The Content and Epistemology of Phenomenal Belief

David J. Chalmers
Philosophy Program
Research School of Social Sciences
Australian National University

1 Introduction

Experiences and beliefs are different sorts of mental states, and are often taken to belong to very different domains. Experiences are paradigmatically phenomenal, characterized by what it is like to have them. Beliefs are paradigmatically intentional, characterized by their propositional content. But there are a number of crucial points where these domains intersect. One central locus of intersection arises from the existence of phenomenal beliefs: beliefs that are about experiences.

The most important phenomenal beliefs are *first-person* phenomenal beliefs: a subject's beliefs about his or her own experiences, and especially, about the phenomenal character of the experiences that he or she is currently having. Examples include the belief that one is now having a red experience, or that one is experiencing pain.

These phenomenal beliefs raise important issues, in the theory of content and in epistemology. In the theory of content, analysing the content of phenomenal beliefs raises special issues for a general theory of content to handle, and the content of such beliefs has sometimes been taken to be at the foundations of a theory of content more generally. In epistemology, phenomenal beliefs are often taken to have a special epistemic status, and are sometimes taken to be the central epistemic nexus between cognition and the external world.

My project here is to analyse phenomenal beliefs in a way that sheds some light on these issues. I will start by focusing on the content of these beliefs, and will use the analysis

---

Published in (Q. Smith & A. Jokic, eds) *Consciousness: New Philosophical Perspectives*. Oxford University Press, 2003. The analysis presented here is a development of a brief discussion in Chalmers 1996, pp. 203-8. I have presented versions of this material (starting in 1997) at Antwerp, ANU, Arizona, Delaware, Fribourg, Miami, Munich, Princeton, Sydney, UC Santa Cruz, the APA (Pacific Division), Metaphysical Mayhem (Syracuse), and the World Congress in Philosophy (Boston). Thanks to many people in audiences there, and elsewhere. Special thanks to Mark Johnston, Martine Nida-Rümelin, Susanna Siegel, and Daniel Stoljar for lengthy discussions.

developed there to discuss the underlying factors in virtue of which this content is constituted. I will then apply this framework to the central epistemological issues in the vicinity: incorrigibility, justification, and the dialectic over the "Myth of the Given".

1.1 Phenomenal realism

The discussion that follows is premissed upon what I call "phenomenal realism": the view that there are phenomenal properties (or phenomenal qualities, or qualia), properties that type mental states by what it is like to have them, and that phenomenal properties are not conceptually reducible to physical or functional properties (or equivalently, that phenomenal concepts are not reducible to physical or functional concepts). On this view, there are truths about what it is like to be a subject that are not entailed a priori by the physical and functional truth (including the environmental truth) about that subject.

The phenomenal realist view is most easily illustrated with some familiar thought-experiments. Consider Frank Jackson's case of Mary, the neuroscientist who knows all relevant physical truths about color processing, but whose visual experience has been entirely monochromatic (Jackson 1982). On the phenomenal realist view, Mary lacks factual knowledge concerning what it is like to see red. Views that deny this deny phenomenal realism. Or consider cases in which a hypothetical being has the same physical, functional, and environmental properties as an existing conscious being, but does not have the same phenomenal properties. Such a being might be a zombie, lacking experiences altogether, or it might be an inverted being, with experiences of a different character. On the phenomenal realist view, some such duplicates are coherently conceivable, in the sense that there is no a priori contradiction in the hypothesis in question. Views that deny this deny phenomenal realism.

(What if someone holds that functional duplicates without consciousness are coherently conceivable, but that physical duplicates without consciousness are not? Such a view would be in the spirit of phenomenal realism. This suggests that we could define phenomenal realism more weakly as the thesis that the phenomenal is not conceptually reducible to the functional, omitting mention of the physical. I do not define it this way, for two reasons. First, I think if functional duplicates without consciousness are conceivable, physical duplicates without consciousness must be conceivable too, as there is no reasonable possibility of a conceptual entailment from microphysical to phenomenal that does not proceed via the functional. Second, it is not easy to give a precise account of what functional duplication consists in, and stipulating physical identity finesses that question. But if someone disagrees, everything that I say will apply, with appropriate changes, on the weaker view.)

Phenomenal realism subsumes most varieties of dualism about the phenomenal. It also subsumes many varieties of materialism. In particular it subsumes what I have called "type-B" materialism (see Chalmers 2002*a*): views that hold that there is an a posteriori necessary entailment from the physical to the phenomenal, so that there is an epistemic or conceptual gap between the physical and phenomenal domains, but no ontological gap. Views of this sort typically allow that Mary gains factual knowledge when she sees red for the first time, but hold that it is knowledge of an old fact known in a new way; and they typically hold that the duplication cases mentioned above are conceptually coherent but not metaphysically possible.

Phenomenal realism excludes what I have called "type-A" materialism: views that hold that all phenomenal truths are entailed a priori by physical truths. Such views include eliminativism about the phenomenal, as well as analytical functionalism and logical behaviorism, and certain forms of analytic representationalism. Views of this sort typically deny that Mary gains any knowledge when she sees red for the first time, or hold that she gains only new abilities; and they typically deny that the duplication cases mentioned above are coherently conceivable.

Those who are not phenomenal realists might want to stop reading now, but there are two reasons why they might continue. First, although the arguments I will give for my view of phenomenal beliefs will presuppose phenomenal realism, it is possible that some aspects of the view itself may be tenable even on some views that deny phenomenal realism. Second, some of the most important arguments *against* phenomenal realism are epistemological arguments that centre on the connection between experience and belief. I will be using my analysis to help rebut those arguments, and thus indirectly to support phenomenal realism against its opponents.

A note on modality: because I am assuming phenomenal realism but not property dualism, all references to necessity and possibility should be taken as invoking conceptual necessity and possibility. Similarly, talk of possible worlds can be taken as invoking conceivable worlds (corresponding to the epistemically constructed scenarios of Chalmers (forthcoming); see also the appendix to this chapter), and talk of constitutive relations should be taken as invoking conceptually necessary connections. If one accepts a certain sort of link between conceptual and metaphysical possibility (e.g. the thesis that ideal primary conceivability entails primary possibility), then these references can equally be taken as invoking metaphysical possibility and necessity.

A note on phenomenal properties: it is natural to speak as if phenomenal properties are instantiated by mental states, and as if there are entities, experiences, that bear their phenomenal properties essentially. But one can also speak as if phenomenal properties are

directly instantiated by conscious subjects, typing subjects by aspects of what it is like to be them at the time of instantiation. These ways of speaking do not commit one to corresponding ontologies, but they at least suggest such ontologies. In a *quality-based* ontology, the subject-property relation is fundamental. From this one can derive a subject-experience-property structure, by identifying experiences with phenomenal states (instantiations of phenomenal properties), and attributing phenomenal properties to these states in a derivative sense. In a more complex *experience-based* ontology, a subject-experience-property structure is fundamental (where experiences are phenomenal individuals, or at least something more than property instantiations), and the subject-property relation is derivative. In what follows, I will sometimes use both sorts of language, and will be neutral between the ontological frameworks.

2 The Content of Phenomenal Concepts and Phenomenal Beliefs

2.1 Relational, demonstrative, and pure phenomenal concepts

Phenomenal beliefs involve the attribution of phenomenal properties. These properties are attributed under phenomenal concepts. To understand the content of phenomenal beliefs, we need to understand the nature and content of phenomenal concepts.

I look at a red apple, and visually experience its color. This experience instantiates a phenomenal quality R, which we might call phenomenal redness. It is natural to say that I am having a red experience, even though of course experiences are not red in the same sense in which apples are red. Phenomenal redness (a property of experiences, or of subjects of experience) is a different property from external redness (a property of external objects), but both are respectable properties in their own right.

I attend to my visual experience, and think *I am having an experience of such-and-such quality*, referring to the quality of phenomenal redness. There are various concepts of the quality in question that might yield a true belief.[1]

We can first consider the concept expressed by 'red' in the public-language expression 'red experience', or the concept expressed by the public-language expression 'phenomenal redness'. The reference of these expressions is fixed via a relation to red things in the external

[1] I take concepts to be mental entities on a par with beliefs: they are constituents of beliefs (and other propositional attitudes) in a manner loosely analogous to the way in which words are constituents of sentences. Like beliefs, concepts are tokens rather than types in the first instance. But they also fall under types, some of which I explore in what follows. In such cases it is natural to use singular expressions such as 'the concept' for a concept-type, just as one sometimes uses expressions such as 'the belief' for a belief-type, or 'the word' for a word-type. I will use italics for concepts and beliefs throughout.

world, and ultimately via a relation to certain paradigmatic red objects that are ostended in learning the public-language term 'red'. A language learner learns to call the experiences typically brought about by these objects 'red' (in the phenomenal sense), and to call the objects that typically bring about those experiences 'red' (in the external sense). So the phenomenal concept involved here is *relational*, in that it has its reference fixed by a relation to external objects. The property that is referred to need not be relational, however. The phenomenal concept plausibly designates an intrinsic property rigidly, so that there are counterfactual worlds in which red experiences are never caused by red things.

One can distinguish at least two relational phenomenal concepts, depending on whether reference is fixed by relations across a whole community of subjects, or by relations restricted to the subject in question. The first is what we can call the *community relational concept*, or red_C. This can be glossed roughly as *the phenomenal quality typically caused in normal subjects within my community by paradigmatic red things*. The second is what we can call the *individual relational concept*, or red_I. This can be glossed roughly as *the phenomenal quality typically caused in me by paradigmatic red things*. The two concepts red_C and red_I will co-refer for normal subjects, but for abnormal subjects they may yield different results. For example, a red/green-inverted subject's concept red_C will refer to (what others call) phenomenal redness, but his or her concept red_I will refer to (what others call) phenomenal greenness.

The public-language term 'red' as a predicate of experiences can arguably be read as expressing either red_C or red_I. The community reading of 'red' guarantees a sort of shared meaning within the community, in that all uses of the term are guaranteed to co-refer, and in that tokens of sentences such as 'X has a red experience at time t' will have the same truth-value whether uttered by normal or abnormal subjects. On the other hand, the individual reading allows a subject better access to the term's referent. On this reading, an unknowingly inverted subject's term 'red' will refer to what she think it refers to (unless the inversion was recent), while on the community reading, her term 'red' may refer to something quite different, and her utterance 'I have had red experiences' may even be unknowingly quite false.[2] In any case, we need not settle here just what is expressed by phenomenal predicates in public language. All that matters is that both concepts are available.

[2] These cases may not be entirely hypothetical. Nida-Rümelin (1996) gives reasons, based on the neurobiological and genetic bases of colorblindness, to believe that a small fraction of the population may actually be spectrum-inverted with respect to the rest of us. If so, it is natural to wonder just what their phenomenal expressions refer to.

Phenomenal properties can also be picked out indexically. When seeing the tomato, I can refer indexically to a visual quality associated with it, using a concept I might express by saying 'this quality' or 'this sort of experience'. These expressions express a demonstrative concept that we might call *E*. *E* functions in an indexical manner, roughly by picking out whatever quality the subject is currently ostending. Like other demonstratives, it has a "character", which fixes reference in a context roughly by picking out whatever quality is ostended in that context; and it has a distinct "content", corresponding to the quality that is actually ostended — in this case, phenomenal redness. The demonstrative concept *E* rigidly designates its referent, so that it picks out the quality in question even in counterfactual worlds in which no one is ostending the quality.

The three concepts red_C, red_I, and *E* may all refer to the same quality, phenomenal redness. In each case, reference is fixed relationally, with the characterized in terms of its relations to external objects or acts of ostension. There is another crucial phenomenal concept in the vicinity, one that does not pick out phenomenal redness relationally, but rather picks it out directly, in terms of its intrinsic phenomenal nature. This is what we might call a *pure phenomenal concept*.

To see the need for the pure phenomenal concept, consider the knowledge that Mary gains when she learns for the first time what it is like to see red. She learns that seeing red has such-and-such quality. Mary learns (or reasonably comes to believe) that red things will typically cause experiences of such-and-such quality in her, and in other members of her community. She learns (or gains the cognitively significant belief) that the experience she is now having has such-and-such quality, and that the quality she is now ostending is such-and-such. Call Mary's "such-and-such" concept here *R*. (Note that the phenomenal concept *R* should be distinguished from the phenomenal quality R (unitalicized) that it refers to.)

Mary's concept *R* picks out phenomenal redness, but it is quite distinct from the concepts red_C, red_I, and *E*. We can see this by using cognitive significance as a test for difference between concepts. Mary gains the belief $red_C = R$ — that the quality typically caused in her community by red things is such-and-such — and this belief is cognitively significant knowledge. She gains the cognitively significant belief $red_I = R$ in a similar way. And she gains the belief $E = R$ — roughly, that the quality she is now ostending is such-and-such.

Mary's belief $E = R$ is as cognitively significant as any other belief in which the object of a demonstrative is independently characterized: e.g. my belief *I am David Chalmers*, or my belief *that object is tall*. For Mary, $E = R$ is not a priori. No a priori reasoning can rule out the hypothesis that she is now ostending some other quality entirely, just as no a priori reasoning can rule out the hypothesis that I am David Hume, or that the object I am pointing to is short.

Indeed, nothing known a priori entails that the phenomenal quality R is ever instantiated in the actual world.

It is useful to consider analogies with other demonstrative knowledge of types. Let $this_S$ be a demonstrative concept of shapes ("this shape"). Jill might tell Jack that she is about to show him her favorite shape. When she shows him a circle, he might form the thought *Jill's favorite shape is* $this_S$. This is a demonstrative thought, where this instance of $this_S$ picks out the shape of a circle. He might also form the thought *Jill's favorite shape is circle*. This is a non-demonstrative thought: instead of a demonstrative concept, the right hand side uses what we might call a *qualitative* concept of the shape of a circle. Finally, he might form the thought $this_S$ *is circle*. This is a substantive, nontrivial thought, taking the form of an identity involving a demonstrative concept and a qualitative concept. Here, as in the examples above, one conceives the object of a demonstration *as* the object of a demonstration ("this shape, whatever it happens to be"), and at the same time attributes it substantive qualitative properties, conceived non-demonstratively.

Of course Jack's concept *circle* (unlike Mary's concept *R*), is an old concept, previously acquired. But this is inessential to Jack's case. We can imagine that Jack has never seen a circle before, but that on seeing a circle for the first time, he acquires the qualitative concept of circularity. He will then be in the position to think the qualitative thought *Jill's favorite shape is circle*, and to think the substantive demonstrative-qualitative thought $this_S$ *is circle*.

Mary's situation is analogous. Where Jack thinks the substantive thought $this_S$ *is circle*, Mary might think the substantive thought $E = R$ ("this quality is R"). Like Jack's thought, Mary's thought involves attributing a certain substantive qualitative nature to a type that is identified demonstratively. This qualitative nature is attributed using a qualitative concept of phenomenal redness, acquired upon having a red experience for the first time. Her thoughts $red_C = R$ and $red_I = R$ are substantive thoughts analogous to Jack's thought *Jill's favorite shape is circle*. Her crucial thought $E = R$ is a substantive thought involving both a demonstrative and a qualitative concept, and is as cognitively significant as Jack's thought $this_S$ *is circle.*

So the concept *R* is quite distinct from red_C, red_I, and *E*. We might say that unlike the other concepts, the pure phenomenal concept characterizes the phenomenal quality *as* the phenomenal quality that it is.

The concept *R* is difficult to express directly in language, since the most natural terms, such as 'phenomenal redness' and 'this experience', arguably express other concepts such as red_C and *E*. Still, one can arguably discern uses of these terms that express pure phenomenal concepts; or if not, one can stipulate such uses. For example, Chisholm (1957) suggests that

there is a "non-comparative" sense of expressions such as 'looks red'; this sense seems to express a pure phenomenal concept, whereas his "comparative sense" seems to express a relational phenomenal concept.[3] And at least informally, demonstratives are sometimes used to express pure phenomenal concepts. For example, the belief that E = R might be informally expressed by saying something like "this quality is *this* quality".

It may be that there is a sense in which *R* can be regarded as a "demonstrative" concept. I will not regard it this way: I take it that demonstrative concepts work roughly as analysed by Kaplan (1989), so that they have a reference-fixing "character" that leaves their referent open. This is how *E* behaves: its content might be glossed roughly as "this quality, whatever it happens to be". *R*, on the other hand, is a substantive concept that is tied a priori to a specific sort of quality, so it does not behave the way that Kaplan suggests that a demonstrative should. Still, there is an intimate relationship between pure and demonstrative phenomenal concepts that I will discuss later; and if someone wants to count pure phenomenal concepts as "demonstrative" in a broad sense (perhaps regarding *E* as 'indexical'), there is no great harm in doing so, as long as the relevant distinctions are kept clear. What matters for my purposes is not the terminological point, but the more basic point that the distinct concepts *E* and *R* exist.

The relations among these concepts can be analysed straightforwardly using the two-dimensional framework for representing the content of concepts. A quick introduction to this framework is given in an appendix; more details can be found in Chalmers (2002*c*). The central points in what follows should be comprehensible if matters involving the two-dimensional framework are skipped, but the framework makes the analysis of some crucial points much clearer.

According to the two-dimensional framework, when an identity $A = B$ is a posteriori, the concepts *A* and *B* have different epistemic (or primary) intensions. If *A* and *B* are rigid concepts and the identity is true, *A* and *B* have the same subjunctive (or secondary) intensions. So we should expect that the concepts red_C, red_I, *E*, and *R* have different epistemic intensions, but the same subjunctive intension. And this is what we find. The subjunctive intension of each picks out phenomenal redness in all worlds. The epistemic intension of red_C picks out, in a given centred world, roughly the quality typically caused by certain paradigmatic objects in

[3] The distinction also roughly tracks Nida-Rümelin's (1995; 1997) distinction between "phenomenal" and "non-phenomenal" readings of belief attributions concerning phenomenal states. "Phenomenal" belief attributions seem to require that the subject satisfies the attribution by virtue of a belief involving a pure phenomenal concept, while "non-phenomenal" attributions allow that the subject can satisfy the attribution by virtue of a belief involving a relational phenomenal concept.

the community of the subject at the centre of the world. The epistemic intension of red_I picks out roughly the quality typically caused by those objects in the subject at the centre.

As for the demonstrative concept *E*: to a first approximation, one might hold that its epistemic intension picks out the quality that is ostended by the subject at the centre. This characterization is good enough for most of our purposes, but it is not quite correct. It is possible to ostend two experiences simultaneously and invoke two distinct demonstrative concepts, as when one thinks *that quality differs from that quality*, ostending two different parts of a symmetrical visual field (see Austin 1990). Here no descriptive characterization such as the one above will capture the difference between the two concepts. It is better to see *E* as a sort of indexical, like *I* or *now*. To characterize the epistemic possibilities relevant to demonstrative phenomenal concepts, we need centred worlds whose centres contain not only a "marked" subject and time, but also one or more marked experiences; in the general case, a sequence of such experiences.[4] Then a concept such as *E* will map a centred world to the quality of the "marked" experience (if any) in that world. Where two demonstrative concepts E_1 and E_2 are involved, as above, the relevant epistemic possibilities will contain at least two marked experiences, and we can see E_1 as picking out the quality of the first marked experience in a centred world, and E_2 as picking out the quality of the second. Then the belief above will endorse all worlds at which the quality of the first marked experience differs from the quality of the second. This subtlety will not be central in what follows.

The epistemic intension of *R* is quite distinct from all of these. It picks out phenomenal redness in all worlds. I will analyse this matter in more depth shortly; but one can see intuitively why this is plausible. When Mary believes *roses cause R experiences*, or *I am currently having an R experience*, she thereby excludes all epistemic possibilities in which roses cause some other quality (such as G, phenomenal greenness), or in which she is experiencing some other quality: only epistemic possibilities involving phenomenal redness remain.

The cognitive significance of identities such as $red_C = R$, $red_I = R$, and $E = R$ is reflected in the differences between the concept's epistemic intensions. The first two identities endorse all epistemic possibilities in which paradigmatic objects stand in the right relation to experiences of R; these are only a subset of the epistemic possibilities available a priori. The third identity endorses all epistemic possibilities in which the marked experience at the centre (or the ostended experience, on the rough characterization) is R. Again, there are many epistemic possibilities (a priori) that are not like this: centred worlds in which the marked

[4] In the experience-based framework: if experiences do not map one-to-one to instances of phenomenal properties, then instances of phenomenal properties should be marked instead.

experience is G, for example. Once again, this epistemic contingency reflects the cognitive significance of the identity.

(Phenomenal realists (e.g. Loar 1997; Hawthorne 2001) analysing what Mary learns have occasionally suggested that her phenomenal concept is a demonstrative concept. This is particularly popular as a way of resisting anti-materialist arguments, as it is tempting to invoke the distinctive epistemic and referential behavior of demonstrative concepts in explaining why an epistemic gap does not reflect an ontological gap. But on a closer look it is clear that Mary's central phenomenal concept *R* (the one that captures what she learns) is *distinct* from her central demonstrative concept *E*, as witnessed by the non-trivial identity $E = R$, and is not a demonstrative concept in the usual sense. This is not just a terminological point. Those who use these analyses to rebut anti-materialist arguments typically rely on analogies with the epistemic and referential behavior of ordinary (Kaplan-style) demonstratives. In so far as these analyses rely on such analogies, they mischaracterize Mary's new knowledge. Something similar applies to analyses that liken phenomenal concepts to indexical concepts (e.g. Ismael 1999; Perry 2001). If my analysis is correct, then pure phenomenal concepts (unlike demonstrative phenomenal concepts) are not indexical concepts at all.)

2.2 Inverted Mary

We can now complicate the situation by introducing another thought experiment on top of the first one. Consider the case of *Inverted Mary*, who is physically, functionally, and environmentally just like Mary, except that her phenomenal color vision is red/green inverted. (I will assume for simplicity that Inverted Mary lives in a community of inverted observers.) Like Mary, Inverted Mary learns something new when she sees red things for the first time. But Inverted Mary learns something different from what Mary learns. Where Mary learns that tomatoes cause experiences of (what we call) phenomenal redness, Inverted Mary learns that they cause experiences of (what we call) phenomenal greenness. In the terms given earlier, Mary acquires beliefs $red_C = R$, $red_I = R$, and $E = R$, while Inverted Mary acquires beliefs $red_C = G$, $red_I = G$, and $E = G$ (where *G* is the obvious analogue of *R*). So Mary and Inverted Mary acquire beliefs with quite different contents.

This is already enough to draw a strong conclusion about the irreducibility of content. Recall that Mary and Inverted Mary are physical/functional and environmental twins, even after they see red things for the first time. Nevertheless, they have beliefs with different contents. It follows that belief content does not supervene conceptually on physical/functional

properties. And it follows from this that intentional properties are not conceptually supervenient on physical/functional properties, in the general case.

This is a non-trivial conclusion. Phenomenal realists often hold that while the phenomenal is conceptually irreducible to the physical and functional, the intentional can be analysed in functional terms. But if what I have said here is correct, then this irreducibility cannot be quarantined in this way. If the phenomenal is conceptually irreducible to the physical and functional, so too is at least one aspect of the intentional: the content of phenomenal beliefs.

At this point, there is a natural temptation to downplay this phenomenon by reducing it to a sort of dependence of belief content on reference that is found in many other cases: in particular in the cases that are central to externalism about the content of belief. Take Putnam's case of Twin Earth. Oscar and Twin Oscar are functional duplicates, but they inhabit different environments: Oscar's contains H_2O as the clear liquid in the oceans and lakes, while Twin Oscar's contains XYZ (which we count not as water but as twin water). As a consequence, Oscar's *water* concept refers to water (H_2O), while Twin Oscar's analogous concept refers to twin water (XYZ). Because of this difference in reference, Oscar and Twin Oscar seem to have different beliefs: Oscar believes that water is wet, while Twin Oscar believes that twin water is wet. Perhaps the case of Mary and Inverted Mary is just like this?[5]

The analogy does not go through, however. Or rather, it goes through only to a limited extent. Oscar and Twin Oscar's *water* concepts here are analogous to Mary and Inverted Mary's relational phenomenal concepts (red_C or red_I), or perhaps to their demonstrative concepts. For example, the relational concepts that they express with their public-language expressions 'red experience' will refer to two different properties, phenomenal redness and phenomenal greenness. Mary and Inverted Mary can deploy these concepts in certain beliefs, such as the beliefs that they express by saying 'Tomatoes cause red experiences', even before they leave their monochromatic rooms for the first time. Because of the distinct referents of their concepts, there is a natural sense (Nida-Rümelin's "non-phenomenal" sense) in which we can say that Mary believed that tomatoes caused red experiences, while Inverted Mary did not; she believed that tomatoes caused green experiences. Here the analogy goes through straightforwardly.

The pure phenomenal concepts *R* and *G*, however, are less analogous to the two *water* concepts than to the chemical concepts H_2O and *XYZ*. When Oscar learns the true nature of water, he acquires the new belief *water* = H_2O, while Twin Oscar acquires an analogous

[5] This sort of treatment of phenomenal belief is suggested by Francescotti (1994).

belief involving *XYZ*. When Mary learns the true nature of red experiences, she acquires a new belief $red_C = R$, while Inverted Mary acquires an analogous belief involving *G*. That is, Mary and Inverted Mary's later knowledge involving *R* and *G* is fully lucid knowledge of the referents of the concepts in question, analogous to Oscar and Twin Oscar's knowledge involving the chemical concepts H_2O and *XYZ*.

But here we see the strong disanalogy. Once Oscar acquires the chemical concept H_2O and Twin Oscar acquires *XYZ*, they will no longer be twins: their functional properties will differ significantly. By contrast, at the corresponding point Mary and Inverted Mary are still twins. Even though Mary has the pure phenomenal concept *R* and Inverted Mary has *G*, their functional properties are just the same. So the difference between the concepts *R* and *G* across functional twins is something that has no counterpart in the standard Twin Earth story.

All this reflects the fact that in standard externalist cases, the pairs of corresponding concepts may differ in reference, but they have the same or similar *epistemic* or *notional* contents. Oscar and Twin Oscar's *water* concepts have different referents (H_2O vs. XYZ), but they have the same epistemic contents: both intend to refer to roughly the liquid around them with certain superficial properties. Something like this applies to Mary's and Inverted Mary's relational phenomenal concepts, which have different referents but the same epistemic content (which picks out whatever quality stands in a certain relation), and to their demonstrative concepts (which pick out roughly whatever quality happens to be ostended).

In terms of the two-dimensional framework, where epistemic contents correspond to epistemic intensions: Oscar's and Twin Oscar's *water* concepts have the same epistemic intension but different subjunctive intensions. A similar pattern holds in all the cases characteristic of standard externalism. The pattern also holds for Mary's and Inverted Mary's relational phenomenal concepts, and their demonstrative phenomenal concepts.

But Mary's concept *R* and Twin Mary's concept *G* have *different* epistemic contents. In this way they are analogous to Oscar's concept H_2O and Twin Oscar's concept *XYZ*. But again, the disanalogy is that *R* and *G* are possessed by twins, and H_2O and *XYZ* are not. So the case of Inverted Mary yields an entirely different phenomenon: a case in which *epistemic* content differs between twins.

This can be illustrated by seeing how the concepts in question are used to constrain epistemic possibilities. When Oscar confidently believes that there is water in the glass, he is not thereby in a position to rule out the epistemic possibility that there is XYZ in the glass (unless he has some further knowledge, such as the knowledge that water is H_2O). The same goes for Twin Oscar's corresponding belief. For both of them, it is equally epistemically possible that the glass contains H_2O and that it contains XYZ. Any epistemic possibility

compatible with Oscar's belief is also compatible with Twin Oscar's belief: in both cases, these will be roughly those epistemic possibilities in which a sample of the dominant watery stuff in the environment is in the glass.

Epistemic content reflects the way that a belief constrains the space of epistemic possibilities, so Oscar's and Twin Oscar's epistemic contents are the same. Something similar applies to Mary and Inverted Mary, at least where their pairwise relational and demonstrative phenomenal concepts are concerned. When Mary confidently believes (under her relational concept) that her mother is having a red experience, for example, she is not thereby in a position to rule out the epistemic possibility that her mother is having an experience with the quality G. Both Mary's and Inverted Mary's beliefs are compatible with any epistemic possibility in which the subject's mother is having the sort of experience typically caused in the community by paradigmatic red objects. So their beliefs have the same epistemic contents.

But Mary's and Inverted Mary's pure phenomenal concepts do not work like this. Mary's concept *R* and Inverted Mary's concept *G* differ not just in their referents but also in their epistemic contents. When Mary leaves the monochromatic room and acquires the confident belief (under her pure phenomenal concept) that tomatoes cause red experiences, she is thereby in a position to rule out the epistemic possibility that tomatoes cause experiences with quality G. The only epistemic possibilities compatible with her belief are those in which tomatoes cause R experiences. For Inverted Mary, things are reversed: the only epistemic possibilities compatible with her belief are those in which tomatoes cause G experiences. So their epistemic contents are quite different.

Again, the epistemic situation with *R* and *G* is analogous to the epistemic situation with the concepts H_2O and *XYZ*. When Oscar believes (under a fully lucid chemical concept) that the glass contains H_2O, he is thereby in a position to rule out all epistemic possibilities in which the glass contains XYZ. For Twin Oscar, things are reversed. This is to say that H_2O and *XYZ* have different epistemic contents. The same goes for *R* and *G*.

So in the case of the pure phenomenal concepts, uniquely, we have a situation in which two concepts differ in their epistemic content despite the subjects being physically identical. So phenomenal concepts seem to give a case in which even epistemic content is not conceptually supervenient on the physical.

Using the two-dimensional framework: the epistemic intension of a concept reflects the way it applies to epistemic possibilities. We saw above that the epistemic intensions of Oscar's and Twin Oscar's *water* concepts are the same, as are the epistemic intensions of Mary's and Inverted Mary's relational and demonstrative phenomenal concepts. But *R* and *G* differ in the way they apply to epistemic possibilities, and their epistemic intensions differ

accordingly: the epistemic intension of *R* picks out phenomenal redness in all worlds, and the epistemic intension of *G* picks out phenomenal greenness in all worlds. When Mary thinks *I am having an R experience now*, the epistemic intension of her thought is true at all and only those worlds in which the being at the centre is having an R experience.

Something very unusual is going on here. In standard externalism, and in standard cases of so-called "direct reference", a referent plays a role in constituting the subjunctive content (subjunctive intension) of concepts and beliefs, while leaving the epistemic content (epistemic intension) unaffected. In the pure phenomenal case, by contrast, the quality of the experiences plays a role in constituting the *epistemic* content of the concept and of the corresponding belief. One might say very loosely that in this case, the referent of the concept is somehow present inside the concept's sense, in a way much stronger than in the usual cases of "direct reference".

We might say that the pure phenomenal concept is *epistemically rigid*: its epistemic content picks out the same referent in every possible world (considered as actual). By contrast, ordinary rigid concepts are merely *subjunctively rigid*, with a subjunctive content that picks out the same referent in every possible world (considered as counterfactual). Epistemically rigid concepts will typically be subjunctively rigid, but most subjunctively rigid concepts are not epistemically rigid. Pure phenomenal concepts are both epistemically and subjunctively rigid.[6]

One might see here some justification for Russell's claim that we have a special capacity for direct reference to our experiences.[7] Contemporary direct reference theorists hold that Russell's view was too restrictive, and that we can make direct reference to a much broader class of entities. But the cases they invoke are "direct" only in the weak sense outlined above: the subjunctive content depends on the referent, but the epistemic content of the concept does

---

[6] Further: epistemically rigid concepts will usually be subjunctively rigid *de jure*, which entails that they are what Martine Nida-Rümelin calls (in a forthcoming article) *super-rigid*: they pick out the same referent relative to all pairs of scenarios considered as actual and worlds considered as counterfactual. When represented by a two-dimensional matrix, super-rigid concepts have the same entry at each point of the matrix.

[7] Russell also held that direct reference is possible to universals, and perhaps to the self. It is arguable that for at least some universals (in the domains of mathematics or of causation, perhaps), one can form an epistemically rigid concept whose epistemic content picks out instances of that universal in all worlds. So there is at least a limited analogy here, though it seems unlikely that in these cases the content of such a (token) concept is directly constituted by an underlying instance of the universal, in the manner suggested below.

There is no analogous phenomenon with the self. There may, however, be a different sense in which we can make "direct reference" to the self, to the current time, and to particular experiences: this is the sort of direct indexical reference that corresponds to the need to build these entities into the centre of a centred world. We can refer to these "directly" (in a certain sense) under indexical concepts; but we cannot form concepts whose epistemic contents reflect the referents in question. This suggests that direct reference to particulars and direct reference to properties are quite different phenomena.

not. In the phenomenal case, the epistemic content itself seems to be constituted by the referent. It is not hard to imagine that some such epistemic requirement on direct reference is what Russell had in mind.

3 The Constitution of Phenomenal Beliefs

3.1 Direct phenomenal concepts and beliefs

We have seen that the content of phenomenal concepts and phenomenal beliefs does not supervene conceptually on physical properties. Does this content supervene conceptually on some broader class of properties, and if so, on which? I will offer an analysis of how the content of pure phenomenal concepts is constituted. I will not give a knockdown argument for this analysis by decisively refuting all alternatives, but I will offer it as perhaps the most natural and elegant account of the phenomena, and as an account that can in turn do further explanatory work.

To start with, it is natural to hold that the content of phenomenal concepts and beliefs supervenes conceptually on the combination of physical and phenomenal properties. Mary and Inverted Mary are physical twins, but they are phenomenally distinct, and this phenomenal distinctness (Mary experiences phenomenal redness, Inverted Mary experiences phenomenal greenness) precisely mirrors their intentional distinctness (Mary believes that tomatoes cause R experiences, Inverted Mary believes that tomatoes cause G experiences). It is very plausible to suppose that their intentional distinctness holds in virtue of their phenomenal distinctness.

The alternative is that the intentional content of the phenomenal concept is conceptually independent of both physical and phenomenal properties. If that is so, it should be conceivable that two subjects have the same physical and phenomenal properties, while having phenomenal beliefs that differ in content. Such a case might involve Mary and Mary' as physical and phenomenal twins, who are both experiencing phenomenal redness for the first time (while being phenomenally identical in all other respects), with Mary acquiring the belief that tomatoes cause R experiences while Mary' acquires the belief that tomatoes cause G experiences. It is not at all clear that such a case is conceivable.

Another possibility is that the intentional content of Mary's phenomenal concept in question might be determined by phenomenal states *other* than the phenomenal redness that Mary is visually experiencing. For example, maybe Mary's belief content is determined by a faint phenomenal "idea" that goes along with her phenomenal "impression", where the former is not conceptually determined by the latter, and neither is conceptually determined by the

physical. In that case, it should once again be conceivable that twins Mary and Mary' both visually experience phenomenal redness upon leaving the room, with Mary acquiring the belief that tomatoes cause R experiences while Mary' acquires the belief that tomatoes cause G experiences, this time because of a difference in their associated phenomenal ideas. But again, it is far from clear that this is conceivable.

There is a very strong intuition that the content of Mary's phenomenal concept and phenomenal belief is *determined* by the phenomenal character of her visual experience, in that it will vary directly as a function of that character in cases where that character varies while physical and other phenomenal properties are held fixed, and that it will not vary independently of that character in such cases. I will adopt this claim as a plausible working hypothesis.

In particular, I will take it that in cases such as Mary's, the content of a phenomenal concept and a corresponding phenomenal belief, is partly *constituted* by an underlying phenomenal quality, in that the content will mirror the quality (picking out instances of the quality in all epistemic possibilities), and in that across a wide range of nearby conceptually possible cases in which the underlying quality is varied while background properties are held constant, the content will co-vary to mirror the quality. Let us call this sort of phenomenal concept a *direct phenomenal concept*.

Not all experiences are accompanied by corresponding direct phenomenal concepts. Many of our experiences appear to pass without our forming any beliefs about them, and without the sort of concept formation that occurs in the Mary case. The clearest cases of direct phenomenal concepts arise when a subject attends to the quality of an experience, and forms a concept wholly based on the attention to the quality, "taking up" the quality into the concept. This sort of concept formation can occur with visual experiences, as in the Mary case, but it can equally occur with all sorts of other experiences: auditory and other perceptual experiences, bodily sensations, emotional experiences, and so on. In each case we can imagine the analogue of Mary having such an experience for the first time, attending to it, and coming to have a concept of what it is like to have it. There is no reason to suppose that this sort of concept formation is restricted to entirely novel experiences. I can experience a particular shade of phenomenal redness for the hundredth time, attend to it, and form a concept of what it is like to have that experience, a concept whose content is based entirely on the character of the experience.

Direct phenomenal concepts can be deployed in a wide variety of beliefs, and other propositional attitudes. When Mary attends to her phenomenally red experience and forms her direct phenomenal concept *R*, she is thereby in a position to believe that tomatoes cause R

experiences, to believe that others have R experiences, to believe that she previously had no R experiences, to desire more R experiences, and so on.

Perhaps the most crucial sort of deployment of a direct phenomenal concept occurs when a subject predicates the concept of the very experience responsible for constituting its content. Mary has a phenomenally red experience, attends to it, and forms the direct phenomenal concept *R*, and forms the belief *this experience is R*, demonstrating the phenomenally red experience in question. We can call this special sort of belief a *direct phenomenal belief*.

We can also cast this idea within an experience-free ontology of qualities. In this framework, we can say that a direct phenomenal concept is formed by attending to a quality and taking up that quality into a concept whose content mirrors the quality, picking out instances of the quality in all epistemic possibilities. A direct phenomenal belief is formed when the referent of this direct phenomenal concept is identified with the referent of a corresponding demonstrative phenomenal concept, e.g. when Mary forms the belief that *this quality is R*. The general form of a direct phenomenal belief in this framework is $E = R$, where *E* is a demonstrative phenomenal concept and *R* is the corresponding direct phenomenal concept.

3.2 Some notes on direct phenomenal beliefs

1. For a direct phenomenal belief, it is required that the demonstrative and direct concepts involved be appropriately "aligned". Say that Mary experiences phenomenal redness in both the left and right halves of her visual field, forms a direct phenomenal concept *R* based on her attention to the left half, forms a demonstrative concept of phenomenal redness based on her attention to the right half, and identifies the two by a belief of the form $E = R$. Then this is not a direct phenomenal belief, even though the same quality (phenomenal redness) is referred to on both sides, since the concepts are grounded in different instances of that quality. The belief has the right sort of content, but it does not have the right sort of constitution.

To characterize the required alignment more carefully we can note that all direct phenomenal concepts, like all demonstrative phenomenal concepts, are based in acts of attention to instances of phenomenal qualities. A direct phenomenal concept such as *R* does not characterize a quality *as* an object of attention, but it nevertheless requires attention to a quality for its formation. The same act of attention can also be used to form a demonstrative phenomenal concept *E*. A direct phenomenal belief (in the quality-based framework) will be a belief of the form $E = R$ where the demonstrative phenomenal concept *E* and the direct phenomenal concept *R* are *aligned*: that is, where they are based in the same act of attention.

One can simplify the language by regarding the act of attention as a demonstration. We can then say that both demonstrative and direct phenomenal concepts are based in demonstrations, and that a direct phenomenal belief is a belief of the form $E = R$ where the two concepts are based in the same demonstration.[8]

2. As with all acts of demonstration and attention, phenomenal demonstration and attention involves a cognitive element. Reference to a phenomenal quality is determined in part by cognitive elements of a demonstration. These cognitive elements will also enter into determining the content of a corresponding direct phenomenal concept.

Consider two individuals with identical visual experiences. These individuals might engage in different acts of demonstration — e.g. one might demonstrate a red quality experienced in the right half of the visual field, and the other a green quality experienced in the left half of the visual field — and thus form distinct direct phenomenal concepts. Or they might attend to the same location in the visual field, but demonstrate distinct qualities associated with that location: e.g. one might demonstrate a highly specific shade of phenomenal redness, and the other a less specific shade, again resulting in distinct direct phenomenal concepts. These differences will be due to differences in the cognitive backgrounds of the demonstrations in the two individuals. I will be neutral here about whether such cognitive differences are themselves constituted by underlying functioning, aspects of cognitive phenomenology, or both.

One can imagine varying the visual experiences and the cognitive background here independently. Varying visual experiences might yield a range of cases in which direct phenomenal concepts of phenomenal redness, greenness, and other hues are formed. Varying the cognitive background might yield a range of cases in which direct phenomenal concepts of different degrees of specificity (for example) are formed.

Along with this cognitive element comes the possibility of failed demonstration, if the cognitive element and the targeted experiential elements mismatch sufficiently. Take Nancy, who attends to a patch of phenomenal color, acting cognitively as if to demonstrate a highly specific phenomenal shade. Nancy has not attended sufficiently closely to notice that the patch has a non-uniform phenomenal color: let us say it is a veridical experience of a square colored with different shades of red on its left and right side.[9] In such a case, the

---

[8] Gertler (2001) has independently developed a related account of phenomenal introspection, according to which a phenomenal state is introspected when it is "embedded" in another state, and when the second state constitutes demonstrative attention to the relevant content by virtue of this embedding. On my account, things are the other way around: any "embedding" holds in virtue of demonstrative attention, rather than the reverse.

[9] This sort of case was suggested to me by Delia Graff and Mark Johnston.

demonstrative phenomenal concept will presumably refer to no quality at all: given its cognitive structure, it could refer only to a specific quality, but it would break symmetry for it to refer to either instantiated quality, and presumably uninstantiated qualities cannot be demonstrated.

What of any associated direct phenomenal concept? It is not out of the question that the subject forms *some* substantive concept where a direct phenomenal concept would normally be formed; perhaps a concept of an intermediate uninstantiated shade of phenomenal red, at least if the instantiated shades are not too different. Like a direct phenomenal concept, this concept will have a content that depends constitutively on associated qualities of experience (Inverted Nancy might form a concept of an intermediate phenomenal green), but it will not truly be a direct phenomenal concept, since its content will not directly mirror an underlying quality.

This possibility of cognitive mismatch affects the path from demonstration to a demonstrated phenomenal quality, but given that a phenomenal quality is truly demonstrated, it does not seem to affect the path from demonstrated phenomenal quality to a direct phenomenal concept. That is, as long as a phenomenal quality is demonstrated, and the cognitive act typical of forming a direct phenomenal concept based on such a demonstration is present, a direct phenomenal concept will be formed.

We might call a concept that shares the cognitive structure of a direct phenomenal concept a *quasi-direct* phenomenal concept; and we can call a belief with the same cognitive structure as a direct phenomenal belief a *quasi-direct* phenomenal belief. Like a direct phenomenal concept, a quasi-direct phenomenal concept arises from an act of (intended) demonstration, along with a characteristic sort of cognitive act. Unlike a direct phenomenal concept, a quasi-direct phenomenal concept is not required to have a content that is constituted by an underlying quality. Nancy's concept above is a quasi-direct phenomenal concept but not a direct phenomenal concept, for example.

We can call a quasi-direct phenomenal concept that is not a direct phenomenal concept a *pseudo-direct* phenomenal concept, and we can define a pseudo-direct phenomenal belief similarly. If the suggestion above is correct, then the only pseudo-direct phenomenal concepts are like Nancy's, in involving an unsuccessful demonstration. As long as a quasi-direct phenomenal concept is grounded in a successful demonstration, it will be a direct phenomenal concept. I will return to this claim later.

3. All direct phenomenal concepts are pure phenomenal concepts, but not all pure phenomenal concepts are direct phenomenal concepts. To see this, note that Mary may well retain some knowledge of what it is like to see tomatoes even after she goes back into her

black-and-white room, or while she shuts her eyes, or while she looks at green grass. She still has a concept of phenomenal redness than can be deployed in various beliefs, with the sort of epistemic relations to relational and demonstrative phenomenal concepts that is characteristic of pure phenomenal concepts. Inverted Mary (still Mary's physical twin) has a corresponding concept deployed in corresponding beliefs that *differ* in content from Mary's. As before, their corresponding beliefs *differ* in epistemic content, including and excluding different classes of epistemic possibilities. Mary's concept is still a concept of phenomenal redness as the quality it is, based on a lucid understanding of that quality, rather than on a mere relational or demonstrative identification. So as before, it is a pure phenomenal concept. But it is not a direct phenomenal concept, since there is no corresponding experience (or instantiated quality) that is being attended to or taken up into the concept. We can call this sort of concept a *standing* phenomenal concept, since it may persist in a way that direct phenomenal concepts do not.

There are some differences in character between direct and standing phenomenal concepts. Direct phenomenal concepts may be very fine-grained, picking out a very specific phenomenal quality (a highly specific shade of phenomenal redness, for example). Standing phenomenal concepts are usually more coarse-grained, picking out less specific qualities. One can note this phenomenologically from the difficulty of "holding" in mind specific qualities as opposed to coarser categories when relevant visual experiences are not present; and this is also brought out by empirical results showing the difficulty of reidentifying specific qualities over time.[10] It usually seems possible for a direct phenomenal concept to yield a corresponding standing phenomenal concept as a "successor" concept once the experience in question disappears, at the cost of some degree of coarse-graining.

As with direct phenomenal concepts, the content of standing phenomenal concepts does not conceptually supervene on the physical (witness Mary and Inverted Mary, back in their rooms). A question arises as to what determines their content. I will not try to analyse that matter here, but I think it is plausible that their content is determined by some combination of (1) non-sensory phenomenal states of a cognitive sort, which bear a relevant relation to the original phenomenal quality in question — e.g. a faint Humean phenomenal "idea" that is relevantly related to the original "impression"; (2) dispositions to have such states; and (3) dispositions to recognize instances of the phenomenal quality in question. It is not implausible that Mary and Inverted Mary (back in their rooms) still differ in some or all of these respects,

[10] See Raffman 1995 for a discussion of these results in an argument for an anti-representationalist "presentational" analysis of phenomenal concepts that is very much compatible with the analysis here.

and that these respects are constitutively responsible for the difference in the content of their concepts.

One might be tempted to use the existence of standing phenomenal concepts to argue against the earlier analysis of direct phenomenal concepts (that is, of concepts akin to those Mary acquires on first experiencing phenomenal redness) as constituted by the quality of the relevant instantiated experience. Why not assimilate them to standing phenomenal concepts instead, giving a unified account of the two? In response, note first that it remains difficult to conceive of the content of direct phenomenal concepts varying independently of the phenomenal quality in question, whereas it does not seem so difficult to conceive of the content of standing phenomenal concepts varying independently. And second, note that the difference in specificity between direct and standing phenomenal concepts gives some reason to believe that they are constituted in different ways.

The lifetime of a direct phenomenal concept is limited to the lifetime of the experience (or the instantiated quality) that constitutes it. (In some cases a specific phenomenal concept might persist for a few moments due to the persistence of a vivid iconic memory, but even this will soon disappear.) Some might worry that this lack of persistence suggests that it is not a concept at all, since concepthood requires persistence. This seems misguided, however: it is surely possible for a concept to be formed moments before a subject dies. The concepts in question are still predicable of any number of entities, during their limited lifetimes, and these predications can be true or false (e.g. Mary may falsely believe that her sister is currently experiencing R). This sort of predicability, with assessibility for truth or falsehood, seems sufficient for concepthood; at least it is sufficient for the uses of concepthood that will be required here.

4. As with pure phenomenal concepts generally, we do not have public language expressions that distinctively express the content of direct phenomenal concepts. Public reference to phenomenal qualities is always fixed relationally, it seems: by virtue of a relation to certain external stimuli, or certain sorts of behavior, or certain demonstrations. (Recall Ryle's remark that there are no "neat" sensation words.) Of course Mary can express a pure phenomenal concept by introducing her own term, such as 'R', or by using an old term, such as 'red', with this stipulated meaning. But this use will not be public, at least in the limited sense that there is no method by which we can ensure that other members of the community will use the term with the same epistemic content. One can at best ensure that they pick out the same quality by picking it out under a different epistemic content (e.g. as the quality Mary is having at a certain time), or by referring through semantic deference (as the quality that Mary picks out with 'R'). In this sense it seems that any resulting language will be "private":

it can be used with full competence by just one subject, and others can use it only deferentially. (An exception may arguably be made for terms expressing *structural* pure phenomenal concepts — e.g. phenomenal similarity and difference and perhaps phenomenal spatial relations — which arguably do not rely on relational reference-fixing.)

Of course the view I have set out here is just the sort of view that Wittgenstein directed his "private language" argument against. The nature of the private language argument is contested, so in response I can say only that I have seen no reconstruction of it that provides a strong case against the view I have laid out. Some versions of the argument seem to fall prey to the mistake just outlined, that of requiring a strong sort of "repeatability" for concept possession (and an exceptionally strong sort at that, requiring the recognisability of correct repeated application). A certain sort of repeatability is required for concept possession, but it is merely the "hypothetical repeatability" involved in *present* predicability of the concept to actual and hypothetical cases, with associated truth-conditions. Another reconstruction of the argument, that of Kripke (1981), provides no distinctive traction against my analysis of direct phenomenal concepts: any force that it has applies to concepts quite generally.

(One might even argue that Kripke's argument provides *less* traction in the case of direct phenomenal concepts, as this is precisely a case in which we can see how a determinate application-condition can be constituted by an underlying phenomenal quality. Kripke's remarks about associated phenomenal qualities (41-51) — e.g. a certain sort of "headache" — being irrelevant to the content of concepts such as addition apply much less strongly where *phenomenal* concepts are concerned. Of course there is more to say here, but in any case it is a curiosity of Kripke's reconstruction of the argument that it applies least obviously to the phenomena at which Wittgenstein's argument is often taken to be aimed.)

4 The Epistemology of Phenomenal Belief

4.1 Incorrigibility

A traditional thesis in the epistemology of mind is that first-person beliefs about phenomenal states are *incorrigible*, or *infallible* (I use these terms equivalently), in that they cannot be false. In recent years such a thesis has been widely rejected. This rejection stems from both general philosophical reasoning (e.g. the suggestion that if beliefs and experiences are distinct existences, there can be no necessary connection between them) and from apparent counterexamples (e.g. a case where someone, expecting to be burnt, momentarily misclassifies a cold sensation as hot). In this light, it is interesting to note that the framework outlined so far supports an incorrigibility thesis, albeit a very limited one.

Incorrigibility Thesis: A direct phenomenal belief cannot be false.

The truth of this thesis is an immediate consequence of the definition of direct phenomenal belief. A direct phenomenal concept by its nature picks out instances of an underlying demonstrated phenomenal quality, and a direct phenomenal belief identifies the referent of that concept with the very demonstrated quality (or predicates the concept of the very experience that instantiated the quality), so its truth is guaranteed.

If we combine this thesis (which is more or less true by definition) with the substantive thesis that there are direct phenomenal beliefs (which is argued earlier), then we have a substantive incorrigibility thesis, one that applies to a significant range of actual beliefs.[11]

The thesis nevertheless has a number of significant limitations. The first is that most phenomenal beliefs are not direct phenomenal beliefs, so most phenomenal beliefs are still corrigible. The most common sort of phenomenal belief arguably involves the application of a *pre-existing* phenomenal concept (either a relational phenomenal concept or a standing pure phenomenal concept) to a new situation, as with the beliefs typically expressed by claims such as 'I am having a red experience' or 'I am in pain'. These are not direct phenomenal beliefs, and are almost certainly corrigible.

There are also cases in which a direct phenomenal concept is applied to a quality (or an experience) other than the one that constituted it, as when one forms a direct phenomenal concept *R* based on a quality instantiated in the left half of one's visual field, and applies it to a quality instantiated in the right half. These are also not direct phenomenal beliefs, and are again almost certainly corrigible.

(The second sort of case brings out a further limitation in the incorrigibility thesis: it does not yield incorrigibility in virtue of content. If the left and right qualities in the case above are in fact the same, then the resulting non-direct phenomenal belief will arguably have the same content as the corresponding direct phenomenal belief, but the incorrigibility thesis will not apply to it. The domain of the incorrigibility thesis is constrained not just by content, but also by underlying constitution.)

It is plausible that all the standard counterexamples to incorrigibility theses fall into classes such as these, particularly the first. All the standard counterexamples appear to

---

[11] Pollock (1986: 32-3) entertains a version of this sort of view as a way of supporting incorrigibility, discussing a "Containment Thesis" according to which experiences are constituents of beliefs about experiences. He rejects the view on the grounds that (1) it does not support incorrigibility of negative beliefs about experiences (e.g. the belief that one is not having a given experience), which he holds to be required for incorrigibility in general, and that (2) that having an experience does not suffice to have the relevant belief, so having the belief also requires thinking about the experience, which renders the incorrigibility thesis trivial. I discuss both of these points below.

involve the application of pre-existing phenomenal concepts (*pain*, *hot*, *red experience*). So none of the standard counterexamples apply to the incorrigibility thesis articulated here.

There is a natural temptation to find further counterexamples to the incorrigibility thesis. For example, one might consider a case in which a subject's experience changes very rapidly, and argue that the corresponding direct phenomenal concept must lag behind. In response to these attempted counterexamples, the most obvious reply is that these cannot truly be counterexamples, since the truth of the incorrigibility thesis is guaranteed by the definition of direct phenomenal belief. If the cases work as described, they do not involve direct phenomenal beliefs: they either involve a concept that is not a direct phenomenal concept, or they involve a direct phenomenal concept predicated of a quality other than the one that constitutes it. At best, they involve what I earlier called pseudo-direct phenomenal beliefs: beliefs that share the cognitive structure of direct phenomenal beliefs (and thus are quasi-direct phenomenal beliefs) but that are not direct phenomenal beliefs.

One need not let matters rest there, however. I think that these counterexamples can usually be analysed away on their own terms, so that the purported pseudo-direct phenomenal beliefs in question can be seen as direct phenomenal beliefs, and as correct. In the case of a rapidly changing experience, one can plausibly hold that the content of a direct phenomenal concept co-varies immediately with the underlying quality, so that there is no moment at which the belief is false. This is just what we would expect, given the constitutive relation suggested earlier. We might picture this schematically by suggesting that the basis for a direct phenomenal concept contains within it a "slot" for an instantiated quality, such that the quality that fills the slot constitutes the content. In a case where experience changes rapidly, the filler of the slot changes rapidly, and so does the content.

Something similar goes for many other examples involving quasi-direct phenomenal beliefs. Take a case where a subject attends to two different visual qualities (demonstrating them as E_1 and E_2), and mistakenly accepts $E_1 = E_2$. In this case, someone might suggest that if the subject forms specific quasi-direct phenomenal concepts R_1 and R_2 based on the two acts of attention, these must have the same content, leading to false quasi-direct phenomenal beliefs (and thus to pseudo-direct phenomenal beliefs). But on my account, this case is better classified as one in which R_1 and R_2 are direct phenomenal concepts with different contents, yielding two correct direct phenomenal beliefs $E_1 = R_1$ and $E_2 = R_2$. The false beliefs here are of the form $E_1 = R_2$, $E_2 = R_1$, and $R_1 = R_2$. The last of these illustrates the important point that identities involving two direct phenomenal concepts, like identities involving two pure phenomenal concepts more generally, are not incorrigible.

Other cases of misclassification can be treated similarly. In the case in which a subject expecting to be burnt misclassifies a cold sensation as hot, someone might suggest that any quasi-direct phenomenal concept will be a concept of phenomenal hotness, not coldness. But one can plausibly hold that if a quasi-direct phenomenal concept is formed, it will be a concept of phenomenal coldness and will yield a correct direct phenomenal belief. The subject's mistake involves misclassifying the experience under standing phenomenal concepts, and perhaps a mistaken identity involving a direct and a standing phenomenal concept.

It is arguable that most cases involving quasi-direct phenomenal beliefs can be treated this way. The only clear exceptions are cases such as Nancy's, in which no phenomenal quality is demonstrated and so no substantive direct phenomenal concept is formed. It remains plausible that as long as a quality is demonstrated, the cognitive act in question will yield a direct phenomenal concept with the right content, and a true direct phenomenal belief. If that is correct, one can then accept a broader incorrigibility thesis applying to any quasi-direct phenomenal belief that is based in a successful demonstration of a phenomenal quality. I will not try to establish this thesis conclusively, since I will not need it, and since the incorrigibility thesis for direct phenomenal beliefs is unthreatened either way. But it is interesting to see that it can be defended.

One might suggest that the incorrigibility thesis articulated here (in either the narrower or the broader version) captures the *plausible core* of traditional incorrigibility theses. A number of philosophers have had the sense that there is something correct about the incorrigibility theses, which is not touched by the counterexamples. This is reflected, for example, in Chisholm's distinction between "comparative" and "non-comparative" uses of "appears" talk, where only the non-comparative uses are held to be incorrigible. I think that this is not quite the right distinction: even non-comparative uses can be corrigible, when they correspond to uses of pure phenomenal concepts outside direct phenomenal beliefs. But perhaps a thesis restricted to direct phenomenal beliefs might play this role.

Certainly the analysis of direct phenomenal beliefs shows why the most common general philosophical argument against incorrigibility does not apply across the board. In the case of direct phenomenal beliefs, beliefs and experiences are *not* entirely distinct existences. It is precisely because of the constitutive connection between experiential quality and belief that the two can be necessarily connected.

Another limitation: sometimes incorrigibility theses are articulated in a "reverse" or bidirectional form, holding that all phenomenal states are incorrigibly known, or at least incorrigibly knowable. Such a thesis is not supported by the current discussion. Most

phenomenal states are not attended to, and are not taken up into direct phenomenal concepts, so they are not the subjects of direct phenomenal beliefs. And for all I have said, it may be that some phenomenal states, such as fleeting or background phenomenal states, *cannot* be taken up into a direct phenomenal concept, perhaps because they cannot be subject to the right sort of attention. If so, they are not even incorrigibly knowable, let alone incorrigibly known.

Incorrigibility theses are also sometimes articulated in a "negative" form, requiring that a subject cannot be mistaken in their belief that they are *not* having a given sort of experience. No direct phenomenal belief is a negative phenomenal belief, so the current framework does not support this thesis, and I think the thesis is false in general.

A final limitation: although direct phenomenal beliefs are incorrigible, subjects are not incorrigible about whether they are having a direct phenomenal belief. For example, if I am not thinking clearly, I might misclassify a belief involving a standing phenomenal concept as a direct phenomenal belief. And in the Nancy case above, if Nancy is philosophically sophisticated she might well think that she is having a direct phenomenal belief, although she is not.

One could argue that this lack of higher-order incorrigibility prevents the first-order incorrigibility thesis from doing significant epistemological work. The matter is delicate: higher-order incorrigibility is probably too strong a requirement for an epistemologically useful incorrigibility thesis. But on the other side, *some* sort of further condition is required for a useful thesis. For example, any member of the class of true mathematical beliefs is incorrigible (since it is necessarily true), but this is of little epistemic use to a subject who cannot antecedently distinguish true and false mathematical beliefs. A natural suggestion is that some sort of higher-order accessibility is required.

Intermediate accessibility requirements might include these: for the incorrigibility of a direct phenomenal belief to be epistemologically significant, a subject must know that it is a direct phenomenal belief, or at least be justified in so believing; or a subject must be capable of so knowing on reflection; or direct phenomenal beliefs must be cognitively or phenomenologically distinctive as a class relative to non-direct phenomenal beliefs.

I am sympathetic with the sufficiency of a requirement appealing to cognitive or phenomenological distinctiveness, if properly articulated. Whether such a requirement holds of direct phenomenal beliefs turns on questions about quasi-direct and pseudo-direct phenomenal beliefs. If there are many pseudo-direct phenomenal beliefs, and if there is nothing cognitively or phenomenologically distinctive about direct phenomenal beliefs by comparison, then direct phenomenal beliefs will simply be distinguished as quasi-direct phenomenal beliefs with the right sort of content, and the incorrigibility claim will be

relatively trivial. On the other hand, if pseudo-direct phenomenal beliefs are rare, or if direct phenomenal beliefs are a cognitively or phenomenologically distinctive subclass, then it is more likely that incorrigibility will be non-trivial and carry epistemological significance.

If pseudo-direct phenomenal beliefs are restricted to cases in which no phenomenal quality is demonstrated, such as the case of Nancy (as I have suggested), then the incorrigibility thesis will hold of a class of beliefs that can be distinctively and independently characterized in cognitive and phenomenological terms: the class of quasi-direct phenomenal beliefs which are based in a successful demonstration. This would render the incorrigibility claim entirely non-trivial, and it would make it more likely that it could do epistemological work. But I will not try to settle this matter decisively here, and I will not put the incorrigibility thesis to any epistemological work in what follows.

It might be thought that the incorrigibility thesis suffers from another problem: that direct phenomenal beliefs are incorrigible because they are *trivial*. After all, beliefs such as *I am here* or *this is this* are (almost) incorrigible, but only because they are (almost) trivial. ('Almost' is present because of the arguable non-triviality of my existence and spatial locatedness in one case, and because of the possibility of reference failure for the demonstrative in the other.)

The analogy fails, however. The trivial beliefs in question are (almost) cognitively insignificant: they are (almost) a priori, containing (almost) no cognitively significant knowledge about the world. This is reflected in the fact that they hardly constrain the class of a priori epistemic possibilities: they are true of (almost) all such possibilities, considered as hypotheses about the actual world. (Two-dimensionally: these beliefs have an epistemic intension that is (almost) conceptually necessary.) A direct phenomenal belief, by contrast, is cognitively significant: it heavily constrains the class of a priori epistemic possibilities, and is false in most of them (considered as actual). For example, Mary's direct phenomenal belief, on leaving her room, is false of all worlds (considered as actual) in which the subject is not experiencing phenomenal redness. (Two-dimensionally: the epistemic intension of a direct phenomenal belief is conceptually contingent.) So direct phenomenal beliefs, unlike the beliefs above, are entirely non-trivial.

So: the incorrigibility thesis articulated here has a number of limitations, but it nevertheless applies to a significant class of non-trivial phenomenal beliefs.

4.2 Acquaintance and justification

At this point is natural to ask: if we can form this special class of incorrigible, distinctively constituted beliefs where phenomenal states and properties are concerned, why

cannot we do so where other states and properties are concerned? Why cannot we form direct height concepts, for example, whose epistemic content is directly constituted by our height properties, and which can be deployed in incorrigible direct height beliefs? Or similarly for direct chemical beliefs, direct age beliefs, direct color beliefs, and so on?

At one level, the answer is that we simply cannot. If one tries to form a direct height concept — one whose content depends constitutively on an instantiated height — the best one can do is form a relational height concept (*my height*, *the height of my house*) or a demonstrative height concept. But these are not pure height concepts at all. They are analogous only to red_C or *E*, in that their subjunctive content may depend on the property in question but their epistemic content does not.

It is arguable whether pure height concepts exist at all: that it, whether there is any concept whose epistemic content picks out a certain height (say, two metres) in any epistemic possibility. But even if there are pure height concepts, they are not direct height concepts. Perhaps one can independently form a pure height concept of a given height (two metres), which might coincide with an instantiated height, but it will not depend constitutively on an instantiated height. The best one can do is attend to an object, have an experience or judgment concerning its height, and use this experience or judgment as the epistemic content of a "pure" height concept. But here the instantiated height property is not constitutively relevant to the concept's content, but only causally relevant: it is the height experience or judgment that is constitutively relevant, and the experience or judgment is only causally dependent on the height. In no case does the epistemic content of a height concept depend constitutively on a demonstrated height property, or on any instantiated height property at all.

Proponents of certain direct realist views may hold that it is possible to form a direct concept of a height property (or other perceivable external properties), by demonstrating it and taking it up into a concept in a manner analogous to the manner suggested for phenomenal properties. I think that this is implausible. In a case where an object is two metres tall but appears to be one metre tall, any "pure" height concept formed as a result will be a concept of one metre, not of two metres. There may be a demonstrative concept of two metres, but that is not enough. More generally, considering a range of cases in which height and experience are varied independently, we can see that any contribution of the height to a pure concept is "screened off" by the contribution of the experience. This suggests that if

anything is playing a constitutive role in the concept's content, it is the experience and not the external property.[12]

The same goes for chemical concepts, age concepts, and external color concepts. Although we can form many such concepts, in no case is it possible to form a direct concept: that is, a concept whose epistemic content depends constitutively on a demonstrated property. It seems that only phenomenal properties can support direct concepts.

This conclusion is apparently revealed by an examination of cases; but it would be preferable not to leave it as a brute conclusion. In particular, it is natural to suggest that the conclusion holds because we bear a special relation to the phenomenal properties instantiated in our experience: a relation that we do not bear to the other instantiated properties in question, and a relation that is required in order to form a direct concept of a property in the manner described. This relation would seem to be a peculiarly intimate one, made possible by the fact that experiences lie at the heart of the mind rather than standing at a distance from it; and it seems to be a relation that carries the potential for conceptual and epistemic consequences. We might call this relation *acquaintance*.

As things stand, acquaintance has been characterized only as that relation between subjects and properties that makes possible the formation of direct phenomenal concepts; so it is not yet doing much explanatory work. But having inferred the relation of acquaintance, we can put it to work. As characterized, acquaintance is a relation that makes possible the formation of pure phenomenal concepts, and we have seen that pure phenomenal concepts embody a certain sort of lucid understanding of phenomenal properties. So acquaintance is a relation that makes this sort of lucid understanding possible. As such, it is natural to suppose that the relation can also do work in the epistemic domain. If so, the result will be an attractive picture in which the distinctive conceptual character and the distinctive epistemic character of the phenomenal domain have a common source.

It is independently plausible to hold that phenomenal properties and beliefs have a distinctive epistemic character. Many have held that phenomenal properties can (at least

---

[12] There may be further moves available to the direct realist. For example, a direct realist might hold that the constitutive role of external properties is restricted to cases of veridical perception, and that non-veridical perception must be treated differently. I think that this sort of restriction threatens to trivialize the constitution thesis, as any causal connection might be seen as a "constitutive" connection by a relevantly similar restriction. (If A causes B which necessitates C, then A is contingently connected to C; but if we restrict attention to cases where A causes B, then A necessitates C relative to this restriction.) And the case remains formally disanalogous to the case of direct phenomenal concepts, in which there is no factor distinct from the quality that even looks like it screens off the contribution of the quality to the concept. But there is undoubtedly more to say here. In what follows I will assume that the direct realist view is incorrect, but direct realists are free to hold that what I say about phenomenal properties applies equally to the relevant external properties.

sometimes) be known with a distinctive sort of justification, or even with certainty; and many have held that phenomenal beliefs have a special epistemic status. Even those who explicitly deny this will often tacitly concede that there is at least a prima facie case for this status: for example, it is striking that those who construct sceptical scenarios almost always ensure that that phenomenal properties are preserved. So it is arguable that simply having a phenomenal property provides the potential for a strong sort of phenomenal knowledge. Something similar is suggested by the Mary case: Mary's experience of the phenomenal property R allows her to have not just a distinctive phenomenal belief, but also distinctive phenomenal knowledge. Some element of this distinctive epistemic character can be captured in the present framework.

One natural suggestion is the following: direct phenomenal beliefs are always justified. Certainly Mary's belief on leaving her room seems to be justified, and most other examples seem to fit this thesis. This thesis has to be modified slightly. There are presumably subjects who are so irrational or confused that none of their beliefs qualify as justified, so that their direct phenomenal beliefs are not justified either. And perhaps there could be subjects who are so confused about phenomenology that they accept not just direct phenomenal beliefs but their negations, casting doubt on whether either belief is truly justified. To meet this sort of case, we might adjust the thesis to say that all direct phenomenal beliefs have some prima facie justification, where prima facie justification is an element of justification that can sometimes be overridden by other elements, rendering a belief below the threshold for "justification" *simpliciter*. Something similar presumably applies to other features of a belief that might seem to confer justification, such as being inferred from justified beliefs by a justified rule of inference.

Assuming that something like this is right: it is nevertheless one thing to make the case that direct phenomenal beliefs are (prima facie) justified, and another to give an account of what this justification consists in. It may be tempting to appeal to incorrigibility; but incorrigibility alone does not entail justification (as the mathematical case shows), and while certain higher-order accessibility theses might close the gap, it is not obvious that they are satisfied for direct phenomenal beliefs.

A better idea is to appeal to the acquaintance relation, thus unifying the distinctive conceptual and epistemic character of phenomenal beliefs. In particular, one might assert the following:

> **Justification Thesis:** When a subject forms a direct phenomenal belief based on a phenomenal quality, then that belief is prima facie justified by virtue of the subject's acquaintance with that quality.

Certainly many philosophers, including especially sense-datum theorists and more recent foundationalists, have appealed to a relation of acquaintance (or "direct awareness") in supporting the special epistemic status of phenomenal beliefs. The current account offers a more constrained version of such a thesis, suggesting that it holds for a special class of phenomenal beliefs (on which the epistemic content of a predicated concept is required to mirror and be constituted by the acquainted quality, to which it is applied), and on the basis of a relation whose existence we have made an independent case for.

Some philosophers (e.g. Russell 1910; Fumerton 1995) have held that we are "acquainted with acquaintance", and have made the case of its existence that way. I think there is something to the idea that our special epistemic relation to experience is revealed in our experience, but I note that the proponent of acquaintance is not forced to rely on such a thesis. It is equally possible to regard acquaintance as a theoretical notion, inferred to give a unified account of the distinctive conceptual and epistemic character that we have reason to believe is present in the phenomenal domain.

Acquaintance can be regarded as a basic sort of epistemic relation between a subject and a property. Most fundamentally, it might be seen as a relation between a subject and an *instance* of a property: I am most directly acquainted with *this instance* of phenomenal greenness. This acquaintance with an instance can then be seen to confer a derivative relation to the property itself. Or in the experience-based framework, one might regard acquaintance as most fundamentally a relation between a subject and an experience, which confers a derivative relation between the subject and the phenomenal properties of the experience. But I will usually abstract away from these fine details. What is central will be the shared feature that whenever a subject has a phenomenal property, the subject is acquainted with that phenomenal property.

Even if acquaintance is a theoretical notion, it clearly gains some pre-theoretical support from the intuitive view that beliefs can be epistemically grounded in experiences, where experiences are not themselves beliefs but nevertheless have an epistemic status that can help justify a belief. One might view acquaintance as capturing that epistemic status.

In certain respects (though not in all respects), the justification of a direct phenomenal belief by an experience can be seen as analogous to the justification of an inferred belief by another belief. For an inferred belief to be prima facie justified, there are three central requirements: one concerning the content of the belief in relation to the justifying state, one concerning the natural connection between the belief and the justifying state, and one concerning the epistemic status of the justifying state. First, the epistemic content of the belief must be appropriately related to that of the belief that it is inferred from. Second, the belief

must be appropriately caused by the justifying belief. Third, the justifying belief must itself be justified.

In the prima facie justification of a direct phenomenal belief by an experience, there are three factors of the same sort. First, content: the epistemic content of the direct phenomenal belief must mirror the quality of the experience. Second, a natural connection: the phenomenal belief must be appropriately constituted by the experience. And third, epistemic status: the subject must be acquainted with the justifying quality. The details of the requirements are different, as befits the difference between belief and experience, but the basic pattern is very similar.

It is plausible that a subject can have phenomenal properties without having corresponding concepts, or corresponding beliefs, or corresponding justification.[13] If so, the same goes for acquaintance. Acquaintance is not itself a conceptual relation: rather, it makes certain sorts of concepts possible. And it is not itself a justificatory relation: rather, it makes certain sorts of justification possible. Phenomenal concepts and phenomenal knowledge require not just acquaintance, but acquaintance in the right cognitive background: a cognitive background that minimally involves a certain sort of attention to the phenomenal quality in question, a cognitive act of concept formation, the absence of certain sorts of confusion and other undermining factors (for full justification), and so on. But it is acquaintance with the quality or the experience itself that does the crucial justifying work.

Some philosophers hold that only a belief can justify another belief. It is unclear why this view should be accepted. The view has no pre-theoretical support: pre-theoretically, it is extremely plausible that experiences (e.g. a certain experience of phenomenal greenness) play a role in justifying beliefs (e.g. my belief that there is something green in front of me, or my belief that I am having a certain sort of experience), even though experiences are not themselves beliefs. And the view has no obvious theoretical support. Perhaps the central

---

[13] Nothing I have said so far requires that experiences can exist without concepts; at most, it requires that experiences can exist without phenomenal concepts. So what I have said may be compatible with views on which experiences depend on other concepts in turn. Still, I think it is independently plausible that experiences do not require concepts for their existence, and I will occasionally assume this in what follows. This is not to deny that *some* experiences depend on concepts, and it is also not to deny that experiences have representational content.

My own view is that at least for perceptual experiences (and perhaps for all experiences), experiences have representational content by virtue of their phenomenology, where this content is sometimes conceptual and sometimes non-conceptual. This yields an interesting possibility (developed in forthcoming work): the constitutive relation between phenomenal states and phenomenal concepts might be extended to yield a similar constitutive relation between perceptual phenomenal states and a special class of perceptual concepts, by virtue of the phenomenal states' representational content. Such an account might yield some insight into the content and epistemology of perceptual belief.

motivation for the view comes from the idea that inference is the only sort of justification that we understand and have a theoretical model for, and that we have no model for any other sort of justification. But this is obviously not a strong reason, and the account I have just sketched suggests a theoretical model of how experiences can justify beliefs that fits well with our pre-theoretical intuitions. So it seems that the cases of justification of beliefs by other beliefs and by experiences are on a par here.

Another motivation for the view comes from the thesis that for a state to justify another state, it must itself be justified (along with the claim that only beliefs can be justified. But again, it is unclear why this thesis should be accepted. Again, it is pre-theoretically reasonable to accept that beliefs are justified by experiences, and that experiences are not themselves the sort of states that can be justified or unjustified. And there is no obvious theoretical reason to accept the thesis. It may be that for a state to justify, it must have *some* sort of epistemic status, but there is no clear reason why the status of acquaintance should be insufficient.

(BonJour (1978) suggests that the denial that justifying states must be justified is an ad hoc move aimed at stopping the regress argument against foundationalism. But considerations about foundationalism and about regress arguments have played no role in my claims: the claims are independently supported by observations about the epistemic and conceptual relations between belief and experience. BonJour also claims that a justifying state must involve assertive content; but again, there is no clear pre-theoretical or theoretical reason to accept this. Pre-theoretically: experiences can justify beliefs without obviously involving assertive content. Theoretically: acquaintance with a property makes the property available to a subject in a manner that makes concepts and assertions involving the property possible, and that enables these assertions to be justified. There is no reason why this requires acquaintance to itself involve an assertion.)

A number of epistemological issues remain. One concerns the strength of the justification of phenomenal beliefs. It is often held that phenomenal beliefs are (or can be) *certain*, for example. Can the present framework deliver this? It can certainly deliver incorrigibility, but certainty requires something different. I think that the relevant sense of certainty involves something like *knowledge beyond scepticism*: intuitively, knowledge such that one's epistemic situation enables one to rule out all sceptical counterpossibilities. There is an intuition that phenomenal belief at least sometimes involves this sort of knowledge beyond scepticism, as the standard construction of sceptical scenarios suggests.

This epistemic status might be captured by a claim to the effect that acquaintance with a property enables one to eliminate all (a priori) epistemic possibilities in which the property is absent. If so, then in the right cognitive background (with sufficient attention, concept

formation, lack of confusion, and so on), the justification of a direct phenomenal belief *P* by acquaintance with a property will sometimes enable a subject not just to know that *P* by the usual standards of knowledge, but to eliminate all sceptical counterpossibilities in which *P* is false. This matter requires further exploration, but one can see at least the beginnings of a reasonable picture.[14]

A second further issue: can the justification thesis be extended to all pure phenomenal concepts, including standing phenomenal concepts? There is some intuitive appeal in the idea that application of a standing phenomenal concept to an instantiated quality may also carry some justification by virtue of acquaintance with the quality (perhaps under the restriction that the content of the standing concept match the quality, and that there be an appropriate natural connection between the quality and the belief). If this belief were justified directly by acquaintance, however, we would need an account of justification by acquaintance that does not give a central role to constitution. Such an account is not out of the question, but it is worth noting that justification for beliefs involving standing phenomenal concepts can also be secured indirectly.

---

[14] I argued in *The Conscious Mind* (1996) that something like acquaintance is required to secure certainty, and that a mere causal connection or reliable connection cannot do the job. If the justification of a belief is based solely on a reliable or causal connection, the subject will not be in a position to rule out sceptical scenarios in which the connection is absent and the belief is false, so the belief will not be certain. In response, a number of philosophers, including Bayne (2001), have argued that acquaintance accounts can be criticized in a similar way. Bayne notes that acquaintance alone is compatible with the absence of certainty (e.g. in conditions of inattention), so certainty requires background factors in addition to acquaintance; but we cannot be certain that these factors obtain, so we cannot rule out sceptical scenarios in which they fail to obtain, so a phenomenal belief cannot be certain.

This argument stems from a natural misreading of my argument against reliabilist accounts. The argument is not: certainty requires certainty about the factors that enable certainty, and a reliabilist account cannot deliver this sort of certainty. That argument would requires a strong version of a CJ thesis, that certain justification requires certainty about the basis of certain justification (analogous to the KJ thesis that justification requires knowledge of justification). I think such a thesis should clearly be rejected. The argument is rather: certainty about P requires (first-order) "knowledge beyond scepticism", or an epistemic state that enables a subject to rule out all sceptical scenarios in which P is false. Reliabilism by its nature cannot do this: there will always be sceptical scenarios in which the reliable connection fails and in which P is false.

Bayne's argument against acquaintance gives an analogue of the invalid CJ argument. At most this establishes that we cannot rule out scenarios in which the belief is uncertain. Even this is unclear, as it is not obvious that certainty about certainty requires certainty about the factors enabling certainty. But even if this point is granted, the existence of sceptical scenarios in which the belief is uncertain does not entail the existence of sceptical scenarios in which P is false. Acquaintance yields certainty about experiences, not about beliefs: it enables one to directly rule out sceptical scenarios in which P is false, whether or not it enables one to rule out sceptical scenarios in which a belief is uncertain. In cases of justification by a reliable connection, there are separate reasons to hold that sceptical scenarios in which P is false cannot be ruled out, but in the case of acquaintance, these reasons do not apply. (Note: the published version of Bayne's article takes these points into account and offers some further considerations.)

Indirect justification for such beliefs can be secured by virtue of the plausible claim that any belief of the form $S = R$ is (prima facie) justifiable, where S and R are standing and direct phenomenal concepts with the same epistemic content. This is an instance of the more general claim that any belief of the form $A = B$ is justifiable when A and B have the same epistemic content. (This thesis may need some restriction to handle cases of deep hyperintensionality, but it is plausibly applicable in this case.) Such beliefs are plausibly justifiable a priori: experience may enter into a grasp of the concepts involved in such a belief, but it does not enter into the belief's justification. If so, then beliefs involving standing phenomenal concepts can inherit justification by a priori inference from direct phenomenal beliefs, which will be justified in virtue of the Justification Thesis.

Finally, a note on ontology: talk of acquaintance often brings sense-datum theories to mind, so it may be worth noting that a commitment to phenomenal realism and to acquaintance does not entail a commitment to sense-data. First, the picture is entirely compatible with an "adverbial" subject-property model, and with other quality-based ontologies on which there are phenomenal properties but not phenomenal individuals. Second, even if one accepts the existence of phenomenal individuals such as experiences, one might well reject a sense-datum model of perception, on which one perceives the world by perceiving these entities.

It is also worth noting that one need not regard the acquaintance relation that a subject bears to a phenomenal property as something ontologically over and above the subject's instantiation of the property, requiring a subject-relation-quality ontology at the fundamental level. It is arguable that it is a conceptual truth that to have a phenomenal quality is to be acquainted with it (at least in so far as we have a concept of acquaintance that is not wholly theoretical). Certainly it is hard to conceive of a scenario in which a phenomenal quality is instantiated but no one is acquainted with it. If so, then the picture I have sketched is combined with a simple subject-quality ontology, combined with this conceptual truth. The ontological ground of all this might lie in the nature of phenomenal qualities, rather than in some ontologically further relation.

4.3 Epistemological Problems for Phenomenal Realism

Phenomenal realism, especially property dualism, is often thought to face epistemological problems. In particular, it is sometimes held that these views make it hard to see how phenomenal beliefs can be justified or can qualify as knowledge, since the views entail that phenomenal beliefs do not stand in the right sort of relationship to experiences. If

what I have said so far is right, this cannot be correct. But it is worth looking at the arguments more closely.[15]

The most influential arguments of this sort have been put forward by Sydney Shoemaker (1975). Shoemaker's arguments are intended as an argument against a view that admits the conceptual possibility of "absent qualia": an experience-free functional duplicate of an experiencing being. The view under attack is slightly stronger than phenomenal realism (a phenomenal realist could admit inverted qualia without absent qualia), is slightly weaker than a view on which zombies (experience-free physical duplicates) are conceptually possible, and is weaker than property dualism. But for the purposes of the argument, it will not hurt to assume a property dualist version of the view on which zombies are metaphysically possible. This has the effect of making Shoemaker's arguments harder to answer, not easier. The answers can easily be adapted to weaker versions of phenomenal realism.

The starkest version of Shoemaker's epistemological argument runs as follows:

(1) If phenomenal realism is true, experiences are causally irrelevant to phenomenal beliefs.

(2) If experiences are causally irrelevant to phenomenal beliefs, phenomenal beliefs are not knowledge.

---

(3) If phenomenal realism is true, phenomenal beliefs are not knowledge.

Some phenomenal realists might deny the first premiss: a type-B materialist could hold that experiences have effects on beliefs by virtue of their identity with physical states, and a property dualist could hold that these effects proceed through a fundamental causal connection between the phenomenal and physical domains, or through a fundamental causal connection among non-physical mental states. But for the purposes of the argument, I will assume the version of phenomenal realism that makes answering the argument as hard as

[15] I discussed these arguments at length in ch. 5 of *The Conscious Mind*, on "The Paradox of Phenomenal Judgment". I now think that discussion is at best suboptimal. The final section of the chapter put forward a preliminary and sketchy version of the view of phenomenal concepts I have discussed here, but I did not give it a central epistemological role (except in a tentative suggestion on pp. 207-8). I now think that this view of phenomenal concepts is central to the epistemology. So the present discussion can be viewed in part as a replacement for that chapter.

possible, so I will rule out these responses. In particular, I will assume epiphenomenalism, according to which the phenomenal has no effects on the physical domain.[16]

The view I have outlined makes it easy to see why this argument fails, even against an epiphenomenalist. Whatever the status of the first premiss, the second premiss is false. The second premiss assumes that a causal connection between experience and phenomenal belief is required for the latter to count as knowledge. But if what I have said is correct, the connection between experience and phenomenal belief is tighter than any causal connection: it is constitution. And if a causal connection can underwrite knowledge, a constitutive connection can certainly underwrite knowledge too.

Even without appealing to constitution, the epiphenomenalist can respond reasonably to this argument by appealing to the notion of acquaintance, and arguing that a subject's acquaintance with experience can non-causally justify a phenomenal belief. (I used this strategy in *The Conscious Mind*.) But when the role of constitution is made clear, the reply becomes even stronger. Acquaintance and constitution together enable a theoretical model of the justification of phenomenal belief (as above), a model that is compatible with epiphenomenalism. And any residual worries about the lack of an appropriate connection between the experience and the belief are removed by the presence of a constitutive connection.

This first argument is only a subsidiary argument in Shoemaker's discussion. Shoemaker's main argument specifically concerns the possibility of absent qualia. His argument involves functional duplicates and conceptual possibility, but as before I will modify these details to involve physical duplicates and metaphysical possibility, thus making the argument harder to answer. The modified argument runs roughly as follows:

(1) If phenomenal realism is true, then every conscious being has a possible zombie twin.

(2) If zombies are possible, they have the same phenomenal beliefs as their conscious twins, formed by the same mechanism.

(3) If zombies are possible, their phenomenal beliefs are false and unjustified.

(4) If it is possible that there are beings with the same phenomenal beliefs as a conscious being, formed by the same mechanism, where those phenomenal

[16] I am not endorsing epiphenomenalism, but I regard it as one of the three serious options that remain once one accepts phenomenal realism and rules out type-B materialism and idealism. The other two are interactionism and a Russellian "panprotopsychism". See Chalmers (2002*a*).

beliefs are false and unjustified, then the conscious being's phenomenal beliefs are unjustified.

---

(5) If phenomenal realism is true, every conscious being's phenomenal beliefs are unjustified.

Some phenomenal realists could respond by denying premiss 1 and holding that zombies are impossible. But even the conceptual possibility of functional duplicates with absent qualia is arguably enough to make an analogous argument go through, if there are no other problems. Premiss 3 is relatively unproblematic. Perhaps one could argue that a zombie's phenomenal beliefs have some sort of justification, but the conclusion that our phenomenal beliefs are no more justified than a zombie's would be strong enough for an opponent. Disputing premiss 4 holds more promise. If one accepts an acquaintance model of justification, one might hold that the justification of a phenomenal belief does not supervene on its mechanism of formation. (I used this strategy in *The Conscious Mind*.) But given what has gone before, by far the most obvious reply is to dispute premiss 2. There is no reason to accept that zombies have the same phenomenal beliefs as their conscious twins, and every reason to believe that they do not.

It is by no means obvious that zombies have beliefs at all. The basis of intentionality is poorly understood, and one might plausibly hold that a capacity for consciousness is required for intentional states. But even if we allow that zombies have beliefs, it is clear that a zombie cannot share a conscious being's phenomenal beliefs. The content of a conscious being's direct phenomenal beliefs is partly constituted by underlying phenomenal qualities. A zombie lacks those qualities, so it cannot have a phenomenal belief with the same content.

Let us take the case of Zombie Mary, where we recombine thought experiments in the obvious way. Assuming that Zombie Mary has a belief where Mary has a direct phenomenal belief, what sort of content does it have? Mary has a belief with the content $E = R$, and Inverted Mary has a belief with the content $E = G$. Let us focus on the direct phenomenal concepts R and G, and their zombie counterpart. It is obvious that Zombie Mary's concept is neither R nor G: if it has content at all, it has a different content entirely. I think that the most plausible view is that the zombie's concept is *empty*: it has no content. On the view I have been outlining, a phenomenal quality can be thought of as filling a slot that is left open in the content of a direct phenomenal concept, and thus contributing its content. If there is no phenomenal quality to fill the slot, as in Zombie Mary's case, the concept will have no content at all.

What about Zombie Mary's analogue of Mary's direct phenomenal belief $E = R$? It is not obvious that a zombie can possess a demonstrative phenomenal concept: for a start, a concept whose content is that of 'this experience' seems to require a concept of experience, which a zombie may lack. But even if a zombie could possess a demonstrative phenomenal concept, any such concept would fail to refer (like failed demonstratives in other domains). And more importantly, the other half of the identity (the zombie's analog of *R*) would be empty. So Zombie Mary's belief would be entirely different from Mary's belief.

It is natural to wonder about the truth-value of Zombie Mary's belief. Clearly her belief is not true. I would say that it is either false or empty, depending on one's view about beliefs involving empty concepts. The latter view is perhaps the most plausible, since it seems that Zombie Mary's belief has no propositional content to evaluate. As for Zombie Mary's "new knowledge": it is clear that she gains no propositional knowledge (though she may think that she does). One might see her as in the position that type-A materialists, and in particular proponents of the "ability hypothesis", hold that we are in the actual world. When Zombie Mary first sees a flower, she may gain certain abilities to recognize and discriminate, although even these abilities will be severely constrained, since they cannot involve experiences.

This is enough to see that the epistemological argument against phenomenal realism does not get off the ground. A zombie clearly does not have the same phenomenal beliefs as its conscious twin in general; and its corresponding beliefs are not even formed by the same mechanism, since constitution by a phenomenal quality plays a central role in forming a direct phenomenal belief. So the second premiss is false, and there is no bar to the justification of direct phenomenal beliefs.[17]

What about other phenomenal beliefs? We have seen that standing phenomenal concepts differ between twins, and that their content is plausibly constituted either by phenomenal properties or by dispositions involving those properties. A zombie lacks all phenomenal properties, so it is plausible that its analogs of standing phenomenal concepts will be empty, too. So beliefs involving standing phenomenal concepts are also immune from this argument.

What about the standing concept of *experience* (or *qualia*, or *phenomenal consciousness*) generally? In this case there is no difference in content between conscious twins. But it remains plausible that phenomenal properties and the capacity to have them play a crucial role in constituting its content, just as they do for specific standing phenomenal concepts. And it is

[17] Conee (1985) and Francescotti (1994) also respond to Shoemaker's argument by denying the equivalent of premiss 2, although for somewhat different reasons.

equally plausible that the zombie's analog of this standing concept is empty.[18] So beliefs involving the standing concept of experience (such as *I am conscious*) are equally unthreatened by this argument. The same goes for beliefs involving concepts in which the concept *experience* plays a part, such as relational phenomenal concepts, and perhaps demonstrative phenomenal concepts.

How are these beliefs justified? For beliefs involving standing phenomenal concepts, such as $E = S$, we have seen that one reasonable model involves inference from $E = R$ and $R = S$. Here, the former belief is justified by acquaintance and constitution, and the second belief is justified a priori by virtue of its content. These two beliefs combine by virtue of the common element R to justify the belief $E = S$. (One can also hold that $E = S$ is justified directly by acquaintance, at cost of losing the special contribution of constitution.) One can justify general beliefs of the form *E is a phenomenal property* in much the same way, given that *R is a phenomenal property* is a priori.

From here, beliefs such as *I am conscious* are a short leap away. The leap is non-trivial, as there are distinctive problems about the epistemology of the self: witness Hume's scepticism about the self, and Lichtenberg's point that in the *cogito*, Descartes was entitled only to *there is thought*, not to *I think*. I have nothing special to say about these epistemological problems. But assuming that these problems can be solved, it is not implausible that a belief such as *if E exists, I have E* is justified (perhaps a priori). Then the whole range of first-person phenomenal beliefs lies within reach.

(If one takes direct phenomenal beliefs as truly foundational, one might even suggest that the *cogito* should have a three-stage structure: from $E = R$ (or some such), to *I have E*, to *I exist*!)

As for beliefs involving relational phenomenal concepts: presumably beliefs such as $S = red_I$, where S is a standing pure concept of phenomenal redness, will be justified a posteriori,

---

[18] This is relevant to an argument against conceivability arguments for property dualism given by Balog (1999). Balog maintains that a zombie could make a conceivability argument with the same form, with true premisses and a false conclusion, so the argument form must be invalid. Balog's argument requires as a premiss the claim that a zombie's assertion 'I am phenomenally conscious' (and the like) expresses a truth. But the discussion here suggests that it is much more plausible that the assertion is false or truth-valueless. This is plausible on independent grounds: in a zombie world, when a zombie realist asserts (an analog of) 'Qualia exist', and a zombie eliminativist asserts 'Qualia do not exist', it seems clear that the zombie eliminativist is closer to being correct. If so, Balog's argument fails.

Balog also discusses "Yogis", creatures that make a form of direct reference to brain states without this being mediated by phenomenology. I think it is clear that Yogis have at most a sort of demonstrative concept (roughly: "*this* inner state"), and do not have the analog of pure phenomenal concepts. For these concepts, no analogous epistemic gap arises. For example, given full physical and indexical information, Yogis will be in a position to know all truths involving the concepts in question.

perhaps by inference from the observation that the relevant paradigmatic objects typically cause one to experience instances of S. And beliefs of the form $S = red_C$ will be justified at least in so far as $red_I = red_C$ is justified. Of course for the first sort of belief to be justified, sceptical problems about the external world (and about the self) must be overcome, and for the second sort of belief to be justified, sceptical problems about other minds must be overcome. I have nothing special to say about these problems here. But assuming that these problems can be dealt with, then both general relational phenomenal beliefs (e.g. $S = red_C$) and particular relational phenomenal beliefs (e.g. $E = red_C$) will be justified straightforwardly.

It seems, then, that a wide range of phenomenal beliefs can be justified by inference from direct phenomenal beliefs (such as $E = R$), a priori phenomenal beliefs (such as $R = S$ and perhaps *If E exists, I have E*), and a posteriori phenomenal beliefs such as ($S = red_I$ and $S = red_C$). I have given a model for the justification of direct phenomenal beliefs. Phenomenal realism, and even epiphenomenalism, seems to pose no particular problem for the justification of the a priori phenomenal beliefs (or at least no distinctive problem that does not arise for a priori justification on any view). And the same goes for the justification of the a posteriori phenomenal beliefs. Even if experience plays no causal role, this does not matter. Experiences have no special role in justifying the a priori beliefs, and the justification of the a posteriori beliefs can be seen as derivative on beliefs of the form $E = S$ (which are already accounted for), plus general methods of external observation and inductive inference.

So all we need to justify all these beliefs is the justification of direct phenomenal beliefs, the justification of a priori beliefs in virtue of their content, and the justification involved in inference, observation, and induction. There are no special problems in any of these matters for the phenomenal realist. One might think that inference poses a problem for the epiphenomenalist: how do $E = R$ and $R = S$ justify $E = S$ if the content of R is partly constituted by an epiphenomenal quality, and if inference requires causation? But this is no problem: R acts as a middle term and its content is not required to play any special causal role. We can think of the inference in question as being *E is R, which is S, so E is S*. Here the content of R is inessential to the validity of the inference: as long as the premisses are justified, the conclusion will be justified.

Perhaps the main residual epistemological issue concerns the persistence of standing phenomenal concepts. One might worry if S is partly constituted by an element that is epiphenomenal, then even if one acquires a justified belief — say of the form *roses cause S* — at one time, it is not clear how this justification carries over to instances of a belief with that content at a later time. It is plausible that more than a match in content is required for justification: the later belief must be in some sense the "same" belief, or at least a

"descendant" belief, involving the "same" (or "descendant") concepts. The same sort of issue arises with inference of the sort in the previous paragraph. Whether or not *E is S* is wholly distinct from the two premisses, we certainly want later beliefs of the form *that was S* to be justified, and to play a role in further inferences in turn. But this arguably requires that the later concept be a "descendant" of the earlier concept in a sense that allows beliefs involving the later concept to inherit justification from beliefs involving the earlier concepts.

In response: I have no good account of what it is for one token of a concept to be a "descendant" of another, in a manner that allows it to inherit justification.[19] Nor, I think, does anyone. Clearly more than sameness in content is required: if a new concept with the same content were to be formed *de novo*, no justification would be inherited. So some sort of natural connection between concept tokens is required. But it is plausible that this sort of connection need only require an appropriate causal connection between the physical vehicles of the concept, along with an appropriate match in content: it is not required that the elements constituting the content of the initial concept do any distinctive causal work.

To see this, consider the persistence of concepts on an externalist view, where content is constituted by external factors that may lie in the distant past. Here, the factors that constitute the content of two tokens of the concept will play no distinctive role in causally connecting the tokens, since those factors lie in the distant past. The persistence will instead be supported by appropriate connections between the tokens' physical vehicles. It is plausible that the phenomenal realist, and the epiphenomenalist, can say something similar: conceptual persistence is underwritten by natural connections among vehicles, perhaps along with an appropriate match in content. Of course it would be desirable to have a full positive account of this sort of conceptual persistence, but it seems that there is no distinctive problem for the phenomenal realist here.

Further questions concern the justification of beliefs about the representational content of experiences, and the role phenomenal beliefs might play in justifying beliefs about the external world. I will not say anything about these issues here. But it is plausible that these issues pose mere challenges for the phenomenal realist to answer, rather than posing distinctive arguments against it. The distinctive epistemological problems for phenomenal realism have been removed.

---

[19] This sort of persistence relation among tokens is central to our use of concepts and beliefs, but has received less discussion than it might have. In effect, it introduces a "typing" of concepts and beliefs that is more fine-grained than a mere typing by content, but less fine-grained than a typing by numerical identity of tokens. This sort of typing was already tacit in my earlier discussion, when I said that direct phenomenal concepts do not persist beyond the lifetime of an experience, but that standing phenomenal concepts do.

4.4 "The Myth of the Given"

A traditional view in epistemology and the philosophy of mind holds that experiences have a special epistemic status that renders them "given" to a subject. This epistemic status is traditionally held to give phenomenal beliefs a special status, and sometimes to allow experiences to act as a foundation for all empirical knowledge. In recent years, this sort of view has often been rejected. The *locus classicus* for this rejection is Wilfrid Sellars's "Empiricism and the Philosophy of Mind" (1956), which criticized such views as involving "The Myth of the Given". Sellars's (deliberately abusive) term for the view has caught on, and today it is not uncommon for this label to be used in criticizing such views as if no further argument is necessary.

I do not know whether my view is one on which experiences are "given". It does not fit Sellars's official characterization of the given (as we will see), and there are other characterizations that it also does not fit. But the term "given" (and in particular "myth of the given") often shifts to encompass many different views, and it may well be that my view shares something of the spirit of the views that were originally criticized under this label. So rather than trying to adjudicate the terminological issue, we can simply ask: are any of the arguments that have been put forward against the "given" good arguments against the view I have put forward here?

Here one runs up against the problem that clear arguments against the "given" are surprisingly hard to find. There are many suggestive ideas in Sellars's paper, but few explicit arguments. When arguments appear, they often take the form of suggesting alternative views, rather than directly criticizing an existing view. But there is at least one clear argument against the "given" in Sellars's paper. This is his famous "inconsistent triad". This was intended as an argument against sense-datum theories, but it clearly applies to a wider class of views.

> It is clear from the above analysis, therefore, that classical sense-datum theories ... are confronted by an inconsistent triad made up of the following three propositions:
>
> A. x senses red sense content s entails x non-inferentially knows that s is red.
>
> B. The ability to sense sense contents is unacquired.
>
> C. The ability to know facts of the form x is phi is acquired.
>
> A and B together entail not-C; B and C entail not-A; A and C entail not-B. (Sellars 1956, section 6)

It is clear how the view I have put forward should deal with this inconsistent triad: by denying A. I have said nothing about just which mental capacities are acquired or unacquired, but on the view I have put forward, it is clearly possible to have experiences without having

phenomenal beliefs, and therefore without having knowledge of phenomenal facts. On my view, phenomenal beliefs are formed only rarely, when a subject attends to his or her experiences and makes judgments about them. The rest of the time, the experiences pass unaccompanied by any phenomenal beliefs or phenomenal knowledge.

Underlying Sellars's critique is the idea that knowledge requires concepts, and that experiences do not require concepts, so that having experiences cannot entail having knowledge. The view I have put forward is compatible with all this. On my view, experiences require little cognitive sophistication, and in particular do not require the possession of concepts. There may be some experiences that require concepts (for example, the experience of a spoon as a spoon), but not all experiences do. No concepts are required to experience phenomenal redness, for example. Knowledge of facts requires belief, however, and belief requires the possession of concepts. So experience does not entail knowledge.

Sellars associated the "given" most strongly with the acceptance of (A), and the denial of (A) is what he argues for himself. In discussing the possibility that a sense-datum theorist might deny (A), all he says is the following.

> He can abandon A, in which case the sensing of sense contents becomes a non-cognitive fact — a non-cognitive fact, to be sure which may be a necessary condition, even a logically necessary condition, of non-inferential knowledge, but a fact, nevertheless, which cannot constitute this knowledge.

On my view, all this is correct. Experiences do not, on their own, constitute knowledge. They play a role in *justifying* knowledge, and they play a role in *partly* constituting the beliefs that qualify as knowledge, in combination with other cognitive elements. But experiences themselves are to be sharply separated from beliefs and from items of knowledge. So none of this provides any argument against my view.

(On my reading, a number of the sense-datum theorists also deny (A), making clear distinctions between the sort of non-conceptual epistemic relation that one stands in by virtue of having an experience and the sort of conceptual epistemic relation that one has when one knows facts. Such theorists clearly avoid the conflation between experience and knowledge that Sellars accuses sense-datum theorists of making.)

Curiously, Sellars never discusses the possibility that experiences could justify knowledge without entailing knowledge. It seems clear that he would reject such a view, perhaps because he holds that only conceptual states can enter into justification, but this is never made explicit in his article.[20]

---

[20] The one further part of Sellars's article that may be relevant to the view I have put forward is part VI (sections 26-29), where he addresses the traditional empiricist idea that experience involves awareness of determinate

Although Sellars does not argue explicitly against this sort of view, such arguments have been given by a number of later philosophers writing in the same tradition. In particular, there is a popular argument against any view on which experiences are non-conceptual states that play a role in justifying beliefs. This argument, which we might call the *justification dilemma*, has been put forward by BonJour (1969), Davidson (1986), and McDowell (1994), among others. We can represent it as follows.

(1) There can be no inferential relation between a non-conceptual experience and a belief, as inference requires connections within the conceptual domain.

(2) But a mere causal relation between experience and belief cannot justify the belief; so

---

(3) Non-conceptual experiences cannot justify beliefs.

The first premiss is plausible, as it is plausible that inference is mediated by concepts. The status of the second premiss is much less clear. While it is plausible that the mere existence of a causal connection does not suffice to justify a belief, it is far from clear that the right *sort* of causal connection could not serve to justify a belief. McDowell says that a causal connection "offers exculpation where we wanted justification". But clearly causal connections cannot involve mere exculpation simply by virtue of being causal connections, as the case of inference shows: here a causal connection of the right kind between states can be seen to justify. So further argument is required to show that no other sort of causal connection (perhaps with subtle constraints on the content of a belief and on the relationship between belief and experience) can provide justification.

But in any case, even if the two premisses are accepted, the conclusion does not follow. An option has been missed: inference and causation do not exhaust the possible justifying relations between non-conceptual experiences and beliefs. On my view, the relation in question is not inference or causation, and neither is it identity or entailment, as on the views that Sellars criticized. Rather, the relation is partial constitution.

---

repeatables. This is closely related to my claim that experience involves acquaintance with properties. Sellars does not provide any direct argument against this view, however. He simply notes (sections 26-28) that Locke, Berkeley, and Hume take this thesis as a presupposition rather than a conclusion (they use it to give an account of how we can be aware of determinable repeatables). And then he asserts (section 29) that this awareness must either be mediated by concepts (e.g. through the belief that certain experiences resemble each other, or that they are red) or be a purely linguistic matter. He gives no argument for this claim, which I think should be rejected. On my view, our acquaintance with qualities requires neither concepts nor language.

I have already given a model of how the justification of a direct phenomenal belief by an experience works, involving three central elements that parallel the three central elements in the case of inference. The analog of the causal element is the constitutive connection between experience and belief; the analog of the content element is the match between epistemic content of belief and quality of experience; and the analog of the epistemic element is the subject's acquaintance with the phenomenal quality. If the model of justification by inference is accepted, there is no clear reason why this model should be rejected.

Some philosophers hold that only a conceptual state can justify another conceptual state. But as with the thesis that only a belief can justify another belief, it is not clear why this thesis should be accepted. It is not supported pre-theoretically: pre-theoretically, there is every reason to hold that experiences are non-conceptual and can justify beliefs. And there is no clear theoretical support for this claim, either. Proponents sometimes talk of "the space of reasons" in this context, but the slogan alone does not convert easily into an argument. McDowell suggests that justifications for our beliefs should be *articulable*, which requires concepts; but as Peacocke (2001) points out, we can articulate a justification by referring to a justifying experience under a concept, whether or not the experience itself involves concepts. Perhaps the central motivation for the thesis lies in the fact that we have a clear theoretical model for conceptual justification, but not for other sorts of justification. But again, this is a weak argument, and again, the exhibition of a theoretical model ought to remove this sort of worry.

In any case: the view I have put forward avoids Sellars's central version of the given (an entailment from experience to knowledge), and BonJour's, Davidson's, and McDowell's central version of the given (a mere causal connection), along with the arguments against those views. It may be that the view I have put forward accepts a "given" in some expanded sense. But the substantive question remains: are there good arguments against the given that are good arguments against this view? I have not been able to find such arguments, but I would welcome candidates.

5 Further Questions

I have drawn a number of conclusions about the content and epistemology of phenomenal beliefs. It is natural to ask whether these conclusions apply more generally.

First, regarding content: I have argued that the content of pure phenomenal concepts and phenomenal beliefs is conceptually irreducible to the physical and functional, because this content itself depends on the constitutive role of experience. Does this sort of irreducibility

extend to other concepts or beliefs? Is the content of concepts and beliefs irreducible to the physical and functional quite generally?

There is one class of concepts for which such a conclusion clearly follows. This is the class of concepts that have phenomenal concepts as constituents. Such concepts might include *the tallest conscious being in this room*, *the physical basis of consciousness*, and *the external cause of R*, where *R* is a pure concept of phenomenal redness. More generally, in so far as a concept has conceptual ties with phenomenal concepts, so that claims involving that concept conceptually and non-trivially entail claims involving pure phenomenal concepts, then the content of such a concept will be irreducible in a similar way.

It is arguable that many or most of our perceptual concepts have this feature. At least some concepts of external colors can be analysed roughly as *the property causally responsible for C in me*, where *C* is a pure concept of a phenomenal color. Things are more complex for community-level concepts. Here it is more plausible that an external color concept might be analysed in terms of community-wide relations to a non-specific phenomenal concept: perhaps *the property causally responsible for the dominant sort of visual experience caused by certain paradigmatic objects in this community*, or something like that. But this still has the concept of visual experience as a constituent, and so will still have functionally irreducible content. The alternative is that external color concepts might be analysed in terms of their relations to certain *judgments* or other non-experiential responses, in which case the reducibility or irreducibility will not be so clear. I will not adjudicate this matter here, but my own view is that while there may be some perceptual concepts without an obvious phenomenal component, many or most of the perceptual concepts that we actually possess have such a component.

One might try to extend this further. In the case of theoretical concepts from science, for example, one can argue that these have conceptual ties to various perceptual concepts (as the Ramsey-Lewis analysis of theoretical concepts suggests). If so, and if the perceptual concepts in question have irreducible content, it is arguable that these concepts have irreducible content. And one might argue for conceptual ties between intentional concepts and phenomenal concepts, and between social concepts and intentional concepts, so that a wide range of social concepts will turn out to have irreducible content. If this is right, then a being without consciousness could have at best impoverished versions of these concepts, and perhaps no such concepts at all.

This sort of argument will not work for all concepts. Many mathematical or philosophical concepts have no obvious tie to phenomenal concepts, for example. And in fact there is good reason to think that some concepts do not have a phenomenal component. If all concepts have

a phenomenal component, it would be hard to avoid the conclusion that all concepts are *entirely* constituted by phenomenal concepts, which would lead naturally to phenomenalism or idealism. My own view is that certain central concepts, such as that of causation, have no deep phenomenal component at all. Once this is recognized, it becomes clear that even if a wide range of concepts have a phenomenal component, only a small number of them are entirely phenomenal.

Even if some concepts have no phenomenal component, it is not out of the question that their content might still be irreducible. One intriguing possibility is that something about a subject's phenomenal states could be central to a subject's *possessing* a concept such as that of causation, or certain mathematical concepts, even though these concepts do not refer to phenomenal states as part of their content. (Compare a reductive view on which neural states might constitute the content of concepts that do not refer to neural states.) There is at least some intuition that a capacity for consciousness may be required to have concepts in the first place; and it is not obviously false that phenomenology plays a role in the possession of even non-phenomenal concepts.

Such a thesis would require much further argument, of course, and I am not certain whether it is true. But even if it is false, the more limited thesis that phenomenology plays a role in constituting the content of phenomenal concepts, and that phenomenal concepts play a role in determining the content of a wide range of other concepts, has significant consequences. If even the more limited thesis is true, then the project of giving a functional analysis of intentionality cannot succeed across the board, and a central role must be given to phenomenology in the analysis of intentional content.

Second, epistemology: I have in effect argued for a sort of limited foundationalism within the phenomenal domain. Direct phenomenal beliefs are in a certain sense foundational: they receive justification directly from experience, and their prima facie justification does not rely on other beliefs. And I have argued that direct phenomenal beliefs can justify at least some other phenomenal beliefs in turn, when aided by various sorts of a priori reasoning. Does this give any support to foundationalism about a broader class of empirical beliefs, or about empirical knowledge in general?

Nothing I have said implies this. This gap between phenomenal knowledge and knowledge of the external world remains as wide as ever, and I have done nothing to close it. The framework here is compatible with various standard suggestions: that phenomenology might justify external beliefs through inference to the best explanation, or through a principle that gives prima facie justification to a belief that endorses an experience's representational content. But so far, the framework outlined here does nothing special to support these

suggestions or to answer sceptical objections. And the framework is equally compatible with many alternative non-foundationalist accounts of our knowledge of the external world.[21]

Still, this framework may help to overcome what is sometimes taken to be the largest problem for foundationalism: bridging the gap between experience and belief. I have argued that an independently motivated account of the role of experience in phenomenal belief, and of subject's epistemic relations to them, has the resources to solve this problem, by exploiting the paired notions of constitution and acquaintance.

Any plausible epistemological view must find a central role for experience in the justification of both beliefs about experience and beliefs about the world. If what I have said here is correct, then we can at least see how experience gains a foothold in this epistemic network. Many other problems remain, especially regarding the relationship between experience and beliefs about the external world. But here, as in the case of phenomenal belief, a better understanding of the relationship between experience and belief may take us a long way.

Appendix

What follows is a brief and simplified introduction to the two-dimensional semantic framework as I understand it. See also Chalmers (2002*c*; forthcoming).

Let us say *S* is epistemically possible in the broad sense if the hypothesis that *S* is the case is not ruled out a priori. Then there will be a wide space of epistemic possible hypotheses (in the broad sense; I will usually omit the qualifier in what follows). Some of these will conflict with each other; some of them will be compatible with each other; and some will subsume each other. We have a systematic way of evaluating and describing epistemic possibilities that differs from our way of evaluating and describing subjunctive counterfactual possibilities. It is this sort of evaluation and description that is captured by the first dimension of the two-dimensional framework.

---

[21] A particular problem in extending this account to a general foundationalism is that we do not usually form direct phenomenal beliefs associated with a given experience, so such beliefs are not available to help in justifying perceptual beliefs. (Thanks to Alvin Goldman for discussion on this point.) Here there are a few alternatives: (1) deny that perceptual beliefs are usually justified in the strongest sense, but hold that such justification is available; (2) hold that the mere availability of justifying direct phenomenal beliefs confers a sort of justification on perceptual beliefs; or (3) extend the account so that perceptual experiences can justify perceptual beliefs directly, through a constitutive connection to perceptual concepts analogous to the connection to phenomenal concepts. I explore the third possibility in forthcoming work on the content of perceptual experience and perceptual belief.

It is epistemically possible that water is not H_2O, in the broad sense that this is not ruled out a priori. And there are many specific versions of this epistemic possibility: intuitively, specific ways our world could turn out such that if they turn out that way, it will turn out that water is not H_2O. Take the XYZ-world, one containing superficially identical XYZ in place of H_2O. It is epistemically possible that our world is the XYZ-world. When we consider this epistemic possibility — that is, when we consider the hypothesis that *our* world contains XYZ in the oceans, and so on — then this epistemic possibility can be seen as an instance of the epistemic possibility that water is not H_2O. We can rationally say "if our world turns out to have XYZ in the oceans (etc.), it will turn out that water is not H_2O". The hypothesis that the XYZ-world is actual rationally entails the belief that water is not H_2O, and is rationally inconsistent with the belief that water is H_2O.

Here, as with subjunctive counterfactual evaluation, we are considering and describing a world, but we are considering and describing it in a different way. In the epistemic case, we consider a world *as actual*: that is, we consider the hypothesis that our world is that world. In the subjunctive case, we consider a world *as counterfactual*: that is, we consider it as a way things might have been, but (probably) are not. These two modes of consideration of a world yield two ways in which a world might be seen to make a sentence or a belief true. When the XYZ-world is considered as actual, it makes true 'water is XYZ'; when it is considered as counterfactual, it does not.

In considering a world as actual, we ask ourselves: what if the actual world is really that way? In the broad sense, it is *epistemically* possible that Hesperus is not Phosphorus. This is mirrored by the fact that there are specific epistemic possibilities (not ruled out a priori) in which the heavenly bodies visible in the morning and evening are distinct; and upon consideration, such epistemic possibilities are revealed as instances of the epistemic possibility that Hesperus is not Phosphorus.

When we consider worlds as counterfactual, we consider and evaluate them in the way that we consider and evaluate subjunctive counterfactual possibilities. That is, we acknowledge that the character of the actual world is fixed, and say to ourselves: what if the world *had been* such-and-such a way? When we consider the counterfactual hypothesis that the morning star might have been distinct from the evening star, we conclude not that Hesperus would not have been Phosphorus, but rather that at least one of the objects is distinct from both Hesperus and Phosphorus (at least if we take for granted the actual-world knowledge that Hesperus is Phosphorus, and if we accept Kripke's intuitions).

Given a statement *S* and a world *W*, the *epistemic intension* of *S* returns the truth-value of *S* in *W* considered as actual. (Test: if *W* actually obtains, is *S* the case?) The *subjunctive*

intension of *S* returns the truth-value of *S* in *W* considered as counterfactual. (Test: if *W* had obtained, would *S* would have been the case?) We can then say that *S* is *primarily possible* (or 1-possible) if its epistemic intension is true in some world (i.e. if it is true in some world considered as actual), and that *S* is *secondarily possible* (or 2-possible) if its subjunctive intension is true in some world (i.e. if it is true in some world considered as counterfactual). Primary and secondary necessity can be defined analogously.

For a world to be considered as actual, it must be a *centred* world — a world marked with a specified individual and time — as an epistemic possibility is not complete until one's "viewpoint" is specified. So an epistemic intension should be seen as a function from centred world to truth-values. For example, the epistemic intension of 'I' picks out the individual at the centre of a centred world; and the epistemic intension of 'water' picks out, very roughly, the clear drinkable (etc.) liquid in the vicinity of the centre. No such marking of a centre is required for considering a world as counterfactual, or for evaluating subjunctive intensions.

Epistemic and subjunctive intensions can be associated with statements in language, as above, and equally with singular terms and property terms. The intension of a statement will be a function from worlds to truth-values; the intension of a term will be a function from worlds to individuals or properties within those worlds. (In some cases, intensions are best associated with linguistic tokens rather than types.)

Epistemic intensions can also be associated in much the same way with the (token) concepts and thoughts of a thinker, all of which can be used to describe and evaluate epistemic possibilities as well as subjunctive counterfactual possibilities. In "The Components of Content" (2002*c*) I argue that the epistemic intension of a concept or a thought can be seen as its "epistemic content" (a sort of internal, cognitive content), and that the subjunctive intension captures much of what is often called "wide content".

A crucial property of epistemic content is that it reflects the rational relations between thoughts. In particular, if a belief *A* entails a belief *B* by a priori reasoning, then it will be epistemically impossible (in the broad sense) for *A* to be true without *B* being true, so the epistemic intension of *A* entails the epistemic intension of *B*. Further, if an identity $a = b$ is a posteriori for a subject, then it is epistemically possible for the subject that the identity is false, and there will be an epistemic possibility in which the referents of the two concepts involved differ, so the subject's concepts *a* and *b* will have distinct epistemic intensions. This applies even to beliefs expressed by a posteriori necessities such as 'water is H_2O" and 'Hesperus is Phosphorus': the epistemic intensions of these beliefs are false at some worlds, so the concepts involved have different epistemic intensions. So epistemic intensions behave

something like Fregean senses, individuating concepts according to cognitive significance at least up to the level of a priori equivalence.

(A complication here is that on some philosophical views, there may be "strong necessities" whose epistemic intension is false at no world. An example might be 'A god exists', on a theist view on which a god exists necessarily but not a priori, or 'Zombies do not exist', on a type-B materialist view on which zombies are conceivable but metaphysically impossible. These necessities go well beyond Kripkean a posteriori necessities, and I have argued elsewhere (Chalmers 2002*b*) that there are no such necessities. If they exist, however, the present framework can accommodate them by moving to a broader class of conceptual or epistemic possibilities, which need not correspond to metaphysical possibilities (see Chalmers (forthcoming) for more details. In the cases above, for example, there will be at least a conceptually possible world (or "scenario") in which there is no god, and one in which there are zombies. More generally, any a posteriori belief will have an epistemic intension that is false at some such world.)

In the work presented here, the two-dimensional framework is being applied rather than being discussed or justified in its own right. The discussion here indicates important distinctions among phenomenal concepts whose analysis requires the idea of epistemic content. And importantly, there are epistemological distinctions that turn on these distinctions in content. This reflects a more general phenomenon: the sort of possibility that is most crucial in epistemology is epistemic possibility, and the sort of content that is correspondingly most crucial is epistemic content.

Bibliography

Austin, D. F. 1990. *What's the Meaning of "This"?* Ithaca, NY: Cornell University Press.

Balog, K. 1999. 'Conceivability, Possibility, and the Mind-Body Problem', *Philosophical Review*, 108: 497-528.

Bayne, T. 2001. 'Chalmers on Acquaintance and Phenomenal Judgment', *Philosophy and Phenomenological Research*, 62:407-19.

Bonjour, L. 1969. 'Knowledge, Justification, and Truth: A Sellarsian Approach to Epistemology', Ph.D. Dissertation, Princeton University. http://www.ditext.com/bonjour/bonjour0.html.

Bonjour, L. 1978. 'Can Empirical Knowledge Have a Foundation?' *American Philosophical Quarterly* 15:1-13.

Chalmers, D. J. 1996. *The Conscious Mind: In Search of a Fundamental Theory*, New York: Oxford University Press.

Chalmers, D. J. 2002*a*. 'Consciousness and its Place in Nature', in S. Stich and F. Warfield (eds.), *The Blackwell Guide to the Philosophy of Mind* Oxford: Blackwell. http://consc.net/papers/nature.html.

Chalmers, D. J. 2002*b*. 'Does conceivability entail possibility?', in T. Gendler and J. Hawthorne (eds.) *Conceivability and Possibility* Oxford: Oxford University Press. http://consc.net/papers/conceivability.html.

Chalmers, D. J. 2002*c*. 'The Components of Content', in D. Chalmers (ed.) *The Philosophy of Mind: Classical and Contemporary Readings*, New York: Oxford University Press. http://consc.net/papers/content.html.

Chalmers, D. J. (forthcoming), 'The Nature of Epistemic Space'. http://consc.net/papers/espace.html.

Chisholm, R. 1957. *Perceiving: A Philosophical Study* Ithaca: Cornell University Press.

Conee, E. 1985. 'The Possibility of Absent Qualia', *Philosophical Review*, 94: 345-66.

Davidson, D. 1986. 'A Coherence Theory of Truth and Knowledge', in E. Lepore (ed.) *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson* Oxford: Blackwell.

Francescotti, R. M. 1994. 'Qualitative Beliefs, Wide Content, and Wide Behavior', *Nous*, 28: 396-404.

Fumerton, R. 1995. *Metaepistemology and Skepticism* Lanham: Rowman and Littlefield.

Gertler, B. 2001. 'Introspecting Phenomenal States', Philosophy and *Phenomenological Research*, 63: 305-28.

Hawthorne, J. (2002), 'Advice to Physicalists', *Philosophical Studies*, 101:17-52.

Ismael, J. 1999. 'Science and the Phenomenal', *Philosophy of Science*, 66: 351-69.

Jackson, F. 1982. 'Epiphenomenal Qualia', *Philosophical Quarterly*, 32: 127-136.

Kaplan, D. 1989. 'Demonstratives', in J. Almog, J. Perry, and H. Wettstein (eds.), *Themes from Kaplan* New York: Oxford University Press.

Kripke, S. A. 1981. *Wittgenstein on Rules and Private Language*. Cambridge, Mass.: Harvard University Press.

Loar, B. 1997. 'Phenomenal States (Second Version)', in N. Block, O. Flanagan, and G. Güzeldere (eds.) *The Nature of Consciousness*. Cambridge, Mass.: MIT Press.

McDowell, J. 1994. *Mind and World* Cambridge, Mass.: Harvard University Press.

Nida-Rümelin, M. 1995. 'What Mary Couldn't Know: Belief about Phenomenal States', in T. Metzinger (ed.) *Conscious Experience* Exeter: Imprint Academic.

Nida-Rümelin, M. 1996. 'Pseudonormal Vision: An Actual Case of Qualia Inversion?' *Philosophical Studies*, 82: 145-57.

Nida-Rümelin, M. 1997. 'On Belief about Experiences: An Epistemological Distinction Applied to the Knowledge Argument'. *Philosophy and Phenomenological Research*, 58: 51-73.

Peacocke, C. 2001). 'Does Perception Have a Non-Conceptual Content?' *Journal of Philosophy*, 98: 239-64.

Perry, J. 2001. Knowledge, Possibility, and Consciousness Cambridge, Mass.: MIT Press.

Pollock, J. 1986. *Contemporary Theories of Knowledge* Lanham: Rowman and Littlefield.

Raffman, D. 1995. ‘On the Persistence of Phenomenology’, in T. Metzinger (ed.) *Conscious Experience* Exeter: Imprint Academic.

Russell, B. 1910. ‘Knowledge by Acquaintance and Knowledge by Description’, *Proceedings of the Aristotelian Society*, 11: 108-128.

Sellars, W. 1956. ‘Empiricism and the Philosophy of Mind’, *Minnesota Studies in the Philosophy of Science*, 1: 253-329. Repr. as *Empiricism and the Philosophy of Mind* Cambridge, Mass: Harvard University Press, 1997.

Shoemaker, S. 1975. ‘Functionalism and Qualia’, *Philosophical Studies*, 27: 291-315.

Wittgenstein, L. 1953. *Philosophical Investigations* Oxford: Blackwell.