

Transcending Zombies

Pete Mandik (petemandik@petemandik.com)

Draft 6.0. Comments welcome.

Abstract

I develop advice to the reductionist about consciousness in the form of a transcendental argument that depends crucially on the sorts of knowledge claims concerning consciousness that, as crucial elements in the anti-reductionists' epistemic-gap arguments, the anti-reductionist will readily concede. The argument that I develop goes as follows.

P1. If I know that I am not a zombie, then phenomenal character is (a certain kind of) conceptualized egocentric content.

P2. I know that I am not a zombie.

P3. Phenomenal character is (a certain kind of) conceptualized egocentric content.

P4. Fixing my physical properties fixes my conceptualized egocentric contents.

C. Fixing my physical properties fixes my phenomenal properties.

§0. Introduction

Any physicalism worthy of the name, that is, any not so thin as to be indistinguishable from dualism (e.g. non-reductive physicalism) or idealism (e.g. neo-panpsychism), faces a choice on the topic of phenomenal consciousness between reductionism (whereby consciousness is physical) or eliminativism (whereby

consciousness is nothing at all).¹ I favor reductive physicalism and this paper constitutes advice for the reductive physicalist on how best to win arguments against anti-reductionists about phenomenal consciousness. The advice will take the length of the paper to develop, but the gist of it will involve strategies for supporting reductionism based on premises concerning certain kinds of knowledge that even the anti-reductionist will be eager to grant. Such pieces of knowledge will include my knowledge that I am not a zombie, that is, my knowledge that I am currently undergoing mental states with phenomenal character.

In this introductory section, I will further set up the relevant issues by reviewing two familiar dialectics. The first concerns the reductionist about consciousness and her opponent, the anti-reductionist. The second concerns the skeptic about the external world and her opponent, the anti-skeptic.

In the familiar dialectic concerning phenomenal consciousness, the first move is the reductionist's move, and it involves advancing a reductive theory of consciousness. There are many such theories, and it is of course only scratching the surface to mention the following: global workspace theories (Baars, 2006), first-order representational theories (Fred Dretske, 1995; Tye, 1995), higher-order representational theories (Carruthers, 2000; Rosenthal, 2005), Dennettian "cerebral celebrity" theories (D.

¹ I hold, but cannot argue for it here, that so-called non-reductive physicalism is an unstable position that must collapse into either full-blown dualism or reductive physicalism. See AUTHOR "Supervenience and Neuroscience" ms.

Dennett, 1998; D. C. Dennett, 1996, 2001), and intermediate-level theories (Jackendoff, 1987; Mandik, 2005; Prinz, 2000).

For present purposes, we may characterize a reductive theory of consciousness in the following manner. Being a theory of *consciousness* it must have outlines discernible from the first person point of view. Being *reductive*, it must have outlines discernible from the third person point of view. Let us say then, the a reductive theory of consciousness entails that what it's like to be my physical doppelganger is just like what it's like to be me.

The second move in the dialectic, the anti-reductionist's move, advances an objection to the reductive theory, or perhaps, reductive theories in general. The anti-reductionist has several arguments to choose from. Prominent examples include the knowledge argument (Jackson, 1982), the explanatory gap argument (Levine, 1983), and the conceivability or zombie argument (Chalmers, 1996). Chalmers (D. J. Chalmers, 2003) illuminatingly describes these three arguments as instances of a general form of what he calls the epistemic gap argument, for all three involve inferring from an epistemic gap (a gap concerning what we can know, conceive, or explain about consciousness on the one hand and physical things and processes on the other) to an ontological gap between consciousness and the physical.

The gap pointed out by the anti-reductionist may be regarded as a question raised for reductive theories of consciousness—the question of what, if anything, attaches the outlines discerned from the first-person point of view to the outlines discerned from the third-person point of view. After the reductive theory has been described the worry arises that the separate portions may be implemented separately. This worry can, of course, be

expressed in terms zombies: creatures that constitute implementations of the aspects of the third person portions of a theory without simultaneously constituting implementations of the first person portions of the theory.²

Now, what should the third move in the dialectic be? What move is best for the reductionist at this point? My advice, which will be developed further, is for the reductionist to attempt to bridge the gap from the first-person to the third-person by discerning *epistemic features* accessible from the first-person point of view that necessitate certain elements accessible from the third-person point of view.

Before developing these points further, I want to turn to the second dialectic, the external world skeptic vs. the anti-skeptic. The first move in the dialectic is the skeptic's move. According to the style of skeptical argument I want to focus on, the skeptic (1) identifies our evidence for the existence of an external world with the *appearance* of the external world and (2) asserts the separability of the *appearance* of the external world

² Of course, another way in which one might contemplate the implementation of the one portion of the reductionist's theory without the other is by imagining ghosts instead of zombies: beings phenomenally similar to us without being physically similar (perhaps by lacking physical properties altogether). It's an interesting question that I will not pursue much further here to ask why the typical dialectic between reductionists and anti-reductionists gets played out in terms of zombies and not ghosts. Perhaps relevant is that dualists agree with physicalists that there is indeed a physical world and thus would not grant to idealists that a so-called ghost world phenomenally indiscernible from our own would be indiscernible *tout court*.

and the *reality* of the external world. In other words, the skeptical line under consideration is that all of our best evidence for the existence of the external world is consistent with the inexistence of the external world: it may very well fail to exist in reality without anything *appearing* different than it already does.

The second move goes to the anti-skeptic and it involves denying the kind of slippage between appearance and reality supposed by the skeptic. The kind of anti-skeptic I'm interested in argues transcendently by assuming that we do know that there's an external world and so there must not be the kind of slippage postulated by the skeptic. Further, this anti-skeptic feels compelled to explain how this lack of slippage comes to be. Now, there are two ways for the anti-skeptic to go here, that is, two ways to fill the appearance-reality gap postulated by the skeptic. The first is a kind of *idealism* whereby so-called external objects are reconstrued as constructs of antecedently understood appearances. The second is a kind of *externalism* whereby appearances are parasitic upon an antecedently existing external reality. On one kind of externalism—epistemological externalism—our justified beliefs about H₂O depend for their being *justified* on certain relations to H₂O. On another kind of externalism—semantic externalism—our justified beliefs about H₂O depend for their being *about H₂O* on certain relations to H₂O.

At this point I would like to return to the first dialectic, for where we are leaving off in the second dialectic gives some indication of the kind of advice I wish to give to the reductionist. The advice is to emulate the kind of move that the anti-skeptic makes. In particular, I think the idealist version is more promising for the reductionist than the

externalist version.³ I further develop my advice to the reductionist in the form of a transcendental argument that depends crucially on the sorts of knowledge claims concerning consciousness that, as crucial elements in the anti-reductionists' epistemic gap arguments, the anti-reductionist will readily concede.

The argument that I will be developing in the remainder of this paper goes as follows.

P1. If I know that I am not a zombie, then phenomenal character is (a certain kind of) conceptualized egocentric content.

P2. I know that I am not a zombie.

P3. Phenomenal character is (a certain kind of) conceptualized egocentric content.

P4. Fixing my physical properties fixes my conceptualized egocentric contents.

C. Fixing my physical properties fixes my phenomenal properties.

A few remarks are needed about this argument and the rest of the paper. P1 is the premise that does the bulk of the heavy lifting and §1 is dedicated to its defense. Also, in §1, I will un-pack the notions of conceptualized and egocentric contents utilized in P1, P3, and P4.

³ Elsewhere (Mandik, 2009a) I develop a case against the externalist version, especially as it is manifest in semantic externalism as applied to consciousness. I will not here rehearse arguments against the externalist version and instead develop the case for the idealist version.

P3 follows straightforwardly from P1 and P2, so I'll say nothing further in its defense.

P4 will be defended and spelled out after conceptualized egocentric contents have been unpacked, thus further discussion of P4 is postponed until §4.

The employment of the parenthetical phrase “a certain kind of” in P1 and P2 deserves comment. I use the phrase to signal my intent to *not* identify phenomenal character with just any old conceptualized egocentric contents. I spell out which conceptualized egocentric contents constitute phenomenal character in §2.

I devote the rest of this section to a discussion and defense of P2. First, I note that there are two potential ways one might contemplate undermining P2 and we can introduce these potential ways via two different ways of applying emphases in the statement of P2. The first way of distributing emphases is “I know that I am *not a zombie*” and the second is “I know that *I* am not a zombie”.

The first way of distributing emphases invites contemplation of potential defeaters that would suggest that whatever things I might know about myself, that I'm not a zombie doesn't count among them. The second way of distributing emphases invites contemplation of potential defeaters that would suggest that regardless of whether I know that someone or other is not a zombie, I cannot know that *I* am not a zombie. These points about the two distinct ways of construing P2 will also be useful in understanding P3.

I intend P2 to be equivalent to a claim of *certainty* regarding my current state, namely that I am certain that I currently have states with phenomenal character. Thus, what is known is known with certainty. Further, what is known with certainty here

concerns *current* states of affairs. Such certainty is unlikely to attach to claims about my past and future states—my memory may be unreliable and there’s little about the future of which I can be certain.

While claims along the lines of P2 have been defended in the recent literature (Horgan & Kriegel, 2007; Lynch, 2006), I will here assume it and point out that it is an assumption that my dualist opponents will readily grant. Further, it is an assumption that they will need to grant. If the claims of first-person knowledge of non-zombie-hood are denied, then this puts dualists at a serious disadvantage concerning the current debate. If they find themselves denying that they know that they have qualitative states, then it is difficult to see how they could have any basis whatsoever to complain that proposed reductions are not, as Chalmers (1996, p. xiii) puts it “taking consciousness seriously”. Chalmers’s accusations that the reductionist is not explaining consciousness but, instead, illicitly redefining it to swap a hard problem for an easy one are (the accusations) predicated on Chalmers *knowing* what consciousness is in virtue of a first-personal acquaintance that is not mediated by grasping any description or definition of consciousness. As is developed at greater length in Mandik and Weisberg (2008), if Chalmers and other dualists are not in a special position to know that they have conscious states, they aren’t in any position to claim that the reductionists are actually changing the topic.

§1. Defending P1

P1 makes a relatively strong claim and it will help in appreciating it to begin by discussing some related weaker claims.

First, consider the relatively non-controversial claim schema, Schema K, that for any knower, s , any individual x , and any property P ,

S knows that $Px \rightarrow (s \text{ believes that } Px \ \& \ Px)$.

It is a relatively widespread view that it is a requirement on knowing that P that one believes it and that the belief be true. Schema K encodes this widespread view as it applies to singular knowledge, singular belief, and singular truth—knowledge, belief, and truth as it pertains to singular items e.g., knowing, believing, and it being true that *Jones jogs*.

Focusing on the example, *Jones jogs*, an instance of the schema would be

Smith knows that *Jones jogs* \rightarrow (Smith believes that *Jones jogs* & *Jones jogs*).

Unpacking this further, we may spell out what Smith's singular belief amounts to by saying that it involves Smith's having a concept of jogging, applying that concept to Jones and thus, in addition to having a general representation of *jogging*, Smith has a singular representation of *Jones*.

It will help to appreciate the thrust of P1 by seeing how far we will fall short of P1 by simply plugging things into Schema K. Consider as one such approximation, K^* :

Smith knows that Smith has qualia → (Smith believes that Smith has qualia & Smith has qualia)

There are three key ways in which K* falls short of capturing the thrust of P1. The first is that it does insufficient justice to the concepts that Smith must bring to bear on any particular occasion of knowing that he has qualia. The second is that K* insufficiently captures the egocentric nature of phenomenal knowledge. The third is that K* leaves out a crucial additional requirement on phenomenal knowledge, namely the identity of the having of the relevant representations and their being true. I unpack these three further immediately below.

In regards to the first point, consider, for example, that the limits of George's conceptual repertoire limit what George can know about rocks. If George lacks the concept of *being igneous*, then he cannot know of any rock that it is igneous. This is not to say that George is incapable of *learning* that some rocks are igneous. This is instead to say that he could not do so without acquiring at that or some prior time the concept of being igneous.

On any particular occasion in which George knows that there is a rock in the room, there will be some set of properties of the rock that he will have concepts for. There may even be occasions in which the property and corresponding concept are simply that of *being a rock*. Things are close but not *exactly* parallel when we turn to George's knowledge of his own occurrent conscious states. While there must be some set of phenomenal properties and a corresponding set of concepts, it simply fails to do justice

to the *phenomenology* of phenomenal knowledge to say that the relevant phenomenal property and corresponding concept are simply that of *being phenomenal*. To sum up, on any particular occasion in which one is certain of currently having phenomenally conscious states, there is some specific set of qualia and some specific set of corresponding concepts brought to bear. The jointly applied concepts accurately represent the qualia present. The certainty in question, then, rules out the qualia fading, going absent, being inverted, or going dancing.

I turn now to the second key way in which K* falls short of capturing P1. In K* it is insufficiently spelled out *how* it is that Smith's belief comes to be a belief about Smith. Not just any representation of Smith by Smith will do. The representation needs to be *egocentric*.

Egocentric representations are distinctive not only in *that* they represent the creature that has them but also in *how* they do so. The most frequently discussed kind of egocentric representations are egocentric *spatial* representations, representations of the spatial location of objects and features in a frame of reference defining locations relative to the representing subject. Egocentric representations do not *only* represent the representing subject but do always represent things in relation to the representing subject.

There are non-spatial examples of egocentric representations as well, including egocentric representations of time (R. Grush, 2009) and temperature (Mandik, 2001). Egocentric representations are oft found at low levels of sensory processing hierarchies and Prinz (2005, pp. 384-385) urges that such representations may be found in audition, touch, taste, olfaction, and interoceptive perception, including the interoceptive perception of bodily states forming the basis for emotional experience.

Several authors, the current one included, suggest that egocentric representations are action-oriented representations (Rick Grush, 2001; R. Grush, 2009; Hurley, 1998; Mandik, 1999; Noe, 2004). In contrast to abstract conceptual representations which are detached from action, the “here and “now” aspects of egocentric representation have immediate connections to the motor abilities of the representing subject. Unlike a conceptual representation of the representing subject, like when Pete Mandik thinks “Pete Mandik’s pants are on fire” my egocentric representation of my pants being on fire is connected to a host of pants-extinguishing action dispositions that do not require for their triggering a mediating step of identification along the lines of the italicized middle step in the inference “Pete Mandik’s pants are on fire. I am Pete Mandik. Therefore, I ought to go jump in a lake” (Kaplan, 1989).

We are now in a position to appreciate what’s being required of phenomenal knowledge in requiring that it have *egocentric* content. Without egocentric content, zombie-related skeptical hypotheses become live options for me. If my experience has any aspect that is not egocentric then even if I have an experience with that aspect and I know that someone or other has an experience with that aspect, then I still wouldn’t know whether *I* am the one that has an experience with that aspect. I may know that someone or other was a non-zombie, but I wouldn’t know whether the non-zombie was me.

Relatedly, I might, while gazing at a red rose and enjoying a red quale, know that *someone or other* is having a red quale, but be in the dark about whether that someone was *me*.

I turn now to the third key way in which K* falls short of P1. When I first introduced this key way, I described it in terms of requiring an identity of the relevant

representations and their being true. Now that I've made the relevant remarks about the conceptual and egocentric requirements on phenomenal knowledge, I am in a position to formulate the third way in which K* falls short of P1, namely that it leaves out the identification of phenomenal character with certain kinds of conceptualized egocentric contents.

If we modified K* to accommodate only the points made so far concerning the first two of the three ways in which K* falls short of P1, we would have only spelled out as a requirement on phenomenal knowledge that there is an *isomorphism* between the relevant representations and what it is that they represent. However, the remarks so far do not suffice to show that phenomenal character must be *identical* to certain conceptualized egocentric contents. Consider an analogous point made about George and the facts he's able to know about rocks. There must be an isomorphism between facts George is able to know about rocks and conceptual contents that George is able to have. But this alone doesn't suffice to establish that rocks are made of concepts or conceptualized contents. Concepts and their contents are mental. Rocks are extra-mental.

I want to argue that, unlike rock facts which are not reducible to any set of conceptual contents, phenomenal character is so reducible. There are four general lines of thought in favor of viewing phenomenal character and rocks as disanalogous with respect to the question of reduction to conceptual content. The first is that we have reason to believe that at least some phenomenal character is constituted by conceptual contents, while we have no such analogous reasons for believing in the conceptual constitution of rocks. The second is that attributing concept-independence to rocks is needed to explain and organize our conceptualizations about rocks in a way that attributing concept-

independence to phenomenal character is not. The third is that the very concept of a rock is a concept of a thing that has a reality that outstrips its appearance, whereas the concept of phenomenal character is an appearance concept. The fourth is that with respect to rocks, we have mere knowledge but not certainty. The certainty that we have with respect to qualia could not be achieved if the representations were distinct from what they represent.

Regarding the first point, there are relatively clear cases in which the acquisition and then subsequent application of a concept to experiences contributes to what it's like to have that experience. One relatively familiar case concerns the way wine may taste quite differently to a novice and to an expert in virtue of the differential between the concepts each can bring to bear on their respective tasting experiences. Such cases lend at least prima facie support to the claim that at least some phenomenal character is conceptually constituted. An analogous case cannot be made for rocks: there's no prima facie reason for believing that rocks are conceptually constituted.

Of course, one might pursue a plausible case that the pet rocks of the short-lived 1970's fad were partially conceptually constituted on the grounds that no rock may be anyone's pet without the rock being conceived of as a pet. So rocks *qua* pets may admit of partial conceptual constitution. But I hope to be granted the scientific-realistic presumption that rocks, *qua* geological kind, are in no way conceptually constituted.

Regarding the second point, the reason we believe that rocks exist independently of our rock judgments is that it helps to *explain* certain patterns in our judgments. The problem of how we know the external world is genuinely external is a huge problem and I certainly do not pretend to have a solution to it. However, I'm aware of no good reason

for being a realist about rocks that doesn't appeal to the explanatory or other theoretical utility of positing them as existing independent of our conceptualizations (see Mandik and Weisberg 2008 esp pp. 225-227). In contrast, there is no explanatory burden that phenomenal character bears that cannot be borne by a reduction of phenomenal character to a certain kind of conceptual content. This sort of point will be developed further in §§5 and 6.

Regarding the third point, consider the following. If phenomenal character outstrips conceptual content, then phenomenal consciousness would be noumenal for the person who has them. But if anything should be phenomenal as opposed to noumenal, it should be phenomenal consciousness. Phenomenal properties are appearance properties. Perhaps, one might object, there are two senses of appearance, an epistemic sense and a phenomenal sense. I address this suggestion later in section 3 wherein I discuss the question in terms of whether the way e.g. colors appear to us in experience outstrips our ability to conceptualize them.

Regarding the fourth point, there is a degree of certainty that attaches to our knowledge of our own phenomenal states that does not attach to our knowledge of rocks. If, however, our representations of phenomenal character were non-identical to what they are representations of, then it would be possible to have the representations without their being *true*. But as long as *that* is a possibility, then I cannot be *certain* that I have accurate phenomenal representations.

§2. Premise 4: My Physical Properties Fix My Conceptualized and Egocentric

Contents

Prima facie, it looks like physical similarity would entail conceptual similarity, that my physical doppelganger is my conceptual doppelganger. He and I have all of the same sorting and discriminatory abilities. We have all the same problem-solving capabilities. We can process all of the same information and do so in virtue of the same information-processing computational/functional architecture. We would perform exactly as well in conversation. On the face of it, such similarities entail conceptual similarity.

While the nature of concepts is far from a settled matter we can sketch some details about them that help to further show that my physical doppelganger would be my conceptual doppelganger.

First, having a concept is having a representation that is general, abstract, or allocentric. Second, concepts are mental dispositions (or the categorical bases thereof). Concepts are abeyant (as opposed to occurrent) representations that allow the possessor to satisfy the re-identifiability criterion, the requirement on possessing concept C that the possessor is able to re-identify objects falling under C as such. Thus, George has the concept of dogs only if George is able to identify dogs as such on multiple occasions.⁴

Third, concepts are endogenously triggerable (Weiskopf, 2007); their deployment is not strictly stimulus-bound or conditioned to current environmental conditions but may be deployed in various “off-line” mental processes such as inference and imagination.

⁴ See, e.g., (Kelly, 2001; Raffman, 1995)

We may summarize these three features by saying that concepts are allocentric, abeyant, endogenously triggerable representations. It seems clear that my physical doppelganger would duplicate all of my allocentric, abeyant, endogenously triggerable representations.

It is worth mentioning that the current argument does not depend on construing my physical doppelganger as being only *intrinsically* similar, we may allow various environmental and historical similarities as well. Thus the argument in this paper is consistent with semantic externalism. This is not to say, however, that I'm a great fan of semantic externalism. The key to note about what counts as my physical doppelganger is that we are careful to not simply assume *phenomenal* similarity.

There should be no objection to the claim that physical properties fix my concepts since neither possessing nor applying a concept by itself entails phenomenal consciousness. Regarding possession, I have been in possession of the concept of cats for a long time, including all day today. However, as I look up and see my cat for the first time today, this is the first time today that my cat concept has had anything to do with my conscious experience. Regarding application, arguably, blind-sight patients who identify objects in their blind field under forced-choice guessing conditions (Weiskrantz, 1996) are applying concepts without thereby having phenomenally conscious experiences of the objects thereby identified.

Some, like Chalmers (2003), may object that fixing my physical properties does not fix *all* of my concepts, since phenomenal character is non-physical and there are some concepts, so-called *direct phenomenal concepts*, which can be possessed only if one has had or is having a state with phenomenal character. My main complaint against this

response is that I don't think there are such concepts as concepts one can only have if one has had or is having a state with phenomenal character. I address this issue at greater elsewhere (Mandik, 2009b). One brief point to make here, though, is that *really* direct phenomenal concepts, concepts had *only while* one is currently having a state with phenomenal character seem not to be concepts at all for their violation of both the re-identifiability criterion and the criterion of endogenous triggering (Prinz 2007 pp.207-208 makes a similar point).

I turn now to consider egocentric contents. Unlike the points made immediately above concerning conceptualized contents, the relevant points about egocentric contents cannot be made in terms of my similarities to any numerically distinct entities. I assume that whatever egocentric contents are, numerically distinct entities capable of having egocentric contents, have contents concerning different entities. Thus no matter how similar to me a being is that is nonetheless numerically distinct from me, it will still be thinking about itself, not me, when it thinks the thoughts it expresses with the first-person pronoun. Thus the relevant question to ask is not whether my physical doppelganger must have the same egocentric contents as me (it can't) but whether fixing *my* physical properties fixes *my* egocentric contents.

A positive response to this question is quite natural. Further, it is not one that should ignite much controversy. Even people who think that phenomenal consciousness is non-physical can grant that fixing my physical properties fixes my egocentric contents since egocentric contents seem not to alone suffice for phenomenal consciousness. Consider, in connection with this, that visual form agnosics can fail to be phenomenally conscious of the shape and orientation of objects, yet still, based on unconscious visual

information, orient their body parts appropriately to, e.g use their hand to insert a card into a slot (Milner & Goodale, 1995). Arguably the unconscious information relied on concerns egocentric spatial contents. Additionally, while the neural activations in LGN mentioned previously have egocentric content, few take seriously the proposal that such sub-cortical neural activations suffice for phenomenal consciousness. Fewer still take seriously the thought that neural activations in LGN would be non-physical.⁵

Now, dualists might object along lines discussed in the previous section that at least some egocentric content is non-physical insofar as direct phenomenal concepts have egocentric contents concerning non-physical qualia. But as this is essentially the same objection as already discussed, it will receive the same treatment as previously mentioned. I'll say nothing further on the matter here.

The question naturally arises of *which* conceptualized egocentric contents fix phenomenal character. Do all of them? If only some of them do, which ones? It is worth noting that nothing particularly precise need be spelled out for the argument to go through. This is because, whatever conceptualized egocentric contents are the ones that constitute phenomenal character, if the remarks from earlier in this section are correct, then the appropriate conceptualized egocentric contents will be fixed by my physical properties because *all* of my conceptualized egocentric contents will be fixed by my physical properties.

It may nonetheless serve to enhance both the intelligibility and the plausibility of the offered argument to provide further detail concerning *which* conceptualized

⁵ For a similar view to the one expressed in this paragraph, see (Neisser, 2006)

egocentric contents fix phenomenal character and under what conditions they do so. To sketch such an account, it will be useful to begin by noting that the knowledge in question involves the occurrent deployment of concepts. It cannot just be, e.g., the kind of knowledge—like my knowledge that dogs are mammals—that I have even when I’m knocked out or distracted and not currently thinking about or perceiving any mammals.

A similar sort of point applies to the egocentric representations involved. Such representations, if they are to help underwrite knowledge claims like *I am currently having conscious experiences*, cannot have the kind of detached status that non-occurrent autobiographical knowledge has. Compare my standing belief that my phone number is 555-5555. It entails that my phone number is currently 555-5555. But the egocentric content of experience cannot be assimilated to such standing or abeyant representations.

According to the Allocentric-Egocentric Interface theory of consciousness (AEI) (Mandik, 2005), conscious states are states of activation in relatively intermediate levels of sensory processing hierarchies. More specifically, conscious states decompose into pairs of states wherein one member of the pair is a state of activation at a relatively low level of the hierarchy, the other member of the pair is a state of activation at a relatively high level of the hierarchy, and both members are reciprocally interacting with each other meaning that the higher-level member of the pair is triggered by the feed-forward flow of information from the lower member of the pair and the lower member’s activation is sustained in part by the feedback or recurrent flow of information from the higher level. This pair of reciprocally interacting states of activation constitute an “allocentric-egocentric” interface insofar as states lower in the hierarchy constitute relatively egocentric representations and representations found at progressively higher levels have

increasingly allocentric contents. The phenomenal character of conscious states is identical to the contents of the hybrid allocentric-egocentric states. Such states have the conceptualized egocentric contents that figure in this paper's central argument.

I will not here review the empirical evidence for AEI, will make do with the brief sketch just supplied. My aim here is to provide enough detail to convey the gist of the kind of *explanatory burden* a theory of this sort can bear.

I turn now to complete the case for construing phenomenal character as conceptual by showing how conceptual contents can bear explanatory burdens oft thought to require construing phenomenal character as non-conceptual. The thought I develop is designed to deal with the claim that there is a kind of *appearance* that cannot be explained in terms of conceptual states.⁶ More specifically, I will be examining the claim that there are certain patterns of success and failure in perceptual discrimination that cannot be explained in terms of conceptual states.

§3. Colors perceived but not remembered: Raffman's rainbow unraveled

There exist color pairs sufficiently similar to be indiscriminable across a memory delay while sufficiently distinct to be discriminable when presented simultaneously (Perez-Carpinell et al., 1998; Raffman, 1995). So, for example, two paint chips presented

⁶ For defenses of such a distinction, see (Chisholm, 1957; Jackson, 1977) and (F. Dretske, 1969) For criticisms, see (Gibbons, 2005)

side by side will be clearly and correctly distinguished as having distinct colors, but if presented one after the other, the viewer will be uncertain whether they have distinct colors. Though, for simplicity, I'll just be focusing here on color, the point generalizes to aspects of vision other than color and also to other sensory modalities besides vision. There are thus a wide variety of stimulus pairs that are discriminable in simultaneous presentations but indiscriminable in serial presentations.

As Raffman (1995) argues, if we make certain natural assumptions concerning the relations of concepts to memory, then the existence of such stimulus pairs puts pressure on the suggestion that conceptual contents exhaust the contents of experience. If the conceptualized is to be equated with the remembered and the recognized, then the existence of such stimulus pairs suggests that experience outstrips our concepts. Whatever constitutes the awareness of the chip that is not sufficiently remembered, that awareness fails to count as the application of a concept since that awareness fails to satisfy the re-identifiability criterion on concept possession.

I want to attack Raffman's argument by calling into question what seems to be one of its key assumptions. The conclusion that conscious experience has non-conceptual content seems to depend on assuming that the colors are present in consciousness in the same way regardless of mode (simultaneous vs serial) of presentation. The assumption seems to be that in every case in which the paint chips are different there must be corresponding elements in consciousness that are different and in every case in which the paint chips are the same there must be corresponding elements in consciousness that are the same.

The assumption works in the context of an argument for nonconceptual contents of consciousness as follows. If I am not able to correctly conceptualize, that is, correctly judge that the second of a pair of serially presented chips is a different color, even though I can distinguish the pair members in simultaneous presentations, then how can this serve as a basis for the conclusion that there is a non-conceptual consciousness of the distinct colors? Such a conclusion would follow if it were further assumed that in spite of the colors of the chips not being available to conceptualization they were available to consciousness. Putting this in terms of qualia, the simultaneously presented and distinguishable chips, chip 1 and chip 2, give rise to corresponding qualia, quale 1 and quale 2. When the chips are presented serially, the subject is unable to correctly judge/conceptualize the difference between the chips, but the chips nonetheless make a corresponding difference in consciousness by triggering, serially this time, quale 1 and quale 2.

(Indeed, in a version of the argument due to Kelly (2001, see especially p. 398, fn. 2), it is experiences, not paint chips (or emulating Kelly's lingo "shades as the subject experiences them" not "shades that the subject experiences") that are distinct and serially presented.)

However, such an assumption is questionable.⁷ We may begin to appreciate what's questionable about it by noting that differences in presentation often result in differences in color perception. Context effects are well known in the literature on color

⁷ Such an assumption looks to be what other authors have called "the sense datum fallacy". See p. 397 of (Evans, 1985) and p. 440 of (Millikan, 1991)

perception.⁸ In normal lighting conditions, one and the same paint chip may seem gray or bright yellow depending on what else is present in the visual field. And these context effects need not involve a difference in what light arrives at the eye from the paint chip in question. Nor are they explained by interactions between retinal cells. The perceptual effects of context depend on higher levels of the visual processing hierarchy than the retina.

We may model an explanation of the failure to serially discriminate simultaneously discriminable chips as due to different perceptions arising from the same chips presented in different contexts. Presenting a chip by itself on one occasion and with another chip on another occasion is to present the chip in two different contexts, contexts that give rise to differences in the perception of the color of one and the same chip.

It is open, then, for the conceptualist to explain the relevant cases as follows. Serially presented paint chips are experienced/conceptualized simply as e.g., *blue* regardless of whether they differ in reality with respect to shade. Simultaneously presented paint chips are experienced/conceptualized as one being e.g., *a darker shade of blue* than the other.⁹ Of course, it is in no conflict with the account I am defending in this paper to posit sub-personal and/or unconscious intermediaries that are non-conceptual. So perhaps it is the case that presenting the same color on different occasions or in multiple locations results in the color being present to the sub-personal or unconscious mind as the same, regardless of whether the color is presented in the simultaneous or the serial

⁸ See (Lotto & Purves, 2002)

⁹ See Rosenthal 2005 p. 189 for a similar account.

context. However, what I am keen to deny is that what makes it into *consciousness* will be the same regardless of simultaneous versus serial context.

Raffman (1995) presents an argument designed to block the sort of move I am here trying to make. She argues that it won't do to say that our experience is only as determinate as we have determinate concepts for (we do have determinate concepts of the unique hues green, blue, red, and yellow), and merely determinable otherwise (we have only determinable concepts for non-unique hues like dark-reddish-orange). Raffman points out that there's no introspectible difference between the ways in which unique and non-unique hues appear with respect to their 'determinateness' despite the radically different ways we have to conceptualize them. (Raffman 1995 pp. 301-302).

Raffman's argument concerning determinateness seems to overlook a powerful resource available to the conceptualist. Raffman overlooks the possibility that the failure of seeming differences with respect to determinateness may simply be due to a failure to apply a *concept of determinateness*. Just as the conceptualist will model differences in apparent *darkness* in terms of the application of a relational concept of one color being darker than another, so may the conceptualist model differences in apparent *determinateness* in terms of the application of a relational concept of one hue or one experience of hue as being more determinate than another. Thus, the failures of appearance with respect to determinateness that Raffman refers to may be regarded by the conceptualist as due to normal perceivers simply failing to apply any such concept of determinateness to their experiences.

One line of resistance to the current argument will come from people who attribute phenomenal consciousness to babies and non-human animals but do not attribute concepts to them. However, such a line overlooks the evidence that lots of non-adults and non-humans have states that satisfy the criteria for conceptualized egocentric contents. I will not review that evidence here. (Some of it will be reviewed in chapter 8). I turn instead to a different response.

My response is that while I grant that babies and non-humans no doubt lack the very concepts I use in my experiences of chairs and coffee mugs, I refuse to grant that babies and animals lack concepts altogether. While lacking concepts of chairs and mugs as such, they likely have, e.g., a concept of an object. And such concepts will suffice for the Allocentric portions of implementations of the Allocentric-Egocentric Interface theory.

My opponents would have a serious objection on their hands if they could *show* that, despite applying different concepts than the concepts that I apply in my experience, babies and non-human animals can have experiences phenomenally identical to my own. Another way they could have a serious objection would be if they could show that babies and non-human animals have absolutely no concepts whatsoever but nonetheless have phenomenally conscious states. However, I really don't see how they could show either of these things. And if my imagined objector is simply *asserting* that such things are possible, then the imagined objection is simply question-begging.

Acknowledgements [[insert from p. 29 of TZ 5.0]]

References

- Baars, B. J. (2006). The Global Workspace Theory of Consciousness. In M. Velmans & S. Schneider (Eds.), *The Blackwell Companion to Consciousness* (pp. 236-246). Oxford: Blackwell.
- Carruthers, P. (2000). *Phenomenal Consciousness: A Naturalistic Theory*. Cambridge University Press.
- Chalmers, D. (1996). *The Conscious Mind: In Search of a Fundamental Theory*. New York: Oxford University Press.
- Chalmers, D. (2003). The content and epistemology of phenomenal belief. In Q. Smith & A. Jokic (Eds.), *Consciousness: New Philosophical Essays* (pp. 220-272). Oxford: Oxford University Press.
- Chalmers, D. J. (2003). Consciousness and its place in nature. In S. Stich & T. Warfield (Eds.), *Blackwell Guide to the Philosophy of Mind*. Blackwell.
- Chisholm, R. (1957). *Perceiving: A Philosophical Study*. Ithaca: Cornell University Press.
- Dennett, D. (1998). The myth of double transduction. In S. Hameroff, A. W. Kaszniak & A. C. Scott (Eds.), *International Consciousness Conference, Toward a Science of Consciousness II, The Second Tucson Discussions and Debates* (pp. 97-107). Cambridge, MA: MIT Press.
- Dennett, D. C. (1996). Consciousness: More like Fame than Television (Bewusstsein hat mehr mit Ruhm als mit Fernsehen zu tun). In C. Maar, E. Pöppel & T. Christaller (Eds.), *Die Technik auf dem Weg zur Seele*. Rowohlt.
- Dennett, D. C. (2001). Are we explaining consciousness yet? *Cognition*, 79, 221-237.
- Dretske, F. (1969). *Seeing and Knowing*. Chicago: The University of Chicago Press.
- Dretske, F. (1995). *Naturalizing the Mind*. Cambridge, MA: MIT Press.
- Evans, G. (1985). Molyneux's Question. In *Collected Papers* (pp. 364-399). Oxford: Clarendon Press.
- Gibbons, J. (2005). Qualia: They're not what they seem. *Philosophical Studies*, 126, 397-428.
- Grush, R. (2001). Self, world and space: on the meaning and mechanisms of egocentric and allocentric spatial representation. *Brain and Mind*, 1(1), 59-92.
- Grush, R. (2009). Space, Time, and Objects. In J. Bickle (Ed.), *The Oxford Handbook of Philosophy and Neuroscience*. Oxford: Oxford University Press.
- Heck Jr, R. G. (2007). Are There Different Kinds of Content? In B. McLaughlin & J. Cohen (Eds.), *Contemporary Debates in Philosophy of Mind* (pp. 117-138). Oxford: Blackwell.

- Horgan, T., & Kriegel, U. (2007). Phenomenal epistemology: What is consciousness that we may know it so well? *Philosophical Issues*, 17(1), 123-144.
- Hurley, S. (1998). *Consciousness in Action*. Cambridge, MA: Harvard University Press.
- Jackendoff, R. (1987). *Consciousness and the Computational Mind*. Cambridge, MA: MIT Press.
- Jackson, F. (1977). *Perception: A Representative Theory*. London: Cambridge University Press.
- Jackson, F. (1982). Epiphenomenal Qualia *Philosophical Quarterly*, 32, 127-136.
- Kaplan, D. (1989). Demonstratives: An Essay on the Semantics, Logic, Metaphysics, and Epistemology of Demonstratives and Other Indexicals. In J. Almog, J. Perry & H. Wettstein (Eds.), *Themes from Kaplan* (pp. 481-563). Oxford: Oxford University Press.
- Kelly, S. D. (2001). Demonstrative Concepts and Experience. *The Philosophical Review*, 110, 397-420.
- Levine, J. (1983). Materialism and Qualia: The Explanatory Gap. *Pacific Philosophical Quarterly*, 64, 354-361.
- Lotto, R., & Purves, D. (2002). The empirical basis of color perception. *Consciousness and Cognition* 11, 609-629.
- Lynch, M. (2006). Zombies and the Case of the Phenomenal Pickpocket. *Synthese*, 149, 37-58.
- Mandik, P. (1999). Qualia, Space, and Control. *Philosophical Psychology*.
- Mandik, P. (2001). Mental Representation and the Subjectivity of Consciousness. *Philosophical Psychology*, 14(2), 179-202.
- Mandik, P. (2005). Phenomenal consciousness and the allocentric-egocentric interface. In R. Buccheri, A. C. Elitzur & M. Saniga (Eds.), *Endophysics, Time, Quantum and the Subjective* (pp. 463-485): World Scientific.
- Mandik, P. (2009a). Beware of the Unicorn: Consciousness as Being Represented and Other Things that Don't Exist. *Journal of Consciousness Studies*, 16(1), 5-36.
- Mandik, P. (2009b). The Neurophilosophy of Subjectivity. In J. Bickle (Ed.), *The Oxford Handbook of Philosophy and Neuroscience*. Oxford: Oxford University Press.
- Mandik, P., & Weisberg, J. (2008). Type-Q Materialism. In C. B. Wrenn (Ed.), *Naturalism, Reference and Ontology: Essays in Honor of Roger F. Gibson*. New York: Peter Lang Publishing Group.
- Millikan, R. G. (1991). Perceptual Content and Fregean Myth. *Mind*, 100(4), 439-459.
- Milner, A., & Goodale, M. (1995). *The Visual Brain in Action*. New York: Oxford University Press.
- Neisser, J. (2006). Unconscious Subjectivity. *Psyche*, 12(3).
- Noe, A. (2004). *Action in Perception*. Cambridge, MA: MIT Press.
- Perez-Carpinell, J., Baldovi, R., de Fez, M. D., & Castro, J. (1998). Color Memory Matching: Time Effect and Other Factors. *Color Research and Applications*, 23(4), 234-247.
- Prinz, J. (2000). A Neurofunctional Theory of Visual Consciousness. *Consciousness and Cognition*, 9(2), 243-259.

- Prinz, J. (2005). A Neurofunctional Theory of Consciousness. In A. Brook & K. Akins (Eds.), *Cognition and the Brain: The Philosophy and Neuroscience Movement* (pp. 381-396). Cambridge: Cambridge University Press.
- Prinz, J. (2007). Mental Pointing: Phenomenal Knowledge Without Concepts. *Journal of Consciousness Studies*, 14(9-10), 184-211.
- Raffman, D. (1995). On the Persistence of Phenomenology. In T. Metzinger (Ed.), *Conscious Experience*. Munich: Imprint Academic Verlag.
- Rosenthal, D. (2005). *Consciousness and Mind* Oxford: Clarendon Press.
- Selvin, S. (1975a). A problem in probability (letter to the editor). *American Statistician*, 29(1), 67.
- Selvin, S. (1975b). On the Monty Hall problem (letter to the editor). *American Statistician*, 29(3), 134.
- Tye, M. (1995). *Ten Problems of Consciousness*. Cambridge, MA: MIT Press.
- Weiskopf, D. (2007). Concept empiricism and the vehicles of thought. *Journal of Consciousness Studies*, 14, 156–183.
- Weiskrantz, L. (1996). Blindsight revisited. *Curr. Opin. Neurobiol.*, 6(2), 215-220.

Postscript on Diachronic Discrimination Failure

Would my objection to the Raffman-style case against conceptualism be defeated by an experimental design that tried to better control for possible context effects of the presentations of the colors? The sort of redesign I here have in mind might go as follows. The stimuli presented in each distinct presentation in the diachronic discrimination case would be one of figures 1 and 2.

**figure 1.****figure 2.**

The task put to the subject is to make a “same as before, yes or no?” judgment about colors appearing on the right side of each display. Synchronic discrimination tasks could use just one of figures 1 and 2 and ask, say of figure 1, if the left and right regions contain the same color.

Such an experimental design is aimed at avoiding the accusation that the colors presented in the synchronic and diachronic contexts are colors presented in different contexts and it thus may not be assumed that there is a color-appearance that is constant across contexts. In this new experiment, the color context of the right-hand color in figure 1 is arguably the same as the color context of the left-hand color in figure 2 since figures 1 and 2 are just spatial rotations of each other.

Does such an experimental design help to defeat the conceptualist? One point in favor of the conceptualist is that in the experiments using figures 1 and 2, there may no longer be a failure of diachronic discrimination. The subject, in being presented with figure 1, is in a position to conceptualize the color on the right as the lighter of the two.

Further, the subject may re-conceptualized the diachronic task as, in seeing fig 2 after fig 1, judging whether the lighter of the two has changed its relative spatial location.