Sensory Representation and Cognitive Architecture: An alternative to phenomenal concepts

Péter Fazekas

School of Philosophy, Psychology and Language Sciences
University of Edinburgh

Zoltán Jakab

Department of Cognitive Science
Budapest University of Technology and Economics

Word count
Abstract: 148
Main text: 11523
Notes: 1943
References: 1756
Total: 15433

Abstract

We present a cognitive-physicalist account of phenomenal consciousness. We argue that phenomenal concepts do not differ from other types of concepts. When explaining the peculiarities of conscious experience, the right place to look at is sensory/ perceptual representations and their interaction with general conceptual structures. We utilize Jerry Fodor's psychosemantic theory to formulate our view. We compare and contrast our view with that of Murat Aydede and Güven Güzeldere, who, using Dretskean psychosemantic theory, arrived at a solution different from ours in some ways. We have suggested that the representational atomism of certain sensory experiences plays a central role in reconstructing the epistemic gap associated with conscious experience, still, atomism is not the whole story. It needs to be supplemented by some additional principles. We also add an account of introspection, and suggest some cognitive features that might distinguish representational atoms with phenomenal character from those without it.

Keywords: phenomenal consciousness, explanatory gap, physicalism, representational atomism, functional unanalyzability, informational psychosemantics, asymmetric causal dependence

1. Introduction

This paper is about phenomenal consciousness and cognitive architecture. In what follows we shall argue that acknowledging the epistemic gap between conscious experience and the corresponding physical processes does not force us to abandon physicalism. To this extent we agree with the so-called *Phenomenal Concept Strategy* (Stoljar, 2005). However, we think that by focusing at phenomenal *concepts*, that approach misidentifies the cognitive underpinnings of the epistemic gap. Even though epistemic and cognitive principles can explain the puzzling phenomena that have traditionally been taken to support dualism, the peculiar characteristics of conscious experience are due to certain cognitive features of the experiences themselves, and not to the concepts (phenomenal or other) built upon them.

1.1 Physicalism and the Phenomenal Concept Strategy

According to physicalism, there is nothing over and above the physical. As it is often formulated, physicalism is the thesis that every fact is *metaphysically* supervenient on the physical facts. In other words, the physical facts determine all facts. Two possible worlds cannot differ in their higher-level (e.g., biological, mental, economic, social) properties without differing in their physical properties.

Anti-physicalist arguments begin by establishing an *epistemic gap* between the physical world and the phenomenal character of conscious experience claiming that phenomenal truths are not necessitated by the complete physical truth. Then follows an inference from the epistemic gap to an ontological gap, namely that phenomenal properties are independent of, and irreducible to, physical-functional properties (Chalmers, 1996, 2003; Levine, 1983; Jackson, 1998).

The Phenomenal Concept Strategy is in fact a collection of several different accounts (e.g. Balog, 2006; Hill & McLaughlin, 1998; McLaughlin, 2003; Loar, 1990; Papineau, 2002, 2007; Perry, 2001; O'Dea, 2002; Tye, 2000, 2003), all of which aim to avoid ontological dualism by invoking *conceptual dualism*. Representatives of the Phenomenal Concept Strategy (PCS for short) are committed to ontological monism, while claiming that there exists a special type of concepts, the so-called *phenomenal concepts*, which differ importantly from other concepts of ours (physical-functional concepts[1]).

PCS claims that the epistemic gap is due to a conceptual gap: a gap between our concepts of physical-functional processes on the one hand, and those of phenomenal consciousness on the other. That is, PCS acknowledges the existence of the epistemic gap but then it stops short of concluding that this epistemic gap indicates an ontological gap between physical-functional processes and phenomenal consciousness. According to PCS, the referents of both phenomenal and physical-functional concepts are physical-functional states upon which we can take two utterly different conceptual perspectives—we can conceive of these states as objects of neurological or abstract (causal, functional, structural) descriptions, and we can also conceive of them in terms of how they feel to us upon undergoing them.

Phenomenal concepts are the vehicles of thought when one thinks of what one's experiences are like or how it feels to have them. Phenomenal concepts refer to the experiences they are concepts of. Their crucial feature is that they are *conceptually irreducible* to physical-functional concepts (Loar, 1997, p597; Tye, 2000, p29): phenomenal concepts neither a priori imply, nor are a priori implied by physical-functional concepts.

1.2 Our objectives

All versions of PCS agree that conceptual irreducibility must be accounted for in terms of some special features of phenomenal concepts. Different versions of PCS point out slightly

[1] Throughout this paper, the term 'physical-functional concepts' stands for what is often called "physical concepts" or "material concepts" (cf. Papineau, 2002, 2007). These concepts are either role concepts, which refer by describing some causal or other role, or directly physical concepts which identify their referents in terms of some intrinsic physical feature (mass, charge, etc.). (Papineau, 2002, p.48.)

different features as responsible for the epistemic gap. In recent years, PCS has received some criticism (Stoljar 2005, Chalmers, 2007; Levine, 2007). Here, however, we shall *not* claim that it is unsuccessful. In fact, we find some of the answers to the objections (Papineau, 2007; Levin, 2008) quite convincing.

Still, this paper is not a defence of PCS, nor is it a new version of it. Instead, we offer an alternative approach to the very same conclusion—namely that cognitive-physicalistic explanations can account for the occurrence of the epistemic gap. We shall shift the focus from phenomenal concepts to the properties they are concepts of: we think it is some features of the experiences (i.e., sensory-perceptual representations) themselves that explain the epistemic gap, and other peculiar features of consciousness.

Those familiar with the current standing of the debate about phenomenal consciousness might wonder what novelty our proposal might offer. Murat Aydede and Güven Güzeldere have recently proposed an information-theoretic approach to phenomenal consciousness (Aydede & Güzeldere, 2005) which achieved something quite similar to what we have just proposed. In fact, at first glance the reader might discover a striking similarity between their claims and ours.

Despite some similarities however, there are important differences between their view and ours. In understanding conceptual irreducibility, they focus on the conceptual faculty, whereas we focus on the level of sensory representations which form the interface between the low-level input modules and the conceptual faculty. In their paper, Aydede and Güzeldere mention two independent factors: first that certain sensory experiences represent their stimuli as primitive; second, that certain sensory *concepts* represent the corresponding stimuli as primitive. Then, relying on Dretske's information-theoretic framework, they reach the conclusion that sensory concepts are concepts of a special kind, and this observation subsequently serves as their chief explanatory principle.

We think that the conceptual features they investigate are not the genuine causes of the special characteristics of phenomenal consciousness but only symptoms which can themselves be explained in terms of the features of sensory representational states. For a long time, we have had the impression that trying to explain the peculiarities of conscious (perceptual) experience by reference to concepts of experience is misguided. In our opinion, it is in the nature of sensory representation where the ultimate explanation of consciousness is to be sought. As it turns out, Fodor's theory of concepts has the expressive power to make this point.

We think that Aydede and Güzeldere have noticed the important role that experiences might play in the explanation, but their Dretskean approach yielded a different focus for them. Now we change the focus back to sensory representation by utilizing a theory of concepts on which sensory or phenomenal concepts are not in any way special. As we see it, conceptual irreducibility does not play any role in accounting for the epistemic gap. In accordance with this, Fodor's framework leaves one no reason to think that concepts of experience are any more irreducible to concepts of neurology than the concept of water is irreducible to that of H_2O. We consider this as an advantage of Fodorian psychosemantics. Accounting for the epistemic gap in terms of the features of sensory representations alone vindicates our claim that in the context of phenomenal consciousness, conceptual features are at best explananda but not the explanans.

2. Psychosemantic Frameworks

Our explanation proceeds in terms of cognitive features of experiences and their relation to other forms of mental representation. In doing so we shall rely on Fodor's view of representation (Fodor, 1983, 1987, 1990, 1994, 1998, 2008). As we have pointed out, Fodorian psychosemantics has some advantages for us over Dretske's one. However, since Aydede &

Güzeldere's account, which will be our chosen framework of compare and contrast, employs Dretske's psychosemantics, we feel it appropriate to introduce Dretske's view briefly.

2.1 Aydede, Güzeldere, and Dreatskean psychosemantics

According to the Dretskean information-theoretic framework, a source *s* by being F (i.e., some entity *s* having an attribute F) generates a certain amount of information which is proportional to the (number of) alternative possibilities eliminated. A signal carrying the information that *s* is F carries as much information about *s* as would be generated by *s*'s being F (Dretske, 1981, p.63).

A crucial distinction within the Dretskean framework is the one between digital and analog information. Dretske says that a signal carries the information that *s* is F in *digital form* if and only if the signal carries no additional information about *s*. That is, it carries no information that is not already nested[2] in *s*'s being F. Contrary to this, if the signal carries additional information about *s* (which is not nested in *s*'s being F) then the signal carries this information in *analog form* (cf. Dretske,1981, p.137). As Dretske puts it: "Every signal carries information in both analog and digital form. The most specific piece of information the signal carries (about *s*) is the only piece of information it carries (about *s*) in digital form. All other information (about *s*) is coded in analog form." (Dretske,1981, p.137.)

Information-theoretic psychosemantics works with the following picture of cognitive architecture. The sensorium is the imput system of cognition. The sensory system consists of transducers and pre-conscious, low-level processors. Transducers convert the physical input (different forms of energy) into signals usable by the rest of the sensory system which processes the information output by transducers. Sensory representations or experiences form the interface between the sensory system and the cognitive system—they are the outputs produced by the sensory system and the inputs consumed by the central cognitive/conceptual sys-

[2] The information that *t* is G is nested in *s*'s being F iff *s*'s being F carries the information that *t* is G.

tem (which in turn controls behavior). Particular sensory representations deliver information about distal layouts (object represented) to the conceptual system.

Suppose one sees a red apple on a kitchen table. One's eye converts electromagnetic waves into neurochemical signals which are further processed by visual input modules. The output of this process is a sensory representation. The information it carries about the red apple in digital form is very detailed and rich (presenting the specific form of the apple, the texture and shades of its surface, etc.). Nested in this very detailed information there is less specific information implied by it (like the shape and predominant color of the apple). The conceptual system virtually always utilizes some analog information nested in the digital information carried by sensory representations. That is, the conceptual system *extracts* information nested in the information carried by sensory representations in digital form.

Aydede and Güzeldere distinguish three classes of concepts: *observational concepts* like APPLE, *perceptual concepts* like ROUND, and *sensory concepts* like RED, based on the *abstraction/ digitalization distance* between sensory representations and concepts based on them. The digitalization distance between observational concepts and their sensory base is the greatest of the three because these concepts deliver information that can be extracted from many sensory representations with widely different digital informational contents. Abstraction distance is minimal for sensory concepts because their digital informational content is also part of the digital informational content of the sensory representations they are acquired from (cf. Aydede and Güzeldere, 2005, pp.206-208).

On this account, the *semantic content* of a concept is identified with the most specific information it carries, i.e. the information it carries in digital form, whereas the *informational content* of a concept is the information carried in digital form plus all the information nested in the digital content (cf. Aydede and Güzeldere, 2005, p.205). Consider our previous example again. When one sees a red apple, one part of one's complex sensory representation, the experience of red (hereafter E-red), carries information about the color of the apple. The complex sensory representation enters one's conceptual system, which extracts information nested in the digital content of this representation. The sensory concept RED is acquired from E-red. Since sensory representations are the sources of information entry to the conceptual system with multiple possible venues (different modalities, different dimensions, etc.), they themselves are information generating sources (Aydede & Güzeldere, 2005, p.223) whose information is picked up by the conceptual system. Thus the tokening of RED carries information about the specific sensory representation E-red. RED, then, carries information about *redness* by carrying information about E-red. The informational content of RED is the ordered pair {*redness*, E-red} (Aydede & Güzeldere, 2005, p.216). One of the fundamental claims by Aydede & Güzeldere is that the very same structure with this dual informational content gets utilized when one attends to one's environment (as a sensory concept), and when one introspects (as a phenomenal concept). They argue that the semantic content of the sensory concept is *rendess*, whereas the semantic content of phenomenal concepts is E-red.

2.2 Fodorian psychosemantics for our background

According to Fodor, the semantic content, or meaning, of concept ***C*** is the property ***C*** is asymmetrically causally dependent on. Suppose that two different types of causes A and B can activate a mental representation ***R***. ***R*** is asymmetrically causally dependent on A, as opposed to B, just in case breaking the A → ***R*** causal link also breaks the B → ***R*** link as well, whereas breaking the B → ***R*** link leaves the A → ***R*** link unaffected (Fodor, 1990). In this case A is the primary cause, whereas B a secondary cause. Later Fodor refers to asymmetric causal dependence as (nomic) locking of concepts to the properties they represent (Fodor, 1998, 2008).

In what follows, a notion of key importance will be *representational atomism*. A representational state **R** is atomic iff its vehicle does not have proper parts that are interpreted as representations by the processor that processes it. In other words, **R** is an unstructured representation. Thus representational atomism is a syntactic property. Representational atoms may have complex contents, for instance, by indicating some complex state of affairs. That an airplane is about to land on an airport may be indicated by a red light on an instrument panel (an unstructured representation) as well as being recorded by a camera and displayed on a screen (a complex representation).[3]

Mentalese consists of concepts and compositional formulas. Propositional knowledge *relates concepts to one another*, e.g., *All bachelors are unmarried*; *Dogs bark*, etc. However, complex concepts must be allowed as well, within this framework (see Fodor, 2008, pp 18, 20, 43-45; 59-60). It is easy to see why this is so. Even though LOT endorses representational atomism (for the vast majority of concepts), due to the compositionality it also endorses, it must also allow for the combination of concepts. If one possesses the concepts HEAVY and BOOK, plus the ability to think, nothing can prevent one from combining the two thus arriving at HEAVY BOOK. Something is a heavy book iff it is heavy and it is a book[4], and it is possible to store these two predicates linked together in one's mental lexicon. Similarly for many other cases of conceptual combination. The moral is, if it is possible to build up propositional representations in LOT using atomic concepts, then complex concepts will be an unavoidable consequence.

Possessing an atomic concept all by itself does not license any inference. BACHELOR all by itself does not afford the subject the inference that bachelors are unmarried. Possessing both BACHELOR and UNMARRIED still does not afford such an inference. One needs to

---

[3] Fodor (1987, 1998, 2008) is not the only author who emphasized the importance of representational atomism. A version of concept atomism has been developed by Margolis (1999). In the context of consciousness, the role of atomic representations (concepts as well as experiences) has recently been explored by Kulvicki (2005) and Aydede & Güzeldere, 2005).

[4] In fact, the semantic analysis here is a little more complex. HEAVY BOOK is reasonably analyzed not as HEAVY & BOOK, but as HEAVY *FOR* A BOOK (just like RED HAIR is better analyzed as RED FOR HAIR, and not as RED & HAIR). Something is a heavy book iff it is a book and it is on the heavy side *relative to books*. Similarly, something is red hair iff it is hair and its color is more similar to prototypical red than most other hair samples.

relate the two concepts by means of propositional knowledge; for instance one needs to learn independently that bachelors are unmarried. This latter item of knowledge is not part of the possession conditions of either BACHELOR or UNMARRIED, according to Fodor. (For the possession conditions of BACHELOR is for it to be nomically locked to bachelors, and the possession conditions for UNMARRIED is for it to be nomically locked to unmarried people. That these locking relations obtain does not confer the subject with the disposition to infer UNMARRIED from BACHELOR.)

Fodor also holds that reference is the only parameter of content (Fodor, 1990a, b;1998; 2008, ch. 3). He is also optimistic about the compositionality of reference (Fodor, 2008, p56). Whether or how reference composes is a difficult issue that we cannot settle in this paper; we simply note that Fodor supposes so, and that the direct reference theory of meaning has a long history (see e.g., Stainton, 1996, 29-61; Neale, 1990; Kripke, 1979, 1980, Larson and Ludlow, 1993; May and Fiengo, 1996, 1998; Putnam, 1975).

In Fodor's theory, concepts are atomic symbols of Mentalese, but they are associated with a file, which is a placeholder with two slots: one for a perceptual prototype (we shall call it the P-slot), and the other for abstract knowledge (let it be the A-slot).[5] However, files are not constitutive of concepts; concepts are the atomic symbols locked to the property they are the concepts of. If one wants to explain what a given concept means, what one can do is report the contents of the associated file. If the A-slot in the concept's file is not empty, this can help explaining the concept's meaning. Thus, the representational content of most concepts is complex, and thanks to the associated file, the complexities of content are often readily accessible to cognitive processing. However, this is not necessarily so on Fodor's theory, since the associanted files may in some cases be next to empty.

[5] See also Margolis, 1999.

2.3 Concepts and experiences

Fodor (2008 Ch 6) agrees that there exist nonconceptual representations, and that they carry Dretskean information about their cause. In his view, it is possible to register the Dretskean information that *s* is F without possessing the concept F. For example, one can undergo an experience of red without tokening RED (op.cit, 186). However, whereas for Dretske it is information extraction or digitalization that gives rise to concepts, Fodor maintains that one needs to possess concepts in the first place to extract information from experience. Recovering information from preconceptual representation presupposes concept possession, he claims (p180).[6]

It is also arguable that sensory experiences, just like concepts, are asymmetrically causally dependent on their stimuli (Tye, 2000, p140; 142 note 25). For example, E-red in some subject's mind can be tokened by TOMATO tokenings (and other secondary causes) only if it can be reliably tokened by red surfaces (the primary cause) in the first place. In cases of rod monochromacy or cortical achromatopsia, TOMATO tokenings can never token E-red, exactly because red surfaces cannot do so either. In the color-sighted, red surfaces can token E-red, and TOMATO tokenings may or may not be able to do so.

Percepts and concepts are two distinct types of representation on the present account. According to Fodor (2008, Ch 6) iconic (i.e., low-level perceptual) representations are such that each and every part of them is also one of their constituents, whereas for discursive representations this is not the case. He later adds (192) that our conscious percepts are not to be thought of as iconic representations. For a conscious visual percept is a product of subpersonal encapsulated (S/E) inference. For example, the operation of perceptual constancy principles that help to identify distal stimulus properties like shape, size, and color, presupposes S/E inference.

[6] The Dretskean notion of information *extraction* is parallelled by information *recovery* in Fodor's usage (2008, p180).

Tye (2008) holds a slightly different view of perceptual representation. He claims (pp12-15) that phenomenally conscious states are such that they *directly, that is, non-inferentially* enable the subject to wonder ”What that is” with respect to some entity. In other words, conscious seeing makes it possible for the subject to form *de re* propositional attitudes about an entity, and do so without using any collateral information (i.e., information not supplied by the perceptual experience in question. The mere possibility of attitude formation is enough; actual wondering is not required. Conscious seeing, according to Tye, need not be an *inferential* process. This is a minor difference between Fodor's and Tye's views, which is interesting because Tye, holds a counterfactualist view of representational content similar to Fodor's one (Tye, 2000, p140, 142 n 25).[7] We need not settle this issue here. All we are going to suppose is that focusing attention, and object recognition/ identification (which is closely related to attentional focus), already presuppose concept deployment. Conscious percepts thus either involve concept tokenings, or predispose the subject to token concepts and to form propositional attitudes. Preconceptual representations in turn supply information that can be recovered by concepts.

Seeing a red triangle can token both RED and TRIANGLE, plus the *de re* judgment that *That triangle is red*. Seeing a particular dog on a given occasion can also token (i) a number of concepts (syntactic atoms): DOG, FURRY, BIG, WHITE, SCARY etc.; (ii) *de re* judgments such as *Sure that scares me*; and (iii) *de dicto* judgments like *Dogs are furry*, and so on. Moreover, the concept tokenings in these examples count as *acts of recognition* (i.e., activation of the concepts by their primary cause, that is, instances of the type of thing that they mean), as opposed to *accidental association* (i.e., activation by other causes that are asymmetrically dependent on the primary cause). When the sight of a red triangle tokens RED, by this tokening the subject recognizes that the object seen is red. Similarly, on seeing a big, white, furry, dog growling at one, the tokening of DOG, FURRY, BIG, WHITE, and

[7] Although Tye (2008) has changed his views about content to some extent, his new theory is still.

SCARY count as recognizing the relevant properties of the object of perception. However, seeing a dog may also token CAT in one's mind, which does not count as recognition, but rather as an accidental association – that is, activation by one of the concept's secondary causes.

We think some sensory representations are representationally atomic, similarly to Fodorian concepts. Somewhat loosely speaking, E-red is an atomic sensory representation (we shall refine this notion in the next section). By contrast, seeing a scene with a dog as in the above example gives rise to a highly complex visual representation. Atomic percepts like E-red can token much less concepts by way of recognition than complex visual representations. A token of E-red can activate RED, SATURATED, and DARK, for instance, all three being attributes that the experience veridically represents its object as having. Even though the surface that is being veridically represented by E-red has a complex surface reflectance property, due to its unstructured nature E-red does not make these complexities in the stimuli available to cognition in any way.

This is the idea of information extraction translated to Fodorian terminology. By this exercise we wish to emphasize both the generality of some of the claims made by both Aydede & Güzeldere and ourselves, and some important differences between the two views. The similarities turn on the fact that representational atomism is a syntactic phenomenon in the first place. Syntactic atoms need not be semantically primitive, still they do not make the complexities in their semantic content accessible to further processing. The differences between the two views are more subtle, however, they are quite important. It is these differences to which we now turn.

2.4 Fodorian psychosemantics: just for fun or is there a point?

According to Aydede and Güzeldere (2005, 212), whereas sensory concepts are necessarily acquired from the experiences sensorially representing the properties they denote, observational concepts are typically, *but not necessarily*, acquired from experiences of their objects. They can be acquired from descriptions as well. The reason for this asymmetry, according to Aydede and Güzeldere, is that the information about stimuli contained in atomic experiences (for example, the information about the property redness contained in E-red) cannot be completely digitalized by the conceptual system. In other words, sensory concepts like RED cannot carry information about the environmental properties they represent without thereby carrying information about the corresponding sensory experience (such that the information about the distal cause is nested in that about the experience). Not so for observational concepts, however: TRUCK, or APPLE can be acquired from a whole host of quite different experiences of trucks and apples respectively, which all carry more specific information about the object than the two generic concepts mentioned (e.g., red apple with tiny spots from one side, on a kitchen table). A lot of this more specific information is ignored when APPLE is acquired, therefore a tokening of APPLE by an apple will not carry information about any particular apple experience (because there is no lawful covariation between APPLE and any particular type of apple experience).

This leads to a problem with Dretskean psychosemantics. According to Dretske, a necessary condition for a piece of information to count as the semantic content of a concept is for it to be *completely digitalized*. According to the definition (Aydede and Güzeldere, 2005, 213; Dretske, 1981, 185), a concept C has as its semantic content the fact that *s* is F (and so the information that *s* is F is completely digitalized by S) just in case (i) C carries the information that *s* is F, and (ii) C carries no other piece of information *t* is G in which the information that *s* is F is nested.[8] But, as we have seen, information about the stimuli of

[8] Applied to this case, nesting means that S carries information about *b* being G because *b* being G itself carries information about *s* being F.

sensory concepts (e.g., colors) cannot be completely digitalized by sensory concepts, because sensory concepts carry information about the corresponding experiences, and only via those experiences do they carry information about their stimuli. As a result, Dretskean psychosemantics cannot supply the right kind of content assignment for sensory concepts, namely information about the corresponding environmental stimulus.

As Aydede and Güzeldere remark (243, n35), Fodor's asymmetric dependency condition also fails to achieve the correct content assignment for sensory concepts. For, they contend, breaking the law between the sensory representation of red and the concept RED will break the law between the property of redness and RED, but not vice versa.

We disagree. On closer scrutiny, Fodor's account does turn out to provide the right semantics for sensory concepts. Here is the argument. RED is supposed to be asymmentrically dependent on redness, but not on E-red. Now, breaking the E-red to RED link clearly need not break the redness-RED link. Consider the case of a rod monochromat (or a subject with cortical achromatopsia). Since, in these kinds of mind, RED may still be nomically locked to redness via spectrophotometer use or expert aid, even though E-red cannot token RED, redness still can. RED in this case will not be a *sensory* concept (as defined by Aydede and Güzeldere), but it will still be a color concept.[9]

However, breaking the redness-RED counterfactual dependence does break the E-red to RED link as well. Suppose a rod monochromat has neither access to color experts (including people blessed with normal color vision), nor to spectrophotometers. Nor does she have abstract knowledge of red as a particular surface color, to support the redness-RED locking. However, a microelectrode is permanently inserted in her V4 such that she can stimulate her own brain, which results in certain interesting experiences, namely tokenings of E-red. She likes those experiences, and wonders what they might be for; however, all by

[9] In this case the file associated with RED contains abstract knowledge (its A-slot is filled properly), but no perceptual prototype (the P-slot is empty).

itself, E-red will never token (or trigger[10]) RED in her mind. If her color vision gets restored, then her E-red tokenings will be integrated in her visual percepts of shape, space and motion, which supplies her a lot of additional information about the function of E-red. *That* (i.e., E-red *plus perceptual integration*; seeing *red surfaces*) will lock RED to redness in her mind. Thus if the redness-RED counterfactual link is not in place, E-red cannot token RED either.

Now consider Mary in her chamber. The redness-RED link *is* in place in her mind (e.g., she could use remote-control spectrophotometers to measure the color of objects outside her chamber). If presented with a red patch on white paper (Nida-Rümelin, 1995), she likely would not be able to tell what color it is. However, on seeing a red tomato in color for the first time in her life, Mary would be able to recognize it by shape, and infer that the color it has is red.[11]

Moral: if the redness-RED link is not in place, then E-red cannot token RED. If, on the other hand, RED is properly locked to redness, then E-red *may or may not* be able to trigger RED, depending on the circumstances. This is exactly what Fodor's theory requires. Locking is supposed to be multiply realizable (Fodor, 2008, Ch 6, 7; 1998, Ch 6).

Ultimate moral: on Fodorian psychosemantics, there is no special class of sensory concepts. RED is no different from TRUCK or APPLE.[12] The consequence is that we better

---

[10] Triggering is the event of locking an innate proto-concept to the property it represents.

[11] The concept of an experience of red (E-RED) works similarly to concepts of color. Fodorian concepts of color experience are nomically locked to color experiences. Mary's concepts of color experience (while in her chamber) satisfy this constraint as well as anyone else's. Her only difference from the color-sighted is that the locking of E-RED to E-red in her case is not mediated by a perceptual prototype. It is mediated by her abstract knowledge and use of brain imaging equipment (see above).

[12] Here is an interesting technical problem. Suppose Mary's concept E-RED is locked to E-red via abstract knowledge and use of equipment: she recognizes occurrences of E-red in other people's brains via brain imaging. Now suppose that, at a later date, she sees a red spot on a sheet of white paper, for the first time in her life. As we have seen, she is unlikely to recognize her experience evoked by the spot as an instance of E-RED. Even though this first occurrence of E-red in her own brain is an instance of the primary cause of E-RED, and E-RED is already locked to E-red in her mind, still this particular instance of the primary cause *cannot* token E-RED in her mind. Does not this mean that sensory concepts like E-RED are special after all? Fortunately it does not. Consider locking via deference. I cannot recognize elm trees or tell gold from brass, but experts can help me doing so. Suppose that, on a few occasions, a goldsmith has sorted jewelry for me in two classes: brass and gold. I witnessed the process, but this did not enable me to subsequently do it alone. That is, my GOLD concept is locked to gold via deference, but not perceptual recognition. Locking entails that gold can, *in certain ways*, token GOLD in my mind; *but it is not a requirement that gold can token GOLD in my mind via any of the relevant psychological mechanisms*. For instance, if GOLD is locked to gold via deference in my mind, then it is not a

attempt a cognitive analysis of sensory experiences themselves, instead of concepts of experiences, in trying to account for the peculiarities of phenomenal consciousness. For instance, when we wish to explain why Mary in her chamber cannot come to know what it is like to see colors, it is not concepts that play the lead. It is sensory experiences that do so.

3. Special Features of Experiences and Cognitive Architecture

3.1 Structured versus unstructured representations

Someone who has visual memories of horses and straight horns with a thread-like pattern on their surface can visually imagine, thereby come to know, how a unicorn looks, based on the description "An unicorn is like a horse with a straight horn on its forehead". Mary in her chamber, however, cannot imagine what seeing red is like, no matter what other descriptions and experiences she might recall from her memory (she cannot recall any memories of color). Seeing a unicorn is a complex experience that carries a lot of information about spatial structure that can also be recovered by concepts. One could generate lengthy and relevant descriptions of what unicorns are like, on the basis of a unicorn picture.

Take another example. The experience of seeing a day-old chick could not be like what it is without certain information about spatial structure being conveyed in the experience. Moreover, this structure can be adequately captured by descriptions. The stimulus is complex (e.g., the chick has two legs, a beak, etc.) and the experience makes this complexity available to cognition. Now compare E-red. Its stimulus is complex too (it is a type of surface reflectance), still E-red does not make this complexity available to cognition in any way.

We say that a perceptual representation possesses constituent structure iff (1) it has discernible aspects or attributes (discernible at least for trained subjects) which (2) can occur

requirement in Fodor's theory that I must be able to recognize gold all by myself. *Mutatis mutandis* for Mary's concept E-RED and her first experience of red.

as experiences on their own as well as in other perceptual experiences.[13] Constituent structure in most cases has the function of mapping some structure of the represented object: the pattern of relations among representational constituents conveys information about a certain external pattern of relations in the environment. So for example it is possible to discern the experience of seeing the leg of a day-old chick as a constituent part of the experience of seeing the day-old chick, and it is also possible to experience an appropriately shaped leg on its own without necessarily experiencing the day-old chick.

Shape experiences, in general, have constituent structure. First, visual experiences of virtually all shapes have discernible parts or aspects that can be undergone on their own, that is, independently of other parts of the same experience. Second, at least a number of current accounts of visual shape representation are based on the notion of constituent structure, that is, on a generative system of representational primitives and rules of combination for them (Marr, 1982; Biederman, 1987, 1990; Wallis and Bülthoff, 1999; see also Palmer, 1999a, Ch. 8).[14]

On the other hand, some perceptual representations are unstructured. These atomic sensory represnetations possess no structure, nor do they map the structure of the represented object. Atomic sensory representations only indicate the presence presence of the represented object — they are 'inner light bulbs' which reliably go on whenever a complex state of affairs obtains in the environment.

Atomic sensory representations have no discernible parts which could occur as standalone experiences. Still, some such experiences have dimensional positions, or *Primitive Sensory Features* (PSFs). In general, PSFs of a perceptual representation are discernible attributes of it which (i) cannot occur as standalone experiences, and (ii) have no discernible

[13] That is, they are *shareable* in Fodor's sense (Fodor, 1987).
[14] On Biederman's account, for instance, visual representations of complex shapes are based on 36 different geons – primitives of shape representation – and their combinations. According to Palmer (op.cit., pp. 406-7) those models that encode shape in terms of hierarchically organized structural descriptions based on some set of shape primitives, have more power than others to explain phenomena of three-dimensional shape perception.

experiential features other than themselves; in other words, they have no representational structure whatsoever.

Color experiences, at least unique hue experiences, are reasonably regarded as representationally atomic.[15] Unique hue experiences have dimensional positions (hue, saturation, and lightness[16]) which are discernible but none of them can occur as stand-alone experiences (for instance, one cannot have an experience of saturation without some hue and lightness). Hue, saturation and lightness are the PSFs of color experiences.[17]

3.2 Encapsulation

So far we have seen that some of our experiences are unstructured; lack of constituent structure characterizes atomic sensory representations, whereas lack of more than one discernible attribute characterizes PSFs. A refinement needs to be added to this picture.

Consider the following objection. Unique hue experiences as they are accessed by higher cognition—that is, attention, recall from episodic memory, making choices based on color perception, verbal report, or silent introspection—are representationally unstructured, and their chromatic contents are primitive sensory features. However, unique hue experiences are unstructured only *for certain cognitive processors*—faculties that we attribute to higher cognition. From a bottom-up perspective these experiences come about as a result of highly complex processing which includes the generation of opponent signals, mechanisms of color constancy, and additional transformations which produce perceived color similarity (Maloney,

---

[15] Although it is possible to argue that the experience of orange is a perceptual mixture of two other color experiences which are plausibly its constituents, namely red and yellow (Boynton, 1997; Tye, 2000, pp. 162-165), this view is not generally accepted (Hardin, 1988, p. 43; Thompson, 2000, p. 171; see also Jakab, 2000, 2006).

[16] An alternative characterization of the dimensions is lightness, redness-greenness, and yellowness-blueness.

[17] In the auditory domain pure sine-wave tones are good candidates for being atomic sensory representations. They have features like loudness, pitch and timbre. None of these are constituents as they cannot occur as standalone experiences. Loudness and pitch are dimensional positions. Timbre seems more complex than a single sensory dimension. It arises from representing the overtones of a fundamental frequency, still all the overtones that contribute to timbre are not discernible one by one. A trained musician may be able to list the keys that sound in a piano chord, but not those frequencies that determine the characteristic timbre of a violoncello as opposed to a piano.

1986, 1999, 2002, 2003; Maloney and Wandell, 1986; Wuerger et al., 1995). These processes involve a number of different stages in the visual system from the retina to area V4 of the visual cortex (DeValois and DeValois, 1997; Wooten and Miller, 1997; Abramov, 1997; Davidoff, 1997; Zeki, 1999, Ch18). Viewed from this perspective, unique hue experiences appear rather structured states of the mind. So what exactly does it mean to be *an unstructured* perceptual representation?

Reply. The levels of visual processing at which singular color experiences are structured are modular and encapsulated, therefore inaccessible to conscious reflection. Mary, for instance, might well know about the firing patterns and information flow at low levels of visual processing. Still her *knowledge of these processes*—propositional representations in higher cognition—cannot interact with *the corresponding processes themselves* (representations in early visual processing) therefore it cannot elicit, or influence, those processes in any way. For such an interaction to take place, a sufficiently direct and detailed information transfer would be required between these two levels of representation. This kind of information transfer certainly does not obtain between color processing and propositional representation. Think of the formation of opponent signals that happens between the retina and the primary visual cortex (Mollon, 2000); mechanisms of color constancy that seem to take place between the primary visual cortex and area V4 (Zeki, 1999, Ch. 18); and the perceived similarity relations of the colors which come about as a result of intricate and encapsulated processing – a transformation of the so-called LMS space (McLeod and Boynton, 1979) into color space (Werner and Wooten, 1979; Hunt, 1982; Hardin, 1988; Wuerger et al., 1995; Matthen, 1999). The reason why encapsulation in these cases is obvious is that it took substantial empirical inquiry to uncover these aspects of color processing. Making these discoveries required scientists to look at the visual system from a third-person perspective. What experiments and abstract theorising did reveal about color vision, armchair reflection alone did not, because color-processing in the brain is inaccessible to conscious reflection. Nor can top-down conceptual influence build up these low-level, intra-modular visual processes of color vision which result in color experiences. There is no "top-down tampering" with the complex events inside an encapsulated low level process no matter how rich the conceptual level gets. Conclusion: representational atomism is *processor-relative*.

4. Our Account at work

Now we are in a position to see how our account unfolds. The next four subsections summarize our proposed solution to the problem of conscious experience. Subsection 4.1 begins with some technical details the importance of which is summarized at the end. The remaining three subsections get straight down to the central problems.

4.1 Atomism and causal role exchange

In general, an atomic mental representations R_A can assume different causal/functional roles in the same cognitive system at different times. Also, different tokens of R_A can assume different causal roles in cognitive systems of the same type.[18] This claim is far from obvious, therefore it will be argued in this section. The conclusion we shall draw from it is that representational atoms cannot be uniquely characterized by their causal/functional role in a given type of system. That is, if two representational atoms A1 and A2 in system S are such that A1 plays causal role R1 whereas A2 plays role R2, then S could be redesigned so that A1 plays R2 and A2 plays R1, while S's input-output relations are unaffected and its inner structure and functioning also remains the same except the causal hookup of A1 and A2 to the rest of the system.

Our argument for this claim is based on discussing two cases of causal role exchange: color experience inversion, and role exchange between conceptual atoms vs. complex

[18] In speaking about causal/ functional role exchange, we mean narrow functional individuation in Harman's sense (Harman, 1988).

formulas of Mentalese.

(1) *Color experience inversion*. Suppose someone wears red-green inverting corneal implants (permanent contact lenses) from birth. To such a subject ripe tomatoes would look the way grass looks to the rest of the trichromat population, and this would also be reflected in their inverted neural responses to stimuli. Moreover, this kind of inversion might be *approximately* behaviorally undetectable (that is, detectable only in the laboratory, but not in everyday life: see Byrne, 2007, esp. note 15; Palmer, 1999b). Here is how color experience inversion should be understood. Color experience inversion has to be systematic, in order to be approximately undetectable: a whole dimension of color space has to be inverted, or re-mapped onto color stimuli. If two singular points (or small regions) of color space exchange their stimuli while leaving the rest of the points intact, that would make this sort of inversion easily detectable. Figure 1 helps to understand this point.

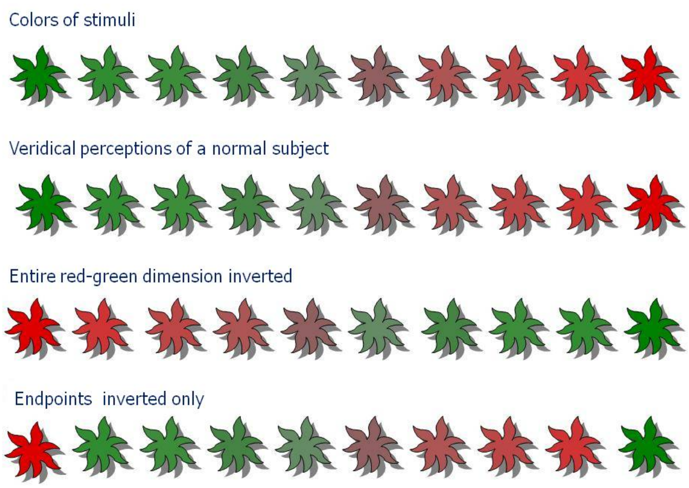

Figure 1. The detectability of two-point inversion. The top row shows the "real colors" of stimuli in a hypothetical color perception experiment. The second row shows the veridical percepts of a subject with normal color vision. The third row shows how a red-green invert would see the stimuli. The fourth row presents the perceptions of a point-invert: only the most saturated reds and greens are inverted; the rest of the R-G dimension is perceived normally. Notice that in a color discrimination experiment the point invert (bottom row) sees a big

color difference between the leftmost stimulus and the second one from left (the same is true of the rightmost one and its neighbor), whereas the color-normal subject and the standard red-green invert subject agree that there is only a little color difference between these stimuli. That is, point inversion is much easier to detect behaviorally than standard (whole-dimension) inversion.

Question: can we interpret standard color experience inverison as a role exchange between *two representationally atomic states*? It seems that the answer is no, since in standard inversion many color experiences (in the example, all reds and greens) change their roles in indicating stimuli, not just two of them. But we can look at this case in a different way. Sensed redness and sensed greenness are two antagonistic channel signals in color processing that can be activated to different degrees and whose different activation levels constitute the chromatic values of particular red and green sensations. (Above we also called sensed redness and greenness primitive sensory features, or PSFs.) For instance, low activity of the redness-signal characterizes unsaturated reds, and blues with a tinge of red; perceptions of saturated red are due to high activity *of the same PSF*. *Mutatis mutandis* for green sensations. Thus in red-green color experience inversion it is the sensed-red-PSF and the sensed-green-PSF that exchange roles: their activity distributions over color stimuli switch places. In addition, their connections with memory and behavioral dispositions must also change to complete the role inversion.[19]

(2) *Role switch between conceptual atoms*. Take two atomic symbols of Mentalese, X and Y. Suppose that, in Paige's mind, X is locked to spoons, and has an associated file containing a

[19] Color experience inversion is possible. First, we might be able to design permanent red-green inverting contact lenses in the future that can be applied to babies who then grow up never taking it off (not that designing such an equipment will ever be a priority of humankind, but that's irrelevant). Second, complete behavioral undetectability of inversion may not be possible due to asymmetries of color space (Hilbert and Kalderon, 2000). This is a debated issue: Palmer (1999b) suggests that behaviorally undetectable red-green spectrum inversion may be possible; Hardin (1998) disagrees. However, even if completely undetectable color experience inversion is not possible, nearly undetectable versions (i.e., undetectable in everyday life, only detectable in the laboratory) do seem possible (Byrne, 2007, esp. note 15).

perceptual prototype of spoons and relevant abstract knowledge. Also in Paige's mind, Y is locked to knives and has an associated file with proper contents. However, in Quillan's mind, X is locked to knives and is associated with the knife-file, whereas Y is locked to spoons and has the spoon file associated with it. Point: if the Fodorian view of concepts (in which atomic concepts are locked to the properties they represent and are associated with relevant knowledge) is right, then the role switch just sketched seems straightforwardly possible.

What complicates this picture a bit is Fodor's claim that concepts are innate. On Fodor's most recent, and admittedly metaphorical formulation of concept nativism (Fodor, 2008, 159), (i) we can learn stereotypes and (ii) we have an innate attractor space populated with point attractors (this is the core metaphor). Innate concepts are these attractors (p160; 162), and learned stereotypes are positioned in this attractor space. If a given stereotype is sufficiently close to a point attractor A_P, then it becomes the stereotype of the concept that A_P is; thereby A_P gets locked to the environmental property of which it is the innate (proto-)concept.

This picture raises some well-known questions like the DOORKNOB/doorknob problem which we now set aside.[20] All that we want to show here is that a role switch between two concepts is conceivable even in this nativist framework. Concepts have two aspects. First, they are points in the attractor space; their identity is conferred by their location. Second, concepts are the syntactic entities with distinctive shape properties that enter cognitive processing.[21] Perhaps these two aspects of concepts constitute a coherent view; so we assume in this paper. But one thing needs to be noted: syntactic shape on the one hand, and location in the attractor landscape on the other need to *correspond* to one another. Location is crucial in locking; syntactic shape is crucial in processing. If a symbol is locked to dogs, it has to be processed accordingly. However, such a correspondence is surely contingent. Suppose that, in little Paige's brain, symbol D gets locked to dogs: it is anchored

[20] The problem is, how is it that a good stereotype of environmental property P will *ipso facto* be close to the attractor (i.e., concept) to which it belongs?
[21] "*Very* roughly: one thinks in file names", says Fodor (2008, p143, note 15; see also Ch 3).

to the dog-point in the innate attractor space as she learns the dog-prototype. Subsequently she acquires the belief that dogs bark, which is coded in her mind by the Mentalese formula ∀x(Dx → BARKx). Quillan, however, has slightly different genetics: in his mind it is symbol C that gets anchored to the dog-point; therefore, when he acquires the belief that dogs bark, ∀x(Cx → BARKx) will be the mentalese formula which represents the corresponding proposition. The role switch generalizes to all other beliefs of theirs, which is just to say the syntactic atoms C and D have swapped causal roles. DOG will be D in Paige's mind and C in Quillan's.[22]

By contrast, complex concepts and mentalese formulas cannot in general switch causal roles. Suppose a complex Mentalese formula FC_1 has the logical form A&B&C whereas FC_2 has the form A&BorC. In this case FC_1 entails entails each and every one of its conceptual constituents, whereas FC_2 does not. This is a structural constraint which prevents them from playing identical computational/inferential roles in different minds of roughly the same type.[23]

Although causal role exchange between representational atoms within the same system seems to be a coherent idea, it does not follow that in adult subjects' minds such an exchange can easily happen. For two atomic concepts to actually exchange roles, all their causal/ infer-

---

[22] Margolis (1999) agrees with Fodor about concept atomism, but he thinks that concepts need not be innate; they can be learned. Such a view makes causal role inversion between different concepts even easier to understand than Fodor's nativism.

[23] What remains true is that two complex but isomorphical Mentalese formulas can switch causal roles if all their atomic conceptual constituents switch causal roles. Two structured concepts or Mentalese formulas CS1 and CS2 are syntactically isomorphic just in case they differ only in their atomic conceptual constituents (names or predicates), but the ordering of their constituents is the same, including identical logical constants in corresponding positions. Identity of logical form is a guide to syntactic isomorphism: isomorphism obtains iff there is a one-to-one mapping between the atomic conceptual constituents of CS1 and CS2 such that replacing the constituents of CS1 according to the mapping scheme yields CS2 (and vice versa). For example, suppose that CS1 is (A&D) → G and CS2 is (B&F)→ G. In this case the mapping is {A ⇔ B; D ⇔ F; G ⇔ G}. Now take two syntactically isomorphic formulas like DOGS BARK and CATS MEOW (say, ∀x[DOG(x)→ BARK(x)] and ∀x[CAT(x)→ MEOW(x)] in Mentalese respectively). Suppose that in Paige's mind DOG is Mentalese symbol D, and CAT is symbol C, BARK comes from locking symbol B to barking things, and MEOW from locking M to meowing things. In Quillan's mind, however, DOG is C locked to dogs, CAT is D, locked to cats, BARK is M locked to barkers, and MEOW is B locked to meowers. That is, the replacement mapping is {D ⇔ C; B ⇔ M}. And the result is: ∀x[D(x)→ B(x)] means *dogs bark* in Paige's mind, and it means *cats meow* in Quillan's mind. At the same time, ∀x[C(x)→ M(x)] means CATS MEOW in Paige's mind, whereas it means DOGS BARK in Quillan's. That is a developmental role switch between the two formulas.

ential connections to other concepts, perception, and behavior need to change one by one. For minds in which a large number of learned connections are already firmly in place this seems to be a fantastic possibility at best. The same is true of color experience inversion in adult minds.

By the possibility of causal role exchange we mean that the role of an atomic representation in a cognitive system could have been filled by some other representational atom. We call this phenomenon *developmental role exchange*, since, as we have seen, the least fantastic versions assume that the causal role exchange happens in the course of ontogenesis. In fact, color experience inversion is not far from practical feasibility (see note 19 above).

Due to the possibility of causal role exchange, representational atoms cannot be uniquely characterized by their causal/functional role. What does this mean? Although the causal role that a representational atom actually fills is characteristic of it within the actual system, still the same role could have been filled by different representational atoms in the same system at different times (or in different systems of the same type at the same time).

Since a representational atom does not fill the role it actually fills necessarily, knowing what role it fills (no matter how detailed the description is) does not help in specifiying what representational atom it is that actually fills the role. This is what we call the functional unanalyzability of representational atoms.

To summarize:

(1) For representational atoms developmental causal role exchange is possible.

(2) Thus representational atoms cannot be uniquely characterized by their causal/functional role in a give type of system.

(3) Sensory experiences that are representationally atomic will be functionally unanalyzable *for this very reason*.

However, with respect to atomism, sensory experiences do not differ from Fodorian concepts. But intuitively there *is* a difference – Fodorian concepts, despite being atomic, do not raise peculiar problems comparable to atomic sensory experiences. Thus representational atomism alone cannot provide a complete explanation of the peculiarities of conscious experience. What else do we need?

Here is a first hint. LOT symbols are theoretical posits. If we accept concept atomism, then we may *infer* from our occurrent thoughts about dogs that DOG, the atomic LOT symbol, must have been tokened in our mind. When DOG is tokened, all that the subject is aware of, and can report, is the contents of the associated file.[24] The tokenings of atomic LOT symbols themselves are never reportable; they are never access conscious in any way. Sensory experiences, on the other hand, are not theoretical posits: their tokenings in cases of attended sensation are the standard examples of access conscious states and reportability. Since atomic LOT symbols, unlike sensory experiences, are never access conscious, subjects never wonder about their peculiar nature. The first step on the road from representational atomism *in general* to representational atoms *with phenomenal character* is access consciousness.[25]

4.2 Words, concepts and percepts: what Mary did not know and why

We are now in a position to see why Mary cannot come to know what it is like to see the colors based on her abstract knowledge alone – that is, why she cannot token color experiences in her mind before release. Let us look at Figure 2.

[24] Plus current perceptual states if DOG is tokened via the senses.
[25] Another difference is that sensory experiences *per se* do not have associated files. When they occur, they might be introspected, which involves concept deployment (see below). The deployed concepts of experiences in turn might have associated files with knowledge about experience.

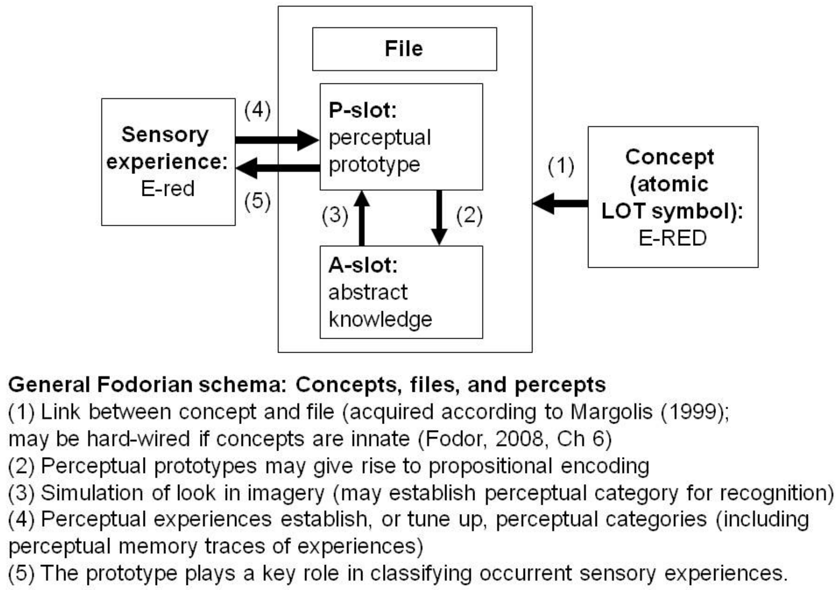

Figure 2. The relation between concepts and experiences in Fodor's view of cognitive architecture. Note that the locking of concepts to their objects is not denoted by any of the arrows in this figure (the same is true of the next two figures). What are denoted are some cognitive processes that may participate in locking.

On the Fodorian scheme, concepts are associated with files. Their P-slots contain perceptual prototypes which are typically formed on the basis of perception, although in certain cases they may also be formed from abstract knowledge. Abstract knowledge (represented by Mentalese) can in some cases guide imagery resulting in perceptual prototypes. Such prototypes may subsequently be deployed in recognition. How does this happen?

Combining concepts with prototypes may be accompanied by combining their perceptual prototypes. For instance, when PURPLE SAINT-BERNARD is entertained, it is typically accompanied by visual memories of some shade of purple, plus that of some big dog (depending on the subject's prior experience with Saint-Bernards and colors). In addition, the two stereotypes are typically combined such that, in the recalled dog prototype, the representation of brown and white colors typical of Saint-Bernards is replaced by the episodic memory trace of some shade of purple. Roughly speaking, the visual image we typically construct on entertaining PURPLE SAINT-BERNARD is one that could arise from remembering actual perceptions of Saint-Bernards painted purple. That is, propositional representations can, and often do, trigger visual images or impressions. A word about Pylyshyn's views of imagery will help to understand how this might happen. as Pylyshyn has repeatedly argued (Pylyshyn, 2002, 2003, 2006), mental images are in many respects unlike the visual representations we undergo when actually seeing, and imagination is simulation based on tacit knowledge, instead of cognition "drawing" pictures and "reperceiving" them. According to Fodor and Pylyshyn, we never think in such internally generated pictures. Thought and reasoning is realized by propositional representations, still such representations are often accompanied by knowledge of how things look. Knowledge of the looks of things can be supplied by perceptual prototypes. Based on this knowledge, when we imagine something, we simulate its look. We do so by deploying concepts, which in turn guide the process of this simulation. Imagery works for simple sensory experiences as well, as long as they have at least dimensional structure. For instance, supposing that I have seen reds, blues and purples, but no magentas so far, the description that "magenta is slightly bluish red, less bluish than purple" can be very helpful for me to imagine, and subsequently recognize, magentas. We shall call this case Humean simulation (from Hume's example of the missing shade of blue).

Prototypes may also enrich abstract knowedge when we describe earlier perceptions (i.e., encode them in propositional format).[26] Now look at Figure 3, which is a representation of Mary's case.

[26] We have a sketch of how propositional representations may be generated from perceptual ones, but it did not fit in the present paper. Here it is enough to mention that such a transformation is possible in one way or other. When I see a house, I can count its windows, and memorize their number. The resulting semantic memory trace can be quite helpful later, when I need to recall the number of the windows. Similarly, I can name the color of the house, describe the shape of the roof (perhaps by using geometrical concepts), and so on. Semantic encoding of episodic memories is a common process in human cognition.

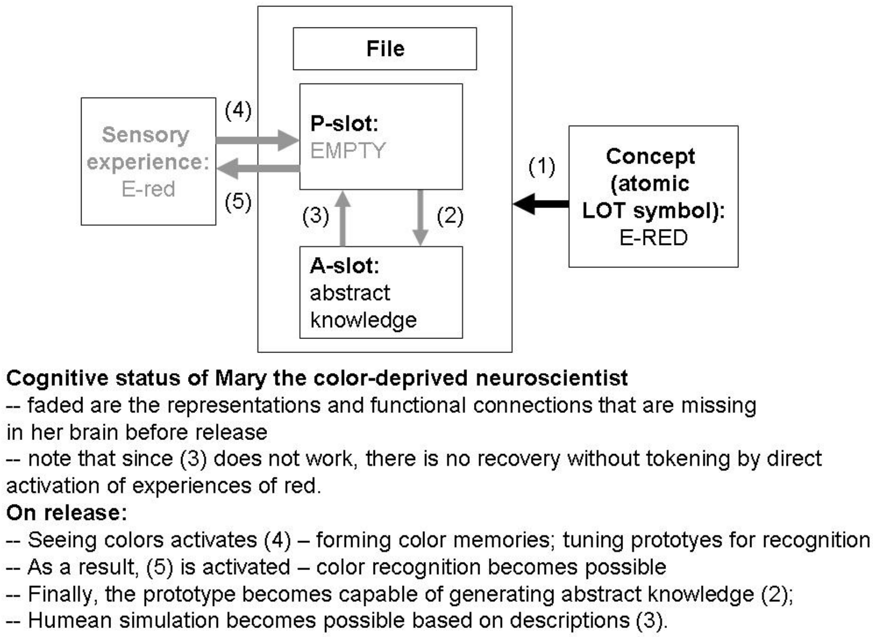

Figure 3. Mary's color concepts and her recovery on release from her chamber.

While in her chamber, Mary's concepts of color and color experience are such that their P-slots are empty. Since color experiences are representationally atomic, abstract knowledge about them (accompanied by perceptual prototypes of other objects and properties) cannot simulate their prototype. Humean simulation cannot work because in Mary the chromatic dimensions of color space (which are PSFs) have never been activated, therefore she has no memory traces of them.[27] On release, however, color perception builds up her color prototypes, which in turn restore the related cognitive capacities. Figure 4 shows how we might acquire complex perceptual prototypes from abstract knowledge. Here the order of acquisition of the missing processes and representations is different from that in Mary's case: it begins with the simulation of complex prototypes from abstract knowledge.

[27] Humean simulation is interpolation within a sensory dimension with which the subject is already familiar. Mary, however, is not at all familiar with the chromatic dimensions of color space from the first-person perspective.

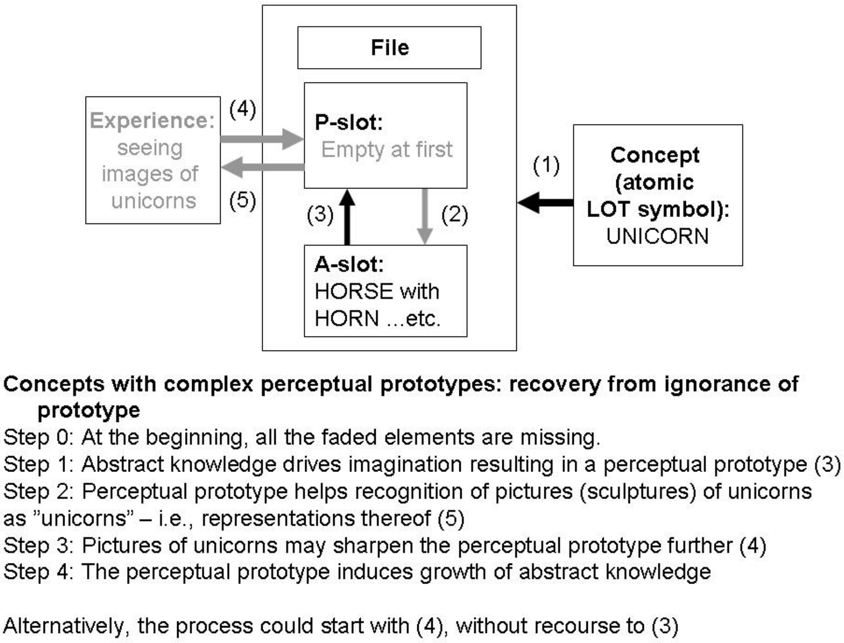

Figure 4. Acquisition of complex prototypes from abstract knowledge. Note that the P-slots of HORSE and HORN (in the A-slot of UNICORN) – must be filled up in order for abstract knowledge to generate the unicorn prototype (see above in this section).

4.3 Knowing experience: the real epistemic barrier

We think the relation between concepts of experience on the one hand, and concepts of neuroscience on the other is by no means more peculiar than that between, say, WATER and H_2O (Fazekas, 2009; Marras, 2005, Tye, 2008, Ch 3). The crucial epistemic barrier that leads to the explanatory gap lies somewhere else.

Each and every concept has an associated file with a P-slot and an A-slot, and for any concept both these slots can be loaded with relevant information.[28] Thus any two concepts can be related to one another using information in their A-slots. As we have argued, the strict limitation is that, within the file of certain concepts, namely those of atomic experiences, the prototype cannot be generated from the abstract knowledge – even though these concepts have a highly relevant prototype that can be acquired by other means. This limitation explains

[28] The only exception is that for some abstract concepts like TRUTH or ELECTRON, no particularly relevant perceptual prototype can be found.

why Mary cannot come to know what it is like to see the colors while in her chamber.

A note about the relation between imagery and knowledge is in order. We think imagination does in some cases *confer one knowledge*, somewhat similarly to the way inference does (cf. Jackson, 1997, 567). If Quillan believes the premises of an argument, but he has not yet thought them through, therefore he does not believe (has not even entertained) the conclusion, then he does not know the conclusion. Subsequently he assesses the premises, reaches the conclusion, thereby he acquires knowledge of the conclusion. Knowledge acquisition in this case is a matter of transforming mental representations. If Paige knows what it is like to feel cold and warm, see the red of live coal and the interior of a steam engine's driving cab, feel the chilly wind on her face, and has powers of imagination, then she can imagine what it is like to drive a steam engine on a stormy day. In this case, exercise of imagination gives her knowledge of what it is like. In Roger Shepard's mental rotation experiments subjects needed to imagine certain objects being rotated in space, and based on this exercise, they were able to choose from a number of pictures the one that depicted an object of the same shape as the sample. In this case again, it is the transformation of mental representations that gives one knowledge of the required sort. However, Mary is not in a position to carry out a cognitive exercise that could give her knowledge of what it is like to see colors – nor does she know it in any other, more direct ways, based on her knowledge. We are not claiming that it is *logically impossible* for her to imagine what it is like to see colors; but it is nomically impossible. Given the way human brains all operate, there is no mental skill, or cognitive process that could yield this sort of knowledge for her. Looking at colors would; stimulating her own V4 via a microelectrode might; and if encapsulation were not true, then perhaps some act of imagination could do so as well.

The other epistemic limitation whose cognitive underpinnings we presented above is that no meaningful analysis of atomic experiences can be given in terms of structure or functional role. The notion of causal role exchange was cited in support of functional unanalyzability, and the representational atomism of sensory experiences explained structural unanalyzability. This supplies an understanding of why the explanatory gap arises, why any reasonable form of physicalism has to be a posteriori, and also why zombies are conceivable. Here is the story.

Suppose that Paige, a color-sighted neuroscientist possesses E-RED, whose P-slot is filled with the perceptual prototype of red, and its A-slot contains contextual information pertaining to what objects and circumstances in the environment activate experiences of this kind. She then takes cerebroscope measurements of some subjects' V4 responses and discovers a salient response pattern that she calls nrp425, and so forms a concept of it (NRP425). She suspects that nrp425 is a sensory experience that, on occurring in someone's V4, has some phenomenal character. Thus the A-slot of Paige's NRP425 is filled with relevant neurological information, but its P-slot is empty – she has no idea of the phenomenal character of occurrent nrp425s. Then she decides to run a correlation study: subjects go in her cerebroscope, and their V4 responses are recorded while they see different shapes and colors. She realizes that subjects' respond with nrp425 to stimuli they classify as looking predominantly red. Paige thus concludes that NRP425 and E-RED are coreferring concepts. That is, nrp425, when occurring in a neurotypical subject's V4, is the experience of red. At this point, a two-way information transfer takes place between the files of NRP425 and E-RED, which consists of three moves: (1) the contextual information in E-RED's A-slot is pasted in NRP425's A-slot, (2) the neurological information in NRP425's A-slot is pasted in E-RED's A-slot, finally, (3) the perceptual prototype in E-RED's P-slot is pasted in NRP425's P-slot. This means, in effect, that the two files get united. Hereafter NRP425 and E-RED will be attached to the same united file (still they remain different concepts, because their atomic LOT symbols remain distinct).

There are a few things to notice here. First, before uniting the files, there is no *a priori* (deducibility, rational derivability, etc.) link between the A-slots of NRP425 and E-RED. This is no surprise for us, since we think that there is no a priori link between the A-slots of WATER and H_2O either (see above). Second, NRP425's P-slot cannot be filled up based on the contents of its A-slot alone (no matter what abstract information we load in it). This is so for three reasons. First, encapsulation prevents the simulation of nrp425 from its low-level constituent processes (see 3.2). Second, NRP425 is atomic from the first-person perspective, therefore no other experiences are of any help in simulating it in imagination (3.1 and 3.2). Third, any atomic sensory experience that the subject *can* entertain is functionally unanalyzable (due to the possibility of causal role exchange), therefore none of these experiences is a better candidate for identification with nrp425 (viewed from the third-person, scientific perspective) than any other.[29] To summarize, no matter how much information is loaded in NRP425's A-slot, its P-slot remains empty (unless united with the file of E-RED in a correlation study). This is the epistemic gap in our view, and also the reason why physicalism must remain *a posteriori*.[30]

4.4 Introspection and concepts of conscious experience

In our view, introspection consists in applying concepts of experiences to our experiences. In this respect we follow Aydede and Güzeldere, and others (Tye, 2000, 51-54; 2008, 189-192; Dretske, 1995, Ch 2). Tye (2008, 189; Ch3) argues that there is nothing special about the concepts that we apply to our conscious experience. As should already be obvious, we are in agreement with this claim as well, and it too follows from Fodorian psychosemantics.

On the other hand, we do not follow phenomenal externalists like Tye in thinking that in the case of introspective visual awareness our attention necessarily goes outside, onto

[29] That is, the functional description that might be found in NRP425's A-slot is no help in such identifications. This is because no unique functional characterization can be found in the A-slot of concepts of experiences. Thus concepts of atomic experiences offer no hint as to what neural state, characterized in functional terms, they might be identical with.
[30] The conceivability of zombies has the same origin: for concepts of experiences, P-slot content (the prototype) cannot be generated from the corresponding A-slot content. Thus we cannot see why *feeling like this*, and also *feeling in any way at all*, should follow from *being like that, neurologically*.

external stimuli (Tye, 2000, p. 51; see also Dretske, 1995, chap. 2; Harman, 1997). Consider the case of having the impression of swirling colors when one wakes up at night, or daydreaming while driving a car. It sounds strange to claim that in these cases attention is directed outside when there is no actual stimulus that one is attending to. The daydreaming driver is prone to making driving errors exactly because he is *not* attending to his *actual* environment.[31] What he attends to is the non-actual (past or merely possible) scenario that he is entertaining in imagery, and he is readily aware of the non-real, fictitious nature of what he is attending to. The object of his attention is properly characterized in terms of the intentional content of his mental states, accompanied by reflective awareness of non-reality. That is, his attention is directed inside (i.e., ”on what's going on in his mind”), *in the sense that it is not directed at the actual environment*. Daydreaming is a type of pretence, which is a voluntary act on the part of the subject.

The case of swirling colors is even more illuminating. Contary to daydreaming, experiencing swirling colors at night is not a willful act; it probably arises from some spontaneous activity in the visual system. It is not even considered by the subject undergoing it as a *scenario* – it is considered simply as experiences of color. If it is reflected upon, then what is thought about it is that it is an experience of some colors. It is readily accompanied by a reflective awareness of non-reality, just like daydreaming. The intentional content of swirling colors is very much impoverished compared to that of a scenario entertained in daydreaming. Still, when one attends to the swirling colors, what attention is directed at is intentional objects that are known to be (i.e., at least tacitly recognized as) non-real. This is how one can make sense of attention directed inward, in a broadly intentionalist framework.[32]

[31] The idea that in being aware of the phenomenal character of our sensations attention goes outside is motivated by representational externalism about phenomenal character (Dretske, 1995; Tye, 1995, 2000, 2008; Byrne and Hilbert, 1997, 2003; Harman, 1997).

[32] Although this story seems to us compatible with phenomenal externalism, we are narrow intentionalists about phenomenal character. In our view what is is like to see the colors (*mutatis mutandis* for other types of experience) supervenes on local states of the brain; still phenononal characters are properties that external objects are represented as having. For more about such views see McLaughlin, 2003; Jakab 2003, 2005, 2006.

Next, consider visual search, where attention is surely directed at the outside environment, even when there is nothing out there corresponding to what one is looking for (like when one is looking for one's keys in the wrong place). Visual search is on as long as the relevant stimulus is not noticed. Seeing the intentional object calls off the search, whereas failure to locate it establishes reflective awareness that is is not in the searched location (in some cases also that it does not exist). Thus in visual search attention is directed at the actual environment, because it is assumed that the intentional object in question exists (and is located nearby).

Daydreaming and the swirling colors case are accompanied by a reflective awareness of non-reality. When we *hallucinate* swirling colors, and act upon our impression as if it indicated something real, it is precisely the reflective awareness of non-reality that is missing. It is somewhat difficult to decide whether attention is directed at the actual environment in hallucination. When we run away from a hallucinated bear, our action is perfectly real and fits the environment in many respects (e.g., we avoid obstacles while running). In sum, in hallucination there is no reflective awareness of non-reality, plus attention is, *at least partially*, directed at the actual environment. (In addition, there is a gross error regarding what is present in that environment.) The last case to quickly consider is that of ordinary perception, where attention unquestionably goes outside, and there is no reflextive awareness of unreality of the intentional object. Summary in total: when undergoing an experience, reflective awareness of non-reality of its intentional object powerfully facilitates inward attention, in the sense specified above.

Reflective awareness of non-reality comes pretty early, namely with children's ability to pretend. According to Alan Leslie (1987; 2002) this capacity is in by two years of age. Joseph Perner, who holds a characteristically different view of mentalization development, suggests that pretence is in by age 4 (Perner, 1994). Attention goes inside as a result of recognizing that experience – the way the world seems to be – is distinct from how it actually is.[33] It is not implausible to hold that before age three, children do not conceptualize much of perceptual experiences. *It is the resulting new attentional focus that will result in suitable mentalese symbols locking on to perceptual experiences*, we propose.

As we have argued, phenomenal concepts are not special in any way; they are ordinary Fodorian concepts that are locked to sensory experiences. What follows from this is that RED and E-RED are two (syntactically) distinct concepts, not one and the same concept with ambiguous semantics, as in Aydede and Güzeldere's theory. In sum, the optimal condition for acquiring concepts of experiences is (i) attention going inside, (ii) while one has perceptual experience with impoverished intentional content. The latter condition also implies that the perceptual representation underlying the experience is not particularly complex (perhaps it is just a representationally atomic perceptual state). In such conditions, proto-concepts of experiences come to be locked to their objects.

It follows from the above that there is a difference between perceptual and introspective awareness. Perceptual awareness is direct or primary; it is awareness of external stimuli. For instance, we are perceptually aware of object colors by means of their being reliably indicated to us by color experience. By contrast, introspective awareness is secondary: it arises from "extra" processing of our occurrent perceptual states; processing that is not part of perceptual awareness. Such processes include (i) attention turning inward due to reflective awareness of non-reality; (ii) the application of concepts of experience, and the formation of relevant belief (e.g., "I am currently undergoing such and such a perceptual experience"), and sometimes (iii) focusing attention (processing efforts) on the recall of perceptual memories, or the maintaining of traces of perception formed a moment ago, as opposed to gathering new information through the senses.

---

[33] Superseding childhood realism is a related milestone in cognitive development. According to Piaget, children come to have an appreciation of mental life after age 6 (see Beilin and Bufall, 1992, p152). See also Flavell, 1999.

4.5 What makes representational atoms phenomenal

We have made the point above that only representational atoms that can become widely accessible in cognition can be phenomenal. The availability of the outputs of sensory systems to the conceptual system, or poisedness, is a key criterion for phenomenality. In this we follow the footsteps of others (Evans, 1982, 158; Tye, 2000, 62; Davies, 1997, 311; Aydede and Güzeldere, 2005). Both factors are important: (i) availability to conceptual processing and that (ii) what's available are the outputs of the sensory systems. Beliefs (propositional representations) do not come with characteristic phenomenology, although when tokened, they are readily available to conceptual processing. What is responsible for this difference? Here we have a little conjecture to add.

First, consider color space. The dimensions along which color experiences are ordered (hue, saturation, and lightness; alternatively, red-green level, blue-yellow level, and lightness) are also the features that these experiences represent their objects as having. Thus these dimensions are *aspects of the perceptual content* of color experiences. In addition, perceived hue, saturation and lightness stand in a very close relationship with *the characteristics of the representational vehicle* of color experiences. This is because the dimensions of perceptual color space arise from neural computations on retinal receptor signals, and so the dimensional values of particular color experiences in color space are physically coded in neural firings in the color vision system. That is, the neural system that underlies the dimensionality of color space (that is, the red-green, blue-yellow, and lightness channels) determines (i) the neuronal state types that are the color experiences (ii) the phenomenal character of those experiences. Therefore, color experiences that are similar in terms of syntacic shape (neuronal type) will be similar in terms of phenomenal character (hence in representational content)[34], that is, they will give rise to similar color appearances.

[34] The idea that representational content and phenomenal character cannot come apart in the case of color experience is not universally accepted, still it has been defended by different authors (e.g., Tye, 2000; Byrne and Hilbert, 1997, 2003; see also McLaughlin, 2003). I accept this view in the present argument.

Compare this with non-sensory representational atoms like Fodorian concepts, or free morphemes in natural languages. There the characteristics of the vehicle are totally independent of its content. The phonological shape or spelling of a word is in most cases contingently associated with its meaning. For instance, 'cat' could come to mean *dog* in an isolated community of English speakers on an island of the Pacific. Similarly, as we saw, an atomic symbol of LOT can in principle be locked to all kinds of different properties.

Still, the syntactic structure of a representational vehicle constrains the content that it can have in a given system; this is true of perceptual and conceptual representations alike. The complex LOT formula A&B&C could not mean *if A and B then C* (assuming an ordinary logical processor); nor does it seem easy to conceive of a mind in which the visual percept of a circle (i.e., the neural response pattern that in our brains visually represents circles) represents rectangles.

So here is our point: even though many sensory representations are atomic, there still is a quite intimate link between their vehicle and content. This distinguishes them sharply from non-phenomenal representational atoms, and thus it may contribute to our understanding of why it is the *sensory* input to conceptual processing that is so characteristically phenomenal. A caveat needs to be added, however: this conjecture is so far based on only one example, namely color vision. Future analysis will be needed to establish whether this principle generalizes to other types of sensory experience.

5. Summary and conclusion

We have presented a cognitive account of phenomenal consciousness. The most important difference between our view and other current physicalist approaches to consciousness is that we think phenomenal concepts do not constitute a special class of concepts. Regarding their cognitive and epistemic characteristics, concepts of experience are exactly like other, "non-phenomenal" concepts. When explaining the peculiar features of conscious experience in terms of cognitive architecture, the right place to look at is sensory/ perceptual representations, and their interaction with general conceptual structures. Recently M. Tye has also proposed that phenomenal concepts are in no way special (Tye, 2008); we agree with him on this point. However, our account differs from his in several ways, although both theories are physicalist.

We found that J. Fodor's theory of concepts (Fodor, 1998, 2008) is apt for expressing our intuitions. We utilized his views of the cognitive architecture of the conceptual system, however, our view is neutral with respect to his concept nativism. We cited E. Margolis's account of concept acquisition, which makes it clear that conceptual architecture may be essentially Fodorian while leaving out the nativist assumption and substituting it with concept learning (Margolis, 1999).

We used M. Aydede and G. Güzeldere's theory of phenomenal consciousness for purposes of comparison. Aydede and Güzeldere chose Dretske's informational psychosemantics to build an account of phenomenal consciousness. This led them to the conclusion that sensory and phenomenal concepts do constitute a special class of concepts. We wish to emhasize that this difference between their view and ours is not an uninteresting byproduct of technical differences between psychosemantic frameworks, but rather, it is very important in the proper understanding of consciousness. We think that if we wish to explain the peculiarities of conscious experience, then we should look at experiences (i.e., sensory or perceptual representations) in the first place. Concepts of experience play a secondary – though by no means unimportant – role in the explanation.

We have suggested that the representational atomism of certain sensory experiences is a key notion in reconstructing the epistemic gap associated with conscious experience, but atomism is not the whole story. It needs to be supplemented by additional principles. This is our reply to David Lewis's contention (Lewis, 1997) that representational atomism is a red herring in explaining consciousness (as he put it, it is one of the ways to miss the point about knowing conscious experience). It is not.

Representational atomism gives rise to the linguistic inexpressibility, or ineffability, of sensory experiences, and also the possibility of their developmental causal role exchange. These two factors account for Mary's epistemic limitations while in her chamber (Jackson, 1997), the explanatory gap (Levine, 1983, 2007), the conceivability of zombies, and thus provide an understanding of why reasonable forms of physicalism must remain *a posteriori*. We also added an account of introspection, and suggested some cognitive features that might distinguish representational atoms with phenomenal character from those without it.

References

Abramov, I. (1997). Physiological mechanisms of color vision. In C. L. Hardin & Luisa Maffi (Eds.), *Color categories in thought and language*. pp. 89-117. Cambridge University Press

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. (2006). *Acquaintance and the Mind-Body Problem*.
URL = <http://pantheon.yale.edu/Web publications/Acquaintance.pdf>.

Beilin, H., Bufall, P., B. (1992). *Piaget's Theory: Prospects and Possibilities*. Lawrence Erlbaum

Biederman, I. (1987). Recognition by components: a theory of human image understanding. *Psychological Review*, **94**, 115-147.

Biederman, I. (1990). Higher-Level Vision. In D. N. Osherson, S. M. Kosslyn & J. M. Hollerbach (Eds.), *An Invitation to Cognitive Science*, Vol. 2. (Visual Cognition and Action). pp. 41- 72. Cambridge, MA: MIT Press

Boynton, R. (1997). Insights gained from naming the OSA colours. In C. L. Hardin and L. Maffi (Eds.), *Color categories in thought and language*. Cambridge, MA: MIT Press.

Byrne, A. (2007). Inverted Qualia, *The Stanford Encyclopedia of Philosophy (Summer 2007 Edition)*, Edward N. Zalta (Ed.) ,
URL = <http://plato.stanford.edu/archives/sum2007/entries/qualia-inverted/>.

Byrne, A., & Hilbert, D., R. (1997). Colors and Reflectances. In Alex Byrne & David R. Hilbert (Eds.), *Readings on Color*, Vol. 1. (The Philosophy of Color), pp. 263-288 Cambridge Mass. The MIT Press

Byrne, A., & Hilbert, D., R. (2003). Color Realism and Color Science. *Behavioral and Brain Sciences*, *26*, 3-64

Chalmers, D. (1996). *The conscious mind: in search of a fundamental theory*. New York: Oxford University Press.

Chalmers, D. (2003). Consciousness and its place in nature. In S. Stich & T. Warfield (Eds.), *Blackwell Guide to the Philosophy of Mind* (pp. 102-142). Oxford: Blackwell.

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

Davidoff, J. (1997). The neuropsychology of color. In C. L. Hardin and Luisa Maffi (Eds.): *Color categories in thought and language*. pp.118-134 Cambridge University Press.

Davies, M. (1997). Externalism and Experience. In In Ned Block, Owen Flanagan, and Güven Güzeldere eds., *The Nature of Consciousness.* Cambridge, Mass.: MIT Press, 309-328.

DeValois, R. L., DeValois K. K. (1997). Neural Coding of Color. In Alex Byrne and David R. Hilbert (Eds.): *Readings on Color*, Vol. 2. (The Science of Color), pp. 93-140 Cambridge Mass. The MIT Press.

Dretske, F (1981). *Knowledge and the Flow of Information*. Cambridge Mass. MIT Press

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

Evans, G. (1982). *Varieties of reference*. Oxford University Press.

Fazekas, P. (2009) Reconsidering the Role of Bridge Laws In Inter-Theoretical Reductions. *Erkenntnis*, **71,** (3) pp. 303-322

Fiengo, R., & May, R. (1998). Names and Expressions *Journal of Philosophy*, Vol. 95, No. 8 pp. 377-409

Flavell, J. (1999). Development of Children's Knowledge about Unconsciousness. *Cild Development*, **70** (2) 396-412.

Fodor, J. A. (1983). *The modularity of Mind*. Cambridge, Mass.: MIT Press.

Fodor, J.A. (1987). Why there still has to be a language of thought. In *Psychosemantics*, pp. 135-154. Cambridge Mass. MIT Press

Fodor, J.A. (1990). A Theory of Content, I-II. In J. Fodor, *A Theory of Content and Other Essays*, Cambridge Mass. MIT Press

Fodor, J.A. (1994). *The Elm and the Expert; Mentalese and Its Semantics.* Cambridge Mass. MIT Press

Fodor, J. A. (1998). *Concepts*. Cambridge, Mass.: MIT Press.

Fodor, J. A. (2008). *LOT2 – The Language of ThoughtRevisited*. Clarendon Press, Oxford

Hardin, C. L. (1988). *Color for Philosophers: Unweaving the Rainbow*. Indianapolis, MA: Hackett.

Harman, G. (1988). Wide functionalism. In *Cognition and Representation*, S. Schiffer and S. Steele, eds. Boulder, Westview.

Harman, G. (1997). The Intrinsic Quality of Experience. In Ned Block, Owen Flanagan, and Güven Güzeldere eds., *The Nature of Consciousness.* Cambridge, Mass.: MIT Press, 663-676.

Hilbert, D. R., Kalderon M. E. (2000). Color and the Inverted Spectrum. In Steven Davis (Ed.): *Color Perception: Philosophical, Psychological, Artistic and Computational Perspectives*. New York: Oxford University Press.

Hill, C. S., & McLaughlin, B. P. (1998). There are fewer things in reality than are dreamt of in Chalmers' philosophy. *Philosophy and Phenomenological Research,* **59**, 445–454.

Hunt, R. G. W. (1982). A model of colour vision for predicting colour appearance. *Color Research and Application, 7*, 95-112.

Jackson, F. (1997). What Mary Didn't Know. In Ned Block, Owen Flanagan, and Güven Güzeldere eds., *The Nature of Consciousness.* Cambridge, Mass.: MIT Press, 567-570.

Jackson, F. (1998). *From Metaphysics to Ethics: A Defence of Conceptual Analysis*. Oxford University Press.

Jakab, Z. (2000). Ineffability of qualia: a straightforward naturalistic explanation. *Consciousness and Cognition*, **9** (3), 329-351

Jakab, Z. (2003). Phenomenal Projection. *Psyche*, 9(04), January 2003 (http://psyche.cs.monash.edu.au/articles/jakab/index.html)

Jakab, Z. (2005). Opponent processing, linear models, and the veridicality of color perception. In *Cognition and the Brain: The Philosophy and Neuroscience Movement*, edited by Kathleen Akins and Andrew Brook, Cambridge University Press 336-378.

Jakab, Z. (2006). Revelation and normativity in visual experience. *Canadian Journal of Philosophy*, **36** (1) 25-56.

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

Kripke, Saul (1979) "A Puzzle about Belief" in A. Margalit (ed.) (1979) *Meaning and Use* Dordrecht: D. Reidel, pp. 206-242.

Kulvicki, J. (2005). Perceptual Content, Information, and the Primary/Secondary Quality Distinction. *Philosophical Studies*, **122** (2), 103-131.

Larson, R.K. and Ludlow, P. (1993). Interpreted Logical Forms. *Synthese* 95, pp. 305-355.

Leslie, A. (1987). Pretense and Representation: The Origins of "Theory of Mind". *Psychological Review*, **94** (4), 412-426.

Leslie, A. (2002). Prenense and Representation Revisited. In Nancy L. Stein, Patricia J. Bauer, and Mitchell Rabinowitz, eds. *Representation, Memory, and Development: Essays in Honor of Jean Mandler*. Lawrence Erlbaum, London.

Levin, J. (2008). Taking Type-B Materialism Seriously. *Mind & Language,* **23**, 402-425.

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

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

Lewis, D. (1997). What Experience Teaches. In In Ned Block, Owen Flanagan, and Güven Güzeldere eds., *The Nature of Consciousness.* Cambridge, Mass.: MIT Press, 579-596.

Loar, B. (1990). Phenomenal States. In J. Tomberlin (Ed.), *Philosophical Perspectives* (Vol. 4, pp. 81-108.). Northridge: Ridgeview Publishing Company. Reprinted in Ned Block, Owen Flanagan, Güven Güzeldere (Eds.) *The Nature of Consciousness – Philosophical Debates*. Cambridge Mass. The MIT Press, 1997, pp. 597-616.

Maloney, L. T. (1986). Evaluation of linear models of surface spectral reflectance with small numbers of parameters. *Journal of the Optical Society of America* A, *3*, 1673-1683.

Maloney, L. T. (1999). Physics-based models of surface color perception. In Gegenfurtner, K. R. & Sharpe, L. T. (Eds.) (1999), *Color Vision: From Genes to Perception.* pp. 387-418 Cambridge, UK: Cambridge University Press

Maloney, L. T. (2002). Surface Color Perception and Environmental Constraints. In Mausfeld R., and Heyer, D. (Eds.): *Colour Vision: From Light to Object*. Oxford: Oxford University Press

Maloney, L., T. (2003). Illuminant estimation as cue combination. *Journal of Vision*, *2*, 493-504.

Maloney, L. T., Wandell, B. A. (1986). Color constancy: a method for recovering surface reflectance. *Journal of the Optical Society of America* A, *3*, No. 1, 29-33.

Margolis, E. (1999). How to acquire a concept. In Eric Margolic and Stephen Laurence (eds.). *Concepts: Core Readings.* Cambridge Mass. MIT Press

Marr, D. (1982). *Vision. A computational investigation into the human representation and processing of visual information*. New York: Freeman.

Marras, A. (2005). Consciousness and reduction. *The British Journal for the Philosophy of Science*, **56,** pp. 335-361

Matthen, M. (1999). The Disunity of Color. *Philosophical Review*, *108*, 47-84.

McLaughin, B. P. (2003). Color, consciousness, and color consciousness. In Q. Smith and A. Jokic (Eds.) *Consciousness: New philosophical perspectives*. pp. 97-152 Oxford: Oxford University Press

McLeod, D. I. A., Boynton, R. M. (1979). Chromaticity diagram showing cone excitations by stimuli of equal luminance. *Journal of the Optical Society of America A*, *69*, 1183-1186.

Mollon, J. D. (2000). Cherries among the Leaves: The Evolutionary Origins of Color Vision. In Davis, S. (Ed.) *Color Perception: Philosophical, Psychological, Artistic and Computational Perspectives.* pp. 10-30 New York: Oxford University Press

Neale, S. (1990). *Descriptions*. Cambridge Mass. MIT Press

O'Dea, J. (2002). The Indexical Nature of Sensory Concepts. *Philosophical Papers 31* (2), 169-181.

Palmer, S. E. (1999a). *Vision Science – Photons to Phenomenology.* Cambridge, MA: MIT Press.

Palmer, S. E. (1999b). Color, Consciousness and the Isomorphism Constraint), *Behavioral and Brain Sciences*, *22*, 923-943

Papineau, D. (2002). *Thinking about consciousness*. Oxford: Clarendon Press.

Papineau, D. (2007). Phenomenal and Perceptual Concepts. In Torin Alter and Sven Walter (eds). *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*. Oxford University Press. pp. 111-144.

Perner, J. (1994). Prelief: The Conceptual Origins of Belief and Pretence. In C. Lewis and P. Mitchell, eds. *Children's Early Understanding of Mind – Origins and Development*. Lawrence Erlbaum, Hillsdale, USA.

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

Putnam, Hilary (1975) "The Meaning of 'Meaning'" in Gunderson (ed.) (1975) *Language, Mind and Knowledge : Minnesota studies in the Philosophy of Science* Vol. VII, Minneapolis: University of Minnesota Press, pp.131-193.

Pylyshyn, Z. W. (2002). Mental Imagery: In search of a theory. Behavioral and Brain Sciences, 25(2), 157-237.

Pylyshyn, Z. W. (2003). Seeing and visualizing: It's not what you think . Cambridge, MA: MIT Press/Bradford Books.

Pylyshyn, Z. (2006). Imagery. In Gregory, Richard. *Oxford Companion to the Mind (Second Edition, 2006)* Oxford University Press

Shoemaker, S. (1981). The Inverted Spectrum. *Journal of Philosophy,* **74** (7), 357-381.

Stainton, R. (1996). *Philosophical Perspectives on Language*. Broadview Press.

Stoljar, D. (2005). Physicalism and Phenomenal Concepts. *Mind and Language,* **20**, 469-494.

Thompson, E. (2000). Comparative Color Vision: Quality space and Visual Ecology. In Steven Davis (Ed.), *Color Perception: Philosophical, Psychological, Artistic and Computational Perspectives*. New York: Oxford University Press, pp. 163-186.

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

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

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

Tye, M. (2008). *Consciousness Revisited*. Cambridge, MA: MIT Press.

Wallis, G., & Bülthoff, H. (1999). Learning to recognize objects. *Trends in Cognitive Sciences*, *3* (1), 22-31.

Werner, J. S., & Wooten, B. R. (1979). Opponent chromatic mechanisms: Relation to photopigments and hue naming. *Journal of the Optical Society of America, 69*, 422-434.

Wooten, B., & Miller, D. L. (1997). The psychophysics of color. In C. L. Hardin & Luisa Maffi (Eds.): *Color categories in thought and language*. Cambridge University Press, pp. 59-88.

Wuerger, S. M., Maloney, L. T., Krauskopf, J. (1995). Proximity Judgments in Color Space: Tests of a Euclidean Color Geometry. *Vision Research*, **35**, (6), 827-835.

Zeki, S. (1999). *Inner Vision. An Exploration of Art and the Brain.* Oxford: Oxford University Press.