

*Cognitive Architecture, Concepts, and  
Introspection: An Information-Theoretic Solution to  
the Problem of Phenomenal Consciousness*

MURAT AYDEDE  
University of Florida

GÜVEN GÜZELDERE  
Duke University

**Abstract**

This essay is a sustained attempt to bring new light to some of the perennial problems in philosophy of mind surrounding phenomenal consciousness and introspection through developing an account of sensory and phenomenal concepts. Building on the information-theoretic framework of Dretske (1981), we present an informational psychosemantics as it applies to what we call *sensory concepts*, concepts that apply, roughly, to so-called secondary qualities of objects. We show that these concepts have a special informational character and semantic structure that closely tie them to the brain states realizing conscious qualitative experiences. We then develop an account of introspection which exploits this special nature of sensory concepts. The result is a new class of concepts, which, following recent terminology, we call *phenomenal concepts*: these concepts refer to phenomenal experience itself and are the vehicles used in introspection. On our account, the connection between sensory and phenomenal concepts is very tight: it consists in different semantic uses of the same cognitive structures underlying the sensory concepts, such as the concept of red. Contrary to widespread opinion, we show that information theory contains all the resources to satisfy internalist intuitions about phenomenal consciousness, while not offending externalist ones. A consequence of this account is that it explains and predicts the so-called conceivability arguments against physicalism on the basis of the special nature of sensory and phenomenal concepts. Thus we not only show why physicalism is not threatened by such arguments, but also demonstrate its strength in

virtue of its ability to predict and explain away such arguments in a principled way. However, we take the main contribution of this work to be what it provides in addition to a response to those conceivability arguments, namely, a substantive account of the interface between sensory and conceptual systems and the mechanisms of introspection as based on the special nature of the information flow between them.

## 1. Introduction

The current manifestation of the mind-body problem primarily centers around a deep disagreement between materialists and anti-materialists about the ontological nature of phenomenal consciousness. Facing the charge that they lack the conceptual resources to understand the phenomenal, materialists are challenged to accept the existence of non-physical properties, or, at a minimum, admit to being completely in the dark regarding how to bridge the “explanatory gap” between the physical and the phenomenal. Joseph Levine, in his recent book (2001), concludes his lengthy discussion of the explanatory gap with this challenge:

What emerges from our discussion is that the explanatory gap is intimately connected to the special nature of phenomenal concepts. E-type materialists [exceptionalists] try to save materialism from the conceivability argument by arguing that phenomenal concepts are special in some way. Well, I grant that, but then we have the problem of providing an explanation in physicalistic terms of that very specialness, and we don’t seem to have one. If we could explain the explanatory gap, then either it would go away or we would just learn to live with it. But it seems we can’t do that without a good account of phenomenal concepts, and that’s something we don’t have. We lack both an account of phenomenal properties and phenomenal concepts. (p. 86)

In this essay, we provide both. The overarching goal of our project is to present a sustained and thorough information-theoretic argument that sheds new light on the problems surrounding phenomenal consciousness, especially the problem of the explanatory gap.

Building on the information-theoretic framework of Dretske (1981), we first develop an informational psychosemantics for what we call *sensory concepts*, i.e., concepts that apply, roughly, to so-called secondary qualities of objects. We show that these concepts have a special informational and semantic character that ties them closely to the brain states realizing conscious experiences from which they are acquired. We then develop an account of *phenomenal concepts* utilizing this special character of sensory concepts. Phenomenal concepts are those concepts we use in introspecting our experiences and their qualities. It is through these concepts that we conceive of the phenomenal character of our experiences. On our account,

sensory and phenomenal concepts turn out to share the same cognitive structures but their semantics are differently anchored. What makes this possible is the dual informational content of these structures. In the end, it is the special nature of phenomenal concepts that enables us to meet the anti-materialist challenge. Further, contrary to widespread opinion, we show that informational psychosemantics contains the resources to satisfy internalist intuitions about phenomenal consciousness in a principled way, while not offending externalist ones.

Our account contributes to what appears to be a growing convergence of views, sometimes loosely grouped under the label of “perspectivalism,”<sup>1</sup> which in recent years have been developed in response to conceivability arguments against physicalism.<sup>2</sup> In its typical versions, perspectivalism is advanced in three stages. First, it diagnoses the puzzle involved in attempting to conceive the phenomenal in terms of the physical as a Frege puzzle—namely, as one arising from distinct but co-denotational concepts. Second, it points out that the Frege case at hand is special in a way that marks it off from standard Frege cases, and that this specialness needs accounting for. Third, it postulates a group of concepts, typically called “phenomenal concepts,” whose nature is said to be perspectival, a fact that is supposed to reveal what is so special about our epistemic access to the phenomenal qualities of our experiences.

We think that extant perspectivalist accounts are lacking in precisely those respects that are crucial for the acceptability of physicalist responses to the conceivability arguments. At the heart of the debate is a set of intuitions about the conceivability of a range of scenarios (e.g., zombies, inverted spectra). The debate centers around what these intuitions show. The anti-physicalists want to draw a metaphysical conclusion (namely, the falsity of physicalism) from the conceivability of these scenarios. Thus, the question of whether conceivability, which is an epistemic affair, entails metaphysical possibility becomes one of the crucial issues. The perspectivalist materialists with whom we join forces in this paper<sup>3</sup> maintain that no such metaphysical conclusion follows, but they tend to restrict their claim only to those cases where the scenarios involve phenomenal consciousness as it relates to the physical world (after all conceivability seems to be our only guide to possibility). Thus, at a global scale, many perspectivalists grant that given a complete physical description of the world and competence with common concepts that apply to macro phenomena, it is possible in principle to derive *a priori* all true claims about macro phenomena. For instance, given complete physical knowledge (augmented with standard indexical information, etc.), it is not conceivable that water should boil at a temperature other than 0° C at sea level. They also grant to anti-physicalists that facts about phenomenal consciousness (and only these facts) cannot be so derived. But they refuse to draw any metaphysical conclusion from that.

It is the special nature of phenomenal concepts that is supposed to discharge the heavy explanatory burden the perspectivalists incur by this refusal. Understandably, the absence of a principled and independently motivated story about phenomenal concepts that would justify taking this “exceptionalist” route has made perspectivalists vulnerable to accusations of special pleading. On standard accounts, phenomenal concepts turn out to have just those features (simple, primitive, unanalyzable, demonstrative, etc.) that happen to make them immune to absorption into a general logical reduction pattern. But where these features come from and what independent reasons we have to believe that they work this way are left mostly unexplained. Phenomenal concepts are claimed to be special, but the critics of physicalism are justified in their demand that their special character needs to be independently grounded in a general and naturalistic account of concept formation and use. That is, whatever it is about these concepts that are supposed to justify making consciousness an exception to an otherwise perfectly general pattern needs to fall out naturally from a general and independently motivated account of conscious experience and the concepts it gives rise to.

In other words, the account of phenomenal concepts must be part of a more general theory whose explanatory power should go beyond justifying the exception and saving materialism. Levine (2001) rightly points out that perspectivalists haven’t so far produced an independently motivated general and naturalistic account with sufficient and credible detail.<sup>4</sup> This work constitutes an attempt to develop such an account—a principled and independently motivated physicalist story, using the resources of information theory, about the special nature of phenomenal concepts and their role in explaining why there is an explanatory gap of the sort not present in other phenomena.

We derive such an account from a cognitive architectural framework and general informational psychosemantics of sensory concepts together with an information-theoretic account of introspection. Fundamentally, we agree with Gareth Evans on the philosophical utility of construing human beings as “gatherers, transmitters, and storers of information,” for “these platitudes locate perception, communication, and memory in a system—the informational system—which constitutes the substratum of our cognitive lives” (Evans, 1982: 122). As such, our work is in the spirit of previous attempts to provide naturalized information-theoretic accounts of thought and reference (Evans 1982, Barwise and Perry 1983, Israel and Perry 1990, Fodor 1987, 1990). But our greater debt is to Fred Dretske’s seminal work, *Knowledge and the Flow of Information* (1981), which first brought the resources of information theory into debates in epistemology and philosophy of mind in a ground-breaking way.<sup>5</sup> As will be apparent in the coming sections, our account also provides an unexpected synthesis of otherwise quite diverse views. For instance, it has consequences that we believe should

satisfy internalists about phenomenal consciousness, while integrating intuitions that motivate both higher-order perception and higher-order thought accounts of consciousness (even though our view does not itself fall under these labels).

We ask the reader to bear with us while we lay the groundwork for our account before we come to the last section (§ 9) of this rather lengthy essay, where we ultimately respond to the conceivability arguments against physicalism. In the next section (§ 2), in a somewhat reconstructed form and sometimes using different terminology, we will present the bare bones of an information-theoretic account of concept formation (from sensory experiences) based on Dretske (1981).<sup>6</sup> We will be substantially modifying and developing this account in the remaining sections: in particular, in §§ 3–8, we will apply this account to sensory concepts and derive from it an account of phenomenal concepts and introspection. In §§ 6–7, we will pay close attention to bodily sensations (especially, pain) and the sensory concepts they give rise to, because we believe that these sensations and the way we think about them form the microcosm of a more general mechanism of introspection. Although we will discuss the conceivability arguments in the last section (§ 9), we submit that by that time the reader will be in a position to see where the real action is. Indeed, we take the main contribution of this work to be what it provides in addition to a satisfying response to the conceivability arguments, namely, a substantive and detailed account of the interface between sensory and conceptual systems and the mechanisms of introspection based on the special nature of the information flow between them.

## **2. The Architecture of Information Flow in Cognition**

Information-theoretic psychosemantics postulates an *architectural* distinction between sensory systems and a central cognitive system controlling the intentional behavior of the organism. The sensory system has the job of providing information about one's environment to the cognitive system, and normally affects behavior only indirectly, via intermediary cognitive structures. Sensory systems hook up with the environment via transducers whose job is to transform the particular forms of energy impinging on the peripheral sensory organs into forms usable by the internal perceptual systems. The output of transducers and much of the subsequent processing in the sensory and perceptual systems appear to be automatic and unconscious (but see below). The output of the sensory system is a sensory representation of (some aspects of) the distal layout that is made available to the central cognitive system.

In this framework, the sensory representations are *conscious* only insofar as the information they contain is available to the central conceptual system, even if the information is not fully put to use. We will also say that the

information contained in sensory representations is available to the organism *consciously* only insofar as the organism *can* conceptualize this information, i.e., only insofar as the information can be used in the acquisition or deployment of the relevant concepts.<sup>7</sup> Hence, we will use “having an experience” (which is generally but not necessarily a conscious affair) and “tokening a sensory representation” interchangeably.<sup>8</sup> We will come back to this point later on, but for the rest of this paper we will concern ourselves with sensory representations that are conscious in this way. Hence we will not further discuss those modular (pre- or intra-perceptual) processes whose state-transitions and outputs are not consciously accessible—that is, which do not constitute direct inputs to the central cognitive system.<sup>9</sup>

What are the functional determinants of this architectural distinction? We have already touched on one: sensory representations don’t normally affect behavior directly. It is largely the central cognitive system which controls voluntary behavior through motor systems. So a necessary condition of a cognitive structure’s being conceptual as opposed to sensory is its executive connections to behavior. A representation is sensory (as opposed to perceptual or conceptual—see § 2.2 below), on the other hand, only if it makes information about one’s environment (internal/bodily as well as external) available to the conceptual system for further processing, which normally also makes the representation (experience) conscious.<sup>10</sup> Besides this, the most important characteristics underpinning the architectural distinction are to be found in the following five distinctions: vertical vs. horizontal information processing, sensation vs. perception, analog vs. digital encoding of information, extractable vs. non-extractable analog information, and acquisition vs. deployment of concepts. We explain each of these in the remainder of this section.

### 2.1. *Vertical vs. Horizontal Information Processing*

First, sensory experiences are supposed to track changes in the environment. In this they are (non-conceptual) representations whose primary job is to make available to their hosts temporally indexed information about the environment. The crucial point here is that sensory experiences normally carry information about features of the environment: they are *responses* to environmental events. As such their informational value is typically restricted within a time frame sufficient for the organism to act back on the environment on the basis of this information. In short, sensory representations are normally stimulus-driven (*a fortiori* not directly voluntary). We will call this aspect of information processing *vertical* information processing.

By contrast, central cognitive processes, such as thinking, reasoning, remembering, imagining, and daydreaming, are normally *horizontal* forms of information processing. By this we mean that they *can*, and frequently *do*,

occur in the absence of a direct causal (i.e., vertical/informational—see below) relation with the things being thought about. This is perhaps the most important hallmark of human intentionality. In contrast to sensory systems, central cognitive systems harbor representational processes defined over concepts that are not directly prompted by what those concepts represent.<sup>11</sup>

Although all concepts can be informationally decoupled from their referents in horizontal processes, most of them *can* also be used vertically, so that their tokenings carry information about the (instantiation of the) property they apply to. In this *extended* sense of a vertical process, experience is the necessary intermediary.<sup>12</sup>

In brief, conceptual representations *are* the kind of cognitive structures that are capable of being engaged in horizontal processing, whereas sensory representations are not.<sup>13</sup>

## 2.2. *Sensation vs. Perception*

There is a useful sense in which perception, unlike sensation, is the vertical informational process whereby objects of sensation and their sensible qualities are discriminated and recognized, i.e., categorized or classified under concepts.<sup>14</sup> For most perceptual and observational concepts, this normally takes the form of recovering the information already (mostly) in the sensory array by computational processes that result in the tokening of a concept applied to the object of perception. We see this process mainly as one of information extraction by digitalization or abstraction from a rich array of information present in analog form in the experience. The mechanism underlying the formation of primitive sensory concepts and their vertical deployment is probably hard-wired in concept-using organisms like us.

According to this scheme, then, visual object recognition, for instance, however automatic it may be, is mostly a central process,<sup>15</sup> since it involves categorizing an object under a “visual concept.” Although the process itself appears to be unconscious and automatic, many features of the output representation (like variation in light intensities, lines, edges, colors, distance, orientation, texture, relative position, etc.), apparently utilized in the extraction process, are also consciously (hence centrally/globally) available. So perception is a central process in our sense and should be treated as a species of conception.

## 2.3. *Analog vs. Digital Encoding of Information*

Another determinant of the architecture, most important for our purposes, is captured by a distinction between the ways in which information is coded in the representations. Here, we follow Dretske’s original characterization (cf. Dretske 1981: Chapter 3):



- (i) The most specific information a signal  $r$  carries about a source  $s$  is the information  $r$  carries about  $s$  in *digital* form.
- (ii) If  $r$  carries more information about  $s$  [or, about  $t$  ( $\neq s$ )] *in virtue of* carrying this digital information about  $s$ , then this extra information is said to be carried by  $s$  in *analog* form.
- (iii) Analog information is information *nested* (nomologically or analytically) in the information carried in digital form.

Note that according to this characterization a signal always carries information in both digital and analog form: it's just that the most specific information is selected as digital.

The cognitive value of a sensory representation lies largely in the information about the distal layout it carries in analog form. Its digital informational content is the most specific information it carries about this layout, which is very rich not only in detail but also in amount. The conceptual system is mostly keyed to the information nested in this specific and rich information. The analogy here between pictures and sensory representations will be helpful. If we take a color picture of a cubical object, the picture will carry very rich, detailed, and determinate information about the size, texture, and orientation of the object, as well as its position relative to other objects, the illumination conditions, its determinate shades of color and their brightness across its surface, and so forth. We can think of this very specific and detailed information as expressible by a very long conjunction. But nested in this most specific information there will be less specific information implied by it, such as the information that the object is (just) cubical, that it has (just) six faces, that it has eight corners, that it is darkly colored, (just) colored, etc. Normally we are interested in the analog information carried by the picture. We may be interested merely to know that the object depicted is cubical—discarding the more specific information about its color, size, orientation, etc. Or, depending on the situation, we may be interested only in its size or orientation.

Similarly with sensory representations. The conceptual system mostly exploits the analog information nested in the digital information carried by sensory representations. In fact, part of what makes a cognitive structure a conceptual representation is the way it digitalizes the analog information contained in the sensory representations. Concepts are those representations (subject to the above architectural constraints) whose most specific informational content is acquired from information carried (mostly) in analog form by sensory representations. Concepts (except sensory ones—see below) are designed to selectively respond to and utilize the analog information contained in sensory representations. So, for instance, even though we cannot *sensorially* represent a triangle without at the same time representing its determinate size, shape, orientation, etc., we can *conceptually* represent an object simply as a triangle without representing anything more specific or



determinate about it.<sup>16</sup> Concepts on this scheme are those structures that are *acquired* from sensory representations, mostly on the basis of the analog information they carry.

In this framework, the *semantic content* of a concept is identified with the information it carries in digital form. The *informational content* of a concept, however, is not unique in the way the semantic content is supposed to be, since a vertical tokening of a concept will carry *all* the information nested in its digital informational content (i.e., in its semantic content). So, for example, when you identify a geometrical shape as an isosceles triangle, your identification carries more information about the object nested in its being such a triangle, e.g., that it has (just) three sides, that it has (just) three corners, that it is (just) a geometrical shape, that it has a surface area, etc. These separate pieces of information are all carried in analog form.

#### 2.4. Extractable vs. Non-Extractable Analog Information

Both sensory and conceptual representations carry information in both analog and digital form. But they encode analog information in fundamentally different ways. In particular:

- Whereas there is always *some* analog information sensory representations carry in *extractable* format, the (primitive) conceptual representations carry *all* their analog information in *non-extractable* form.<sup>17</sup>

A sensory representation is physically realized in such a way that its complex structure allows the analog information contained in it to be extracted by the conceptual system operating on it. Of course, what information can be extracted from the sensory representation doesn't depend solely on its complex informational structure; it also depends on the capabilities and the sophistication of the conceptual system and what other information is available to the system. But, subject to these constraints, it is necessary for conceptualization that the analog information in the sensory representation is carried in a form that is extractable, and not all information carried in analog form is.

To illustrate, consider the example Dretske uses (1981: 138–39).<sup>18</sup> It is possible to carry *all* the information encoded by a picture of a scene with a simple/primitive signal, say a buzzer system. Suppose the buzzer is activated when and only when a camera attached to the buzzer detects the occurrence of a situation *exactly* like the one depicted in the picture. As Dretske notes, computer recognition programs that rely on whole-template matching procedures approximate this kind of transition from one form of coding to another. Both structures carry exactly the same information, both digital and analog. However, we will say that the buzzer's buzzing carries the

analog information carried by the picture in a way that is not extractable, whereas the picture carries it in an extractable form.

This distinction needs to be developed in more detail in terms of physical constraints on the structures realizing the representations, but what is intuitively obvious—and all we need for present purposes—is that the representational format which allows for information extraction must consist in a structure *complex* enough to be the *only* source for subsequent digitalizations based on it.<sup>19</sup> The activation of the buzzer, though it carries all the information carried by the picture, is structured in such a way that does not allow for digitalization of the information it carries in analog form. Primitive conceptual representations are like the buzzer system: although their vertical tokenings carry analog information nested in their digital content, they are structured in such a way that they cannot serve as the sole basis for digitalization of this information. This is part of the reason why primitive concepts are sometimes characterized as *discrete* representational structures or symbols.

### 2.5. *Acquisition vs. Deployment of Concepts*

Although the distinction between the *acquisition* and *deployment* of concepts is not a functional determinant of the informational architecture, it is important to keep in mind for clarificatory purposes. Both acquisition and deployment can be *vertical* and *horizontal* in some intuitively extended sense. So, for instance, we can acquire concepts by reading, or by being talked to, by looking at pictures, by engaging in inference to the best explanation, etc.<sup>20</sup> This would be horizontal acquisition of concepts. Also note that for the moment we are using the term “acquisition” in a way that is neutral between *triggering* and *learning*.

Now that we have the functional determinants of the architecture of information flow in cognition and the relevant distinctions,<sup>21</sup> we will focus on the nature of the concepts to which this architecture gives rise.

## 3. Concepts and Their Sensory Bases

Among the concepts directly and immediately acquired from sensory experience are what we will call *sensory concepts*. These form a special class of concepts that will be important for what follows. Intuitively and roughly put (to be qualified in a moment), sensory concepts are those concepts whose digital informational content is also part of the digital informational content of the sensory representations from which they are acquired, so that the *abstraction/digitalization distance* between the concepts and these experiences is *minimal*.<sup>22</sup>

The digital informational content of sensory representations is rich along several dimensions. We can think of these dimensions as presenting determinables such that the resolution of our sensory experiences marks the limit

of their most determinate values about which we can gather sensory information. To the extent that we can separate these dimensions, we can speak of that part of the total digital information content of an experience that belongs to one of these dimensions fixed by the modality of the experience. So, for instance, under conditions that are optimal for color vision, seeing a ripe tomato will involve a visual experience whose total digital content contains the most specific information about the color of the tomato: it will represent the tomato as having a determinate shade of red, say,  $\text{red}_{16}$ . This is part of the total digital content of the visual experience containing information about the color of the object seen. As mentioned above, we can conceive of this total digital content as being expressed by a very long conjunction detailing all the most specific information it carries. The particular shade of color that a region in the visual field has, then, would be one of the conjuncts.<sup>23</sup> Sensory concepts are those concepts that are closest (in terms of abstraction distance) along these different dimensions to the digital informational content of experiences from which they are acquired.

If the property of being  $\text{red}_{16}$  is a disjunctive property whose disjuncts are particular spectral reflectances, then the information the sensory representation of the tomato carries about  $\text{red}_{16}$  is about this disjunctive property. Every disjunct would be a particular ratio fixed by the percentage of light that the surface of an object reflects at each of the three characteristic wavelengths determined by the response sensitivity of three retinal cone types.<sup>24</sup> But whatever feature of sensory representation is responsible for carrying this information, it carries it without revealing its complex and disjunctive character. For instance, this feature, by carrying information about a surface's being  $\text{red}_{16}$ , also carries the analog information that it has a spectral reflectance, or that it (just) reflects light at different wavelengths. These are nested in the information that the surface is  $\text{red}_{16}$ . But these pieces of analog information cannot be recovered or extracted from the signal, i.e., from whatever feature of the sensory representation carries the color information in question.

There is, however, still some abstraction/digitalization—some loss of information—in this process. This can be explained in terms of a distinction between concepts used in synchronic discriminatory tasks and concepts used in diachronic recognitional or identification tasks. In fact, we typically reserve the notion of a *concept* for those cognitive structures involved in the latter sort of task. Consider the tomato again. If the conditions are appropriate, it will be possible to discriminate slight variations in the shade of red across the surface of the tomato. But when the same shades of color are shown to us diachronically we may not be able to discriminate among them: most of the time the best we can do is identify and co-classify them as, say, dark red. Both kinds of task involve discrimination and categorization of different color stimuli, and so, in this sense, require conceptual capacities. In what follows, however, when we talk about sensory concepts, we will

have in mind the most specific concepts one can have as revealed by diachronic recognition tasks, which involve memory. It is clear that the abstraction distance between sensory experiences and the sensory concepts conceived in this way is still minimal, although there is still some information lost. Notice that in the case of color concepts this distance can be explained entirely in terms of set-theoretic notion of inclusion. When these concepts are vertically deployed, the information they carry is disjunctive: they say something like “it is either  $\text{red}_1$  or  $\text{red}_2$  or  $\text{red}_3$  or . . .  $\text{red}_n$ ”, where  $n$  is finite and  $\text{red}_1$  is the most determinate shade of red one’s visual experiences can carry information about and thus be synchronically discriminated.

It is important to note that the disjuncts here are still *colors*—determinate shades of red. This is important because the abstraction process here is not based on information about the constituents of colors (whatever objective properties color experiences/concepts detect), which are not themselves colors. So, for example, if color vision detects sets of individual surface spectral reflectances, color sensations don’t represent them *by* representing their constituent properties, say, individual reflectances or whatever further properties constitute these reflectances. Hence, color sensations don’t represent colors as having constituent structure, or as we will say sometimes for convenience, as simple/atomic properties.<sup>25</sup>

Contrast this to the visual representation of shapes. Our visual system happens to be such that we can’t visually represent a geometrical figure (in such a way that we can then recognize it as what it is, say, as a square) without simultaneously representing the lines, angles, curves, edges, and corners that, in some intuitive sense, *constitute* the figure. It is important to note that these constituents are not more determinate instances of the same figure type, so that even the concept of a most determinate geometrical figure of that type will not be minimally close to the sensory base it is directly acquired from—even though these sensory bases are the sole authoritative source of acquisition for such concepts. We will call such concepts *perceptual concepts*. The information necessary and sufficient for the correct application of these concepts, whose abstraction distance is nevertheless not minimal (but shorter than what we will call below observational concepts), is normally contained in the sensory base from which they are directly acquired. Typical perceptual concepts in the case of vision include concepts of spatiotemporal relations, geometrical figures, and shapes.

For the sake of completeness, we can distinguish sensory and perceptual concepts from *observational concepts* like the concept of an apple, a robin, a tree, a lake, and a truck. These concepts are also *typically* acquired from an appropriate sensory base, but they need not be, and often are not. However, the information contained in experiences required in the correct application of these concepts is more impoverished, in the sense that it

always underdetermines correct categorization. In other words, although the information about the denotations of these concepts can be perceptually available, its delivery requires that certain channel conditions *external* to the sensory systems be in place. The abstraction distance between these concepts and the sensory bases from which they may be acquired is considerably greater than in the case of sensory and perceptual concepts. What seems to mark the difference is that (most of) the sensory information used in the acquisition and deployment of observational concepts is typically only contingently related to the objects in their extensions.

It is no accident that thought experiments involving spectrum inversion are carried out in terms of sensory bases of sensory concepts, where the property detected and denoted is represented as simple or atomic.<sup>26</sup> Although we cannot conceive of inversion with respect to the properties denoted by perceptual concepts (e.g., of shapes) and their sensory bases, there is nothing preventing a differently organized cognitive system from performing this feat. We can imagine and even construct devices that “sensorially” detect geometrical shapes (quite abstract from our cognitive point of view) by outputting simple and primitive sensory representations. For instance, we can construct a detector that responds with a green light when it detects a square (*any* square) and with a red light when it detects a circle (*any* circle). Suppose that all the information it uses in making its responses is lost at the final output stage. When this device, a 2D geometrical shape detector, lights up green, its relevant state carries the information that something it is informationally connected to is square. If it lights up red, its state carries the information that something is circular. But even if the “sensory” outputs of the device carry these pieces of information, they are structured in such a way that there is no way to recover any information about the structural relationships holding among the internal constituents of these shapes. Of course, these “sensory” outputs also carry information about the constituent properties (necessarily so), but only in analog form that is not extractable. Nor is it possible to extract any topological information that obtains between these different shapes—if the device carries information about the shapes in a spatial array. For all the device “knows,” whatever is being represented by these colored lights, it is simple and atomic. There are no computational/formal constraints stemming from the representations themselves that would make the thought experiment of an “inverted shape” unintelligible here. For all the device “knows,” circles could look exactly the same to it as squares do now, and vice versa.

If this device is also equipped with a central conceptual system that can acquire concepts from such “sensory” representations, the concept of a circle the device directly acquires from its “experiences” will be a *sensory* (as opposed to *perceptual*) concept in our sense. *Our* concept of a circle is not sensory because the sensory representations from which it is acquired don’t carry the information that something is a circle as part of its total

digital informational content, so that when our conceptual system digitalizes this piece of information there is always more specific information that is lost but nevertheless available to the central cognitive system for digitalization.<sup>27</sup> Furthermore, it is this lost information that seems to be used in the acquisition and vertical deployment of the target concept. What prevents the abstraction distance from being minimal here is the existence of more specific but used-and-then-discarded information that is nevertheless available to the conceptual system for digitalization (which, subject to some further conditions, makes this used-but-then-discarded information contained in the experience *consciously available*).

In contrast to our perceptual system, the architecture of this device is such that the abstraction distance between the “sensory” and “conceptual” representations of circles and squares is minimal. Not surprisingly, we are not such machines. But it is important to keep in mind that there is no *logical* necessity in our having the perceptual and cognitive architecture that we do, including the set of particular abstraction distances it gives rise to—although there are most likely evolutionary and ecological reasons for this architecture.<sup>28</sup>

Another way to see what makes sensory concepts so special is to understand the nature of the abstraction distance between them and their sensory bases. As we have said, this distance is minimal (subject to the qualification we have just introduced), which is what marks these concepts off from the rest. Following Fodor (1990) and Margolis (1998), we will call the mechanisms that mediate the informational link between the vertical tokenings of a concept and the instantiations of the property it applies to *sustaining mechanisms*.<sup>29</sup> The intra-cranial portion of the sustaining mechanisms for sensory concepts is not cognitive: since there is (almost) no loss of information in the acquisition of color concepts, there is nothing further to be made available to the central system for digitalization. Acquisition of sensory concepts is therefore brute and primitive: to acquire these concepts it is enough to occupy the relevant sensory states for an organism equipped with an appropriate conceptual system—i.e., by an information pick-up system operating on the sensory representations. This is why the notion of *learning* is not appropriate for the acquisition of these concepts. Rather, the preferred term for this, for both empiricists and nativists, is “triggering.” So one sense in which the abstraction distance is minimal is that the process underlying the acquisition and vertical deployment of sensory concepts does not involve any loss of information that is nevertheless available to the conceptual system for further digitalization.

Contrast this to the intra-cranial sustaining mechanisms for other concepts, which are (partially but essentially) *cognitive*. The acquisition and deployment of perceptual concepts may be innate and automatic in some sense, but these still involve a digitalization process with considerable loss of information, information that is still available for digitalization. When we

visually recognize shapes of objects or geometrical figures, most of the information about their spatially distributed and organized constituents (illumination gradients, edges, corners, curves, color, etc.) is still consciously available. It isn't that *we* consciously use this information in the acquisition and deployment of such concepts—this is something our perceptual (as opposed to sensory) systems automatically do for us. But what is interesting is that even though this process may be automatic and unconscious, most of the information used in the process (which is then discarded) is available to *us*, to the central cognitive system, and thus is conscious in just that sense. Because of the importance and centrality of perceptual concepts, their acquisition may still be innately determined—i.e., such concepts may be triggered rather than learned. We leave this issue open.

The notion of learning seems most appropriately applied to the acquisition of observational (and for that matter, theoretical) concepts. The sustaining mechanisms for those concepts are heavily cognitive, involving the use and loss of a great amount of information, which is also normally consciously available.<sup>30</sup> Generally, the more cognitive the sustaining mechanisms of a concept are, the greater the abstraction distance between it and the sensory bases from which it is acquired.

Before we move on to examine what makes sensory concepts special, we would like to make a few observations about what is implied by the architecture of the information flow from the sensory to the conceptual. If we are right, then there is a deep point to be made about autonomous representational systems:

- (i) Such systems are nomologically bound to be hooked up to their environments in a way that at some level of abstraction they will *always* harbor sensory representations that represent complex physical properties in their environment as simple or atomic, or rather, do not represent them as having internal complexity.
- (ii) Furthermore: necessarily, if an autonomous intentional organism has concepts at all (or a conceptual system, as opposed to just sensory representations), however primitive or sophisticated, then it has *some* sensory concepts in our sense.

One of the most basic truths about autonomous intentional systems is that they have to interact with their environment informationally. So they have to have information entry mechanisms. These mechanisms cannot deliver every piece of information in analog form, i.e., in a form that is always nested by some further more specific information. There will have to be a cut-off point about the most specific information the mechanism can provide about the environment. If this piece of digital information carries the analog information nested in it in an extractable format, then there will have to be structural features of the output representation carrying the (total) digital information



that nest this information. Then the same question arises about the digital content of *these* features and its format. This process cannot go indefinitely. At some point there will have to be representational features with digital informational content that nests the analog information carried by them in a non-extractable format, at which point the property digitally represented won't be represented as having internal constituents—if the property has internal constituents (this can be a massively disjunctive property like colors). As will become clear as we proceed, it is these necessities that partly create the mystery around phenomenal consciousness.

#### 4. What Makes Sensory Concepts Special

It is not accidental that the distinction we drew between sensory and perceptual concepts is approximately coextensive with the distinction traditionally drawn between concepts of secondary and primary qualities, respectively.<sup>31</sup> Secondary qualities are those which are represented in our experiences in a primitive way: sensory representations carry information about them in a way that makes the information carried about their constituents analog but non-extractable. (It is the job of empirical scientific investigation to reveal the complex nature of secondary qualities, and extract the information about their constituents.) Hence, sensory experiences carry the most specific information about these properties without revealing their internal structure. This is why the abstraction distance between the concepts of secondary qualities and their sensory bases is minimal; equivalently, this is why the acquisition of these concepts is non-cognitive and brute.

Sensory concepts apply, in the first instance, to the objects of perception—to whatever it is that our sensory experiences represent.<sup>32</sup> This is so despite the fact that they are directly and immediately acquired from sensory representations. The flow of information required for their acquisition (and vertical deployment) necessitates the presence of sensory intermediaries that carry information about the properties denoted by these concepts. Indeed, this is one of the main differences between sensory and observational concepts.<sup>33</sup> There is an asymmetry in their acquisition: while sensory concepts are necessarily acquired from the experiences sensorially representing the properties they denote, observational concepts are different. Observational concepts are typically acquired from experiences representing their denotations, but this is not necessary. We can acquire them “horizontally,” i.e., by sensory means (speech perception, seeing pictures, reading books/newspapers, inference, etc.) that are only very indirectly related to, and hence don't carry information about, their denotations.

There is a deep reason for this asymmetry which we haven't touched on so far but will be very important for what follows: the information about the secondary qualities contained in experiences cannot be *completely*

*digitalized* by the conceptual system, whereas the conceptual system can completely digitalize the information contained in experiences about the properties denoted by observational concepts.<sup>34</sup> “Complete digitalization” is a technical term introduced by Dretske that expresses a necessary condition for a piece of information to count as the semantic content of a concept. Recall that the semantic content of a concept is the most specific information its vertical tokenings carry about the objects it applies to, which is equated with its digital informational content. But Dretske eventually refines this definition by requiring that the semantic content be that piece of information which is *completely digitalized*. Here is the definition (1981: 185):

Structure  $S$  has the fact that  $t$  is  $F$  as its semantic content [i.e.,  $S$  is the concept of an  $F$ ] = *df.*

- (a)  $S$  carries the information that  $t$  is  $F$  and
- (b)  $S$  carries no other piece of information,  $r$  is  $G$ , which is such that the information that  $t$  is  $F$  is nested (nominally or analytically) in  $r$ 's being  $G$ .

Condition (b) ensures that if  $S$  carries the information that  $t$  is  $F$ , it does so not by carrying information about any intermediary which nests the information that  $t$  is  $F$ . When the two conditions are satisfied  $S$  carries the information that  $t$  is  $F$  in completely digital form, or equivalently,  $S$  is said to completely digitalize the information that  $t$  is  $F$ , which then becomes  $S$ 's semantic content. More intuitively, the intention is to rule out those cases where concepts carry the most specific distal information about an object by carrying information about their proximal causes, in our case their sensory bases.<sup>35</sup> So, for instance, the concept ROBIN, when acquired from experiences that carry information about robins, should not carry information about the structure of sensory representations that give rise to ROBIN.<sup>36</sup> Since we are working in a naturalistic framework, if concepts carried information about sensory representations from which they are acquired, this information would be information about the instantiation of certain neurophysiological properties (or disjunctive sets of such properties) realizing these sensations. Hence our concepts would be selectively responding to such properties in the first place. And this would imply either that our concepts represent neurophysiological conditions, or that our sensory concepts have dual semantic content, and therefore are systematically ambiguous.

Interestingly, Dretske does not make a point about the empirical impossibility of complete digitalization; nor does he talk about the fact that complete digitalization is routinely violated in the case of sensory concepts.<sup>37</sup> If sensory representations of secondary qualities are realized by a more or less homogeneous set of neurophysiological properties, or by a

finite disjunction of such properties, then vertical tokenings of sensory concepts carry information about their distal causes (instantiations of secondary qualities) by carrying information about the instantiations of these proximal physical properties. Whether or not experiences of such qualities are physically realized in a homogeneous way is ultimately an empirical question, but we think that there is enormous empirical as well as *a priori* evidence that this is the case—certainly intrapersonally, and most probably interpersonally.<sup>38</sup> What is important for our purposes is the claim that the neurophysiological realization bases of sensory representations of such qualities are not indefinitely and arbitrarily varied, but consist of a finite disjunctive set of physical properties, and are more or less homogeneous in just this sense. We think that this claim is true, but we stand ready to be corrected by future empirical evidence.

There are also overwhelmingly strong engineering reasons for this claim: whenever you make an architectural distinction between a sensory buffer and a conceptual system that extracts information about the distal layout from this buffer (and whose behavior is causally sensitive to what this buffer contains), there will be a need to correlate the information carried by concepts and the elements of the buffer in such a way that matches up with the distal layout. If the only way the conceptual system carries information about the distal properties is through a physically realized sensorium, then it had better be the case that the same elements of this sensorium carried the same information, at least in the case of secondary qualities where the abstraction distance is minimal. Otherwise, the informational efforts of the conceptual system will be fooled. From an engineering perspective, it is unclear how such an architectural design can be constructed without making the realization bases of those sensory representations more or less physically homogeneous (i.e., not arbitrarily varied), at least within a single system.<sup>39</sup>

Notice that in the case of observational concepts there is no real problem about complete digitalization. There are indefinitely many ways robins, trucks, etc. can affect our sensory receptors, and thus many ways in which they can be represented in experience. In such cases, the standard information-theoretic remedy is to say that these concepts track their distal causes without tracking proximal sensory representations, since the alternative is to say that they track a massively (probably open-ended) disjunctive proximal property. We believe that the former is indeed more plausible than the latter. But if so, we can now see better why sensory representations carrying information about properties denoted by observational concepts are not necessary for acquiring such concepts. These concepts, though observational, are modality-neutral (amodal), and to that extent not perspectival. But that is not to say that their cognitive sustaining mechanisms don't involve sensory/perceptual channels and concepts; they do. It is to say,

however, that the sustaining mechanisms involved provide information only (mostly) contingently related to the denotation of these concepts.

It is the failure of complete digitalization that makes sensory concepts special by giving them a *perspectival* and *quasi-indexical* character. Their acquisition, semantics, and vertical deployment are essentially host-unique in two senses:

- (i) It matters essentially for whose cognitive system these cognitive structures function as concepts.
- (ii) They track features of the environment (instantiations of secondary qualities however objectively understood) essentially by tracking something *about their host*, namely, the sensory experiences from which they are directly acquired.

In other words, these are concepts which a properly functioning conceptual system *cannot* normally acquire unless suitably hooked up to a properly functioning sensory/perceptual information delivery system *of the same host* that has *actually* delivered the necessary information, i.e., carried information about the properties denoted by these sensory concepts.<sup>40</sup> We also want to emphasize that their acquisition is direct and immediate, by which we mean this: their sustaining/acquisition mechanisms are not cognitive, but primitive and brute—that is, they don't involve the exploitation of consciously available information that is then discarded in the digitalization process. This is roughly to say that the abstraction distance involved in their acquisition is minimal. In the context of our discussion above, this implies that no information about the internal constituents of the properties denoted by sensory concepts is available in an extractable format: they don't represent their denotations as having a complex internal structure. All these points about sensory concepts will be crucially important later on, when we criticize conceivability arguments against physicalism.<sup>41</sup>

### 5. Fixing the Semantic Content of Sensory Concepts

What justifies the claim that, despite the failure of complete digitalization, the semantic content of a sensory concept, say RED, is the secondary quality, *redness*, possessed by the objects of the sensory experiences from which we directly acquire it?<sup>42</sup> Irrespective of what semantic content our theories assign to these concepts, there should be no doubt about what their semantic contents are: they are the qualities that our experiences represent the external objects as having. Our experiences place these qualities in the world of objects external to our bodies. So do our sensory concepts. Given this, the question before us is how to reconcile a Dretskean informational semantics with the failure of complete digitalization. For even if we rightly want to be able to say that RED represents *redness* despite the failure of

complete digitalization, what justifies rejecting the option, which seems to be a consequence of the theory, that the semantic content of RED is the *experience of redness*, i.e., E-red, realized by a certain set of neurophysiological properties?

Here is another way of putting the problem. Informational semantics starts with the information carried by a structure on its way to working out how to determine its *semantic* content (SC). We have seen that Dretske wants to assign the completely digitalized informational content of a concept (*C*) as its semantic content: in other words,

- the semantic content is the most specific information carried by *C* about a source *o* such that there is no separate structure *e* such that *C* carries the most specific information about *o* by carrying the most specific information about *e*.

But this assigns E-red as the semantic content of RED—assuming, as we do, there is no further informational intermediary of the relevant sort between E-red and RED tokenings. The theory gives us the wrong result.

Let us say that the (most specific) *informational* content (IC) of RED, which interests us, can be given by an ordered pair:

$$\text{IC}(\text{RED}) = \langle \text{redness}, \text{E-red} \rangle.$$

The structure of the information flow is such that RED carries information about *redness* by carrying information about E-red.<sup>43</sup> If we want to insist, as we should, that

$$\text{SC}(\text{RED}) = \text{redness}$$

despite this informational alignment, we have to modify the content-assigning mechanisms of a Dretskean informational semantics, and we have to do that in a principled way.

One option, which Dretske himself might be tempted to take, is to invoke teleology: the semantic content of RED is determined by whatever indicator function the tokenings of RED are supposed to serve. It might be plausibly claimed that it is the *redness* of external surfaces that RED has the function of carrying information about, not E-red. Indeed, we think this claim is not difficult to justify on the basis of evolutionary considerations, by appealing to the idea of the adaptiveness of cognitive structures given our practical needs and interests in our environment. In fact, we believe it to be true. But it doesn't solve the problem. For, as Dretske (1986) himself is aware, if we try to determine semantic content in terms of indicator functions, the problem about the indeterminacy of semantic content tends to translate into a problem about the indeterminacy of function. For it is possible to

argue in the following way. The true function of RED is to indicate E-red, but since E-red is perfectly informationally correlated with *redness*, any need or interest that the organism might have related to *redness* will be satisfied by a structure whose job it is to indicate E-red. In other words, we can equally well claim that RED has the function of indicating *redness* in virtue of having the function to indicate E-red. After all, when, in abnormal circumstances, E-red fails to correlate with *redness*, RED's functioning is not to blame; it does its job just fine, it is the world that doesn't cooperate—or so the intuition goes. Dretske's solution to this problem in his (1986) is not applicable to sensory concepts since his proposal acknowledges that the problem of indeterminacy of function can be solved only for those concepts that can completely digitalize the most specific information they receive about the things they are supposed to denote, i.e., only for those concepts whose abstraction distance is great enough to allow them to be acquired from an indefinitely large set of proper sensory bases—and sensory concepts are not among these.

So what anchors the semantic content of RED to *redness*? A crucial part of the answer, we believe, can be gleaned by reflecting on the integration of the information coming from a variety of intra- and inter-perceptual sources. Consider the visual information that our cognitive system uses in the acquisition and vertical deployment of observational concepts, like ROBIN, CAR, and TOMATO. We have said that there is no serious problem about the complete digitalization of information with respect to these concepts. So we can safely claim that these concepts apply to external objects: they are much further away in abstraction space from the sensory experiences that give rise to them. But the acquisition and vertical deployment of these concepts utilize lots of sensory and perceptual information that is also consciously available, which is to say, apt for digitalization by the same central cognitive system. Between the sensory experiences and the vertical tokening of ROBIN, in other words, a lot of more specific information is lost. Even though this process may be inferential/computational, it seems mostly automatic and unconscious. Nevertheless, much of the information used in the process is consciously available, such as the determinate size, shape (even the particular lines, curves, edges, etc.), texture (even the smaller changes in light intensities), orientation, distance, and the varying grades of color of the robin that has occasioned the tokenings of ROBIN.

When we say this information is lost or discarded, we don't mean that it is forever hidden from consciousness; rather, it is lost from the perspective of the tokening of ROBIN, which is to say that the tokening of ROBIN no longer carries information about these more specific values along dimensions just mentioned. It is the loss of this sort of more specific information that enables us to visually recognize this object as a robin, instead of, say, a small-grey-robin-to-my-left, etc. But this information is also the same information used and integrated in the recognitional process and is nevertheless

consciously available, the conceptualization of some of which has a shorter abstraction distance (and, in the case of color concepts, a minimal one). As the abstraction distance gets closer and closer to the sensory experiences, concepts start to lose their completely digitalized character. Now here is the crucial point: if the conceptual system uses and integrates more specific information about external things (e.g., about their determinate color, variations in light intensities, edge here, curve there, etc.) delivered by sensory representations *on its way* to categorizing these things as *external* objects (as robins, cars, tomatoes), then the conceptualization of the former kind of information had better anchor their semantic content outside the organism; otherwise the conceptual system will not present a coherent picture of my immediate environment. It makes no sense—theoretically or biologically—to anchor the semantic content of RED to E-red if we are able to visually classify the thing in front of ourselves *as a red tomato*, especially when E-red involved in the sensory intermediary *actually* delivers the information about the *redness* of the tomato to the central cognitive system. RED *actually* carries this information (by carrying information about E-red), and, precisely because of this, enables us to recognize the object as a red tomato. Indeed, otherwise what would be the semantic content of RED TOMATO? The mind boggles.

Consider the 2D geometrical figure detector introduced before. Logically speaking, we could be like such a device with respect to the recognition of middle-size objects such as tomatoes, robins, and cars. For, if we were like that, the information leading up to our recognition of these objects would not be consciously available to us. This means that we could not acquire any concepts usable in the discrimination of more specific information about these objects—even though our pre-perceptual system might actually make use of this very information in the recognition process, denied to the central system for purposes of conceptualization.<sup>44</sup> If we were like this, we would have the same problem about how to anchor the semantic content of concepts like ROBIN, CAR, TOMATO—these would be sensory concepts that would not represent the systematic distal and proximal causes about which they carry information as having complex internal structures.

Briefly, the partial answer to our original question, then, comes down to the need for *coherent* integration of information. It is the pressure exerted by our practical interests in having a *coherent* global representation of our external environment that forces the conceptual system to pick out the first element, *redness*, in  $\text{IC}(\text{RED}) = \langle \text{redness}, \text{E-red} \rangle$  as the semantic content of RED. We have seen how the processes integrating various sorts of information in the acquisition and vertical deployment of observational concepts generate a need for coherence. We have also seen how evolutionary forces determine where to put this coherence: on the global representation of a reality *external* to one's sensory and cognitive systems. To point this out is, of course, not to specify the mechanisms by which this is accomplished.



But perhaps this latter task is more appropriate for psychologists or neuroscientists to tackle.<sup>45</sup>

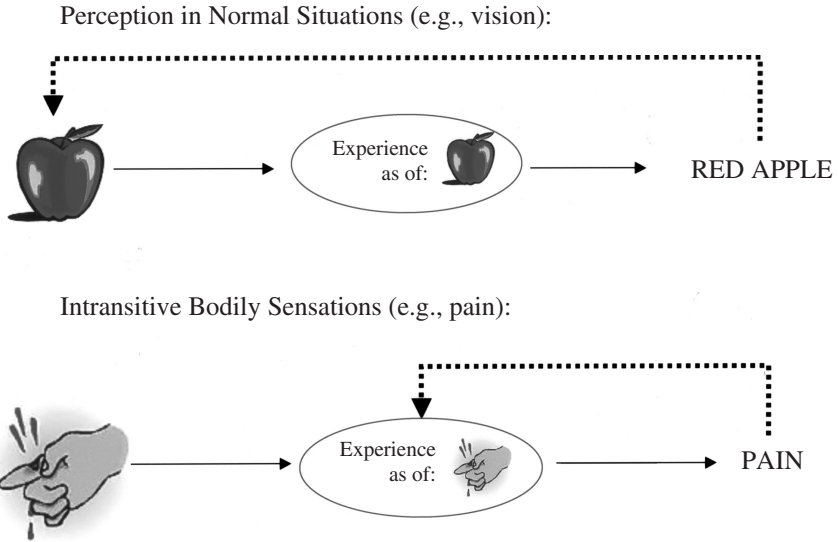
## 6. Concepts of Bodily Sensations

This partial account predicts that the less need there is for coherent integration—say, because of less information, the scarcity of its sources, or a redirection of immediate interest due to the affective/hedonic value of the experienced stimuli, etc.—the less pressure there is to anchor the semantic content of a sensory concept to the outside. We indeed find the gradual change implied by this in all sensory modalities and submodalities. Vision is the paradigm source of information generating sensory concepts whose semantic content is unequivocally external to the subjects. Things get increasingly less clear as we look at other modalities (hearing, smell, taste, and touch) according to how close to the body the detected properties are, how rich the information provided by the experiences in these modalities is, how much information from other channels is used, how impoverished the quality space determined by sensory experiences is, or how many quality spaces each (sub)modality determines along its different dimensions. The limit in this direction are the interoceptive submodalities of touch producing so-called intransitive bodily sensations, such as pains, itches, tickles, and the like.

If we assume that these sensory experiences carry information about, and thus have come to represent, bodily conditions such as tissue damage, then one would naturally expect the same, or at least a very similar, informational division of labor that we find in exteroception, say, vision. We would expect that the job of these bodily sensations is to deliver information about bodily conditions to the central cognitive system for digitalization resulting in the acquisition of sensory concepts which apply in the first instance to aspects of these bodily conditions. But this is not what we find. The sensory concepts PAIN, ITCH, and TICKLE apply to token experiences, to bodily sensations—well ... to pains, itches, and tickles—not to the bodily conditions these sensations represent. The result is a curious asymmetry between sensory concepts (like RED) and concepts of bodily sensations. Despite identical information flow, sensory concepts acquired directly and immediately from the relevant experiences apply to different states (see **FIGURE 1**).<sup>46</sup>

Let us assume that the most specific informational content of PAIN, which interests us now, can be given by an ordered pair consisting of a certain sort of tissue damage and a sensory representation thereof, call it E-damage, more or less homogeneously realized by physical properties in the brain and delivering this information to the conceptual system. So,

$$\text{IC(PAIN)} = \langle \text{tissue damage, E-damage} \rangle.$$



**FIGURE 1.** *Asymmetry in concept application despite identical information flow.*

PAIN carries information about tissue damage by carrying information about E-damage. So PAIN doesn't completely digitalize the most specific information it gets about the tissue damage. In this case, as a matter of fact, the semantic content of PAIN is anchored to E-damage:

$$\text{SC(PAIN)} = \text{E-damage.}$$

Why is this different from the case of RED?

If our partial answer to the parallel question above is right, then we can discern one part of the reason. It relates to the nature of the perceptual object, i.e., the object sensorially represented. Tissue damage is, of course, only one of many causes of pain experiences—we have used it as a stand-in for whatever it is that specific sorts of pains represent. These are mostly internal conditions of the body, normally not open to other sensory channels. So inter-modal sensory information for integration is either non-existent or extremely limited. Not only that; the quality space generated by noxious stimuli is quite impoverished compared to exteroception, especially vision. Although pain experiences sort out the noxious stimuli both temporally and according to a spatially articulated somatosensory map, there is not much information *integration* going on in a way so as to epistemically clue the conceptual system in on what it is that is being perceived. There is certainly information available in the pain experience to sort out different kinds of bodily disorders or damage. But again, this does not help to generate concepts whose abstraction distance is sufficient for complete

digitalization. On the contrary, it appears that the quality space created by pain experiences gives rise to a corresponding set of sensory concepts whose abstraction distance is minimal. There is certainly a lot of analog information about bodily conditions contained in pain experiences, but clearly it is in a form not extractable by the conceptual system so as to generate concepts with greater abstraction distance, which, as we have seen, was necessary for concepts with denotations external to experiences. In other words, pain experiences are informationally impoverished, in that the conceptual system cannot digitalize concepts with sufficient abstraction distance from them with their semantic content focused outside. Pain experiences don't represent bodily conditions about which they carry information as composed of complex properties.<sup>47</sup>

As we have seen, however, the ability to generate concepts with sufficient abstraction distance such that complete digitalization could be obtained, which results in putting the semantic content of these more abstract concepts outside, was the key to the *integrative* processes which resulted in the need for a coherent picture of an external reality. For it is the more specific information about the qualities of the external denotations of these more abstract concepts that is being used in their acquisition. So it is imperative for the purposes of (re)presenting a coherent reality that if this more specific information is also available for conceptualization it be attributed to the very same external objects denoted by more abstract concepts. We don't have sufficiently rich information to yield completely digitalized concepts in the case of bodily sensations: hence, the sensory concepts they give rise to apply to their proximal causes.

There is probably another reason of why E-damage is picked out as the semantic content of PAIN: whether or not E-damage represents tissue damage: it hurts! If pain experiences hurt irrespective of whether they are veridical, then it is not surprising that our immediate epistemic and practical focus is directed, in the first place, onto the experience itself.<sup>48</sup> The affective or hedonic tone of the experience puts a heavy demand on the cognitive centers to urgently re-allocate cognitive and behavioral resources and response priorities for stopping the experience by doing whatever is necessary to get rid of its cause. Again, what we see here is the cognitive adaptations of the conceptual system to align its semantics with the needs of the organism.

Note that there is some room—albeit a small amount—in the folk conception for thinking of pains as pure bodily conditions. There are situations where we find it natural to talk in a way that there may be unfelt pains: when we do that, we are talking of pains as disordered states of our bodies gone unnoticed. We talk about our headaches lasting, say, during a heated discussion even though we have not felt them most of the time. When we talk about the very same pain coming back, we find it natural to conceive of it as if it had already been there, unnoticed and unfelt (in fact, some

cognitive-behavioral therapies for chronic pain utilize this phenomenon). This is certainly not the dominant conception of pain. But that such use has kept a foothold in the folk usage is noteworthy, since it is precisely what our account of sensory concepts predicts. The information about the bodily condition is there: the vertical tokening of PAIN carries it. Our ambivalence about what to say in such rare cases when pressed is generated by this double informational content, making it somehow possible to alter the semantic focus depending on the context (see below).

A proper understanding of bodily sensations and their conceptualization is crucially important for a proper account of the informational architecture of the cognitive mind and the special role sensory concepts play at the interface between sensory and conceptual systems. Notice that “pain perception” is, technically speaking, a form of introspection—if introspection is the means by which we learn about our own mental states from a first-person perspective. Our first-person knowledge of our pains, itches, and tickles is knowledge of our experiences. To know we have them is to know we have experiences. And to come to know that is to engage in introspection.

A proper understanding of bodily sensations is crucial because here we see the basic mechanisms of introspective access to our own experiences in their barest form, being located at one extreme of the spectrum of sensory representations. It is by working from this extreme that we will develop an account of phenomenal introspection in general.

### 7. Information Pick-up in the Introspection of Bodily Sensations

Knowledge, including introspective knowledge, requires discriminative and recognitional capacities, which are conceptual capacities, as we explained previously. This meshes well with the case of sensory concepts of bodily sensations. When PAIN is vertically deployed, i.e., when it is used as a classificatory response to noxious stimuli and their effects on the body in a way that carries information about them by carrying information about the corresponding experience, E-damage, the *semantic content* of the *de re* judgment made is that [the E-damage to which PAIN is informationally connected] is pain.<sup>49</sup> But this judgment, when made from a first-person perspective, is an introspective judgment, i.e., a judgment about a token experience, a mental event. Insofar as the concept of pain is, intuitively, a mental concept, the judgment classifies a certain neurophysiological event in the brain carrying information about tissue damage under a mental concept.

But what exactly is the informational value of this judgment? What information does it convey? We have been talking about sensory concepts carrying information about the sensory representations from which they are acquired. But even if the kind of weak type-type identity theory we assume for sensory experiences of secondary qualities is true, what generates the information? When there is information in a signal about a source, what

makes this possible is the elimination of alternative possibilities at the source that could have occurred, and the nomological dependency of the signal on these. If we want to talk about sensory concepts carrying information about experiences, we have to treat experiences as information-generating sources *on their own*—even when much of the information thus generated at the sensory level nomologically depends on the elimination of possibilities at a source beyond them, i.e., in the world.

This is precisely what we find when we look at experiences from the point of view of *sensory concepts*. Experiences are venues for information entry to the central conceptual system. There is as much information generated at the sensory level for pick-up by the conceptual system as there are different venues (sensory modalities and submodalities), distinct dimensions within these venues (pitch, frequency, amplitude; color, geometry, light intensity, etc.), and different (usually continuous) values each of these dimensions can take (red, orange, yellow, etc.; loud, very loud, even louder, etc.). Not only can we discriminate reds from oranges, oranges from yellows, but we can also discriminate a color from the spatial expanse of which it is the color, as well as discriminate visual experiences from tactile, auditory, gustatory, and olfactory ones.

It is this multiplicity of information entry that allows us to treat sensory experiences and its parameters as information-generating sources. Of course, if there is information, there is no logical guarantee that there will be something carrying this information; but our conceptual system has evolved to pick up this information and use it in the service of guiding behavior. When we token RED in response to a ripe tomato, our concept does carry information about which neurophysiological property<sup>50</sup> is instantiated in the relevant part of our visual cortex. Although the information carried by RED does not represent this property as having a complex structure, the tokening of RED does eliminate other possibilities relevant for the color dimension of our visual experience: any number of neurophysiological states realizing different color experiences could have obtained.

When we make similarity judgments like “*x*’s color is more like *y*’s than *z*’s” we are making similarity judgments about the colors of objects on the basis of a similarity ordering of our color experiences. Indeed, if we look at what color science seems to tell us, because of metamerism phenomena, most of the time we find no similarity in the particular spectral reflectances paralleling the similarities our experiences represent. It is the sensorially represented similarities that the conceptual system picks up in making these similarity judgments.<sup>51</sup> Here it is useful to appeal to a quality space generated by how the experiences represent their objects. The inferential regimen governing our sensory conceptual repertoire reflects or parallels the relational structure of this quality space. But this space must be such that our conceptual system is able to pick it up from the similarity relations among color sensations. And the only way of doing that we can imagine from an

engineering perspective is a structuring and ordering of the physico-functional properties of the relevant brain states in a certain way, i.e., so that it will act as an information-generator of the right sort—the sort that enables the conceptual system to end up with the particular set of sensory concepts and particular inferential structure that it actually has.

It isn't just the different values of a dimension of an experience that generate the kind of information exploited by the conceptual system. As we have mentioned, the conceptual system is also sensitive to variations along intramodal dimensions, as well as activations of the different modalities themselves. There should be no controversy about this: the information is there to be picked up, and our conceptual system does register it.<sup>52</sup>

So what is the informational value of the *de re* judgment “[the E-damage to which PAIN is informationally connected] is pain”? It consists of whatever other possibilities are eliminated by the instantiation of the relevant neurophysiological property constituting E-damage. It is not only *this* kind of pain we are having (different from *this* and *that* . . . ), but also it is *pain* we are having, not an itch or tickle, or a sensation of warmth, or a sound for that matter. It may be that not all this information is being carried by PAIN. When we discriminate stinging pains from pricking ones, we seem to be deploying more specific sensory concepts (with a shorter abstraction distance) than just classifying our sensation as pain, eliminating only, say, the possibility of its being a tickle or a sensation of mild warmth, and so forth. The general point, however, should be clear: the content of such judgments is determined strictly according to information-theoretic principles, that is, elimination of the relevant alternative possibilities.

What needs to be emphasized here—and what may be obscuring this naturalistic picture—is that the way PAIN indicates or represents this neurophysiological property reflects the way E-damage (i.e., pain) indicates or represents the tissue damage. E-damage carries analog information about the tissue damage in a non-extractable format so that this information is not available to the conceptual system for further digitalization. That is to say, whatever most specific information E-damage is sensorially carrying about the tissue damage, none (or, very little) of the information nested in it is available to the conceptual system for extraction. But this amounts to the fact that E-damage does not indicate or represent the tissue damage as having a complex structure. The conceptualization of this information by the central system reflects this condition: PAIN<sup>53</sup> does not represent the neurophysiological property it detects as having a complex internal structure.

But a vertical tokening of PAIN does carry information about what other possibilities are eliminated. When we apply PAIN vertically we don't represent the property to which it is actually applied as physical—how could something be physical if it doesn't seem to have *any* internal complexity to it? But, we don't represent it as non-physical or immaterial

either—how could we locate something non-physical in the body? PAIN is basically topic-neutral on this issue, as J. J. C. Smart (1962) insightfully pointed out a long time ago (but for different reasons). There is nothing peculiar or mysterious about any of this, if we keep information-theoretic principles firmly in mind. Simple signals can carry information about quite complex properties without making this information, on their own, available for further extraction. This is precisely what happens with our sensory experiences and sensory concepts.

### 8. Introspection and Phenomenal Concepts

As we have seen, “pain perception” is, technically speaking, a form of introspection. It does, however, share its information-flow structure with other forms of sensory processing and concepts. How is it, then, that one counts as introspection while the perceptual categorizations accomplished with exteroceptive sensory concepts do not? The latter are, in the first instance, perceptions (i.e., vertical categorizations of external stimuli under concepts) of aspects of an external reality; they are not perceptions of brain states realizing the sensory representations mediating the information flow—although they carry information about these states too. We expressed this asymmetry earlier by saying that although both kinds of sensory concepts have dual informational content, their semantic contents are differently anchored or focused. We have explained why this should be expected given the immediate informational and practical needs of organisms like us shaping our selectional history, and the way the sensory information is integrated at different levels of the abstraction process.

How do we introspect our exteroceptive experiences generally? How do we come to know what it is like to see red, to hear middle C played on a clarinet, to smell sulfur dioxide, to taste dark corn syrup, to feel a warm and soft object touching one’s cheek? As should be clear by now, we propose that introspection of such sensory states involves a different utilization of the very same sensory concepts deployed in the perception of the external properties that these sensory experiences represent. Introspection of exteroceptive experiences is the mechanism or capacity that capitalizes on the second element of the *information* content of a sensory concept by selecting it as the *semantic* content of the concept. Clearly, as we have seen in the case of concepts of bodily sensations, our cognitive system is capable of doing this: *they are the existence proof for such a capacity*—this is why we have spent so much time examining bodily sensations and their concepts. But we do seem to have the introspective capacity to pick up the information generated intra-personally by the multitude of information entry channels and make it the semantic focus of our sensory concepts.

Vertical tokenings of RED carry information about experiences of red. But sensory concepts carry information about brain states without carrying



information about their constituent structure in an extractable format, just as sensory experiences of secondary qualities carry information about, say, colors, even though the information about colors' constituents is not extractable by the conceptual system. In this, vertical tokenings of sensory concepts like RED not only discriminate and classify the relevant range of brain states as simple primitives with respect to their intrinsic nature but also show them as having external relations to other such states paralleling the sensory quality space represented by them. More precisely, sensory representations of secondary qualities carry the most specific information about them by carrying the analog information nested in this information in a non-extractable form. Similarly for vertical tokenings of sensory concepts: they carry the most specific information about the sensory representations of secondary qualities without carrying the analog information nested in it in an extractable form. Introspection is precisely that mechanism which takes the second element in the *information* content of sensory concepts and makes it their *semantic* content.

Here we need to introduce a further distinction for types of concepts, following the dictum "different denotations yield different concepts." Instead of talking of the *sensory concept* of RED (call it s-RED), whose semantic content is the property *redness*, being utilized in a different way, we can talk about the *phenomenal concept* of RED (call it p-RED), whose semantic content is the *experience of redness*.<sup>54</sup> As we have seen, concepts of bodily sensations like PAIN are already phenomenal concepts in this sense: they apply to token experiences.

But how does the semantic switch or shift occur? To answer this question, let us start by noting that the source of the phenomenal concept, p-RED, that introspection utilizes is the very same structure underlying the sensory concept, s-RED, which the perceptual categorization of distal stimuli deploys. But sensory concepts are not simply the *etiological* source of phenomenal concepts. In fact, sensory concepts become phenomenal concepts when the former are used to specify what the experiences they carry information about are experiences of. Thus, sensory concepts are also the *epistemic source* of phenomenal concepts. The significance of this can be captured by:

- (ES) When p-RED is applied to experiences of *red*, it is impossible not to categorize the experiences, by this very application, as the *epistemic source* of the perceptual judgment/categorization of a distal stimulus; but this is just to categorize these experiences as representations of *redness* of a certain kind, i.e., of the kind *this* [sort of brain state] subserves—if and when we have the necessary intentional concepts (see below).

The truth of (ES) is the source of the familiar claim that introspection is “transparent” (at least in exteroceptive modalities): i.e., the reason why the properties we encounter when we introspect our experiences seem all to be the properties that our experiences *detect* rather than *exhibit*, is that introspection uses the same sensory concepts in a different way—by choosing the second element in their *information* content as the *semantic* content of the very *same* concept used to classify *what* in fact is represented by the experience. In other words, the only conceptual resources that we have in our disposal to conceive of our experiences and their qualities in introspection are the very same ones available to us in conceiving what our experiences present.

The extent to which this capacity or mechanism to shift the semantic focus of sensory concepts like RED is innately given, and the extent to which it depends on ontogenetic maturation processes or cognitive development, is an open question. We think the evidence from developmental psychology indicates that this capacity is acquired only after the acquisition of intentional concepts and a modicum of folk psychology. Children seem to acquire these concepts and this mastery fairly early—approximately between the ages of three and four. We believe it is no accident that the acquisition of the capacity to introspect one’s experiences emerges only after this development.

The received view in developmental psychology about young children’s introspective capacities is sometimes called the “Theory Theory” (TT) of self-awareness, according to which introspective knowledge is obtained via the same mechanisms that underlie our ability to attribute mental states to others and to reason about them.<sup>55</sup> According to the received view, this latter capacity is accomplished by an internalized theory of mind, a folk psychology. As Nichols and Stich (forthcoming) note, the TT account is not a fully developed account and remains heavily underdescribed, and it is not clear to what extent any attempt to flesh out the account more fully could seriously retain the idea that one’s psychological state self-attributions rely on *exactly* the same capacities as those involved in detecting and attributing mental states to others. However, we take the importance of the received view to lie in its evidence base. The empirical evidence leaves very little doubt that young children’s ability to make introspective judgments goes hand in hand with their ability to understand others in mentalistic terms.<sup>56</sup> At a minimum, there is strong evidence that having a rudimentary understanding of basic intentional idioms is necessary for young children’s ability to introspect.

Intentional concepts, such as the concepts of information, representation, belief, and so forth, are acquired through third-person channels, not from one’s own case.<sup>57</sup> Indeed, information theory does not allow for their first-person acquisition. The same is true for the concept of experience *qua* representation. However, even though intentional concepts cannot be acquired from a first-person perspective, once acquired, they can be

vertically deployed, which is to say that we can apply them to our own experiences *because* they are experiences.<sup>58</sup>

The exact way in which the acquisition of intentional concepts facilitates the acquisition of introspective capacities (and vice versa?) needs to be worked out in further detail, but we think that this is more or less an empirical job to be left to psychologists and neuroscientists. Our main point is that at some stage in cognitive development we acquire the capacity to selectively focus the semantic content of our exteroceptive sensory concepts (we already pointed out the existence proof of its feasibility), and the acquisition of this capacity draws upon the acquisition of intentional concepts along with a rudimentary understanding of folk psychology.

Interestingly, that we need to possess intentional concepts in order to introspect our exteroceptive experiences and what they are like is the reason why we don't *normally* think of pain perception as a form of introspection, since when we "perceive" our pains, what we "perceive" is a token experience that is *not*, indeed, *need not* be conceived of in representational terms. The concept of pain here already has the token experience as its semantic content as a simple/primitive representation of a certain kind of brain state located within a quality space. No wonder pains have always been thought as paradigm cases of mental objects that don't themselves seem representational at all: we didn't have and didn't need intentional concepts to "perceive" or come to know about them. Indeed, young children, as every parent knows, can think about and communicate their pains even before they have acquired the intentional apparatus of folk psychology.

So our proposal is that when we vertically apply p-RED to our experience of *redness*: the semantic content of the introspective *de re* judgment involved is something like:

- This is how *redness* is [registered]  
(or, experienced, sensorially represented, etc.),

where 'this' picks out a certain brain state primitively (only eliminating the relevant alternative possibilities and thus locating it within a relationally defined quality space—so it is predicative, not just purely indexical), i.e., without revealing its constituent structure. More accurately, we could have expressed it as "p-RED is how *redness* is [registered]," except that 'p-RED' is not English.

One important consequence of this is that we now have a purely naturalistic (partial) explanation of the much-debated "reflexive" and "self-intimating" character of sensory states. As (ES) points out, in the very perceptual recognition of *redness* we also cognize the sensory experience mediating the recognition, and vice versa. The sensory concept RED is necessary for generating a cognitive structure, p-RED, with the semantic

content displayed above, since p-RED is the very same structure as s-RED only used differently *because* it carries information about both the sensory experience of red and *redness*. The “reflexive” and “self-intimating” character of sensory experiences stems in effect not from the experiences themselves but from the dual informational and semantic nature of the sensory concepts directly acquired from them. Put differently, and to relate the point to (ES) that we highlighted above:

(ES’) It is the very “same concept” that is used both in picking out the relevant brain state—thus, eliminating the relevant alternative possibilities and hence locating it within a relationally defined quality space—and simultaneously “commenting” on it as a [sensory registration] of *redness*.

Of course, “sensory registration”<sup>59</sup> is the intentional concept involved in the semantic shift. We conceive of the nature of this job in such a way that the intentional concept at issue can be quite rudimentary and basic—to the extent that the acquisition of folk psychology permits it in its earliest phases. Obviously, if (ES’) is true, there is a curious sense in which it is as if the same cognitive structure were used twice over simultaneously (as p-RED applying to E-red, and as s-RED applying to *redness*) in the introspective judgment about a red experience.

We don’t know any other naturalistic account that integrates so tightly the vehicle of introspection with the vehicle expressing what the introspected state represents without giving up representationalism. We have already explained how closely these conceptual vehicles are informationally related to the target of the introspection (i.e., E-red) and to what it represents (i.e., *redness*). In fact, the so-called transparency of introspection that externalists emphasize so much is simply the other side of the “same coin”: it naturally falls out of our account because of this tight integration. At the same time, as we hope to have shown, this account does justice to internalist intuitions, which we find important.

Another important aspect of this sort of introspective vertical processing is its sensitivity to the temporal window, or duration, of the activation of perceptual channels and its particular values. This is probably one of the major intuitions behind the tradition (found in Locke, Kant, and more recently Armstrong and Lycan) that regards introspection as a sort of internal sensing or monitoring—introspection as *inner sense*. But again it is worth emphasizing that this monitoring eventuates in discrimination and conceptual categorization in the way we have explained. This feature of our account makes it a synthesis of otherwise quite opposite accounts of introspection: introspection as internal monitoring (Armstrong 1968, Lycan 1996), and introspection as higher-order

thought à la Rosenthal (1995, 1997, 2001), Dretske (1995), and Tye (1995).<sup>60</sup>

### 9. Defeating Conceivability Arguments Against Physicalism

Following Dretske's seminal work (1981), we have provided an information-theoretic account of sensory concepts. We have explained how they are directly and immediately acquired from sensory experiences and how they are vertically deployed. We have also provided an account of how experiences and their qualities are introspected through the deployment of phenomenal concepts. On our account, phenomenal concepts are acquired from sensory concepts through the acquisition of intentional idioms and the rudiments of folk psychology. We are now in a position to address some long-standing vexing philosophical problems. We will show how to reconstruct so-called conceivability arguments against physicalism from within the account of sensory and phenomenal concepts we have provided. Once we do that, the proper physicalist response will be self-evident.

Let us start with what has been pointed out thus far. The acquisition of sensory concepts from their sensory bases is not mediated by any consciously available more specific information: the sustaining mechanisms for these concepts are non-cognitive. They can also be vertically applied as such, without cognitive mediation. Notice that none of this implies that sensory concepts don't have conceptual or functional roles. They do—as we have seen when discussing the qualitative space generated by the multitude of their sensory bases and their similarity comparisons. Sensory concepts acquired from a given sensory quality space reflect a rich set of conceptual interrelations. But we have also seen that these conceptual relations reflect the external relations of each quality to others whose conceptualizations they are. They don't say anything (in an extractable form) about the internal nature of the secondary qualities represented by these sensations. The main point, however, is that even though sensory concepts have conceptual/functional roles in this sense, *they are not part of the sustaining mechanisms mediating their acquisition and vertical deployment*. Their semantics is fixed independently of such roles, by a direct and immediate informational link to sensory experiences. To say that this link or sustaining mechanism is direct and immediate is to say that it is non-cognitive, which is to say that these sensory concepts digitalize the most specific information carried by sensory experiences about the relevant values of a secondary quality they apply to. And this is to say that they have an abstraction/digitalization distance that is minimal.

It follows from this that our sensory concepts can pick out the qualities they denote directly and immediately, and that they are independent of any other concepts in this sense. In particular, they are independent of any physical or functional concepts, and therefore not only cannot be defined

in terms of them but also no such concepts are involved even in fixing their reference: i.e., none of them is involved in the sustaining mechanisms that determine their semantics. We have also shown that no concepts except sensory ones work this way—all others involve *cognitive* sustaining mechanisms—and that the rationale for this is a nomologically necessary fact about autonomous intentional creatures like us. This is true even for what we have called perceptual concepts. This means that sensory concepts cannot be derived from any other concepts or theories couched in them. As long as the introspection of sensory states requires redeployment of sensory concepts as phenomenal concepts, the same will be true of phenomenal concepts.

This is why, on our view, the primary and secondary intensions of phenomenal concepts coincide or collapse into one on the two-dimensional semantics found in Chalmers (1996). Note that the primary and secondary intensions of sensory concepts don't coincide despite the fact that the intracranial sustaining mechanisms for these concepts are exactly the same as those for phenomenal concepts: both are non-cognitive and brute. This is because there is still an appearance/reality distinction for sensory concepts: the canonical sensory evidence for their application is still distinct from the properties they apply to, hence they can come apart. But the canonical evidence for phenomenal concepts is provided by the very experiential qualities that they apply to—they cannot come apart.<sup>61</sup> This in turn means that, insofar as conceivability is a matter of concept use, it will be possible to genuinely conceive a zombie replica of a person, a creature with exactly the same physical/functional organization as the person's, but who lacks sensory experiences with conscious phenomenal qualities.

Grant us for a moment that the Naturalistic Story (NS) we have told so far about phenomenal concepts and introspection is more or less true of us, and that when we conceive of counterfactual situations or design thought-experiments we do so by deploying our concepts. In particular we deploy our phenomenal concepts in constructing zombie scenarios in the context of conceivability arguments against physicalism. Then it is easy to see that we cannot derive, *a priori* or otherwise, the existence of phenomenal consciousness from a complete physicalistic description of our world (augmented by indexical facts, and given complete semantic competence with the relevant concepts including phenomenal concepts—plus, a “that's all” close for the completion of physical description—see Chalmers and Jackson, 2001). Since the semantics of phenomenal concepts, unlike all other concepts, is fixed non-cognitively in the way we have explained, there is not only no semantic or conceptual analysis of phenomenal concepts but also no reference-fixing evidence base related to these concepts apart from the phenomenal properties themselves. This straightforwardly implies that there is no derivation from a complete physical description of the world augmented by other premises just mentioned—call this the *physicalist premise base*—to any

facts described in terms of phenomenal concepts. This is just to say that since phenomenal consciousness is not entailed by the physicalist premise base, there is no formal contradiction in entertaining the conjunction of this base and statements positively deploying phenomenal concepts. This in turn makes zombie scenarios genuinely conceivable.

However, if you have granted our Naturalistic Story about the phenomenal concepts and followed our reasoning about how it yields the conceivability of zombies, you should not be tempted to draw the metaphysical conclusion that zombies are therefore metaphysically possible and that phenomenal properties do not metaphysically supervene on physical facts. Drawing this metaphysical conclusion would be fallacious precisely because we have constructed this situation from the very beginning purely from physicalist ingredients in such a way that phenomenal concepts denote physical properties but are not derivable from the physicalist premise base—this is guaranteed by the semantics of these concepts.

Even if you don't grant the claim that the Naturalistic Story is *true of us*, you should be able to concede the possibility of a world populated with entirely physical creatures of which something like NS is true. These creatures will no doubt resemble us in many ways in their informational efforts, but whether or not these creatures will have experiences with genuine phenomenal qualities as we do, they certainly will have informational states that are functionally like our experiences, and they will think of their own "experiences" in much the same way we do. At a minimum, they will find themselves in exactly the same philosophical bind that we do: they will claim to have "experiences" that don't seem to metaphysically supervene on their metaphysical make-up. In other words, they will be able to genuinely conceive physical duplicates of themselves without any "experiences." For they won't be able to derive the existence of "phenomenal" properties conceived through *their* "phenomenal" concepts from the corresponding physicalist premise base true of *their* world.<sup>62</sup>

We take these considerations to show that a crucial premise in conceivability arguments against physicalism cannot be *unqualifiedly* true. We call this premise the Bridging Premise (BP), since if it is true it bridges the gulf between the epistemic and the metaphysical, allowing the anti-physicalist to draw the anti-physicalist metaphysical conclusion from epistemic premises.

(BP) For any proposition *P*, if *P* is genuinely conceivable, then *P* is metaphysically possible.

If the Naturalistic Story we have told about phenomenal concepts is true, then (BP) cannot be read as an exceptionless *logical entailment* claim. At a minimum, the possibility of the kind of scenario we have provided should give pause to anyone who is prepared to embrace (BP) in this strong form. We claim that those who would like to read (BP) as an exceptionless logical



entailment claim must hold that the kind of naturalistic story we have told does not have the consequences we claim it does. In other words, they must deny that our story yields a situation where phenomenal claims cannot be derived *a priori* from a physicalist premise base all the while the truth makers for these phenomenal claims are entirely physical and already captured in the physicalist premise base in non-phenomenal terms. We think that any such attempt will be futile.<sup>63</sup>

But even independently of our argument, such a strong reading of (BP) must be suspect. *Prima facie*, conceivability seems purely a matter of epistemology or psychology, i.e., the capacity of cognitive organisms to represent reality one way or the other. As such, any reflection on what is conceivable and what is not shouldn't have a direct *logical bearing* on the constitution of metaphysical reality (necessity/possibility). To think otherwise is to risk, in our opinion, an unacceptable form of verificationism. So we reject (BP) on this strong reading.<sup>64</sup> But we don't think this rejection should be too controversial, given our naturalistic story about phenomenal concepts.

We use "genuinely conceivable" in a slightly technical sense that we should clarify. Consider the standard way in which the apparent conceivability of H<sub>2</sub>O without water (or vice versa) is explained away. It consists in showing that the conception of such a situation is only apparent—this is why Kripke claims that there is an *air of apparent contingency* in scientific identity statements. This is done by showing that what the conceiver actually conceives is not the situation expressed by the statement

(a) water  $\neq$  H<sub>2</sub>O,

but rather, one expressed by

(b) the watery stuff  $\neq$  H<sub>2</sub>O,

where 'the watery stuff' is a definite description contingently picking out a substance on the basis of the superficial qualities we normally use to identify water (or fix the reference of 'water'). It is in this sense that we would like to claim that (a) is only apparently conceivable. This sense requires the availability of a (commonsense) description/conception associated or connoted by 'water'/WATER that contingently picks out the same substance denoted by the scientific term/concept. If no such description or conception is available, we will say that the statement in question expresses a situation that is genuinely conceivable.<sup>65</sup> The association or connotation relation we have in mind does not require semantic or conceptual connections,<sup>66</sup> but requires the use of those very concepts that are involved in fixing the reference, or as we prefer to put it, the use of information (whether or not actually digitalized) supplied by experiences and used in the sustaining mechanisms for concepts. Notice that the concepts of most superficial

properties cited in the description ‘watery stuff’ are part of the sustaining mechanisms for WATER. The most important point for our purposes is that in the case of non-sensory concepts, there is always consciously available information that mediates their acquisition and vertical application, and most of this information is contingently related to the items denoted by these concepts. It is this information in the sustaining mechanisms that is available for further conceptualization, and thus makes it possible for the conceiver to conceive the denotation as that of which this information is true of, and thus generating a possible world where this evidential information is true of an item that is different than the actual denotation of the concept in question.

It is in this sense we would like to claim that zombie replicas of ourselves are genuinely conceivable. In other words, it is genuinely conceivable

- (Z) that a complex property (our physical replica) expressed by a purely physical/functional predicate or concept is instantiated without the instantiation of an (apparently simple phenomenal) property that we pick out with a phenomenal predicate or concept we possess.<sup>67</sup>

When we fix the fact that the physical/functional property is instantiated, what we have to find out, in order to show that (Z) is not genuinely conceivable, is a description (expressing a complex concept) that is associated or connoted by the phenomenal concept (in the way we specified—i.e., consciously available information involved in its sustaining mechanism) that contingently picks out the same property. But that there is no such description should be clear from the way we analyzed the direct and immediate acquisition and vertical deployment of sensory concepts, and their introspective uses.<sup>68</sup> So (Z) is genuinely conceivable in our sense.

The most interesting aspect of modern conceivability arguments consists in novel attempts to draw metaphysical conclusions from this fact, i.e., the fact that (Z) is genuinely conceivable. For it appears that denying all other standard *a posteriori* identities turn out to be genuinely inconceivable in our sense. Given a complete physical description of the world (augmented by “standard” indexical information and a “that’s all” clause, etc.) and our semantic competence with concepts involved in these identities, it seems that denying the identities would generate inconsistencies.<sup>69</sup> The reason for this is that since the sustaining mechanisms of the concepts involved in these identities are cognitive and the abstraction distance they give rise to is not minimal, it is always possible to conceive of the referent of the concepts involved in terms of the contingently related information mediating their acquisition and vertical deployment. But, as we have just pointed out, conceiving the referent in terms of the information used in fixing the reference of these concepts like WATER always generates the possibility that the canonical reference-fixing information associated with these

concepts can be true of things other than the actual referents of these concepts. Since this is always possible whenever we deal with concepts whose sustaining mechanisms are cognitive, no standard *a posteriori* necessities are genuinely conceivable. This fact, of course, has been taken as a powerful argument for a strong reading of (BP). And this is the novelty of the modern conceivability arguments. The anti-physicalists in effect say this:

In all cases of *a posteriori* necessary reduction of the macro phenomena, we have a derivation from an augmented physicalist base; and accordingly, no denial of these (*a posteriori* necessary) reduction statements turns out to be genuinely conceivable. There appears to be only one exception to this: the denial of the supervenience of phenomenal consciousness on the physical. The reason for this is simple: phenomenal consciousness doesn't metaphysically supervene on the physical, so physicalism is false. The eloquence of this position is that we get to retain (BP) as an exceptionless logical entailment as it applies across the board, which has the nice consequence of yielding a unified epistemology of modality. The physicalist, on the other hand, has to make the *ad hoc* claim that (BP) is true except when it applies to phenomenal reductive statements. This is a desperate attempt to save physicalism by special pleading.

We would agree that holding the phenomenal case as an exception to (BP) would be a bad *ad hoc* maneuver *if* we didn't have *independent* reasons to think that (BP) has exceptions and therefore can at best be true as a reliable but defeasible generalization.

We have already provided an *independently motivated* and *principled* defeater to a strong reading of (BP). The defeater is the fact that the nature of our sensory and phenomenal concepts, in terms of which we conduct thought-experiments about what is conceivable and what is not, is such that they inevitably support the intuitions about the genuine conceivability of (Z) and the zombie worlds.<sup>70</sup> But if our purely naturalistic story is correct, this is to be expected: no metaphysical conclusion follows. Thus, our story in effect says that when conception involves sensory and phenomenal concepts in a certain way,<sup>71</sup> (BP) must be suspended. Our strategy is to neutralize the argumentative evidence that the anti-physicalist appeals to by showing how, within a purely physicalist framework, we can have a situation that supports the same intuitions. These intuitions support the genuine conceivability of zombies and the like but are based on physicalist assumptions. Our physicalist account explains and predicts these intuitions in a principled way, independently of an *ad hoc* motivation to save physicalism, by giving a completely general and unified psychosemantics for the concepts involved.

Of course, as should be clear from our argumentative strategy, our information-theoretic account and the way we use it to block conceivability arguments is no knockdown argument against anti-physicalism. Strictly speaking, our account of concept formation and introspection, with necessary

modifications, may be compatible with anti-physicalism. But this is as it should be. What we have provided is a naturalistic account that will make a non-demonstrative but extremely strong case against views like epiphenomenalism only when combined with general considerations about causality and methodological considerations about explanation and theory building. But before all that, it is certainly good to know—and important to underline—that there is no knockdown argument against physicalism or naturalism in general.

It is interesting to note that if our Naturalistic Story is right, the defeater we present against (BP), and ultimately against *a priori* type conceivability arguments, is empirical in nature. For our story is ultimately a form of philosophically informed theoretical psychology to be vindicated eventually by findings from empirical science. Indeed, consider Jackson's thought experiment about the physically omniscient color scientist, Mary, who has spent all her life in a black and white room until her release one day, when she sees colors for the first time. Before her release, suppose that Mary knows all there is physical to know not only about color vision but also all about introspection and concept formation. Then, supposing that something like our account is true, the complete details of this account are what she would know. But then she would automatically be in a position to know about the curious asymmetry involved in the epistemic access to phenomenal/physical facts. This body of knowledge she has before her release would not of course remove her curiosity (the surprise element) about coming to know in a first-person way facts she already knew under their scientific description. On the contrary, she would be even more curious and intrigued to *instantiate* those phenomenal/physical states herself, which are necessary for acquiring the peculiar perspectival concepts, and thus first-person knowledge. Knowing all the scientific facts would also make her know that she lacks certain kinds of concepts necessary to know facts in a perspectival way, different from the way she already knew them from a third-person stance. We would expect her not to be moved by the familiar conceivability arguments at all. Given her scientific omniscience and her determination to follow scientific methodology in theory building, she would be in a position to know better.

In this revised thought-experiment, where we assume the truth of NS, Mary's situation before her release is a curious one. Mary can derive from the physicalistic premise base the existence of phenomenal concepts and their curious semantics, but since she herself doesn't yet have these concepts about color experiences, she cannot *use* them to attribute *color experiences* to others (she doesn't yet have color experiences herself). But she can *mention* these concepts as given by NS and attribute *them* to others. There is of course an obvious fact about Mary before her release: namely, since she doesn't have the relevant phenomenal concepts yet, she cannot derive claims couched in them from the physicalistic premise base. But of course it isn't this fact that makes derivability impossible. Even

after her release, when she comes to possess all the relevant phenomenal concepts, she still cannot derive them from the physicalist base. Any phenomenal claim couched in phenomenal terms/concepts will make *use* of phenomenal concepts, and she won't be able to derive these claims from the physicalistic base even when this base includes the details of NS, unless she uses a premise in which she identifies the referent of the theory-given phenomenal concepts with the referent of her first-person phenomenal concepts by *using* the latter. But this premise itself is not derivable from the physicalist premise base with NS. The situation here is parallel to the impossibility of deriving an indexical claim from indexical-free premises (cf. Perry 1979). Mary may know all about herself in a third-person way, i.e. representing her knowledge as true of Mary (as opposed to be "true of myself"), and also know that Mary has such and such indexical thoughts and that they work in such and such a way. But unless she identifies *herself* indexically as the subject of these indexical thoughts about Mary, her knowledge about Mary's indexical thoughts will be as foreign to her as Mary's salary.

We may say, then, that there is a *transcendental* sense in which (Z) is not conceivable. As in the case of Mary above, suppose that a completed physicalistic cognitive science vindicates our information-theoretic account of concept acquisition and sensory/phenomenal concepts and we have a complete physical knowledge of our world. Then the physicalist premise base true of our world will entail that we have sensory and phenomenal concepts that we use to denote conscious qualities of our experiences and make claims about them that *we* cannot derive from this physicalist base. Nevertheless, we will know that our sensory/phenomenal concepts as *we* have acquired them from our experience pick out the same physical/functional properties that certain physical/functional description in our theory picks out. This scenario might not strike the reader as a promising route to closing the explanatory gap about phenomenal consciousness, as our scientific knowledge about our mind/brain currently stands. But once we have the details in place and if they turn out to vindicate NS, the puzzle of consciousness might reduce to the same sort of "puzzle" involved in our identifying *ourselves* as the man with the torn sugar sack pushing the shopping cart in the supermarket and wondering who is leaving behind the sugar trail (Perry 1979; see also Perry 2001). Nobody is tempted to embrace bizarre metaphysical conclusions by the puzzle of the essential indexical, but that is because we know all the relevant facts. In this paper we have tried to provide an account of how the mystery surrounding consciousness might be removed if something like our Naturalistic Story were true and available in full detail; for then we would have all the relevant facts.<sup>72,73</sup>

## Notes

<sup>1</sup> Carruthers (2000), Churchland (1985, 1989), Loar (1990/1997, 1999), Lycan (1987, 1996), Papineau (1993, 2002), Pereboom (1994), Sturgeon (1994, 2000), Rey (1997), Hill (1997), Hill & McLaughlin (1999), Tye (1995, 1999, 2000), and Perry (2001). Rosenthal (2001) and Shoemaker (2001) also come close to this line.

<sup>2</sup> Descartes and Locke are the most prominent historical expounders of conceivability arguments. Recent versions can be found in Farrell (1950), Feigl (1967), Kripke (1970), Nagel (1974), Jackson (1982, 1986), Robinson (1982), Hart (1989), McGinn (1991), and Chalmers (1996). For a good critical discussion of such arguments, see Levine (1993, 1998, 2001) and Biro (1991). Levine does not endorse the metaphysical conclusion of these arguments; in fact, he argues against drawing such a conclusion. But his discussion of the explanatory gap between the phenomenal and the physical has contributed significantly to a more articulate development of conceivability arguments. Throughout the essay we will be using the terms “materialism,” “physicalism,” and “naturalism” interchangeably.

<sup>3</sup> What Levine (2001) calls exceptionalists (E-type materialists) who take consciousness as an exception to the global *logical* supervenience of macro phenomena on the microphysical, where this involves *a priori* derivability of the former from the latter. This is a subgroup of what Chalmers (1996, 1999) calls Type-B Materialists (among them are Loar, Sturgeon, Hill and McLaughlin, and possibly Tye—see note 1). The non-exceptionalists divide into two categories: (1) those who claim that almost nothing can be derived *a priori* from the microphysical (Levine 2001, Block and Stalnaker 1999, and possibly Lycan) so that there is nothing special about consciousness; (2) those who claim that consciousness can also be derived *a priori* from the microphysical—so again nothing special about consciousness (these tend to be *a priori* materialists like behaviorists and functionalists who think that phenomenal concepts can be analyzed—what Chalmers called Type-A materialists such as Armstrong 1968; Lewis 1966, 1972; Shoemaker 1981/1997). We take ourselves to be close to exceptionalists for reasons that will become clear as we move on, but we would like to remain neutral about the specifics of a model of scientific reduction, partly because we have some reservations about the notion of *a priori* derivability involved. Luckily the puzzle about the logical supervenience of consciousness can be raised independently of specific models of reduction as Levine (2001: Chaps. 2–3) shows. If Levine is right about there being no *a priori* derivations of macro phenomena including consciousness, then on the account of reduction he prefers, the puzzle about consciousness reduces to the puzzle of why consciousness can be reduced to the physical only via what he calls “gappy identities.” We take ourselves to show why—if we adopt his model of reduction. See § 9 below. So our account of phenomenal concepts is not meant to be tailored only for the exceptionalist line.

<sup>4</sup> For instance, while one might expect to see a very close link between sensory concepts (e.g., of red) and phenomenal concepts (e.g., of experiencing red), so far very little has been said about it; in fact, this distinction sometimes seems to be overlooked—e.g., Churchland (1985, 1989). Tye (1999) mentions the distinction, but then he goes on to write in a way which suggests that he thinks the distinction is not important. Not only that; no detailed perspectivalist account has been developed about how we acquire these phenomenal concepts, what it is exactly to which we apply these concepts, and how. (Carruthers 2000 and Papineau 2002 are recent exceptions to this charge: we have become aware of their work in sufficient detail only after this paper left our hands for the journal—it is too late to discuss their work in relation to ours here, but we hope to do that in our forthcoming book since there are similarities and many connection points between their works and ours worth exploring.) Most perspectivalists commit themselves to the claim that phenomenal concepts somehow apply to brain states realizing conscious qualitative experiences. Again, we believe that very little has been said about why phenomenal concepts don't present to us physical properties of these brain states *as such*, which is at the source of the puzzle. As a result, the postulation of phenomenal concepts has seemed quite *ad hoc*, even though, from a technical point of view, it seems to block some forms of

conceivability arguments (e.g., Jackson's Knowledge Argument) but without producing much conviction. Along with many, we are deeply dissatisfied with the present situation.

<sup>5</sup> While the early 1980s saw a proliferation of independently motivated and fruitful philosophical work in philosophy of mind and language that adopted resources of information theory, information-theoretic accounts have not, so far as we can see, made their way into the more recent physicalist attempts to naturalize phenomenal consciousness. The empirical sciences of the mind, on the other hand, have by and large adopted an information-theory based paradigm since the 1960s, and there has been significant progress in cognitive psychology and psychobiology, and more recently in cognitive neuroscience, towards a deeper understanding of cognition within this paradigm. (See Attneave 1954, Cherry 1957, Broadbent 1958, Sayre 1965, and Uttal 1968, 1973 for the foundations of such work; see Gibson 1966 and Neisser 1963 for historically important critical evaluations.) Our project is motivated by the desire to provide a naturalistic account of our sensory and cognitive lives coupled with the belief that information theory possesses the most promising set of conceptual resources presently available to do so, and consequently to shed light on philosophical puzzles about phenomenal consciousness that have thus far remained without a principled and satisfactory solution.

<sup>6</sup> For the ease of our presentation, we won't, for the most part, make a special effort to tell when and where our story differs from Dretske's.

<sup>7</sup> We think human beings are only one of many species that have a conscious mental life. Here we do *not* demand a sophisticated conceptualization process or the employment of full-fledged concepts of linguistic creatures for the possession of conscious sensory representations, in the way some cognitivists do. Carruthers (1989), for example, claims that non-linguistic animals only have non-conscious sensations since they lack the advanced conceptual structures necessary to form second-order thoughts about those sensations. We revisit the question of the role of higher-order representations in consciousness in §8 below. See also Dretske (1994) and Dennett (1994) for an interesting exchange on this issue.

<sup>8</sup> The distinction between conscious and unconscious experiences or sensory representations is a deeply controversial issue, and the empirical literature in this area is vast. Within the span of differing opinions, it is possible to find psychologists whose careers are built on investigating the utility of this distinction (e.g., Dixon 1971, 1981) as well as those who deny the very existence of the distinction itself (e.g., Holender 1986), on the basis of roughly the same set of data.

Put in broad strokes, the literature examines cases that provide evidence of an architectural dissociation between the subject's representational capacities and her awareness of the content of her representations, in two categories: neuropsychological cases caused by brain damage (as in blindsight and hemilateral neglect), and experimental cases with normal subjects where the intensity or duration of the sensory stimulus is reduced to sub-threshold levels or the subject's attention is manipulated. For an overview of the former, see Farah (1995), and Köhler and Moscovitch (1997); for typical exemplars of the latter, see Marcel (1983a, 1983b). Kihlstrom (1984, 1987) extend the conscious/unconscious distinction from the realm of sensation and perception to cognition; Reingold & Merikle (1990) and Velmans (1991) provide interesting discussions of the meta-theoretical issues involved.

While we note that our account allows for a distinction between conscious and unconscious sensory representations, we will not develop its foundations here.

<sup>9</sup> Availability to a central conceptual system is necessary for a sensory representation to be a conscious state. In addition to availability, sufficiency is probably secured by detailing some constraints on the role the information in the experience can play, after its delivery, within the conceptual system on its way to setting the behavioral parameters, especially about its availability to learning, planning, decision-making, and speech systems. Cases where the unconscious (either subliminal or due to a neuropsychological syndrome) perception of a stimulus affects behavior are probably cases where some information in the experience is made available to the conceptual system influencing behavior but nevertheless is not available to *all* central



cognitive capacities—its processing within the conceptual system is limited. (A view of this kind is defended in Bernard Baars's *Global Workspace Theory of Consciousness*: see Baars 1988.)

This way of thinking suggests that the central system may not be entirely non-modular in the Fodorian sense, and thus may be composed of loosely connected central faculties, or better, processing streams, and that there are conditions under which global availability may be hindered. The evidence has been accumulating in recent years that there are two central streams (the dorsal and the ventral) in the brain where the same sensory information, after its delivery, is used mostly independently of each other. One of the functions of the dorsal stream is to use the incoming sensory information in what we might call mundane behavior in auto-pilot—e.g., as we navigate around objects, the fine-tuning of our behavior seems to be initiated and controlled before the relevant sensory information becomes conscious through the ventral stream). See, for example, Goodale et al. (1994), Milner and Goodale (1995), Weiskrantz (1997), and Kanwisher (2001).

We are aware of the discussion surrounding accounts of state consciousness, and our stipulation here about sensory consciousness follows Dretske's position (1995, 1997) with respect to necessity but not sufficiency. We also think that this formulation as a necessary condition should be acceptable to those who defend a higher-order-representation (HOR—perception or thought) theories of state consciousness. As will become apparent later on, our view accommodates the central insights of HOR theories to some extent, without itself being one.

<sup>10</sup> Needless to say, these are functionally specified information-theoretic distinctions and do not always map on to anatomical or even physiological structures in a clean-cut way. The situation is further complicated by the fact that the brain has many recurrent networks and a considerable proportion of information processing occurs bi-directionally. Nonetheless, it seems reasonable to say that in the human visual system, the primary visual cortex (V1) is a crucial relay station that bridges sensory and conceptual processing in that the representational structures carrying visual information prior to the primary visual cortex (V1) are by and large part of the sensory system, whereas conceptualization continues in information processing past V1. The significance of the role of V1 in visual processing is also evident in cases where it is lesioned, resulting in the phenomenon of blindsight, where a number of visual discriminations can be made under forced-choice conditions while the subject reports no awareness of the presented stimulus whatsoever. See endnote 44 below for further discussion of blindsight.

<sup>11</sup> Horizontal uses of concepts may be either voluntary or involuntary. To give you the flavor, here is a hypothetical situation. You read in the newspaper that a friend you had not seen since college days, now a pop star, has just won the first prize in the Eurovision song contest. You start thinking about the days when you used to take music lessons together. Then you reminisce about your quirky piano teacher and the jokes he used to tell you, which makes you decide that you get in touch with your friend, and look for ways of finding out how to reach her, etc. . . . Just think about the range of things that such an event could prompt you to think about: at the moment of your thinking almost none of the objects of your thoughts has any direct or obvious causal/informational relation to your thinking. (Horizontal prompting of a thought may be an involuntary occurrence, but then you may voluntarily prompt thoughts in yourself.)

<sup>12</sup> Information-carrying tokenings of a concept can arise in two ways: as a direct and immediate response to its "proper" sensory base, or by a reliable inferential process from "improper" sensory bases, including speech perception. If we were to label all information-carrying tokenings as vertical, we would need to distinguish them from horizontal "disengaged" tokenings as in thinking, free association, imagining, etc. in which they can be correctly tokened without carrying information about their denotations or truth-conditions. These two kinds of vertical tokenings complicate the picture (as Dretske insisted in personal communication). However, it is clear that those vertical tokenings based on inference from "improper" sensory bases require that the semantics of these concepts be already in place. Then the question becomes whether their semantics is acquired from "proper" sensory bases or something theoretical or descriptive, like Mary's concept of red before her release. Here we restrict the

information-carrying/vertical uses of *sensory concepts*, which we will define below, to those prompted by “proper” sensory representations.

<sup>13</sup> There is much evidence from brain imaging studies using ERP and fMRI that imagination, which is a central/horizontal capacity, uses the same representations as those involved in some stage in sensory processing; see, for example: Raji (1999) and Senkfor et al. (2002). But we think these representations are special sensory/perceptual concepts which are deployed in turning sensations into perceptions as we characterize this central process below. There is also similar evidence that at least some of the representations used in the sensory systems as well as in imagination are shared by movement execution in the motor cortex; see, for example, Rizzolatti et al. (1997), Lotze et al. (1999), and Kohler et al. (2002).

<sup>14</sup> The contrast between sensation and perception is sometimes described in terms of a distinction between *seeing* and *seeing as* (*hearing* and *hearing as*, etc.; see, for example, Dretske 1995). Having normal vision, you can certainly see an aardvark, but having no idea of what aardvarks are (not having the concept of one), you cannot see it *as* an aardvark. In ‘*S* sees *x*’, ‘*x*’ occurs transparently (i.e., could be replaced by any co-referring expression without changing its truth value), but the occurrence of ‘*F*’ in ‘*S* sees *x* as *F*’ is opaque, reflecting the fact that the truth-value of the statement depends on whether *S* has the concept expressed by ‘*F*’ and applies it to *x* as a consequence of standing to *x* in the *seeing* relation.

However, as noted above, we are not cognitivists about conscious experience in the way Carruthers (1989) is, or as Dretske (1994) characterizes Dennett’s position. Although we think that concept possession is necessary for consciousness in the sense required by our characterization of state consciousness above, we grant that having sensations, strictly speaking, does not require the capacity for or actual deployment of concepts, while perception does. The intuitive idea is that if there is no way for the central conceptual system to recover (conceptualize/digitalize—see below) the sensory information in the sensory array, then the sensory state that carries this information is not conscious.

<sup>15</sup> Or perhaps, it is one of the vertical streams within the central system dedicated to extract information from the visual *sensations* for conceptualization—but central nonetheless in that most of the information in the stream is globally available, even though the *processing itself* is not voluntary or introspectable, and may even be open to top-down cognitive influences.

<sup>16</sup> Note that this issue was the basis of a controversy between Locke and Berkeley on abstract ideas. Berkeley, who argued against the possibility of abstract ideas, puts his challenge to Locke in the following empirical fashion:

If any man has the faculty of framing in his mind such an idea of a triangle as is here described, it is in vain to pretend to dispute him out of it, nor would I go about it. All I desire is that the reader would fully and certainly inform himself whether he has such an idea or no. And this, methinks, can be no hard task for anyone to perform. What more easy than for anyone to look a little into his own thoughts, and there try whether he has, or can attain to have, an idea that shall correspond with the description that is here given of the general idea of a triangle, which is “neither oblique nor rectangle, equilateral, equicrural or scalenon, but all and none of these at once”? (Berkeley 1710/1977: 13–14.)

<sup>17</sup> When we talk about a concept’s carrying information we have of course its vertical tokenings in mind. Most of the time we will omit this qualification in what follows since the context should make it clear what sort of concept tokening we have in mind.

<sup>18</sup> Dretske does not draw a distinction between extractable and non-extractable ways of carrying analog information, although it is implied by what he says. However, this distinction will play an important role in what follows. For an insightful elaboration of this distinction, see Kulvicki (2001, forthcoming-a) who uses this distinction to give an original and general account of isomorphism.

<sup>19</sup> For the purposes of exposition, we are restricting the abstraction base to the one provided only by the sensory representation/picture. Of course, this need not be the case once a threshold is

passed and a certain basic conceptual repertoire is in place. What can be learned from a signal depends not only on the signal itself but also on what is independently known about the source, and this requires inferential deployment of other concepts and *collateral information*.

<sup>20</sup> We argue below that *sensory* concepts in the sense we will introduce cannot be acquired horizontally. However, non-sensory conception of secondary qualities such as colors can be formed through horizontal means. See, for example, Shepard and Cooper (1992) for a study of a conceptually rather sophisticated understanding of colors and, interestingly, their qualitative interrelations in the blind and the color-blind.

<sup>21</sup> Most of what we have said so far can be found in Dretske's (1981). What follows is our way of carefully applying this information-theoretic framework to what we call sensory and phenomenal concepts, which Dretske does not address—he even gives the impression that he would rather bypass talking about concepts very close to the sensory periphery.

<sup>22</sup> By 'minimal' we will mean "shortest possible," or technically speaking, "maximally short." But since this latter expression sounds a little oxymoronic, we will stick with 'minimal'.

<sup>23</sup> Or, alternatively, we can specify the total digital content of a given visual sensation by specifying the information about the hue, brightness, and saturation of every discriminable spatial point in the visual field according to a multi-dimensional coordinate system. Complete specification of this information will then give the total digital content of the experience.

<sup>24</sup> Although it is controversial, we will assume a primary quality view of secondary qualities like color à la Hilbert (1987). We don't think, however, that much hangs on this choice in what follows: any objective property would do.

<sup>25</sup> This is not exactly true, but the way in which it is not true won't be important for what follows. There are at least three further dimensions along which we gather information about colors: hue, brightness, and saturation. Furthermore, there are also phenomena like the one exhibited by so-called binary hues: they are represented in experience in a way that these hues seem to be composed of relevant unique hues (e.g., orange is represented as containing, in some sense, red and yellow, whereas pure red, like other unique hues, is not represented as being constituted by other hues). These complicate the claims made in the main text, but not in a way that alters the main point, which is that there is a limit to what determinables our sensory systems can discriminate; at bottom, some determinate values of these will have to be represented as simple/atomic. Whatever these determinables are, sensory concepts will be those based on these such that the abstraction distance will be minimal. This is true for all sensory modalities. As indicated, we will ignore this complication in what follows. Moreover, we will talk for convenience as if there were no abstraction distance between the sensory representations of, say, red, which represent determinate shades of red, and the most specific concepts of shades of red we can diachronically discriminate. So we will say that this distance is minimal.

The point about how our sensory representations represent certain determinables (as determinate simples) has been made by a number of philosophers before (Armstrong, Shoemaker, Lycan, as well as the British empiricists in general), but for a detailed elaboration of this idea in the context of discussing sensory/pictorial representations, see Kulvicki (2001, forthcoming-b) who tries to give a principled distinction between primary and secondary qualities in terms of how they are represented in conscious sensation, i.e., in epistemic terms, in opposition to the more traditional way of drawing the distinction in metaphysical terms.

<sup>26</sup> Strictly speaking, we should rather say that the sensory representations of secondary qualities do not represent them as having a complex constituent structure. As we have hinted above, this is different from saying that they are represented as simple or atomic. But we will be relaxed about this in what follows. For the distinction, see Armstrong's discussion of the headless woman fallacy in his (1968, 1987).

<sup>27</sup> Lost in the sense that the tokening of the more abstract target concept does not carry it.

<sup>28</sup> Conversely, there can be machines constructed (or organisms evolved, under appropriately availing circumstances) where the abstraction distance between "sensory" and "conceptual" representations in "color perception" (understood functionally as the detection of surface

reflectance profiles through the medium of light) is, unlike in our case, *not* minimal. (It is tempting to speculate that the curious case of the “color-blind” painter described by Sacks 1995 is like this.) Again, there is no logical necessity for organisms to come to possess the kind of neural architecture that they do, and hence no logical necessity about the particular manners in which information about the world is put to use by organisms which utilize that information to their benefit.

<sup>29</sup> We prefer this notion to Dretske’s notion of *channel conditions* because it is more specific and suggests mechanisms internal to the agent, which is what we would like to emphasize here. Although cognitive factors (what is independently known about the information source) can be part of channel conditions, Dretske, with this notion, emphasizes those conditions external to the agent, or at least external to his mind.

<sup>30</sup> See Margolis (1998) and Laurence and Margolis (2002) for a parallel account of concept learning that involves *cognitive* sustaining mechanisms, which is nevertheless not a hypothesis forming and testing model à la Fodor (1975, 1981). They show, within a similar framework, that many lexical concepts may be primitive despite being learned from experience; hence they deny that atomism implies a radical nativism of the sort endorsed by Fodor. This is good news for the Language of Thought Hypothesis (LOTH), and concept atomism in general, because it frees them from one of their main burdens. Their work nicely supports and complements the Dretskean account given here.

<sup>31</sup> “Approximately” because we think that the match may not be perfect. If some spatiotemporal properties/relations (like being a point or an expanse), as we suspect, turn out to be primitively represented in our experiences, then they may turn out to be categorized as secondary, contrary to the tradition. Although we think that there are principled ways to avoid this consequence, still we are happily prepared to live with this consequence if it turns out we can’t avoid it. (See Kulvicki 2001 and forthcoming-b for further discussion.)

<sup>32</sup> Exceptions to this claim are what Armstrong (1962, 1968) called “intransitive bodily sensations” like pains, itches, and tickles, which we will take up later on.

<sup>33</sup> We would like to put aside perceptual concepts for the moment. Their proper treatment requires an empirically informed answer to “Molyneux’s question,” which we currently do not have.

<sup>34</sup> Again the verdict about perceptual concepts is not clear: they fall in between sensory and observational concepts. But our intuition is that they will be closer to sensory concepts in that they can’t completely digitalize the information about their denotations.

<sup>35</sup> There is a parallel condition intended to rule out such cases in Fodor’s version of informational semantics (Fodor 1987, 1990): the asymmetric-dependency condition, which says that a cognitive structure (concept) *C* has the semantic content *X* in virtue of an informational law between them such that if there is another such law between *C* and anything *Y*, then this law asymmetrically depends on the existence of the law between *C* and *X*. In other words, breaking the law between *C* and *X* breaks the law between *C* and *Y* but not vice versa. Like Dretske’s condition, this account, too, fails to assign the distal secondary property to a sensory concept as its semantic content, and ends up assigning, instead, proximal properties underlying the sensory representation of the distal property. For it is clear that breaking the law between the sensory representation of red and the concept RED will break the law between the property of redness and its concept, but not vice versa.

<sup>36</sup> We will use uppercase letters to name concepts and italics to name properties denoted by concepts. To many ears, talking about a property as the denotation of a concept may sound strained. Although nothing very important hangs on this, we not only find this usage convenient, but also are prepared to justify it on the basis of informational semantics, which freely traffics in property instantiations. This makes property realism the preferred reading by informational psychosemantics. (Of course, just as linguistic predicates are not referring expressions, concepts *qua* Mentalese predicates do not refer—we have been using the more neutral

‘denotation’ intending to indicate that concepts represent, and when relevant, what they represent.)

<sup>37</sup> However, in fn. 4 to Chapter 8 of his (1981: 261), Dretske shows signs of being aware of the problem here. He draws a distinction between a *visual concept* of a robin and the concept of a robin *tout court*, and says that the former is not completely digitalized. But the point is not further elaborated. As we already mentioned, Dretske generally seems to eschew talking about sensory concepts.

<sup>38</sup> Interpersonal cases pose special problems (of the sort Shoemaker 1981/97 highlights) that we will discuss elsewhere, in the context of spectrum inversion thought experiments against functionalism about qualia.

<sup>39</sup> We in fact suspect that even a stronger claim is true: sensory concepts are those tokens of the structures realizing the sensory representations of secondary qualities (even perhaps primary ones), but used differently by somehow being recruited by the central system. For sensory concepts we are prepared to accept what we take to be the central claim of Prinz (1997, 2002) and Barsalou (1999), namely that there is no fundamental distinction between percepts and concepts, and that concepts are percepts (only used and organized differently): hence, necessarily, sensory concepts are not amodal. For a “quotational” model incorporating a similar idea, see Papineau (2002).

<sup>40</sup> We want to emphasize the qualifier ‘normally’ here. Although we adopt an actualist psychosemantics in this paper in order not to further complicate exposition, strictly speaking, we would like to remain, in general, neutral between an actualist psychosemantics (à la Dretske) and a purely counterfactualist one (à la Fodor (1987) and (1994)). If we adopt a counterfactual account, then we can cast the point in the text in terms of a narrow content conceived as a partial function from contexts to semantic content, in which case actual delivery of information for genuine sensory concepts is not required, but what is required is a “sensory” state that would track different causes in different contexts, namely secondary qualities of objects. Systematic hallucinations of such a secondary quality would then enable one to acquire a genuine sensory concept.

<sup>41</sup> Also, it is clear that more needs to be said about the nature of the perspectivity involved: in particular, we need to distinguish *pure indexicality* from what we might call *quasi-indexical predication*. Although sensory concepts in our sense are mental predicates, there is obviously something indexical about them: their semantics is hostage to where and when their tokenings occur. Fully developing the notion of quasi-indexical predication will require a paper on its own, which we will take up elsewhere.

<sup>42</sup> We will use the abbreviation ‘E-p’ to denote the experience of property *p*. So, our terminology includes, for example, RED (the concept), *red* or *redness* (the property), and E-red (the experience).

<sup>43</sup> The order in the pair is meant to reflect this dependency relation. The relevant information is carried by the appropriate *tokenings* of RED and E-red about the *instantiations* of the property *redness*. Most of the time, we will forgo talking this way for convenience.

<sup>44</sup> Blindsight subjects *might* be precisely in this position: even though they may recognize and categorize certain features of stimuli presented in their blind field under forced-choice conditions, they can’t voluntarily apply the concepts to objects in their blind field, despite the fact that they possess information about those stimuli and may have (some form of) relevant sensory representation. In other words, the problem might not be a problem of sensory representation, or *only* a problem of sensory representation, but rather it may be related to what information the conceptual system is allowed to vertically pick up from such a representation. So, for instance, it is allowed to pick up the information that there is a horizontal line before one’s eyes, but no further vision-specific information can be directly conceptualized. And since the patient doesn’t know where the information about the horizontal line is coming from, he is not only volunteering this information; in fact, she denies having it.

The fact that lesions in the primary visual cortex (V1)—which is something of a gateway between early sensory processing of visual information and further upstream processing that is believed to involve conceptual structures (e.g., in the prefrontal cortex)—is the main culprit in the deficits exhibited in blindsight also seems to support the hypothesis that the problem centers around the interface of information transmittance between the sensory and the conceptual systems. The most favored view in the present literature is that the “residual functional capacities” for recognition and classification of stimuli under forced-choice conditions are due to the existence of pathways that carry visual information to subcortical areas (even when portions of V1 is unable to receive that information) as well as transmit it upstream. As such, blindsight might be a problem about the integrity of late sensory representations in the vicinity of V1 as well as a problem about information pickup by the conceptual system related to the ventral stream, rather than merely a lack of sensory representation.

The scientific literature on blindsight is large and controversial, and the philosophical treatment of it often lacks a desirable sophistication of the scientific evidence involved. For a brief introduction, see Pöppel *et al.* (1973) for the first study on blindsight, Weiskrantz (1986, 1997) for authoritative overviews of this neuropsychological syndrome, and Ptito *et al.* (2001), Stoerig and Cowey (1992), Cowey (1995), and Cowey and Storeig (1995) for studies that address different questions about blindsight in accord with Wieskrantz’s framework. See Campion *et al.* (1983) for a skeptical view about blindsight as a genuine syndrome, Barinaga (1992) for an overview of controversies in blindsight research, and Fendrich *et al.* (1992) and Gazzaniga, M., *et al.* (1994) for an alternative to the received view. In the philosophical literature, the phenomenon of blindsight was first referred to and utilized by John Searle (1979) and Patricia Smith Churchland (1980). For different and not always compatible philosophical interpretations of the empirical results and their consequences for theories of consciousness, see Heil (1983), Cam (1985), Carruthers (1989), Van Gulick (1989), Dennett (1991), McGinn (1991), Flanagan (1992), Tye (1993, 1995), Güzeldere, Hardcastle, & Flanagan (1999), and Güzeldere (in prep.).

<sup>45</sup> Actually, we can further speculate on this by pointing out that certain syntactic requirements of a *system* of mental representations within the central conceptual system may generate the coherence requirement naturally: for instance, certain syntactic positions in this system may require predication of concepts denoting distal objects. Thanks to Jonathan Weinberg for pointing out this line to us.

<sup>46</sup> For an elaboration of this theme about pain and other bodily sensations in the context of an argument against pure representational theories of qualia, see Aydede (2001, forthcoming-a, and forthcoming-b). Note that even advocates of direct perceptual or representationalist theories of pain grant that our dominant/ordinary concept of pain (and of other bodily sensations) applies to token experiences. They insist, however, that these experiences represent certain bodily conditions, and that their qualitative content is exhausted by their representational content.

<sup>47</sup> There are many puzzling aspects of pain experience, in terms of both its bodily basis and its complex phenomenology. We try to address some of these questions in Aydede and Güzeldere (2002), Aydede (forthcoming-a, and forthcoming-b).

<sup>48</sup> For an elaboration of this theme, see again Aydede (2001, forthcoming-b). We should note, however, that to the extent to which the awareness of pain is also awareness of this affective (hurtful) aspect of pain, introspecting it will involve mechanisms partly different from the ones we will offer below—insofar as this affective quality of pain is non-sensory or non-representational.

<sup>49</sup> Single square brackets will be used in what follows as a way of indicating that the expression occupying the linguistic position marked by them is to be read as occurring transparently. The judgments expressed by such brackets in the subject position will be *de re* judgments, expressing singular propositions. The term ‘pain’ in the expression of this judgment is meant to express the *sensory* concept PAIN. For even people who are congenitally insensitive



to pain (sic!—but this is the technical term standardly used) can have a concept of pain and can intelligibly communicate their pain thoughts. The concept these people have is not a *sensory* concept in our sense.

<sup>50</sup> As we have said, this may be a psychofunctional property, realized by a small disjunctive set of physical properties of the relevant sort.

<sup>51</sup> Cf. Shoemaker (1981/97, 1994a), Clark (1993).

<sup>52</sup> Hilbert (ms.) contains an insightful discussion of why having this information is an extremely useful thing for cognitive organisms like us with sophisticated and peculiar epistemological needs—not just for doing philosophy, but for adaptive behavior that has survival value, since we need information about the sources of our perceptual beliefs to assess their reliability.

<sup>53</sup> Or, whatever most specific concept we can deploy in *recognition* tasks.

<sup>54</sup> But we will be relaxed about the terminology as long as it is clear what we mean by “using the same concepts in a different way.”

<sup>55</sup> See, for instance, Perner (1991), Wimmer & Hartl (1991), Gopnik (1993), Gopnik & Wellman (1994), Gopnik & Meltzoff (1994), and Frith & Happé (1999), among others.

<sup>56</sup> Nichols and Stich (2002, forthcoming) argue that this evidence is not strong and claim to have found counterevidence. We remain unconvinced by their argument and the evidence they provide. We think that their criticism doesn't respect the competence/performance distinction, and once this distinction is made, all the cases they criticize as inconclusive can be attributed to performance failures. In fact, their own examples seem to show that a certain amount of competence with intentional idioms is required to make sense of children's self-attributions despite their failure to perform equally well in other-attribution of mental states.

There is a sense in which the account of introspection we develop here can be seen as supporting Nichols and Stich's tentative proposal about what they call *Percept Monitoring Mechanism* (PMM) account of introspecting one's experiences. Their main account, *Monitoring Mechanism* (MM), is an account about introspecting propositional attitudes. Nichols and Stich do not elaborate on the informational foundations of PMM and they don't relate it to sensory concepts. However, we are confident that they would find our account quite congenial. The point we disagree about is how and when this mechanism becomes on-line in young children. The empirical evidence seems to show that it becomes on line *at least* in parallel with the acquisition of intentional idioms at the early stages of acquiring folk psychology.

<sup>57</sup> This acquisition process may not exactly be *learning*; it can come about by the *triggering* effect of external stimuli.

<sup>58</sup> There is actually a nice account of this available in what is involved in 'k' (the variable standing for what is independently known about the source) in Dretske's original definition of information in his (1981: 65): “A signal *r* carries the information that *s* is *F* = The conditional probability of *s*'s being *F*, given *r* (and *k*), is 1 (but, given *k* alone, less than 1).” The acquisition of the intentional idiom and folk psychology is the acquisition of independent knowledge about what is happening at the source, i.e. at the experiential level in one's own case.

Note the revealing analogy that exists between acquiring this introspective capacity and the way in which acquiring new concepts from a third-person perspective makes one aware of new experiential qualities in wine tasting or listening to classical music (etc.): similarly, acquiring intentional/representational concepts from a third-person perspective makes one vertically aware of one's own experiences, aware of what they are like and what it is like to be in them, by somehow making it possible to shift the semantic content of sensory concepts we had already acquired in the process of perceiving the world around us. The information about experiences, similarly, is already there in the tokening of relevant sensory concepts.

<sup>59</sup> Or its referential equivalent, as indicated by square brackets in the previous sentence.

<sup>60</sup> However, unlike Rosenthal, Dretske (at least in print) thinks of introspection as more like theoretical inference (inferentially mediated displaced perception) rather than perceptual *de re* knowledge—see also Shoemaker (1994a), whose views on introspection are similar to Dretske's in certain respects. See Aydede (2003) for a criticism of Dretske's inferential account



of introspection; see Shoemaker (1984, 1994b), and Güzeldere (1995) for a criticism of the higher-order perception accounts of consciousness and introspection. Sellars (1953) is perhaps the earliest account of introspection as involving inference from folk psychology.

<sup>61</sup> See Sturgeon (1994, 2000) for making the distinction in terms of canonical evidence base for concept application.

<sup>62</sup> Balog (1999) argues in a very similar way against the possibility of phenomenal zombies. She starts with the assumption (for reductio) that zombies are possible. Then she claims that these creatures will have phenomenal concepts that work very much like ours do, except that they will either not refer or if they do they will refer to zombies' physical/brain states. But then they will equally be philosophically puzzled because they won't be able to derive the existence of "phenomenal" properties from their physicalist premise base true of their own world. We have found her argument ingenious and convincing. However, in the absence of a detailed positive and substantive account of the nature of phenomenal concepts, which must ultimately be open to empirical confirmation or falsification on our view, we suspect that her argument preaches only to the converted insofar as it is possible to argue that zombies' "phenomenal" claims are either false or truth-valueless. Our argument in the main text is not about zombies. Our Naturalistic Story is an empirical story *about us* (although presently speculative and sketchy, admittedly). Nevertheless, we are impressed by Balog's insightful treatment of the zombie thought-experiments, which should complement our speculative but ultimately empirical account of phenomenal concepts.

<sup>63</sup> There is another option for the anti-physicalist: deny that our naturalistic story is actually naturalistic. For instance, it can be claimed that an informational psychosemantics cannot naturalize human intentionality. We certainly don't take ourselves to have shown that this claim is false. The information-theoretic approach is a systematic research program which we think has quite a promising potential, which has only started to burgeon in the last two decades. But we cannot discuss this more general question here. Of course, it is possible that NS might be empirically false. But this would still leave open the possibility of a world in which it is true, in which case (BP) would still be vulnerable to our criticism in its strong form as an exceptionless logical entailment claim.

<sup>64</sup> We think that Chalmers' (1999, 2002) arguments for reading (BP) as a logical entailment fail because of the reasons we gave relying on our Naturalistic Story. Others criticized them on other (sometimes similar) grounds, and we also find these criticisms convincing: Yablo (1993), Levine (1993, 1998), and Byrne (1999). But even if Chalmers is right about the damage that would be done to the epistemology of modality unless (BP) is read as a logical entailment, still, we think that this damage is not worse than the damage done to the metaphysics of mind if (BP) is read in this strong way. If necessary, we are prepared to make whatever adjustments are needed in the epistemology of modality to save a physicalist metaphysics of mind—well ... almost (we don't want to be dogmatic about this). We are aware that some (e.g., W.D. Hart, George Bealer, and David Chalmers, of course) order their priorities the other way around. But we don't think this will be necessary: a weaker reading of (BP) can do all the work required by a proper epistemology of modality, i.e., a reading according to which (BP) is a reliable but defeasible rule.

<sup>65</sup> Such statements need not be identities. Whatever form they take, the point concerns the availability of an associated description which contingently picks out the same thing picked out by a term used in the expression of the statement whose conceivability is in question. Here the claim is not that the conceiver should actually associate such descriptions. The point is about whether any such description or conception exists and is potentially available to the conceiver given her sensory/perceptual capacities. The analog information contained in sensory arrays and used in reference fixing should be available for further digitalization or conceptualization, but this does not mean that the conceiver has actually conceptualized every bit of this information used in fixing the reference of a non-sensory concept. Also, not all relevant information is in the sensory array, there is almost always collateral information used in fixing the reference of observational concepts.

<sup>66</sup> So, no definition or semantic/conceptual analysis is needed, contrary to what Chalmers (1996), Jackson (1994), and Levine (1993) assume. In this sense our reconstruction of their conceivability argument makes it even stronger by weakening one of its controversial premises.

<sup>67</sup> (Z) describes a phenomenal zombie. There are various non-equivalent ways of describing zombies (e.g., on the basis of local vs. global supervenience; token vs. type identities; state-based vs. individual- vs. species- based scope; by including propositional attitudes vs. not, etc.). Some of them involve important nuances. But we will not bother to be more specific here.

<sup>68</sup> We don't envision the need for intentional concepts in the introspective use of sensory concepts as part of their reference-fixing or sustaining mechanisms. The case of PAIN and concepts of other bodily sensations show this. Rather the intentional concepts somehow help the semantic content to refocus on the proximal information already there.

<sup>69</sup> Block and Stalnaker (1999) and Levine (2001) argue against this claim maintaining that *a posteriori* identity statements involving phenomenal concepts are no different than standard *a posteriori* necessities: neither can be derived *a priori* from the physicalist premise base. We are sympathetic. But here we would like to give this point to our anti-physicalist opponent since we think that even then our opponent's argument doesn't go through. Also, we think that, as Levine (2001) shows, even if it is true that neither can be derived *a priori*, since there is still a difference between the two cases, this can be turned into a parallel anti-physicalist argument in much the same way Levine does by appealing to the "gappy" identities involved in the mind-brain case.

<sup>70</sup> As we said, we use the terminology of 'apparent/genuine conceivability' in a technical sense which we have characterized above. This choice of terminology is not entirely a happy one, especially in light of the fact that we are committed to the claim that zombie-worlds are *metaphysically impossible*. This entails that the genuine conceivability of zombies, as this notion is developed and used here, must still be a cognitive illusion—albeit a very different one than the ones created by the denial of standard *a posteriori* necessary scientific identities. Levine's (2001: 87ff.) terminology of 'thin/thick conceivability' might have been a better choice here, but his distinction is not exactly the same as ours—although very similar. But having said this, and explained what we mean by the distinction, we will stick with this terminology until we find a better one.

<sup>71</sup> Exactly what this way is needs to be spelled out, for it seems that not all deployments of sensory/phenomenal concepts in thought experiments are such as to require suspension of (BP).

<sup>72</sup> The arguments from absent qualia and spectrum inversion are species of conceivability arguments. To this extent, our response to these should be predictable from what we have said so far. We intend to elaborate on these elsewhere. But very briefly: we obviously deny that absent qualia cases are possible on the ground that any creature that instantiates a certain information-processing architecture (of the sort we have partly specified) will have qualitative sensory states that are conscious. We accept the possibility of inverted spectrum cases with some qualifications but think that our account actually predicts the possibility of such cases, and that therefore they don't threaten physicalism/functionalism of the sort we envision.

The most important qualification to be noted is that inversion scenarios generally involve consequential complications that are often overlooked in sketchy descriptions, and these complications constitute constraints imposed on what is in fact conceivably invertible, significantly limiting the space of possibilities. Among these complications that need to be addressed with care are, for instance, the difficulty involved in the inversion or its lack of the "cool/warm" features of colors, as pointed out by Hardin (1987, 1988), and the difficulty presented in the possibility of interpersonal inversions premised on the possibility of intrapersonal inversions, as addressed by Shoemaker (1994a, 2001).

<sup>73</sup> This paper is part of a larger work in progress, a book manuscript tentatively titled *Information and Experience* (to be published by Oxford University Press), where our information-theoretic account is developed in a broader context and further detail. Some of the points it

addresses have been raised in our previous work that have evolved into the present discussion, including Aydede and Güzelde (2000) and Aydede and Güzelde (2001).

Intellectually this essay owes a lot to the insightful work of Armstrong (1968, 1987), Dretske (1981), Loar (1990/97), Levine (1993, 1998, 2001), Lycan (1987, 1996), Rosenthal (1991, 1997), and Shoemaker (1981/97, 1994a, 1994b). We thank them all for their influence on us. Block's (1980, 1995) insistence over the years that materialism must come to grips with qualia or phenomenal consciousness in a serious way was also influential in our attempt to develop an information-theoretic account that we claim can do justice to internalist as well as the nagging "qualiaphile" intuitions that worry non-materialists and (some) materialists alike. We would also like to express our deep gratitude to Fred Dretske for his help, patience, and generosity for the long hours of discussion of this material. It was no easy task to try to convince him that his own brand of full-fledged externalism is not warranted by his own information-theoretic account, and as far as we can tell, we have failed in that. Also many thanks to Fred Adams, José L. Bermúdez, Paul Castle, David Chalmers, Owen Flanagan, Martin Golding, George Graham, Janine Jones, John Kulvicki, Kirk Ludwig, Brian McLaughlin, John Perry, Tom Polger, Philip Robbins, David Sanford, Wade Savage, William Seager, Brian C. Smith, Leopold Stubenberg, Jonathan Weinberg, Gene Witmer, and Bill Wotjak for their thoughtful comments, criticisms, and discussion. While in progress, portions of this work were delivered at the Eastern APA meeting in New York (December 2000), the 27th SPP meeting in Cincinnati (June 2001), and at the NEH Summer Institute on Consciousness and Intentionality at the University of California, Santa Cruz (July 2002), as well as in graduate seminars in philosophy of mind at the University of Chicago, University of Florida, and Duke University, 1999–2003. We thank the audiences and our students for their comments and questions.

## References

- Armstrong, David (1962). *Bodily Sensations*, London: Routledge and Kegan Paul.
- Armstrong, David (1968). *A Materialist Theory of the Mind*, London: Routledge and Kegan Paul.
- Armstrong, David (1987). "Smart and the Secondary Qualities" in *Metaphysics and Morality: Essays in Honor of J. J. C. Smart*, edited by P. Pettit, R. Sylvan, and J. Norman, London: Blackwell Publishers.
- Aydede, Murat (2001). "Naturalism, Introspection, and Direct Realism about Pain," *Consciousness and Emotion*, 2(1): 29–73.
- Aydede, Murat (2003). "Is Introspection Inferential?" in *Privileged Access: Philosophical Accounts of Self-Knowledge*, edited by Brie Gertler, Ashgate Publishing.
- Aydede, Murat (forthcoming-a). "Pain," *Stanford Encyclopedia of Philosophy*, edited by N. Zalta, California: CSLI Publications.
- Aydede, Murat (forthcoming-b). "A Critical and Quasi-Historical Essay on Theories of Pain" in *Pain: New Essays on the Nature and the Methodology of Its Study*, edited by M. Aydede, Cambridge, MA: MIT Press, 2005.
- Aydede, Murat and Güven Güzelde (2000). "Intelligence, Intentionality, and Consciousness: Foundational Problems in Artificial Intelligence," *Journal of Theoretical and Experimental Artificial Intelligence*, 12: 263–277.
- Aydede, Murat and Güven Güzelde (2001). "Consciousness, Conceivability Arguments, and Perspectivalism: The Dialectics of the Debate," *Communication and Cognition*, 34(1/2): 99–122.
- Aydede, Murat and Güven Güzelde (2002). "Some Foundational Problems in the Scientific Study of Pain," *Philosophy of Science*, 69(Supplement): 265–283.
- Attneave, F. (1954). "Some Information Aspects of Visual Perception," *Psychological Review*, 61: 183–193.

- Baars, Bernard (1988). *A Cognitive Theory of Consciousness*, Cambridge, UK: Cambridge University Press.
- Balog, Katalin (1999). "Conceivability, Possibility, and the Mind-Body Problem," *The Philosophical Review*, 108(4).
- Barinaga, M. (1992). "Unraveling the Dark Paradox of 'Blindsight'," *Science*, 258(27): 1438–1439.
- Baron-Cohen, S. (1995). *Mindblindness*, Cambridge, MA: MIT Press.
- Barsalou, Lawrence (1999). "Perceptual Symbol Systems" in *Behavioral and Brain Sciences*, 22(4).
- Barwise, J. and John Perry (1983). *Situations and Attitudes*, Cambridge, MA: MIT Press.
- Berkeley, G. (1710/1977). *A Treatise Concerning the Principles of Human Knowledge*, C. Turbayne (ed.), Indianapolis: Bobbs-Merrill Publishing.
- Biro, John (1991). "Consciousness and Subjectivity" in *Consciousness, Philosophical Issues*, 1, edited by Enrique Villanueva, Atascadero, California: Ridgeview Publishing.
- Block, Ned (1980). "Troubles with Functionalism" in *Readings in Philosophy of Psychology*, Ned Block (ed.), Cambridge, MA: Harvard University Press.
- Block, Ned (1995). "On a Confusion about a Function of Consciousness," *Behavioral and Brain Sciences*, 18(2).
- Block, Ned, Owen Flanagan, and Güven Güzeldere (1997). *The Nature of Consciousness: Philosophical Debates*, Cambridge, MA: MIT Press.
- Block, Ned, and Robert Stalnaker (1999). "Conceptual Analysis, Dualism and the Explanatory Gap," *Philosophical Review*, 108(1): 1–46.
- Broadbent, D. (1958). *Perception and Communication*, Oxford: Pergamon Press.
- Byrne, Alex (1999). "Cosmic Hermeneutics," *Philosophical Perspectives*, 13, Atascadero, CA: Ridgeview Publishing.
- Cam, P. (1985). "Phenomenology and Speech Dispositions," *Philosophical Studies*, 47: 357–368.
- Campion, John, R. Latto, and Y. M. Smith (1983). "Is Blindsight the Effect of Scattered Light, Spared Cortex, and Near-Threshold Vision?," *Behavioral and Brain Sciences*, 3: 423–486.
- Carruthers, Peter (1989). "Brute Experience," *Journal of Philosophy*, 86: 258–69.
- Carruthers, Peter (2000). *Phenomenal Consciousness: A Naturalistic Theory*, Cambridge, UK: Cambridge University Press.
- Chalmers, David (1996). *The Conscious Mind*, Oxford, UK: Oxford University Press.
- Chalmers, David (1999). "Materialism and the Metaphysics of Modality," *Philosophy and Phenomenological Research*, 59: 473–493.
- Chalmers, David and Frank Jackson (2001). "Conceptual Analysis and Reductive Explanation," *Philosophical Review*, 110: 315–361.
- Chalmers, David (2002). "Does Conceivability Entail Possibility?" in *Imagination, Conceivability, and Possibility*, edited by Tamar Gendler and John Hawthorne, Oxford, UK: Oxford University Press.
- Cherry, C. (1957). *On Human Communication: A Review, a Survey and a Criticism*, Cambridge, MA: The Technology Press of Massachusetts Institute of Technology.
- Churchland, Patricia S. (1980). "A Perspective on Mind-Brain Research," *Journal of Philosophy*, 77: 185–207.
- Churchland, Paul M. (1985). "Reduction, Qualia, and the Direct Introspection of Brain States," *Journal of Philosophy*, 82(1).
- Churchland, Paul M. (1989). "Knowing Qualia: A Reply to Jackson" in *A Neurocomputational Perspective: The Nature of Mind and the Structure of Science*, Cambridge, MA: MIT Press.
- Clark, Austen (1993). *Sensory Qualities*, Oxford, UK: Oxford University Press.
- Cowey, A. (1995). "Visual Perception: Blindsight in Real Sight," *Nature*, 377(290).
- Cowey, A. and P. Stoerig (1995). "Blindsight in Monkeys," *Nature*, 373: 247–249.
- Dennett, D. C. (1991). *Consciousness Explained*, Boston: Little, Brown, and Co.

- Dennett, D. C. (1994). "Dretske's Blind Spot," *Philosophical Topics* (The Philosophy of Daniel Dennett), 22(1/2).
- Dixon, N. F. (1971). *Subliminal Perception: The Nature of a Controversy*, New York: McGraw Hill.
- Dixon, N. F. (1981). *Preconscious Processing*, New York: Wiley and Sons.
- Dretske, Fred (1981). *Knowledge and the Flow of Information*, Cambridge, MA: MIT Press.
- Dretske, Fred (1986). "Misrepresentation" in *Belief: Form, Content, and Function*, edited by R. Bogdan, Oxford: Clarendon Press.
- Dretske, Fred (1994). "Differences that Make No Difference," *Philosophical Topics* (The Philosophy of Daniel Dennett), 22(1/2).
- Dretske, Fred (1995). *Naturalizing the Mind*, Cambridge, MA: MIT Press.
- Dretske, Fred (1997). "What Good Is Consciousness?," *Canadian Journal of Philosophy*, 27(1): 1–15.
- Evans, Gareth (1982). *The Varieties of Reference*, Oxford: Clarendon Press.
- Farah, Martha. (1995). "Visual Perception and Visual Awareness after Brain Damage: A Tutorial Overview" in *Attention and Performance: XV*, edited by C. Umila and M. Moscovitch, Cambridge, MA: MIT Press, pp. 37–75.
- Farrell, B.A. (1950). "Experience," *Mind*, 59: 170–198. Reprinted in *The Philosophy of Mind*, edited by V.C. Chappell, Englewood Cliffs, New Jersey: Prentice-Hall, 1962.
- Feigl, H. (1967). *The "Mental" and the "Physical": The Essay and a Postscript*, Minneapolis: University of Minnesota Press.
- Fendrich, Robert C., Mark Wessinger, and Michael S. Gazzaniga (1992). "Residual Vision in a Scotoma: Implications for Blindsight," *Science*, 258: 1489–1491.
- Flanagan, O. (1992). *Consciousness Reconsidered*, Cambridge: The MIT Press.
- Frith, U. and Happé, F. (1999). "Theory of Mind and Self Consciousness: What Is It Like to be Autistic?," *Mind and Language*, 14: 1–22.
- Flavell, J., Green, F., and Flavell, E. (1986). *Development of Knowledge about the Appearance-Reality Distinction*, Chicago, Illinois: Society for Research in Child Development.
- Flavell, J., Everett, B., Croft, K. and Flavell, E. (1981). "Young Children's Knowledge about Visual Perception," *Developmental Psychology*, 17: 99–103.
- Flavell, J., Flavell, E., and Green, F. (1986). "Young Children's Knowledge about the Apparent-real and Pretend-real Distinctions," *Developmental Psychology*, 23: 816–822.
- Fodor, Jerry A. (1975). *The Language of Thought*, Cambridge, Massachusetts: Harvard University Press.
- Fodor, Jerry A. (1981). "Methodological Solipsism Considered as a Research Strategy in Cognitive Psychology" in *RePresentations: Philosophical Essays on the Foundations of Cognitive Science*, Cambridge, Massachusetts: MIT Press. Originally appeared in *Behavioral and Brain Sciences* 3(1), 1980.
- Fodor, Jerry A. (1987). *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*, Cambridge, Massachusetts: MIT Press.
- Fodor, Jerry A. (1990). "A Theory of Content" (I & II) in *A Theory of Content and Other Essays*, Cambridge, Massachusetts: MIT Press.
- Fodor, Jerry A. (1994). *The Elm and the Expert*, Cambridge, Massachusetts: MIT Press.
- Gazzaniga, M., R. F. Fendrich, and C. M. Wessinger (1994). "Blindsight Reconsidered," *Current Directions in Psychological Science*, 3: 93–96.
- Gibson, J. J. (1966). *The senses considered as perceptual systems*, Houghton Mifflin.
- Goodale, M. A., L. S. Jakobson, and J. M. Keillor (1994). "Differences in the Visual Control of Pantomimed and Natural Grasping Movements," *Neuropsychologia*, 32(10): 1159–1178.
- Gopnik, A. (1993). "How We Know Our Own Minds: The Illusion of First-person Knowledge of Intentionality," *Behavioral and Brain Sciences*, 16: 1–14.

- Gopnik, A. and Astington, J. (1988). "Children's Understanding of Representational Change and Its Relation to the Understanding of False Belief and the Appearance-reality Distinction," *Child Development*, 59: 26–37.
- Gopnik, A. and Graf, P. (1988). "Knowing How You Know: Young Children's Ability to Identify and Remember the Sources of Their Beliefs," *Child Development*, 59: 1366–1371.
- Gopnik, A. and Meltzoff, A. (1994). "Minds, Bodies, and Persons: Young Children's Understanding of the Self and Others as Reflected in Imitation and Theory of Mind Research" in *Self-awareness in Animals and Humans*, edited by S. Parker, R. Mitchell, and M. Boccia, New York: Cambridge University Press.
- Gopnik, A. and Slaughter, V. (1991). "Young Children's Understanding of Changes in Their Mental States," *Child Development*, 62: 98–110.
- Gopnik, A. and Wellman, H. (1994). "The Theory Theory" in *Mapping the Mind*, edited by S. Gelman and L. Hirschfeld, Cambridge, UK: Cambridge University Press.
- Güzeldere, Güven (1995). "Is Consciousness the Perception of What Passes in One's Own Mind?" in *Conscious Experience*, edited by T. Metzinger, Exeter, UK: Imprint Academic, pp. 335–357.
- Güzeldere, Güven (1997). "The Many Faces of Consciousness: A Field Guide," in *The Nature of Consciousness: Philosophical Debates*, edited by Ned Block, Owen Flanagan, and Güven Güzeldere, Cambridge, MA: MIT Press.
- Güzeldere, Güven, Valerie Hardcastle, and Owen Flanagan (1999). "The Nature and Function of Consciousness: Lessons from Blindsight" in *The New Cognitive Neurosciences*, edited by M. Gazzaniga, Cambridge, MA: MIT Press.
- Güzeldere, Güven, Edward Nahmias, and Robert Deane (2002). "Darwin's Continuum: Building Blocks of Deception" in *The Cognitive Animal*, edited by C. Allen, M. Bekoff, G. Burghardt, Cambridge MA: The MIT Press.
- Güzeldere, Güven (2003). "Zombies" in *The MacMillan Encyclopedia of Cognitive Science*, London: Nature Publishing Company.
- Güzeldere, Güven (in prep.). "Blindsight: An Empirical and Philosophical Controversy."
- Hanson, Philip (1990). *Information, Language, and Cognition*, Oxford: Oxford University Press.
- Hardin, C. (1987). "Qualia and Materialism: Closing the Explanatory Gap," *Philosophy and Phenomenological Research*, 48(2): 281–298.
- Hardin, C. (1988). *Color for Philosophers*, Indianapolis: Hackett Publishing.
- Hart, W. D. (1988). *The Engines of the Soul*, Cambridge, UK: Cambridge University Press.
- Heil, J. (1983). *Perception and Cognition*, Berkeley: University of California Press.
- Hilbert, David (1987). *Color and Color Perception*, Stanford, CA: CSLI Publications.
- Hilbert, David (ms.). "Why Have Experiences?" draft, University of Illinois at Chicago.
- Hill, Christopher (1991). *Sensations: A Defense of Type Materialism*. London, UK: Cambridge University Press.
- Hill, Christopher (1997). "Imaginability, Conceivability, Possibility and the Mind-Body Problem," *Philosophical Studies* 87: 61–85.
- Hill, Christopher and Brian McLaughlin (1999). "There Are Fewer Things in Reality Than Are Dreamt of in Chalmers's Philosophy," *Philosophy and Phenomenological Research*, 59(2).
- Holender, D. (1986). "Semantic Activation Without Conscious Identification in Dichotic Listening, Parafoveal Vision, and Visual Masking: A Survey and Appraisal," *Behavioral and Brain Sciences*, 9: 1–23.
- Huettel, Scott, Güven Güzeldere, and Gregory McCarthy (2001). "Dissociating Components of Visual Attention in a Change Blindness Task Using Response-Contingent Event-Related fMRI," *Journal of Cognitive Neuroscience*.
- Israel, D. and John Perry (1990). "What Is Information?" in *Information, Language and Cognition*, edited by P. Hanson, Vancouver: University of British Columbia Press.



- Jackson, Frank (1982). "Epiphenomenal Qualia," *Philosophical Quarterly*, 32: 127–136.
- Jackson, Frank (1986). "What Mary Didn't Know," *Journal of Philosophy*, 83(5): 291–295.
- Jackson, Frank (1994). "Finding the Mind in the Natural World" in *Philosophy and the Cognitive Sciences: Proceedings of the 16th International Wittgenstein Symposium*, edited by R. Casati, B. Smith, and G. White, Vienna: Verlag Holder-Pitcler-Tempsky, pp. 101–112.
- Kanwisher, Nancy (2001). "Neural Events and Perceptual Awareness," *Cognition*, 79: 89–113.
- Kihlstrom, J. (1984). "Conscious, Subconscious, Unconscious: A Cognitive Perspective" in *The Unconscious Reconsidered*, edited by K. Bowers and D. Meichenbaum, New York: John Wiley and Sons, pp. 149–211.
- Kihlstrom, J. (1987). "The Cognitive Unconscious," *Science*, 237(4821): 1445–1452.
- Kohler, E., C. Keyser, M. A. Umiltà, L. Fogassi, V. Gallese, and G. Rizzolatti (2002). "Hearing Sounds, Understanding Actions: Action Representation in Mirror Neurons," *Science*, 297: 846–848.
- Köhler, S. and M. Moscovitch. (1997). "Unconscious Visual Processing in Neuropsychological Syndromes: A Survey of the Literature and Evaluation of Models of Consciousness" in *Cognitive Neuroscience*, edited by M. Rugg, Cambridge, MA: MIT Press, pp. 305–373.
- Kripke, Saul (1970, lectures). *Naming and Necessity*, Cambridge, MA: Harvard University Press, 1980.
- Kulvicki, John (2001). *On Images: Pictures and Perceptual Representations*, Ph.D. Dissertation, University of Chicago, Philosophy Department.
- Kulvicki, John (forthcoming-a). "Isomorphism in Information-carrying Systems" to appear in *Pacific Philosophical Quarterly*.
- Kulvicki, John (forthcoming-b). "Perceptual Content, Information, and the Primary/Secondary Quality Distinction" to appear in *Philosophical Studies*.
- Laurence, Stephen and Eric Margolis (2002). "Radical Concept Nativism," *Cognition*, 86: 22–55.
- Lempers, J., Flavell, E., and Flavell, J. (1977). "The Development in Very Young Children of Tacit Knowledge Concerning Visual Perception," *Genetic Psychology Monographs*, 95: 3–53.
- Lewis, David (1966). "An Argument for the Identity Theory," *Journal of Philosophy*, 63: 17–25. Reprinted in *Philosophical Papers, Vol. 1* (Oxford University Press, 1980).
- Lewis, David (1972). "Psychophysical and Theoretical Identifications," *Australasian Journal of Philosophy*, 50: 249–258. Reprinted in *Readings in the Philosophy of Psychology, Vol. 1* edited by N. Block, Cambridge, MA: MIT Press, 1980.
- Levine, Joseph (1993). "On Leaving Out What It's Like" in *Consciousness*, edited by Martin Davis and Glyn W. Humphreys, Oxford, UK: Basil Blackwell.
- Levine, Joseph (1998). "Conceivability and the Metaphysics of Mind," *Noûs*, 32(4): 449–480.
- Levine, Joseph (2001). *Purple Haze: The Puzzle of Consciousness*, Oxford, UK: Oxford University Press.
- Loar, Brian (1990/1997). "Phenomenal States" in *The Nature of Consciousness: Philosophical Debates*, edited by Ned Block, Owen Flanagan, and Güven Güzeldere, Cambridge, Massachusetts: MIT Press.
- Loar, Brian (1999). "David Chalmers's *The Conscious Mind*," *Philosophy and Phenomenological Research*, 59(2): 465–472.
- Lotze, Martin, Pedro Montoya, Michael Erb, Ernst Hülsmann, Herta Flor, Uwe Klose, Niels Birbaumer, and Wolfgang Grodd (1999). "Activation of Cortical and Cerebellar Motor Areas during Executed and Imagined Hand Movements: An fMRI Study," *Journal of Cognitive Neuroscience*, 11: 491–501.
- Lycan, William G. (1987). *Consciousness*, Cambridge, Cambridge, MA: MIT Press.
- Lycan, William G. (1996). *Consciousness and Experience*, Cambridge, MA: MIT Press.
- Marcel, A. J. (1983a). "Conscious and Unconscious Perception: Experiments on Visual Masking and Word Recognition," *Cognitive Psychology*, 15: 197–237.



- Marcel, A. J. (1983b). "Conscious and Unconscious Perception: An Approach to the Relations between Phenomenal Experience and Perceptual Processes," *Cognitive Psychology*, 15: 238–300.
- Margolis, Eric (1998). "How to Acquire a Concept," *Mind and Language*, 13(3): 347–369.
- McGinn, Colin (1991). *The Problem of Consciousness*, Oxford, UK: Basil Blackwell.
- Milner, A.D. and M.A. Goodale (1995). *The Visual Brain in Action*, Oxford, UK: Oxford University Press.
- Nagel, Thomas (1974). "What Is It Like to Be a Bat?," *Philosophical Review*, 83: 435–450.
- Neisser, U. (1967). *Cognitive Psychology*, Englewood Cliffs: Prentice-Hall.
- Neisser, U. (1976). *Cognition and Reality*, San Francisco: Freeman and Co.
- Nichols, Shaun and Stephen Stich (2002). "How to Read Your Own Mind: A Cognitive Theory of Self-consciousness" in *Consciousness: New Philosophical Essays*, edited by Q. Smith and A. Jokić, Oxford, UK: Oxford University Press.
- Nichols, Shaun and Stephen Stich (forthcoming). "Reading One's Own Mind: Self-awareness and Developmental Psychology" in *Working through Thought*, edited by R. Kukla, R. Manning, and R. Stainton, Boulder, Colorado: Westview Press.
- O'Neill, D. and Gopnik, A. (1991). "Young Children's Understanding of the Sources of Their Beliefs," *Developmental Psychology*, 27: 390–397.
- O'Neill, D., J. Astington, and J. Flavell (1992). "Young Children's Understanding of the Role that Sensory Experiences Play in Knowledge Acquisition," *Child Development*, 63: 474–491.
- Papineau, David (1993). "Physicalism, Consciousness, and the Antipathetic Fallacy," *Australasian Journal of Philosophy*, 71: 169–183.
- Papineau, David (2002). *Thinking about Consciousness*, Oxford, UK: Oxford University Press.
- Pereboom, Derk (1994). "Bats, Brain Scientists, and the Limitations of Introspection," *Philosophy and Phenomenological Research*, 54(2): 315–329.
- Perner, Joseph (1991). *Understanding the Representational Mind*, Cambridge, MA: MIT Press.
- Perry, John (1979). "The Problem of the Essential Indexical," *Noûs*, 13: 3–21.
- Perry, John (2001). *Possibility, Consciousness and Conceivability*, Cambridge, MA: MIT Press.
- Pöppel, E., R. Held, and D. Frost (1973). "Residual Visual Function after Brain Wounds Involving the Central Visual Pathways in Man," *Nature*, 243: 295–296.
- Prinz, Jesse (1997). *Perceptual Cognition*, Ph.D. dissertation, Department of Philosophy, University of Chicago.
- Prinz, Jesse (2002). *Furnishing the Mind: Concepts and Their Perceptual Basis*, Cambridge, Massachusetts: MIT Press.
- Ptito, A., A. Fortin, and M. Ptito (2001). "'Seeing' in the Blind Hemifield Following Hemispherectomy," *Progress in Brain Research*, 134: 367–378.
- Raij, Tommi. (1999). "Patterns of Brain Activity during Visual Imagery of Letters," *Journal of Cognitive Neuroscience*, 11: 282–299.
- Reingold, E. and P. Merikle. (1990). "On the Inter-relatedness of Theory and Measurement in the Study of Unconscious Processes," *Mind and Language*, 5(1): 9–28.
- Rey, Georges (1997). *Contemporary Philosophy of Mind*, Oxford, UK: Blackwell Publishers.
- Rizzolatti, G., L. Fogassi, and V. Gallese (1997). "Parietal Cortex: From Sight to Action," *Current Opinion in Neurobiology*, 7: 562–567.
- Robinson, H. (1982). *Matter and Sense*, Cambridge, UK: Cambridge University Press.
- Rosenthal, David (1991). "The Independence of Consciousness and Sensory Quality" in *Consciousness, Philosophical Issues*, 1, edited by Enrique Villanueva, Atascadero, California: Ridgeview Publishing.
- Rosenthal, David (1997). "A Theory of Consciousness" in *The Nature of Consciousness*, edited by Ned Block, Owen Flanagan, Güven Güzeldere, Cambridge, Massachusetts: MIT Press, 1997.
- Rosenthal, David (2001). "Introspection and Self-Interpretation," *Philosophical Topics* (Special Issue on Introspection), 28(2).

- Sacks, Oliver (1995). *An Anthropologist on Mars*, New York: Vintage.
- Sayre, Kenneth M. (1965). *Recognition: A Study in the Philosophy of Artificial Intelligence*, Indiana: University of Notre Dame Press.
- Searle, J. (1979). "Intentionality of Intention and Action" in *Expression and Meaning*, Cambridge: Cambridge University Press.
- Sellars, W. (1956). "Empiricism and the Philosophy of Mind" in *Minnesota Studies in the Philosophy of Science*, Vol.1, University of Minnesota Press. Reprinted in Sellars' *Science, Perception and Reality*, London, UK: Routledge and Kegan Paul, 1963.
- Senkfor, Ava J., Cyma Van Petten, and Marta Kutas (2002). "Episodic Action Memory for Real Objects: An ERP Investigation With Perform, Watch, and Imagine Action Encoding Tasks Versus a Non-Action Encoding Task," *Journal of Cognitive Neuroscience*, 14: 402–419.
- Shepard, R. and L. Cooper (1992). "Representation of Colors in the Blind, Color-blind, and Normally Sighted," *Psychological Science*, 3: 97–104.
- Shoemaker, Sidney (1981/1997). "The Inverted Spectrum," *Journal of Philosophy*, 74(7): 357–381. Reprinted in *The Nature of Consciousness: Philosophical Debates* (with a postscript), edited by Ned Block, Owen Flanagan, and Güven Güzeldere, Cambridge, MA: MIT Press.
- Shoemaker, Sidney (1984). "Self-reference and self-awareness" in *Identity, Cause, and Mind*, Cambridge, UK: Cambridge University Press, pp. 6–18.
- Shoemaker, Sidney (1994a). "Phenomenal Character," *Noûs*, 28: 21–38.
- Shoemaker, Sydney (1994b). "Self-Knowledge and 'Inner Sense'," *Philosophy and Phenomenological Research*, 54(2): 249–314.
- Shoemaker, Sidney (2001). "Introspection and Phenomenal Character," *Philosophical Topics* (Special Issue on Introspection), 28(2).
- Siewart, Charles (1998). *The Significance of Consciousness*, Princeton, NJ: Princeton University Press.
- Smart, J.J.C. (1959). "Sensations and Brain Processes," *The Philosophical Review*, 68: 141–156.
- Stoerig, P. and A. Cowey (1992). "Wavelength Sensitivity in Blindsight," *Brain*, 115: 425–444.
- Sturgeon, Scott (1994). "The Epistemic View of Subjectivity," *Journal of Philosophy*, 91(5): 221–235.
- Sturgeon, Scott (2000). *Matters of Mind: Consciousness, Reason and Nature*, London: Routledge (International Library of Philosophy).
- Tye, Michael (1993). "Blindsight, the Absent Qualia Hypothesis, and the Mystery of Consciousness" in *Philosophy and Cognitive Science*, edited by C. Hookway and D. Peterson, *Royal Institute of Philosophy Supplement*, 34: 19–40.
- Tye, Michael (1995). *Ten Problems of Consciousness: A Representational Theory of the Phenomenal Mind*, Cambridge, MA: MIT Press.
- Tye, Michael (1999). "Phenomenal Consciousness: The Explanatory Gap as a Cognitive Illusion," *Mind*, 108: 705–725.
- Tye, Michael (2000). *Consciousness, Color, and Content*, Cambridge, MA: MIT Press.
- Uttal, R.W. (1968). *Real-Time Computers: Technique and Applications in the Psychological Sciences*, New York: Harper & Row.
- Van Gulick, R. (1989). "What Difference Does Consciousness Make?," *Philosophical Topics*, 17: 211–230.
- Weiskrantz, L. (1986). *Blindsight: A Case Study and Implications*, Oxford: Oxford University Press.
- Weiskrantz, Larry (1997). *Consciousness Lost and Found*, Oxford: Oxford University Press.
- Velmans, M. (1991). "Is Human Information Processing Conscious?" *Behavioral and Brain Sciences*, 14: 651–726.
- Wimmer, H. and Hartl, M. (1991). "The Cartesian View and the Theory View of Mind: Developmental Evidence From Understanding False Belief in Self and Other," *British Journal of Developmental Psychology*, 9: 125–128.
- Yablo, Stephen (1993). "Is Conceivability a Guide to Possibility?," *Philosophy and Phenomenological Research*, 80(1): 1–42.