Acquaintance and the Mind-Body Problem

Katalin Balog

[Penultimate draft. In: *New Perspectives on Type Identity: The Mental and the Physical*, Christopher Hill and Simone Gozzano (eds.), Cambridge University Press, 2012.]

> I wished to represent, in my own way, according to my own ideas, the material that was given to me, my material, myself.... But there is something that I – perhaps understandably – didn't take into account: that we cannot ever represent ourselves *to* ourselves. (Imre Kertész, *Fiasco*)

In this paper[1] I will lay the groundwork for an account of *acquaintance* and discuss the consequences of the account for the metaphysics of mind. Acquaintance is a unique epistemological relation that relates a person to her own phenomenally conscious states and processes *directly, incorrigibly,* and *in a way that seems to reveal their essence*.[2] When one is aware of a phenomenal state in the process of having it, something essential about it is revealed, directly and incorrigibly – namely, *what it is like to have it.* Such an epistemic relation has struck many philosophers as deeply mysterious. One of the aims of this paper is to dispel some of the mystery by providing an account of direct phenomenal concepts. These are the concepts deployed when a person is acquainted with her own conscious states in introspection, e.g., when I think to myself "I have felt this in my shoulder before" upon noticing a familiar feeling as I throw a frisbee. For reasons that will become clear I call my proposal "the quotational account of direct phenomenal

[1]Special thanks to David Papineau and Michael Della Rocca for very helpful comments on earlier drafts of this paper. I would also like to thank Ned Block, Chris Hill, Joe Levine, Barry Loewer, Mike Martin, Gabriel Rabin, Howard Robinson, and audiences at the NEH Summer Institute on Consciousness and Intentionality at UCSC, 2002, the MIT Philosophy Department, the Summer Workshop of Collegium Budapest, 2004, and the Cognitive Science Group at the CUNY Graduate Center, 2010 for comments and criticism.

[2] I accept this with caveats – see the discussion later in this paper.

concepts." The quotational account is a speculative proposal about human mental architecture. Although it is neutral between physicalist and dualist accounts of qualia in that both metaphysical views are compatible with it, if the general cognitive architecture invoked in it turns out to be correct physicalism scores a strategic victory. It is because the general cognitive architecture invoked in the quotational account has the resources to explain the nature of acquaintance. Therefore it obviates the need to explain acquaintance by way of appealing to the special, irreducibly mental, non-physical nature of phenomenal consciousness.

This paper has two aims. The first is to elaborate on an account of phenomenal concepts that, in my view, yields a satisfying physicalist account of acquaintance. The second, related goal is to show how such an account can be used as a powerful and quite general response to a whole slew of recent arguments against physicalism. I will start by briefly introducing these arguments.

1) Physicalism, Dualism, and the Zombie Argument

According to physicalism, the world's fundamental ontology is physical and the best account of that ontology is provided by fundamental physics. Contemporary physics tells us that this ontology consists of particles, strings and fields of various types that occupy space-time (or bear spatio-temporal relations to one another) and possess a limited number of quantitative properties (mass, charge, electromagnetic potential, and so on). Physics also claims that there are only a few fundamental dynamical and perhaps non-dynamical laws that govern the structure of space-time and evolution of its occupants. Physicalism thus understood is defined as follows: all truths, including truths about phenomenal consciousness, are metaphysically necessitated by the complete physical truth about the world. [3] This is the *Physicalist Entailment Thesis* (Phys)

[3] This formulation is due to Jackson 1993. The first precise formulation of physicalism of this sort comes from Lewis 1983. Subsequent discussions are variations of the same theme. Many philosophers, among them non-physicalists, accept this kind of definition as capturing the intuitive notion of physicalism (see, e.g., Chalmers 1996, p 41-42, Loewer 2001, Melnyk 2003, and Papineau 1993b).

(Phys) For all true statements T, □ (P -> T) [4]

> where *P* is the complete fundamental physical description of the world including the fundamental physical laws and also including a statement to the effect that it is complete.[5]

If there are psychological truths – for example, that Mary knows what it is like to see red – that are not necessitated by *P* then physicalism is false.

According to Dualism, the complete physical description of our world doesn't necessitate all mental truths: *P* leaves something out. Contemporary dualists generally do not think – as Descartes did – that what is left out are basic mental *entities* but they do maintain that there are basic mental (and proto-mental) *properties*, in particular, that there are basic phenomenal properties. They also usually think that there are fundamental laws that link phenomenal properties to certain properties of physical systems.[6] By their lights a complete description of our universe must include truths about where, when, and which conscious states are exemplified.

I won't rehearse the reasons to believe that physicalism is true.[7] But I will discuss some arguments that have persuaded many philosophers that physicalism is not true.[8]

---

[4] More formally, the definition is: (Y)(Y -> □ (P -> Y)), where Y is a sentential substitutional quantifier.

[5] This last clause is needed to deal with the following complication in formulating physicalism. Statements that make reference to special kinds of property—to put it crudely, negative, and global properties—are not necessitated by the full physical description of the world; they are only necessitated by the conjunction of the full physical description of the world together with the statement that it *is* the full fundamental description of the world. However, this issue will not make a difference for the rest of this paper so I will ignore it.

[6] I will assume that these laws are contingent; i.e., not metaphysically necessary. If laws are taken to be metaphysically necessary then it is difficult to state the difference between Physicalism and Dualism since then both would hold that configurations of physical property instantiations metaphysically necessitate mental property instantiations.

[7] For an argument for physicalism, see, e.g., Loewer 1995, and Papineau 1995.

There is a line of argument going back at least to Descartes' argument for the distinctness of mind and body that claims to show that physicalism is indeed false. In fact, these arguments can be understood to conclude, on the basis of *a priori* considerations, that *no* world where phenomenal properties are exemplified can be a purely physical world. The descendent of this argument that has received the most attention in the last decade is David Chalmers' "Zombie Argument."

Chalmers' (Chalmers 2009) most recent formulation of the zombie argument is as follows:

The Zombie Argument

1) *P*&~*Q* is conceivable. [9]

2) If *P*&~*Q* is conceivable then *P*&~*Q* is metaphysically possible (CP principle)

3) If P&~Q is metaphysically possible then physicalism is false.

_____

4) Physicalism is false.

By "statement *S* is conceivable" Chalmers (1996) meant "*S* cannot be ruled out a priori". Later Chalmers 2002 introduces a battery of conceivability concepts. For my present purposes I will bracket the complications these different notions of conceivability introduce into the debate.[10]

---

[8] In the empirical spirit recently gaining traction in philosophy, I would like to point out that according to a recent survey (conducted by David Bourget and David Chalmers in November 2009 at *Philpapers*, http://philpapers.org/surveys/), 27 percent of the sample – consisting mostly of professional philosophers, philosophy Ph.D students, and some others – are dualists.

[9] *P* is the complete fundamental physical description of the world, including the fundamental physical laws, and *Q* is a positive phenomenal truth, e.g., that someone is having a visual experience with a particular phenomenal character at a particular time.

[10] Chalmers (2009) adds some clarifications and emendations to the argument. Since none of these affect my response to the Zombie Argument I will ignore them and stick with the simplified version of the argument.

The Zombie Argument is valid. Premise (3) is entailed by the proposition that Phys is a necessary condition for Physicalism.[11] Philosophers who think that there is a functional or representational analysis of phenomenal consciousness reject (1).[12] But I agree with Chalmers that there is no functional or representational analysis of phenomenal consciousness and that no physical description *a priori* entails any positive phenomenal description. Later I will offer some considerations based on the nature of phenomenal concepts for why this is so but for now I will just assume that (1) is true. So for both Chalmers and myself the crucial premise in the argument is (2).[13]

How can physicalists respond to the Zombie Argument and its ilk? In Balog (1999) I refuted the Zombie Argument by arguing that if it is sound then it follows – given a few plausible assumptions – that a zombie counterpart to this argument is also sound. But it is not, hence the conceivability argument is unsound as well. However, in another paper (Balog ms) I show that Chalmers' zombie conceivability argument can be modified in a way that makes it resistant to this refutation. Here I propose to follow an approach – dubbed by Stoljar (2005) the "phenomenal concepts strategy" – that answers this new version of Chalmer's argument as well as other dualist arguments[14] by proposing a physicalist account of phenomenal concepts.

---

[11] *Phys* states that for all true positive statements *T*, □ (P ⊃ T), so if P&~Q is metaphysically possible then (Phys) is false and therefore physicalism is false.

[12] E.g. Lewis (1966) and Jackson (2003).

[13] Chalmers' defense and development of the two-dimensional framework and of the conceivability-possibility link can be found in Chalmers and Jackson 2001, Chalmers 2002, and Chalmers 2004. There are important discussions in Block and Stalnaker (1999), Yablo (1993, 2002), and Soames (2005). I briefly discuss what I think goes wrong with a related argument by Frank Jackson (Balog 2002).

[14] Similar arguments include, among others, arguments based on conceivability considerations by Kripke (1972), Nagel (1974), Bealer (1994), Chalmers (1996, and 2009), as well as the Knowledge Argument of Jackson (1982), versions of the Property dualism Argument in Robinson (1993), White (2007), and Nida-Rümelin (2007), and the Explanatory Gap Argument in Levine (2001) and (2007). The response to the Zombie Argument I offer via my account of phenomenal concepts can be adapted to respond to these other arguments as well, but in the paper I will directly address only Chalmers' version.

2) Desiderata for an Account of Phenomenal Concepts

Consciousness, appears puzzling for many reasons – not just because the conceivability of zombies. Below is a list of those features that seem most intractable for physicalism. I have gleaned these from the philosophical literature, but they also mostly strike me as what a non-philosopher would say; if not quite in these words. I suggest that a successful account of phenomenal concepts will explain these features, or most of them, since the traditional puzzles about consciousness are mostly *epistemic* in nature.

1) *Only subjects who have undergone or are currently undergoing the relevant phenomenal states can token the corresponding phenomenal concepts*. This underlies Jackson's 1982 Knowledge Argument and is widely accepted.[15]

2) *Asymmetric epistemology. We are directly aware of our own conscious states in ways no one else can be.* One can be aware of one's conscious states simply by attending to them; to be aware of other's conscious states one has to observe their behavior. No one seems to contest this observation except Wittgensteinians and analytic behaviorists.

3) *Transparency: when one turns one's attention to one's own conscious perceptual experience, one can become aware of the features of the objects perceived.* There is a stronger version of the transparency thesis advocated by *representationalists*.[16] Representationalists argue that when one attends to one's conscious experience, one is aware *only* of the representational content of the experience, or alternatively, *only* of features of the objects perceived, and conclude from this that qualia, i.e., intrinsic, qualitative, introspectable features of

---

[15] For a denial of this claim see Tye (2009).

[16] See, e.g., Harman 1990, McDowell 1994, Tye 2000, and Jackson 2004 for transparency arguments.

conscious experience don't exist.[17]

4) *Infallibility/incorrigibility: we seem to be infallible about certain judgments involving certain phenomenal concepts – e.g., my judging 'phenomenal red is occurring right now'*. The reason we tend to believe it is that it doesn't seem as though any belief concerning objective matters of fact can coherently override or correct our own judgment about what we feel when it occurs simultaneously with the experience.[18] I will argue that there are cases for which the thesis will come out true.[19]

5) *Zombies are conceivable*, which means that the scenario in which zombies exist cannot be ruled out on a priori grounds. The main objectors to this are analytic functionalists.[20]

6) *There is an explanatory gap*. No amount of knowledge about the physical facts (brain functioning and so on) is able to explain why a particular brain state/process has a particular feel, e.g. feels giddy. Whatever causal/functional/physical information we have about the brain processes that underlie phenomenal experience – i.e. about the neurophysiological, functional, or representational features of phenomenal experience – the fact that such experience has a distinct phenomenal character might be still *left out*. In contrast, all facts

[17] For discussions of this argument see, e.g., Martin 2002 and Stoljar 2004.

[18] Note the difference between this and perceptual illusions like the Müller-Lyer illusion. We cannot help but see the two lines as differing in length although we can correct the ensuing belief that they differ in length by, e.g., measuring them. On the other hand, no measurement, or, for that matter, no information about our brain states would or should correct our judgment that the lines *appear* to be different in length.

[19] Ryle and Wittgenstein were notable critics of the infallibity claim. See also Schwitzgebel (2008) for a rather pessimistic assessment of the reliability of introspective acquaintance with qualia.

[20] See also Kirk 2005 for an interesting argument whose grounds go beyond analytic functionalism.

about water (that it is transparent, potable, etc.) are explicable in terms of facts about H_2O, together with physical and chemical laws. Nothing seems to be left out by such an explanation. Since we can't explain in the same way why a brain state feels giddy it is held that there is an explanatory gap between the phenomenal and the physical.

7) *Acquaintance. We know our conscious states not by inference but by immediate acquaintance which gives us direct, unmediated, substantial insight into their nature*. This, in opposition to the representationalist strong transparency thesis, commits one to the existence of *qualia*. I believe that qualia exist.[21] I think that we can attend to our experience and form direct, non-inferential concepts of its qualitative character that figure in phenomenal judgments. I also believe this gives us substantial insight into the nature of consciousness, and will shortly give an account of what this substantial insight consists in.

8) *There is something it is like to have conscious states.* This, e.g., that there is something it is like to see a cloudless blue sky, is the most obvious "given" about consciousness. Even (most) representationalists don't question its existence. The denial of 8 qualifies one as an eliminativist about consciousness.

The task of the physicalist is to explain 1-8 in a manner compatible with physicalism. It is important to emphasize that this doesn't mean that the physicalist will have to give a perspicuous physical explanation of qualia, that is, close the explanatory gap. In my view once we understand what the explanatory gap consists in we will see that it cannot be closed. However, a satisfactory physicalist account should explain *this*, the fact that there is an unbridgeable explanatory gap, and show that all the other, puzzling features of consciousness are, far from posing a problem for the physicalist view, features the physicalist will *expect* consciousness to have. Most theorists have attempted to explain 1-8 in terms of the nature of consciousness itself or to *explain away* these

[21] See Block 2003 and Loar 2003 for arguments for qualia. Dualists, naturally, tend to be committed to qualia.

features. It is not surprising that neither physicalist nor dualist accounts of consciousness have been very successful at explaining these features since features 1-7 are entirely *epistemic* features. So it seems reasonable to suppose that the key to their understanding will correspondingly lie in understanding the conceptual apparatus we use to think about them.

I propose focusing on our epistemic relation to consciousness, and especially on *acquaintance,* in trying to account for the puzzles of consciousness. This approach to the problems of consciousness has been aptly dubbed "the phenomenal concept strategy".[22]

3) The constitutional account of phenomenal concept

I will assume in the following that concepts are or can be constituents of thoughts and that concepts and thoughts are representations. I will also assume that concepts are mental representations that are language like – words of *Mentalese.*[23] The important point for the following is that since concepts and experiences are occurrent entities (events, states, processes) they can be constituents of one another and bear causal relations to one another.[24]

Concepts are the words of Mentalese. A particular token of a concept, e.g., DOG possesses a number of different kinds of properties and relations that are relevant to my discussion: i) realization properties, ii) conceptual role, and iii) semantic properties.

i) When one tokens an instance of DOG, say in thinking the thought DOGS BARK, that token is realized by some neural state or process. The neural properties that are relevant to the token's being a token of DOG are its *realization properties*. A

[22] The phenomenal concept strategy has been challenged by Chalmers 2007, Levine 2006 and Stoljar 2005. I respond to this challenge elsewhere (Balog 2012).

[23] There may well be non conceptual mental representations – image-like, map-like representations as well. It is plausible that tokens of phenomenal experience are non-conceptual representations.

[24] There are philosophers who would like to avoid Mentalese or avoid representations altogether. It may be that my account can be made compatible with their ontologies but that is not something that I can do here.

concept's realization properties are analogous to the particular physical type that realizes this (written, or electronic) token of "dog" or the particular sounds that realize a particular utterance of "dog."

ii) A concept's *conceptual role* is the totality of causal relations (and dispositions) that tokens of thoughts containing the concept bear to each other and to perceptual inputs and behavioral outputs. Certain aspects of a concept's conceptual role may be essential to or even individuative of that concept while others are merely accidental; e.g., it is essential to the concept OR that one be inclined to make certain inferences, such as the inference from *P* to *PvQ*. It might also be essential to perceptual concepts, e.g., RED, that they be caused by certain perceptual inputs. Presumably, however, it is not essential to RED that one be caused to believe RED IS MY FAVORITE COLOR by the same perceptual inputs. How exactly to draw the distinction (which may be vague) between a concept's essential and non-essential roles is controversial.

iii) A concept's *semantic features* concern what, if anything, the concept refers to. For example, the concept DOG, refers to the property of being a dog. Exactly what determines the reference of a Mentalese word (with particular realization properties, syntax and role) is a difficult and controversial matter. It is widely (though not universally) held that a concept's role (or the part of it essential to the concept) at least plays a part in determining the concept's reference. This part is the concept's *mode of presentation.* It often, but not always, has the form of a description - i.e., the thinker is disposed to infer the description from the tokening of the concept - i.e., from ARISTOTLE one is disposed to infer THE TEACHER OF ALEXANDER, etc. One can think of these descriptions as contents of a file attached to the concept. It is also widely accepted that reference is determined at least partly by external – causal, informational, or teleosemantic – relations of the concept to its environment.

A thinker typically has only partial epistemic access to features i-iii by introspection. When I attend to my thoughts I typically can obtain introspective knowledge of their semantic contents, e.g., that I am thinking about dogs. It is also plausible, though controversial that one can obtain information about the conceptual roles of one's concepts – and which of these are essential – by intuitions based on thought experiments, e.g., by asking oneself questions like "could one know p if p were false?".

But the realization properties of one's Mentalese words – the "shapes", or "mental ink" they are written in, so to speak –are almost always completely opaque. Almost always, with the exception – I propose – of *phenomenal concepts*.

I would like to propose an approach to phenomenal concepts that fits into this general framework and at the same time explains the epistemic puzzles involving consciousness outlined above. An examination of the features of phenomenal concepts suggests that a successful account of phenomenal concepts will posit an intimate connection between conscious states and the concepts we form of them. Loar[25] suggested the idea that phenomenal concepts are very special, direct demonstrative concepts. Abstracting from some of the details, what he seems to have in mind is that when a person is having a particular experience she can deploy a concept that refers directly to the experience and that in some way the mode of presentation associated with the demonstrative involves the experience itself. How could we understand direct reference via these special modes of presentation? As Papineau[26] points out, the suggestion doesn't help if by 'mode of presentation' we mean a description that we can already think and so we can use that description to think of an entity which has those properties. That would be presupposing phenomenal concepts in the explanation of those very concepts. We have to think about the 'mode of presentation' of phenomenal concepts in some other way.

There is a problem with Loar's account that points the way towards an answer to our question above. Loar thinks of phenomenal concepts as in some way "tracking" their referents. This suggests that he is thinking of the phenomenal concept and its referent as distinct entities related by causation. But it seems that this leaves too much of a distance

[25] See Loar 1990, 1997. The idea that the mind-body problem is a product of the special ways in which we conceive (in the first person) of our phenomenal states is first formulated in this paper. A similar proposal by Scott Sturgeon 1994 appeals to the special epistemology of phenomenal states.

[26] Papineau 2002, ch. 4.

between, e.g., a phenomenal concept P one applies to a particular pain **p**[27] as it occurs and **p** itself, as on this view their occurrence is independent. On a "tracking" view, P, or rather, a concept just like P could be tokened by someone in the complete absence of pain. A person like this would be a partial conceptual zombie; a *conceptual duplicate* of a normal human who, however, fails to have all the qualia the normal human has. But it seems to me that such a zombie is really impossible. Anybody who tokens a direct phenomenal concept as of a presently occurring pain is really in pain. The trouble with Loar's account is that it opens up the possibility of an appearance/reality distinction for direct phenomenal judgment whereas for direct phenomenal judgment there is no such distinction.

There is a way of thinking about phenomenal concepts which avoids these problems. It involves variations on the idea that (certain) phenomenal concepts are partly constituted by the phenomenal experiences they refer to.[28] On this view, a current phenomenal experience is *part* of the token concept currently applied to it, and the experience – at least partly – determines that the concept refers to the experience it contains. Of course, by "part" I do not mean "spatial part" but rather part in the sense that it is *metaphysically* impossible to token the concept without tokening its referent. I will cash this out presently. If this account is right, phenomenal concepts have very special *realization properties*: the neural states realizing these concepts are the very same neural states the concepts refer to!

This account is not intended to apply to all concepts that refer to phenomenal states or properties but only to "direct phenomenal concepts". But of course most of our reference to phenomenal states and qualia do not contain the phenomenal states themselves. What about "indirect phenomenal concepts"? Clearly, a person can token a concept that refers to pain without her literally experiencing pain as when she replies to her dentist's question by "I am not in pain" or when one sees another person stub her toe

[27] The same problem, by and large, arises for type phenomenal concepts as well; however, because of complications having to do with failures of incorrigibility, I won't appeal to the type case here.

[28] Similar ideas are proposed in Papineau 2002, Balog 2006 and Block 2006; David Chalmers 2003 also put forward a variation of this account.

and thinks THAT HURTS. Indirect phenomenal concepts are applied to non-occurrent (e.g., past or future) experiences of oneself or to the experiences of other people.[29] Understanding these is essential for understanding consciousness;[30] but for the rest of the paper I will focus exclusively on direct phenomenal concepts.

Direct phenomenal concepts pick out their referent in virtue of their being partly constituted by a token of their reference.[31] In this they are unique among concepts. On this account, there is an intimate relation between a phenomenal concept and its referent; more intimate than any causal or tracking relation. It is also a way of cashing out the idea that the experience serves as its own mode of presentation.[32] The experience, so to speak, presents itself.

Later on I will fill in the details of my version of the constitutional account; but the core idea is what does the work in terms of explaining 1-7. Let me proceed to actually spelling out those explanations.

1) *Only subjects who have undergone or at least are currently undergoing the relevant phenomenal states can token the corresponding phenomenal concepts*. This is straightforwardly the case for direct phenomenal concepts because of the way they are constituted. In the case of indirect phenomenal concepts, the explanation is a bit more

[29] The relationship between types of phenomenal concept and types of application is actually more complicated, as Kati Farkas has pointed it out to me. It is possible to apply direct phenomenal concepts to another's experience, like when one introspectively focuses on one's own experience of red and judges YOU ARE EXPERIENCING R where R is a direct phenomenal concept formed on the basis of one's experience of red. However, the distinction between direct and indirect phenomenal concepts is not affected by this complication.

[30] My view is that these concepts are individuated in part by conceptual roles that link them to direct phenomenal concepts.

[31] There is a further complication. Direct phenomenal concepts, like the one I form of a buzzing sound as I listen to it can refer either to particular (current) experiences of the thinker, or to phenomenal types exemplified in current conscious experience. I will indicate as a go which kind of concept I have in mind.

[32] Some of Loar's remarks suggest that he might understand "serves as its own mode of presentation" in this way but other remarks suggest that he is thinking of the relation as causal.

complicated.

2) *Asymmetric epistemology.* One's awareness of one's own conscious states constitutively involves those very states. One couldn't be aware of another's states in the same way given the distinctness of the minds/brains involved.

3) *Transparency*: When one turns one's attention to one's own conscious perceptual experience, one is aware of the features of the objects perceived. On the constitutional account, the experience contained within the concept maintains its representational features; I take it that experiences including sensations, afterimages, phosphenes, etc are representational.[33] So, for example, when a visual experience, that is, a phenomenally conscious non-conceptual representation of an object (or objects) and their properties partially constitutes a phenomenal concept representing it, attention directed to it will typically *also* or *primarily* be directed to the way the object is represented to be. I, however, deny the stronger version of the transparency thesis advocated by representationalists, namely the thesis that when one attends to one's conscious experience, one is aware *only* of the representational content of the experience. In my view, one can also direct one's attention to the phenomenal character of the experience, which is not identical to its representational content. (More on this in my explanation of acquaintance, under 7.)

4) *Direct phenomenal judgment is infallible/incorrigible.* On the constitutional account, 4 will come out true for certain kinds of phenomenal judgments. For example, a phenomenal concept may refer to a particular type of visual experience, say the experience typically caused by seeing red objects in ordinary light, etc. – call this type of experience "reddish" – by being constituted in part by a particular token of that type of experience. Then if I form the judgment I HAVE R where R is a direct phenomenal concept of reddish, my judgment cannot fail to be true.

[33] I am not claiming that phenomenal experience can be analyzed in terms of or is exhausted by its representational character as representationalists hold but just that phenomenal experience purports to represent.

This suggestion bears some similarity to Tyler Burge's (1988) account of self-knowledge. According to Burge, certain judgments about the intentional contents of one's states are self-certifying. Take for example, the judgment 'I am thinking that there may be life on another planet'. In order to make the judgment one has to do the thinking so the judgment must be true. From this point of view, Burge's account of our judgments about our thoughts, and the constitutional account of (certain of) our judgments about our experiences, are similar. Burge, however, doesn't offer any specific theory of our subjective concepts of our own thoughts. On my proposal, in order to token a direct phenomenal concept, one has to token the phenomenal state to which it refers, and *this is* what makes some of our phenomenal judgments self-certifying.[34]

5) *The conceivability of zombies* is explained by the *directness* and *substantiality* of our direct phenomenal concepts which, under the constitutional account, is compatible with physicalism. The directness of phenomenal concepts follows from the fact that the reference of a direct phenomenal concept is determined by how it is constituted and *not* by any description that is associated *a priori* with the concept. Phenomenal concepts are supposed to be different in this way from concepts like WATER and even name concepts like CICERO. Chalmers and Jackson (2001) claim that these concepts are associated *a priori* with descriptions (e.g. "the transparent potable liquid…", "the Roman orator who is at the origin of a causal chain culminating in this token") and these connections are *sufficient* to rule out a priori a scenario where, e.g., everything is physically the same but yet there is no water. One doesn't have to commit to this to see that zombies are conceivable; however, the conceivability of zombies is only really significant if this is the case. So the point is that if one allows that this is true with respect to the concept WATER, or CICERO – thereby allowing the Zombie Argument to get off the ground, one *still* has to admit that it is not so with respect to phenomenal concepts; that the existence of zombies cannot be ruled out a priori. Because of the fundamentally different cognitive architecture of phenomenal concepts, there are no a priori connections between

[34] Notice that on Burge's view, judgments about our own experiences are not self-certifying in the way judgments about our own thoughts are. The judgment 'I have a reddish experience' is not self-certifying, at least not on the grounds that the judgment 'I am thinking that there might be life on other planets' is.

phenomenal and physical/functional/structural concepts that are sufficient to rule out a priori the zombie scenario.[35]

6) *The explanatory gap.* Recall that the explanatory gap problem is that no amount of knowledge about the physical facts (brain functioning and so on) is able to explain why a particular brain state/process has a particular feel, e.g., feels giddy. This contrasts with the way the fact that water is composed of H_2O molecules together with physical and chemical laws explains why water is potable, transparent and so on. Once we have an explanation of why H_2O behaves in watery ways (and that it is the only substance that does so) we have an explanation of why water is H_2O. Since we can't explain why a brain state feels giddy in neurophysiological terms, we can't close the physical-phenomenal gap. You can see why this is in the following way. In the case of water and H_2O, the hypothesis that water=H_2O is quite natural in the light of all we know about H_2O and the laws that govern the behavior of H_2O – indeed, the opposite hypothesis doesn't even make sense. The hypothesis that the processes involving H_2O molecules are only nomologically correlated to the non-physical and non-chemical processes involving water is a non-starter.[36] On the other hand, the hypothesis that a phenomenal state is *identical* with a certain neurophysiological/functional state of the brain is just as compatible with our evidence as the opposing view. The hypothesis – endorsed by certain dualists – that phenomenal states and brain states are merely nomologically correlated makes perfect sense.

---

[35] *Nota bene*: I am not denying that there are inferential links between thoughts involving direct phenomenal concepts that are individuative of them. I will argue that there are conceptual links between direct phenomenal concepts on the one hand, and indirect phenomenal concepts, other mental concepts, and behavioral concepts, etc. on the other. My point is that to the extent that these are *a priori* they are not of the sort that enables one to rule out *a priori* the zombie-scenario.

[36] (Block and Stalnaker (1999) discuss the possibility of 'ghost water' – a non-physical kind that exists side by side with being composed of hydrogen and oxygen atoms and has all the same causal roles as the latter. Even if that is a coherent possibility, it would be the case that "water" refers to both H_2O and ghost water and *not* that water refers to ghost water alone. So even in that possibility it wouldn't be the case that H_2O is merely nomologically connected to water.

The difference is that while in the case of water we do not have any special access to its nature and properties that is not based on physical or functional information,[37] in the case of phenomenality we do. We do seem to have a special insight into the ultimate nature of phenomenal experience; and that nature doesn't seem captured or exhausted by any physical or functional description. As far as we know, that nature might elude any physical understanding. Notice that I stated the problem of the explanatory gap in a way that is independent of whether one subscribes to the semantic thesis discussed in the previous subsection that *all* but phenomenal terms have physical/functional analyses. It is significant that this can be done since it demonstrates that not all of the puzzles of consciousness will go away if we simply deny the semantic framework of the Zombie Argument. However, the constitutional account can explain why the explanatory gap arises, and it does so again in a way that is compatible with physicalism.

The constitutional account explains the gap by appealing to the direct and substantial grasp phenomenal concepts afford of their referent. When I focus on the phenomenal state, I have a "substantive" grasp of its nature. I grasp it in terms of *what it's like* to be in that state. Because this grasp is substantive but at the same time independent of any causal or functional information (unlike in the case of WATER), information about the functioning of the brain simply won't explain *what its like to be in that state.*

Since the issue of the substantive nature of phenomenal concepts is very closely connected to the issue of our acquaintance with phenomenal states, I'll take up the question of substantivity in the next section, together with the question of why the existence of the explanatory gap is not a problem for the physicalist.

7) *Acquaintance. We know our conscious states not by inference but by immediate acquaintance which gives us direct, unmediated, substantial insight into their nature*. If phenomenal concepts are partly constituted by phenomenal states, our knowledge of the presence of these states (in the first person, subjective way of thinking of them) is not

[37] Except for water's appearance properties, for example, that its surface looks shiny in a storm, that it presents itself in a particular way to the touch, etc. But I am not going to press this point here.

mediated by something distinct from these states. Rather the state itself serves as its own mode of presentation. Without getting deeply into philosophical issues involving perception, it is clear that this is quite different from visual (and other sensory) perception of external objects. On one account (with which I agree), when I visually perceive a red apple in front of me I token a phenomenal representation of the apple. The phenomenal representation and the apple are distinct existences and that at least leaves room for the possibility of illusion. When I focus on the phenomenal quality of that visual perception – not on what it represents but on the qualitative character of the visual experience – my representation contains that very experience. Thinking about it and simply having the experience will then share something very substantial, very spectacular: namely the phenomenal character of the experience. And acquaintance, on this account, is the special, intimate epistemic relation we have to our phenomenal experience through the *shared phenomenality* of experience and thought. Shared phenomenality produces the sense that one has a direct insight into the nature of the experience. Hence the unique epistemic standing of acquaintance.

This last observation is connected with the explanatory gap. The core feature of phenomenal states that acquaintance reveals, i.e., their phenomenality does not admit of explanation in terms of physical, functional, or structural features of brain states because of the very way we conceive of this feature, *directly, yet substantially* via acquaintance. Is this a problem for physicalism? You can see why not by focusing on what it means to have direct, unmediated insight into the nature of phenomenality. The important point is that this kind of direct insight (via shared phenomenality of thought and experience) does not reveal anything about the metaphysical nature of phenomenality. It is not the same sense as of "insight into the nature of X" as a scientific analysis of a brain state would provide. The one involves *having* the state, the other, analyzing it into its components. Those are very different activities. But there is a strong tendency to think that an insight into the nature of a phenomenon (e.g., via acquaintance) should lead one a priori to any other insights into the nature of the same phenomenon (e.g., via neuro-scientific analysis), and so that any physical account of consciousness is thereby inadequate. This intuition also forms the basis of the conceivability arguments from Descartes on, and I

believe it stems from a mistaken understanding of what it is to have a phenomenal insight into the nature of consciousness.

8) *There is something it is like to have conscious states.* It should be clear by now that the constitutional account does not explain the phenomenality of brain states – it accepts and *explains* the existence of an explanatory gap between phenomenal and physical descriptions. The strategy is to show that all the epistemic features on our list (1-7) can be accounted for by the special cognitive architecture involved in phenomenal concepts, and this special cognitive architecture is neutral with respect to metaphysical nature of the phenomenal states involved. It is thus open to the physicalist to maintain that types of brain state are identical with types of phenomenal state. Of course there is no explanation of why this brain state type (neurophysiologically or functionally characterized) is identical with a phenomenal state type (phenomenally characterized) – hence the explanatory gap – but there is an explanation in terms of the constitutional account of why there is an explanatory gap even if physicalism is true. From this perspective, the puzzle that the explanatory gap presents is rather a trick the mind plays on itself as a result of the peculiar cognitive architecture involved in first person phenomenal thought. This trick is, like a perceptual illusion, effective even in the face of intellectual conviction to the contrary. It is hard, even for the most devout physicalist, to shake the urge to get more of an explanation. [38] It is not unlike the urge, even after accepting Hume's demonstration that a non-question begging justification of induction is not to be found, to still search for a justification.

Since the constitutional account is neutral about the nature of phenomenal properties, it can be adopted by a non-physicalist.[39] The explanations of most of the features will look much the same with the exception that phenomenal concepts are constituted by non-physical states. However, there will be two explanations of why there is an unclosable explanatory gap. The dualist will say that the gap cannot be closed

[38] Papineau 1993a, 2002, 2007 has an explanation he calls the "anti-pathetic fallacy" which he uses to explain what he calls the "intuition of distinctness", that is, our intuition that physicalism cannot be right.

[39] Chalmers (1996, 2003) suggests a constitutional account of phenomenal concepts.

because phenomenal properties are not physical or functional properties. But this explanation is redundant since, as we have seen, the gap can also be explained merely in terms of direct phenomenal concepts.

4) The quotational account of phenomenal concepts

The constitutional account proposes that a certain kind of concept refers to something that (partly) constitutes it, and refers to it *in virtue of* it being so constituted but no actual account has been proposed of how a concept can be like that. How can constitution determine reference? A dualist can attribute this to a primitive relation of acquaintance which doesn't itself require explanation. This seems to be an account of phenomenal reference by *fiat.* Can the physicalist do any better? Can we naturalize phenomenal self-reference?

The problem of naturalizing mental content is the problem of specifying the non-mental properties that determine the content of a particular concept, for example, specifying *in virtue of what* a particular concept refers to water. If the concept is complex, the question can be partially answered by an account of how the content of a concept with that structure is determined by the contents of its constituents. For simple concepts some other kind of account (or perhaps different accounts for different kinds of concepts) must be found. There have been a number of proposals;[40] all of them, in their present form, have problems.[41] I am not going to try to come up with a "solution", much less a general one. Rather, I will try to make it plausible that, in the particular case of direct phenomenal concepts, reference is determined by constitution. I will do this by

[40] E.g., informational accounts (Dretske 1988), nomological accounts (Fodor 1990), teleological accounts (Millikan 1989 and Papineau 1993b), and conceptual role accounts (Block 1987 and Harman 1987).

[41] The inadequacy of physicalist accounts of content suggests that there may be an *explanatory gap* between the intentional and the physical as well as between the phenomenal and the physical. If there is such a gap then it might be due to the failure of physicalism but it also might be due to the nature of the concepts we employ in attributing content.

showing that phenomenal concepts are analogous to quotation expressions and explaining how certain conceptual roles can make an operation mental quotation.

The question I want to shed light on then is this: why does a phenomenal concept (token) refer to a phenomenal experience that constitutes it, or, in the case of type phenomenal concepts, to the type of experience a token of which is constitutive of it, and most importantly, why does it so refer *in virtue* of this very fact of constitution? After all, this is not the case for most concepts. The concept DOG is not constituted by dogs, and the fact that the concept ATOM *is* constituted by atoms has nothing to do with why it refers to atoms. Information accounts and nomological accounts require an external relation between a concept and its referent unlike constitution, which makes them unsuitable candidates for the explanation of self-reference.[42] It seems plausible that one must look to the *conceptual role* of phenomenal concepts for an explanation of their self-referential nature.

The idea of an item partly constituting a representation that refers to that item is reminiscent of how linguistic quotation works. The referent of "__" is exemplified by whatever fills in the blank. In a quotation expression, a token of the referent is literally a constituent of the expression that refers to a type which it exemplifies and that expression has its reference (at least partly) in virtue of being so constituted. So, for example, ""dog"" refers to the word spelled d-o-g, a token of which is enclosed between the quotation marks. Although in English we normally quote only expressions of English we can also quote foreign language representation and non-linguistic representations. We can even imagine, perhaps just as a joke, placing something which is not a representation, e.g., a cat, between quotes and thus produce a representation that everyone can understand refers to the type cat. My proposal is that there is a concept forming mechanism that operates on an experience and turns it into a phenomenal concept that refers to either the token experience, or to a type of phenomenal experience that the token

[42] Teleosemantics doesn't require external relations between a concept and its referent. Papineau 2002, 2006, who advocates a version of the constitutional account, appeals to teleosemantics to explain the reference of phenomenal concepts. But teleosemantics also holds (Papineau 2006) that the fact that phenomenal concepts are constituted by exemplars of their referent can play no direct role in explaining *why* they so refer. I disagree, as will be evident shortly.

exemplifies. Further – and this is the heart of the proposal – the operation, like linguistic quotation, can be explained in terms of its conceptual roles.

A way to account for the semantics of quotation is to appeal to the disposition of competent language users to accept all instances of the disquotational schema on a priori grounds. So what accounts for the fact that "" is quotation in English is that users of "" who understand the meaning of "refers", etc., are disposed to accept all instances of the following schema on a priori considerations:

L1 "x" refers to x.

L2 ""x"" refers to "x".

where x stands in for any word of English. There is a potentially unlimited number of iterations of the schema at higher and higher levels.[43]

In presenting the *mental* disquotational schema, I can't simply offer a sentence schema in English, like I did with respect to linguistic quotation. To explicitly describe the Mentalese sentence schemas in question I will need to use special notation. In talking about Mentalese sentences, I will refer to concepts (Mentalese words) by CAPITALIZED WORDS as before, I will use '*' to refer to the mental quotation operation, and will use **bold font** to refer to the token experiences themselves that I claim to be part of these Mentalese sentences both inside and outside of the "*" operator. Notice in particular, that any expression in the position of "**experience x**", for example, stands for an *experience*, and not a *concept* of Mentalese.

Here is my account of mental quotation. There is some mental operation (which I refer to as "*") that takes an occurrent experience and forms it into a Mentalese concept

[43] This way of spelling out the concept-constituting conceptual role involves idealization. An ideal reasoner could use and understand any number of iterations of the quotation marks. In practice people won't be able to use or understand triple, quadruple and higher order quotation.

referring to that experience.[44] What accounts for the semantics of this operation – i.e., that the resulting representation does refer to the very experience it is constituted by – is that competent thinkers who also have the concept of reference are disposed to accept all instances of the following schemas on a priori grounds:

M1 ***experience x*** REFERS-TO **experience x**.

M2 ****experience x**** REFERS-TO ***experience x*.**

where "**experience x"** ranges over token experiences.[45]

Mental quotation, on this account, is analogous to linguistic quotation, with one difference. The difference is that, unlike linguistic quotation, what is between the mental quotes (*) at the first level is not a mental word but a mental representation that is not itself a word; it is an experience. I do think experiences represent – but not conceptually so they are not in themselves concepts. This means that the expression on the right hand side of M1 has simply experiences, i.e., non-conceptual representations as instances. Some might object that the resulting thoughts will not be well-formed. However, I think there is a case to be made that plain – unquoted – experiences can be parts of thought under special circumstances.[46]

Let's take a closer look at an instance of M1. Suppose, for example, that you are currently having a visual experience of a patch of red, and that you are seeing it long

[44] It is a superbly interesting – and at the moment wide open question – exactly *what* this operation consists in. All that can be plausibly said on the basis of phenomenological evidence that it involves attention of some sort.

[45] In principle there might be further iterations of this schema at higher levels, like with linguistic quotation; but I doubt that our actual cognitive architecture allows us to form phenomenal concepts of phenomenal concepts anywhere beyond the 3rd or 4th level.

[46] R1 is controversial also on the grounds that it presupposes that all phenomenal experience represents. One could in fact omit R1 and establish the quotational account solely on the basis of R2; however, I believe considering R1 adds to the persuasiveness of the quotational account.

enough to reflect on it. I propose that as you focus your attention on your experience you can form the thought

R1 ***reddish experience *e*** REFERS-TO **reddish experience *e***

where **reddish experience *e*** stands for a particular token reddish experience.

The rough "translation" of this into English is the plausible claim that reddish experience refers to red. R1, however, is not in English; it is in Mentalese. And what my account requires is that all competent thinkers possessing phenomenal concepts and the concept REFERS have a disposition to accept R1, with its particular mode of presentation on a priori grounds. That it expresses a truth is not enough. The sentence ""dog" refers to canines", e.g., is true yet it is not the case that all possible competent speakers have a disposition to accept it on a priori grounds.

My claim is that R1 is just an expression of the *Transparency Thesis,* i.e., that when one turns one's attention to one's own conscious perceptual experience, it becomes evident that it represents the objects and/or features perceived.[47] I take the *Transparency Thesis* to be plausible on a priori grounds, and hence I take instances of M1, like, e.g., R1 to be compelling on a priori grounds.

How about the second level mental disquotational schema, M2? Here the analogy with linguistic quotation is even closer. Both of the quotation expressions that appear in M2 are bona fide concepts. Considering our previous example again, I claim that one can reflect on one's direct phenomenal concept of a current reddish experience and realize that it refers to the very phenomenal character (reddish) that is phenomenally *present* in the concept. I suggest that the result of such reflection is the second order judgment:

[47] I would like to point out that this is not exactly what is typically advocated by representationalists when they appeal to the transparency of experience: R1 requires reflection *both* on phenomenal experience *and* the objects and/or properties it refers to. Representationalists think one can *only* reflect on the representational character of experience, i.e., on the external objects and their properties one's experience represents, (and perhaps one's visual relation to these objects and properties), but *not* on the non-relational phenomenal features of experiences (see Loar 2003). As is clear by now, I disagree with this reading of transparency.

R2 ****reddish experience *e***** REFERS-TO * **reddish experience *e**.**

This can be roughly expressed in English as the obvious thought that the concept of reddish refers to reddish. But, as before, notice that R2 is not in English, it is in Mentalese and the concept CONCEPT doesn't appear in it anywhere. Why believe that all possible competent thinkers - possessing phenomenal concepts and the concept REFERS - accept thoughts like this on a priori grounds *when they are conceived in the special way R2 affords?* I propose that this is simply explained by the awareness that phenomenal states are somehow "present" in our concepts of them, embodied in the infallibility/incorrigibility intuition discussed above."[48]

You might have wondered by now about **reddish experience *e***: how can it appear twice in both R1 and R2? As I said, **reddish experience *e*** stands for a token experience; only token experiences can possibly be part of occurrent thoughts in Mentalese. But, just like the sentence "'"red" refers to red" involves two different tokens of the word "red"; it seems that both R1 and R2 also involves two different tokens of reddish experience. This is especially clear when you consider that mental quotation operates on a token experience and turns it into a token concept; a token experience "taken up" into a token concept cannot be identical with a token experience that is not. In other words, **reddish experience *e*, *reddish experience *e**, and **reddish experience *e***** cannot all involve the same token experience but needs to involve different tokens of the same type of experience. We can incorporate this in our schema in the following way:

$M1^+$ ***experience x_1*** REFERS-TO **experience x_2**.

$M2^+$ ****experience x_1**** REFERS-TO ***experience x_2***.

[48] The sense in which instances of M1 and M2 are acceptable on a priori grounds is similar to sense in which we can know a priori that we have phenomenal experience. It seems that the mere possession of phenomenal concepts is sufficient for knowledge of the existence of phenomenal states – though perhaps not in exactly the same way that possessing the concept BACHELOR is sufficient for knowing that bachelors are unmarried. I am not going to explore these issues further here.

where "**experience x_1**" and "**experience x_2**" ranges over pairs of distinct experiences of the same type.

This seems to pose no problem for the quotational account. Just as in the linguistic case one of the necessary competencies of a speaker is to recognize tokens of the same word *as* tokens of the same word, one of the necessary competencies of a thinker in the mental case – involving phenomenal thought – is to recognize tokens of the same experience *as* tokens of the same experience if presented simultaneously or close to simultaneously. $M1^+$ and $M2^+$ will go through as long as **experience x_1** and **experience x_2** are both tokens of the same type and they are close enough in time – as they intuitively are if they appear in the same thought – for this to be evident for the subject entertaining the thought.

This completes my explanation of what makes the concept forming mechanism that operates on phenomenal experience mental quotation. There is a further issue that I need to say more about. The reference of phenomenal type concepts *includes* the particular experience that constitutes the token of the concept, but will not be *exhausted* by it. A token of ""dog"", for example, includes in its reference the particular word between the quotes, but it might also refer to just all tokens of the word printed in small type, or to all tokens written in any type or font, or to all tokens written or spoken, etc. Similarly, my phenomenal concepts "reddish", or "dark reddish", or "scarletish" can all be constituted by the same particular phenomenal experience; they might all be constituted as the concept ***experience *e*** where **experience *e*** happens to fall under all three concepts. What determines the type a phenomenal concept refers to, if the token experience that constitutes it doesn't, or at least doesn't fully determine it? The quotational account is incomplete if it cannot answer this question.

I propose that the answer again has to do with conceptual roles. For example, what determines the reference of ""dog"" on any particular occasion, depends on the conceptual role of that instance of ""dog"". Both small type and capitalized versions of the word fall under this concept if, for example, were I to be confronted with some examples of the word in small type, and some examples of the word in capitals I would be inclined to judge "same word". The case is similar with phenomenal concepts. A

particular token of the concept ***experience *e**** refers to, e.g., reddish experiences if, were I confronted with any kind of reddish experience, I would judge "same kind of experience". It refers to dark reddish experiences if, were I confronted with dark reddish experiences, I would judge "same kind of experience" but not so when I am confronted with light reddish experiences, etc.

5. Conclusion

The quotational account says that there is a cognitive mechanism that takes a phenomenal experience *e* and forms a phenomenal concept out of it that includes *e* in its reference. This cognitive mechanism is concept forming in this way *in virtue of* the conceptual roles of the resulting entities encoded in the schema M1 and M2. Further, what determines the scope of the concept is a further aspect of its conceptual role having to do with dispositions involving phenomenal similarity judgments. When we understand phenomenal concepts in this way the traditional puzzles of consciousness can be resolved.

It is important to see what this theory is not claiming. My theory is *not* that what it is to be a phenomenal state is to be mentally quoted. My view is the inverse, namely, that to be a phenomenal concept, a concept has to be constituted by a *phenomenal experience*. This means that *constitution matters* for phenomenal concepts. Phenomenal concepts are constituted by an instance of their referent, an experience with a phenomenal character, and we cannot help but be aware of the phenomenal character when we token the concept. This explains the sense that we are *acquainted* with phenomenal experience in a way we are not acquainted with the referent of any other concept. It also explains the sense that these concepts seem to allow us direct insight into the nature of their referent. Yet, nothing I have said in this paper about phenomenal concepts is incompatible with physicalism; with the view that both phenomenal states and phenomenal concepts are realized by physical states. As a matter of fact, the supposition that phenomenal states are non-physical would add nothing to the explanatory power of the theory. I consider this as a major argument for physicalism.

Bibliography:

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

Balog, K. 1999. "Conceivability, Possibility, and the Mind-Body Problem", *The Philosophical Review*, 108: 497-528. Reprinted in Volume XXIII of *The Philosopher's Annual*.

Balog, K. 2006. "Ontological Novelty, Emergence, and the Mind-Body Problem", in *Kreativität*, Meiner Verlag, 26-45.

Balog, K. 2002. "The A Priori Entailment Thesis", an article on Frank Jackson's *From Metaphysics to Ethics, Philosophy and Phenomenological Research*, 62: 645-652.

Balog, K. 2012. "In Defense of the Phenomenal Concept Strategy", *Philosophy and Phenomenological Research 84*(1).

Balog, K (ms). "Illuminati, zombies and metaphysical gridlock."

Bealer, G. 1994. "Mental Properties," Journal of Philosophy 91:185-208.

Block, N. 1987, "Functional Role and Truth Conditions," *Proceedings of the Aristotelian Society* LXI, 1987, 157-181.

Block, N. & Stalnaker, R. 1999. "Conceptual Analysis, dualism, and the explanatory gap", *Philosophical Review* 108:1-46.

Block, N. 2002. "The harder problem of consciousness", *The Journal of Philosophy*, 99: 391-425.

Block, N. 2003. "Mental Paint", in (M. Hahn & B. Ranberg, eds.) *Reflections and Replies: Essays on the Philosophy of Tyler Burge*, MIT Press.

Block, N. 2006. Max Black's objection to mind-body identity. in *Oxford Studies in Metaphysics*, II, edited by Dean Zimmerman, Oxford University Press, 3-78.

Burge, T. 1988. "Individualism and Self-Knowledge." *Journal of Philosophy* 85: 649-663.

Chalmers, D, 1996, *The Conscious Mind*, New York: Oxford University Press.

Chalmers, D., and Jackson, F., 2001, "Conceptual Analysis and Reductive Explanation", *Phil Review* 110:315-61.

Chalmers, D., 2002, “Does Conceivability Entail Possibility?” Published in (T. Gendler & J. Hawthorne, eds) *Conceivability and Possibility*, Oxford University Press, pp.145-200.

Chalmers, D., 2003, The Content and Epistemology of Phenomenal Belief, in Q. Smith and A. Jokic (eds), *Consciousness: New Philosophical Perspectives*, Oxford University Press, 220-273.

Chalmers, D., 2004. “The foundation of two-dimensional semantics”, in *Two-Dimensional Semantics: Foundations and Applications*, Oxford Univesity Press, 55-141.

Chalmers, D., 2007, “Phenomenal Concepts and the Explanatory Gap.” in T Alter and S Walter (eds) *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and physicalism*, Oxford University Press, 167-195.

Chalmers, D., 2009, “The Two-Dimensional Argument Against Materialism”, in B. McLaughlin (ed.) *The Oxford Handbook of the Philosophy of Mind,* 313-339.

Cottingham, J., Stoothoff, R. and Murdoch, D. eds. 1984. *The Philosophical Writings of Descartes*. Cambridge: Cambridge University Press.

Dennett, D. 1991. *Consciousness Explained*, Little, Brown and Company.

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

Evans, G. 1982. *The Varieties of Reference,* Oxford UP.

Fodor, J. 1990. *A Theory of Content and Other Essays*. Cambridge: MIT Press.

Fodor, J. 1998. *Concepts: where cognitive science went wrong,* Oxford; New York: Clarendon Press.

Harman, G. 1987, "(Non-solipsistic) Conceptual Role Semantics". In *New Directions in Semantics*, ed. Lepore, E. London: Academic Press, 55-81.

Harman, G. 1990. "The intrinsic quality of experience", Phil Perspectives 4, Action Theory and Philosophy of Mind, 31-52.

Jackson, F. 1982. “Epiphenomenal Qualia”, *Philosophical Quarterly* 32: 127-36.

Jackson, F. 1993. Armchair Metaphysics. In *Philosophy in Mind*, ed. M. Michael and John O’Leary-Hawthorne. Dordrecht: Kluwer, 23-42.

Jackson, F. 1998. *From Metaphysics to Ethics*. New York: Oxford University Press.

Jackson, F. and Chalmers, D. 2001. “Conceptual Analysis and Reductive Explanation”, *The Philosophical Review,* 110:315-61.

Jackson, F. 2003. “Mind and illusion”. In (A. O'Hear, ed) *Minds and Persons*. Cambridge University Press, 251-271.

Jackson, F. 2004. “Representation and Experience”, in (H. Clapin, P. Slezack & P. Staines eds.) *Representation in Mind: New Approaches to Mental Representation*, Elsevier, 18-39.

James, W. 1950. *The Principles of Psychology*, New York: Dover Publications.

Kirk, R. 2005. *Zombies and Consciousness*, Oxford University Press.

Kripke, S. 1972. *Naming and Necessity*. Cambridge: Harvard University Press.

Levine, J. 2001. *Purple Haze*, Oxford University Press.

Levine, J. 2007. “Phenomenal concepts and the materialist constraint”, in (Torin Alter & Sven Walter, eds.) *Phenomenal Knowledge and Phenomenal Concepts*, Oxford, OUP, 145-167.

Lewis, D. 1966. “An argument for the Identity Theory”, *Journal of Philosophy*, LXIII. 1 (January 6, 1966): 17-25.

Lewis, D. 1983. New Work for a Theory of Universals. *Australasian Journal of Philosophy* 61:343-77.

Loar, B. 1990. "Phenomenal states", *Philosophical Perspectives* 4, Action Theory and Philosophy of Mind, 81-108.

Loar, B. 1997. “Phenomenal states”, in *The Nature of Consciousness*, (Block, Flanagan, Güzeldere, eds.), MIT Press, 597-616 (revised version of Loar 1990).

Loar, B. 2003. “Transparent Experience and the Availability of Qualia”, in (A. Jokic &`Q. Smith eds.) *Consciousness*, Oxford University Press, 77-97.

Loewer, B. 1995. "An Argument for Strong Supervenience" In (E. Savellos & U. Yalcin, eds.) *Supervenience: New Essays*. Cambridge University Press, 218-225.

Loewer, B. 2001. “From Physics to Physicalism”, in (K. Gillett, B. Loewer, eds.) *Physicalism and its Discontents*, Cambridge University Press, 37-57.

Martin, M.G.F, 2002. “The Transparency of Experience”, *Mind and Language*, 17: 376-425.

McDowell, J. 1994. "The Content of Perceptual Experience", *Philosophical Quarterly*, 44: 190-206.

Melnyk, A. 2003. *A Physicalist Manifesto: Thoroughly Modern Materialism*, Cambridge UP.

Millikan, R. 1989. Biosemantics. *Journal of Philosophy* 86: 281-97.

Nagel, T. 1974. What Is It Like to Be a Bat? *Philosophical Review* 83:435-50.

Nagel, T. 1979. *Mortal Questions*, Cambridge UP.

Nida-Rümelin, M. 2007. "Grasping Phenomenal Properties", in T Alter and S Walter (eds) *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*, Oxford University Press, 307-339.

Papineau, D. 1993a. "Physicalism, Consciousness, and the Antipathetic Fallacy", *Australasian Journal of Philosophy*, 71: 169-183.

Papineau, D. 1993b. *Philosophical Naturalism*, Oxford: Blackwell Publishers.

Papineau, D. 1995. "Arguments for Supervenience and Physical Realization." In (E. Savellos & U. Yalcin, eds.) *Supervenience: New Essays*. Cambridge University Press, 226-242.

Papineau, D. 2002. *Thinking about consciousness*, Oxford UP.

Papineau, D. 2007. "Phenomenal and Perceptual Concepts", in T Alter and S Walter (eds) *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*, Oxford University Press, 111-145.

Rey, G. 1988. "A Question about Consciousness", in *Perspectives on Mind*, ed. H. Otto and J. Tueidio. Dordrecht: Kluwer, 5-24.

Robinson, H. 1993. "The Anti-materialist Strategy and the Knowledge Argument", in *Objections to Physicalism,* ed. Howard Robinson. Oxford: Oxford University Press, 159-83.

Stoljar, D. 2001. "Two Conceptions of the Physical", *Philosophy and Phenomenological Research* 62: 253-281

Stoljar, D. 2004. "The Argument from Diaphanousness", in (M. Escurdia, R. Stanton & C. Viger) *Language, Mind and World: Special Issue of the Canadian Journal of Philosophy*, University of Calgary Press, 341-90.

Stoljar, D. 2005. "Physicalism and phenomenal concepts", *Mind and Language* 20: 469-495.

Sturgeon, S. 1994. "The Epistemic View of Subjectivity", *Journal of Philosophy*, 91: 221-236.

Schwitzgebel, E. 2008. "The Unreliability of Naive Introspection", *Philosophical Review,* 117: 245-273.

Tye, M. 2000. *Consciousness, Color and Content,* ch. 3., MIT Press.

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

White, S. 2007. "The Argument for the Semantic Premise". In T Alter and S Walter (eds) *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*, Oxford University Press, 210-249.

Yablo, S. 1993. "Is Conceivability a Guide to Possibility?". *Philosophy and Phenomenological Research* 53 (1):1-42.

Yablo, S. 2002. "Coulda, Woulda, Shoulda". In Tamar S. Gendler & John Hawthorne (eds.), *Conceivability and Possibility*. Oxford University Press.

Soames, S. 2005. *Reference and Description: The Case Against Two-Dimensionalism.* Princeton: Princeton University Press.