shown, these arguments are not arguments against partial conceptualism. Rather, they emphasize features of experience that most participants in the debate agree need to be accounted for. The most natural way of accounting for *all* these features—some of which suggest experience is nonconceptual, some of which suggest it is conceptual—is to claim that experience is a bit of both.

It will be helpful to mention one more consideration in favor of partial conceptualism. (A similar case was used in our discussion of situation-dependency, see 3.3.4). Let us focus on Sara's rich experience of, say, orange cliffs. When she sees the cliffs, she sees them *as orange* regardless of the weather or the time of day. That she sees them *as orange* explains some of her actions—including the fact that as she attempts to paint the cliffs, she colors them *orange*. Sara will notice, however, that the painted result most often doesn't look right; the orange drawing does not *look* like the cliffs because, though there is a sense in which the cliffs look orange to us, the way in which they do that is actually by looking to be a variety of different shades of different colors. Being a good painter requires "seeing" these colors—seeing, for instance, that the shadows on the cliff aren't dark orange but *purple*; that the cliffs in

the morning actually look pink in places, beige in others, very light orange in others; etc. “Seeing” in this way requires practice. But what suggests that these nuances are indeed represented in experience is the fact that when they are pointed out, one *does*, of course, *see* them. If all that was represented in Sara’s cliff experience, prior to anyone pointing out the color nuances, was *orange*, then painting the cliffs as simply *orange* would look right enough. After all, the painting at that point would be like her representation, i.e., it would represent a large pane of orange. But, and this is the point, such a painting does *not* look right—though it may not be immediately salient *why* it doesn’t look right. A painting with a mix of orange, purple and pink, in roughly the right places, however, *looks right*. This does strongly suggest that these nuances *are* represented in experience even before attention is drawn to them.

5.2. Some partial conceptualist models

At least a few philosophers are already partial conceptualists in the sense described here, whether they realize it or not. Tye, by accepting *seeing as* arguments, is led into a partial conceptualist view of *experience as a whole*. He writes: “clearly some representation in visual experience is a conceptual matter (e.g., the representation of object types such as car, ball, and telescope)” (2000, 75). But he goes on to say that

as for the question of which levels of representational content in experience metaphysically determine its phenomenal content, my own view (Tye 1995) is that the relevant levels are nonconceptual. (76)

Unlike Tye, I believe that the levels of experiential content that determine its phenomenal character are partially conceptual. The argument from concept acquisition is in part aimed at those, like Tye, who think they can account for

differences in phenomenal characters merely nonconceptually. As claimed above, it seems that they cannot deal with color cases.

Carruthers (2000), moved by *seeing as* arguments and arguments from concept acquisition, explicitly writes that “perceptual contents are often imbued with concepts (whether general or individual), while also containing representations which are analog in relation to those concepts” (2000, 136). Peacocke also has a hybrid view, one according to which the content of experience has *three* layers, two of which are different in kind from the content of thought, together with a conceptual third layer. We’ll start here by taking a closer look at Peacocke’s account.

Levels of content

The first layer of content of experience, on Peacocke’s view, is a *scenario*, that is, a way of “filling up the space around the perceiver” (1995, 61). Specifying a scenario requires two things. First, we need to fix an origin and axes. So,

for instance, one kind of origin is given by the property of being the center of the chest of the human body, with the three axes given by the directions back/front, left/right, and up/down with respect to that center. (ibid)

This is only *one kind* of origin; Peacocke is explicit about the fact that origins “will not be a specific place and set of directions in the real world” (62), a claim which fits nicely with some problem-solving in vision theory. From the first inputs on the retina to an actual grabbing motion, the coordinate frames of the visual mental representations must undergo a number of transformations, from being, first, oculo-centric (with the origin at the center of the eye), to being head-centric, to being torso-centric and finally to joint-space representations.

The second thing needed to fix a scenario, Peacocke claims, involves a filling in of space around the origin and axes:

for each point (strictly, I should say point type), identified by its distance and direction from the origin, we need to specify whether there is a surface there and, if so, what texture, hue, saturation, and brightness it has at that point, together with its degree of solidity. The orientation of the surface must be included. So must much more in the visual case: the direction, intensity, and character of light sources; the rate of change of perceptible properties, including location; indeed, it should include second differentials with respect to time where these prove to be perceptible. (63)

The actual content of one of Sara's visual experiences of a banana will be, on this view, a *positioned* scenario; that is, a way of filling up space *together with* an actual assignment to the origin and labeled axes of given directions and places and an assignment of time.⁶⁸

Peacocke is motivated, for the most part, by concerns about fineness of grain and richness, which his account accommodates well. After all, there are a number of possible positioned scenarios consistent with the belief 'that bananas are yellow' each specifying, for each point in the space, the particular "hue, saturation, and brightness [...] at that point" (1995 63). The hue, saturation, brightness, etc., specified of each point of the positioned scenario space are unlikely to be ones for which I possess concepts. (Actually, most people might not even possess the concepts HUE, SATURATION and BRIGHTNESS, though those might be ways that the visual system presents information independently of concept possession—see 3.2.)

⁶⁸ It is not clear that scenarios that aren't positioned can be contents at all, since the scenario without an assignment of actual origin and axes cannot be correct or incorrect in the same way that the thought *that's scary* without a specification of the demonstration and time can't be truth-evaluated (see 2003, 109).

Peacocke argues that there is another layer of content—*protopropositional* content—layered, so to speak, on top of scenario content. “Protopositions” he says, “are assessable as true and false. A protoposition contains an individual, together with a property or relation” (2003, 118). Protopropositional content seems to be roughly Russellian in character—hence, not yet conceptual.

On top of all this, Peacocke argues, part of the content of experience may be, in some cases, the same kind of content as that of belief: we should “insist on the partially conceptual character of the perceptual content when one sees something to be a dog, or a tree” (115). He says later: “It is not clear that there is good reason for denying the overwhelmingly plausible view that we see things as trees or hear a sound as of a car approaching” (123).

Mental representations

There are several models of mental representation which would be congenial to partial conceptualism. The first is Dretske’s speedometer model (used in Carruthers 2000); second is Tye’s map model.

Dretske likes to draw analogies to speedometers, and here is how the analogy is supposed to work here. A speedometer can represent the speed of a car in a very fine-grained way, while it also represents it in a much coarser way. The hand of the speedometer as it moves will represent speeds finely, but the speedometer can also include marks for each 10 mph speed range. When the hand is somewhere (fine-grained) between 0 and 10 mph, the mark stipulates that this is “Speed 1”; when the hand is somewhere between 10 and 20 mph, the mark stipulates that this “Speed 2”;

etc. In this way, the speedometer represents both fine-grained and coarse-grained information (see Carruthers 2000, 135). Similarly, experience may represent the shades of orange cliffs nonconceptually (using mental representations that are rather more fine-grained than concepts), while at the same time “marking” the similar shades as *orange* (conceptually).

Tye (1995) argues (not quite successfully, as claimed in 3.2.) that experience representations are not conceptual because, he says, experiences are not systematic or productive. He concludes that experience representations are *map-like*. (He sees the fact that “adjacent parts of the cortex represent adjacent parts of the retinal image” as evidence that there is “an orderly topographic projection of the retinal image onto the brain” (1995, 120). He goes on to say that “the obvious suggestion, then, is that [perceptual sensations] themselves have a topographic or map-like structure” (ibid, 121).

Since Tye does grant that experience can be partially conceptual (it is only the level that is relevant to phenomenal characters that is entirely *nonconceptual*), he argues that the map-like representations are representations “to which descriptive labels are attached” (ibid). Though these labels *need not* be concepts, they can be. For someone to see something *as a duck*, one of the labels attached to the topographic map must be the concept DUCK.

Though it is not obvious that we should think of mental representations as *map-like*, the picture that Tye gives is the right kind of picture. So, even if the content of Tye’s representation might be a “map-like content” (in that it represents space around the perceiver in roughly the way Peacocke claims), we can insist that the

representations themselves be symbolic.⁶⁹ Each “simple” representation represents facts about hue, saturation, and brightness at some location (given some origin). And we can add that clusters of these representations can become “attached” to a different kind of a representation, i.e., a concept.

The fact that we can make sense of what a partial conceptualist account of experience might look like is encouraging. However, some issues remain, and we now turn to consider what the implications of adopting partial conceptualism will be for a representationalist trying to deal with Block’s Inverted Earth.

5.3. Back to Block

In the introduction to Part II, I discussed Block’s Inverted Earth thought experiment. Block uses it to argue against the representationalist. After all, if, during her time on Inverted Earth, the content of Sara’s experience of a blue VW bug changes but the phenomenal character of that experience stays the same, the representationalist is in trouble. For the representationalist, a change in experiential content means a change in phenomenal character (and vice versa). The representationalist, as argued in the introduction, needs to worry most about the outcome of the conceptual/nonconceptual debate. Indeed, the representationalist seems committed to saying that Sara’s experiential contents must stay the same, no matter how long she spends on Inverted Earth (after all, her phenomenal characters stay the same). Now, if partial conceptualism is the most plausible account of experience, then it turns out that concepts are constituents of experience. Any attempt to argue that Sara’s experiential

⁶⁹ Tye himself says, while discussing the topographic map involved in pain sensations, that “activity in any given cell [which makes up the map] may be conceived as representing, in the manner of a simple symbol, that there is tissue damage at the body region to which the cell is dedicated” (ibid).

contents stay the same on Inverted Earth will require arguing that Sara's *concepts* also "stay the same", no matter how long she spends on Inverted Earth. But most representationalist seem fine with granting Block that our *beliefs* would change on Inverted Earth—that the content of our color concepts would shift. If I am right about partial conceptualism, then representationalists should not be so quick to grant Block's claim about concepts.

Conclusion

My goal in this dissertation has been to examine the role played by concepts in the defense and elaboration of reductive representationalism. Part I considered the role played by *phenomenal* concepts in the defense of reductive representationalism against anti-physicalism. My focus there was on concepts *of* experience. Part II focused on whether there can be concepts *in* experience—a question which the representationalist must answer to fully spell out her account and defend it adequately against further objections, such as Block’s Inverted Earth.

There is still much to be done. As should have become obvious at the very end of Part II, fully spelling out an account of reductive representationalism requires that representationalists say something more about *concepts*; namely, about whether their contents are wide or narrow. To successfully defend representationalism against Block’s Inverted Earth will require arguing either for some theory of narrow content, or for a theory of wide content unlike the one Block assumes in his article. (For instance, Tye (2000) argues that the nonconceptual layer of experience has wide content but defends a teleological/co-variational theory of wide content, which allows him to say that Sara’s nonconceptual experiential content would not shift on Inverted Earth).

Also, some of the first defenders of the phenomenal concept strategy are now backtracking. Tye is working on a book called “Consciousness Revisited: Materialism without phenomenal concepts” (forthcoming 2008). I doubt that Tye is denying *that there are* phenomenal concepts (at least in the broad sense of concepts that pick out our phenomenal characters). Still, he has argued (at the Pacific APA for

instance) that phenomenal concepts cannot do the job physicalists typically want them to do (especially in the case of Mary). Whether he's right is one important question left to answer.

In any case, let me summarize the progress I have made here. In Part I, I have argued that phenomenal concepts *can*, despite what anti-physicalists believe, explain the relevant data within a physicalist framework. The physicalist can explain what Mary learns in either of two ways—either by making the case that the significant bit of what happens to Mary is her actually *undergoing the experience*, or by claiming that what Mary learns is a form of “seeing as”. The physicalist can explain the core contrast by appealing to pre-theoretical connections between phenomenal concepts and the concept NONPHYSICAL. Finally the physicalist can deal with property dualist arguments, either by arguing that the Semantic Premise is false, or by pointing out that there are other ways of “filling in the blanks”.

In part II, I have argued that experience can be neither entirely conceptual nor entirely nonconceptual. A modified version of the famous argument from richness and fineness of grain shows that experience cannot be entirely conceptual. And the argument from concept-acquisition shows that experience cannot be entirely nonconceptual. So, going forward, any plausible view of representationalism must involve some form of *partial* conceptualism. This will by no means resolve all of the issues concerning representationalist accounts and concepts, but it will be a step in the right direction.

Bibliography

- Balog, K. (manuscript). Acquaintance and the mind-body problem.
- Biederman, I. (1993). Visual Object Recognition. In Goldman, A. (ed) *Readings in Philosophy and Cognitive Science*. MIT Press.
- Biederman, I. & Shiffrar, M. (1987). Sexing day-old chicks: a case study and expert systems analysis of a difficult perceptual learning task. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 13, 640-645.
- Block, N. (2007). Max Black's objection to mind-body identity. In Torin Alter & Sven Walter (eds.), *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*. Oxford University Press.
- Block, N. (2002). The harder problem of *consciousness*. *Journal of Philosophy* 99 (8):391-425
- Block, N. (1995). On a confusion about a function of consciousness. In N. Block, O. J. Flanagan & G. Guzeldere (eds.) *The Nature of Consciousness: Philosophical Debates*. MIT Press.
- Block, N. (1990). Inverted earth. *Philosophical Perspectives* 4:53-79.
- Bering, J. & Bjorklynd, D. (2004). The natural emergence of reasoning about the afterlife as a developmental regularity. *Developmental Psychology*, 40, 217-233.
- Bermúdez, J. and Arnon C. (2008) "Nonconceptual Mental Content", *The Stanford Encyclopedia of Philosophy (Summer 2008 Edition)*, Edward N. Zalta (ed.), URL = <<http://plato.stanford.edu/archives/sum2008/entries/content-nonconceptual/>>.
- Bloom, P. (2004). *Descartes' Baby: How the science of child development explains what makes us human*. Basic Books.
- Brewer, B. (1999). *Perception and Reason*. Oxford University Press.
- Brewer, B. (2005). Do sense experiential states have conceptual content? *Contemporary Debates in Epistemology*, eds. E. Sosa and M. Steup. Blackwell.
- Burge, T. (1977). Belief *De Re*. *Journal of Philosophy*, 74, 338-62.

Burns, B., and Shepp, B. E. (1988). Dimensional interactions and the structure of psychological space: The representation of hue, saturation, and brightness. *Perception and Psychophysics*, 43, 494-507.

Byrne, A. & Pryor, J. (2006) Bad Intensions. In Manuel Garcia-Carpintero & Josep Macia (eds.), *The Two-Dimensionalist Framework: Foundations and Applications*. Oxford University Press.

Byrne, A. (2005). Perception and Perceptual Content. *Contemporary Debates in Epistemology*, eds. E. Sosa and M. Steup. Blackwell.

Carruthers, P. and Veillet, B. (2007). The phenomenal concept strategy. *Journal of Consciousness Studies* 14 (s 9-10):212-236.

Carruthers, P. (2004). Reductive explanation and the "explanatory gap". *Canadian Journal of Philosophy* 34 (2):153-174.

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

Chalmers, D. (2007). Phenomenal concepts and the explanatory gap. In Torin Alter & Sven Walter (eds.), *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*. Oxford University Press.

Chalmers, D. (2004). Phenomenal concepts and the knowledge argument. In Peter Ludlow, Yujin Nagasawa & Daniel Stoljar (eds.), *There's Something About Mary: Essays on Phenomenal Consciousness and Frank Jackson's Knowledge Argument*. MIT Press.

Chalmers, D. (2003). The content and epistemology of phenomenal belief. In Quentin Smith & Aleksandar Jokic (eds.), *Consciousness: New Philosophical Perspectives*. Oxford University Press.

Chalmers, D. (1999). Materialism and the metaphysics of modality. *Philosophy and Phenomenological Research* 59 (2):473-96.

Chalmers, D. (1996). *The Conscious Mind*.

Clark, A. (2004). Feature-placing and proto-objects, *Philosophical Psychology*, 17(4), 451-477.

Crane, T. 1992. The nonconceptual content of experience. *The Contents of Experience*, ed. T. Crane. Cambridge University Press.

Chuard, P. (2006). Demonstrative concepts without reidentification. *Philosophical Studies* 130 (2):153-201.

- Dennett, D. (1991). *Consciousness Explained*. Boston: Little, Brown and Company.
- Dennett, D. (1988). Quining qualia. In A. Marcel & E. Bisiach (eds.), *Consciousness in Contemporary Science*. Oxford University Press.
- Dretske, F. (1995). *Naturalizing the Mind*. MIT Press.
- Dretske, F. (1981). *Knowledge and the Flow of Information*. Blackwell.
- Evans, G. (1982). *The Varieties of Reference*. Oxford University Press.
- Fodor, J. (2000). *In Critical Condition: Polemical Essays on Cognitive Science and the Philosophy of Mind*. MIT Press.
- Fodor, J. (1987). *Psychosemantics*. MIT Press.
- Fodor, J. (1975). *The Language of Thought*. New York: Crowell.
- Heck, R. (2000). Nonconceptual content and the “space of reasons”. *Philosophical Review* 109, 483-523.
- Hurvich, L. M. (1981). *Color Vision*. Sinauer Associates.
- Ismael, J. (1999). Science and the phenomenal. *Philosophy of Science* 66 (3):351-69.
- Jackson, F. (1986). What Mary didn't know. *Journal of Philosophy* 83 (May):291-5.
- Kaplan, D. (1989a). Demonstratives. In Almog Joseph, Perry John & and Wettstein Howard, eds. *Themes from Kaplan*. Oxford University Press.
- Kaplan, D. (1989b). Afterthoughts In Almog Joseph, Perry John & and Wettstein Howard, eds. *Themes from Kaplan*. Oxford University Press.
- Kelly, S. (2001). The non-conceptual content of perceptual experience: situation dependence and fineness of grain. *Philosophy and Phenomenological Research* 62,601-8. Reprinted in Gunther 2003.
- Kirk, R. (2006). "Zombies", *The Stanford Encyclopedia of Philosophy (Winter 2006 Edition)*, Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/win2006/entries/zombies/>.
- Kripke, S. (1972). *Naming and Necessity*. Harvard University Press.
- Kuhlmeier, V. Bloom, P. & Wynn, K. (2004) Do 5-month-old infants see humans as material objects? *Cognition*, 94, 95-103.

Laurence, S. & Margolis, E. (2007) The Ontology of Concepts—Abstract Objects or Mental Representations? *Nous*, 41, 561 – 593

Laurence, S. Margolis, E. (2005). Concepts. *The Stanford Encyclopedia of Philosophy*. Ed, Edward N. Zalta, URL= <
<http://plato.stanford.edu/entries/concepts/>>

Laurence, S. & Margolis, E. (1999). *Concepts: Core Readings*. MIT Press.

Levin, J. (2007). What is a phenomenal concept? In Torin Alter & Sven Walter (eds.), *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*. Oxford University Press.

Levine, J. (2007). Phenomenal concepts and the materialist constraint. In Torin Alter & Sven Walter (eds.), *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*. Oxford University Press.

Levine, J. (2001). *Purple Haze: The Puzzle of Consciousness*. Oxford University Press.

Lewis, D. (1990). What experience teaches. In William G. Lycan (ed.), *Mind and Cognition*. Blackwell.

Livingston, K., Andrews J. & Harnad, S. (1998). Categorical Perception Effects Induced by Category Learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 24(3) 732-753.

Loar, B. (1997/2002). Phenomenal states. In Chalmers, David (ed) *Philosophy of Mind: Classical and Contemporary Readings*. Oxford University Press.

Loar, B. (1990). Phenomenal states. *Philosophical Perspectives* 4:81-108.

Locke, J. (1689/1975). *Essay Concerning Human Understanding*, Oxford: Oxford University Press.

Lycan, W.G. (2006). "Representational Theories of Consciousness", *The Stanford Encyclopedia of Philosophy (Winter 2006 Edition)*, Edward N. Zalta (ed.), URL = <
<http://plato.stanford.edu/archives/win2006/entries/consciousness-representational/>>.

Lycan, W.G. (1996). *Consciousness and Experience*. MIT Press.

Machery, E. (2005). Concepts are not a natural kind. *Philosophy of Science* 72 (3):444-467.

Marr, D. (1982). *Vision*. Freeman.

- McDowell, J. (1994). *Mind and World*. Harvard University Press.
- Melcher, JM. Shooler, JW. (1996). The misremembrance of wines past: Verbal and perceptual expertise differentially mediate verbal overshadowing of taste memory. *Journal of Memory and Language*, 35, 231-245.
- Nida-Rümelin, M. (1996). What Mary couldn't know. In T. Metzinger, ed., *Phenomenal Consciousness*, Schoenigh, Paderborn.
- Nida-Rümelin, M. (1998). On Belief About Experiences: An Epistemological Distinction Applied to the Knowledge Argument. *Philosophy and Phenomenological Research* 58 (1): 51-73.
- Noë, A. (1999). Thought and experience. *American Philosophical Quarterly* 36, 257-265.
- O'Dea, J. (2002). The indexical nature of sensory concepts. *Philosophical Papers* 32 (2):169-181.
- O'Regan, K. and Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences* 24 (5):883-917.
- O'Regan, K. (1992). Solving the "real" mysteries of visual perception: The world as an outside memory. *Canadian Journal of Psychology* 46:461-88.
- Palmer, S. (1999). *Vision Science: Photons to Phenomenology*. MIT Press.
- Papineau, D. (2007). Phenomenal and perceptual concepts. In Torin Alter & Sven Walter (eds.), *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*. Oxford University Press.
- Papineau, D. (2002). *Thinking About Consciousness*. Oxford University Press.
- Peacocke, C. (2003). Scenarios, Concepts, and Perception. In Gunther, Y. (ed) *Essays on Nonconceptual Content*. MIT Press.
- Peacocke, C. (2001). Does perception have a nonconceptual content? *Journal of Philosophy* 98 (5):239-264.
- Peacocke, C. (2001). Phenomenology and nonconceptual content. *Philosophy and Phenomenological Research* 62 (3):609-615.
- Peacocke, C. (1995). Content, computation, and externalism. *Philosophical Issues* 6:227-264.

- Peacocke, C. (1992). *A Study of Concepts*. MIT Press.
- Perry, J. (2001). *Knowledge, Possibility, and Consciousness*. MIT Press.
- Prinz, J. (2007). Mental pointing: Phenomenal knowledge without concepts. *Journal of Consciousness Studies* 14 (9-10):184-211.
- Pylyshyn, Z. (2003). *Seeing and Visualizing*. MIT Press.
- Pylyshyn, Z. (Forthcoming). *Things and Places: How the Mind Connects with the World*. MIT Press.
- Raffman, D. (1995). On the persistence of phenomenology. In Thomas Metzinger (ed.), *Conscious Experience*. Ferdinand Schoningh.
- Rey, G. (2007). Phenomenal Content and the Richness and Determinacy of Color Experience. *Journal of Consciousness Studies* 14 (s 9-10):212-236.
- Rey, G. (1998). 'A Narrow Representationalist Account of Qualitative Experience', in *Philosophical Perspectives, Vol. 12*
- Rey, G. (1997). *Contemporary Philosophy of Mind*. Blackwell.
- Rey, G. & Tetzlaff, M. (forthcoming). Systematicity in Honeybee Navigation. In Lurz (Ed.) *Philosophy of Animal Minds*. Cambridge University Press.
- Rosenthal, D. (1993). Thinking that one thinks. In M. Davies & G. W. Humphreys (eds.), *Consciousness: Psychological and Philosophical Essays*. Blackwell.
- Sedivy, S. (1996). Must Conceptually-informed Experience Involve Non-conceptual Content?" *Canadian Journal of Philosophy*, xxvi, 413-31.
- Siegel, S. (2006). Which properties are represented in perception? In Tamar S. Gendler & John Hawthorne (eds.), *Perceptual Experience*. Oxford University Press.
- Siegel, S. (2005). The Contents of Perception. *Stanford Encyclopedia of Philosophy*. Ed. Edward N. Zalta, URL= < <http://plato.stanford.edu/entries/perception-contents/>>
- Simons, D.J. & Chabris, C.F. (1999). Gorillas in our midst: sustained inattention blindness for dynamic events. *Perception*, 28, 1059-1074.
- Smart, J. J. C. (1959). Sensations and brain processes. *Philosophical Review* 68 (April):141-56.
- Speaks, J. (2005). Is there a problem about nonconceptual content? *Philosophical Review* 114:3,

- Stalnaker, J. (forthcoming). *Locke Lectures*. Oxford University Press.
- Stoljar, D. (2005). Physicalism and phenomenal concepts. *Mind and Language* 35 (July):296-302.
- Stoljar, D. (2001). Two conceptions of the physical. *Philosophy and Phenomenological Research* 62 (2):253-81.
- Tanaka, J., & Taylor, M. (1991). Object categories and expertise : is the basic level in the eye of the beholder ? *Cognitive Psychology*, 23, 457-482.
- Thagard, P. (2007) Cognitive Science. *The Stanford Encyclopedia of Philosophy*. E. Zalta (ed.), URL = <<http://plato.stanford.edu/archives/sum2007/entries/cognitive-science/>>.
- Thagard, P. (2005). *Mind: Introduction to Cognitive Science*, second edition. MIT Press.
- Tye, M. (2008). Qualia. *The Stanford Encyclopedia of Philosophy*. E. Zalta (ed.), URL = <<http://plato.stanford.edu/archives/sum2008/entries/qualia/>>.
- Tye, M. (2005). Non-conceptual content, Richness, and Fineness of Grain. In T. Gendler and J.Hawthorne (eds). *Perceptual Experience*. Oxford University Press.
- Tye, M. (2000). *Consciousness, Color, and Content*. MIT Press.
- Tye, M. (1997) The Problem of Simple Minds: Is There Anything it is Like to be a Honey-bee? *Philosophical Studies*.
- Tye, M. (1995). *Ten Problems of Consciousness: A Representational Theory of the Phenomenal Mind*. MIT Press.
- White, S. (2006). Property dualism, phenomenal concepts, and the semantic premise. In Torin Alter & Sven Walter (eds.), *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*. Oxford University Press.
- White, S. (1986). Curse of the qualia. *Synthese* 68 (August) : 333-68.
- Zalta, E. (2001). Fregean senses, modes of presentation, and concepts. *Noûs* : 335-359.