Does the claim that phenomenology overflows cognitive access rest on an illusion?[1]

Pierre Jacob
Institut Jean Nicod,
Ecole Normale Supérieure,
Pavillon Jardin,
29, rue d'Ulm,
75005 Paris,
France

Abstract

Is the phenomenology of visual perception limited to what is accessible to the cognitive mechanisms of attention and working memory, which are necessary for report, as advocates of the global neuronal workspace model of consciousness have argued? Or else does the phenomenology of visual perception overflow cognitive access, as Ned Block maintains? In this paper, I argue that the current evidence from the science of vision supports Block's thesis that visual phenomenology overflows cognitive access. I further try to rebut several attempts by advocates of the global neuronal workspace model of consciousness, who claim that Block's thesis rests on a cognitive illusion. Finally, I point out that acceptance of Block's thesis raises in turn the deep puzzle that people might have experiences of which they are not aware.

[1] I dedicate this paper to the memory of Hilary Putnam. I first met him and attended his lectures in the Fall of 1973. I was baffled by his brilliance.

Introduction

In the present paper, I address an issue that lies at the interface between the scientific investigation of visual experience and the philosophy of perception. I borrow the terminology of my title from the recent work of Ned Block. Like Ned Block, but a few years after him, I too was Hilary Putnam's student — for five years between 1973 and 1978. The issue to be addressed here is whether the phenomenology of an individual's visual perception should be identified with the content of what she can report verbally or otherwise.[2]

In his (1995) paper, Block has argued that the words 'conscious' and 'consciousness' are ambiguous or express a *mongrel* or *hybrid* concept. He further argued that failure to distinguish between *phenomenal* consciousness and *access* consciousness is likely to generate confusions in cognitive science similar to the confusions caused e.g., in early mechanics by Aristotle's failure to distinguish between two distinct senses of 'velocity': *instantaneous* and *average* velocity.

To cut a long story short and focusing on visual perception, while Block (1995, 2005) identified *phenomenal consciousness* with what Nagel (1974) has famously called *what it is like* to have a visual experience, Block's notion of *access consciousness* is best captured by what advocates of the *global neuronal workspace* model of consciousness mean by 'conscious information' — a model widely credited to Baars (1988).[3] Not all information processed in the visual cortex is consciously processed. As Block (2005) observes, one of the basic assumptions of the global neuronal workspace model of consciousness is that there is a "winner-take-all competition" among visual stimuli to be broadcast in a global workspace. In Dennett's (2001) apt political metaphor for the global workspace model, visual stimuli

[2] In addressing this issue, I am, I think, faithful to Putnam's (1975, p. xvii) nuanced meta-philosophical picture, according to which philosophy is a "science" to the extent that it is the task of philosophers to formulate "rival research programmes" and to modify their hypotheses "by trial and error."

[3] For further elaboration, cf. Dehaene and Naccache (2001), Dehaene and Changeux (2004), Dehaene et al. (2006), Kouider et al. (2007).

compete for 'fame in the brain': "only few visual contents become elevated to this political power… hang around, monopolizing time in the limelight."[4] On this model, what makes a piece of visual information conscious is that it is made available (i.e., broadcast) to a global neuronal workspace, which is itself constituted by a network of *consumer* systems (including attention, working memory, perceptual categorization, reasoning, planning, evaluation of alternatives, decision-making, and more generally, rational control of action). Block's notion of access-consciousness is meant to capture the availability of visual information to a variety of cognitive mechanisms necessary for report.

In his (2007a, 2008) papers, Block has argued that the current evidence from the cognitive neuroscience of vision is best accommodated by the hypothesis that the neural machinery underlying cognitive accessibility is not a constitutive part of the neural machinery underlying visual phenomenology. Block's claim has been hotly disputed, in particular by philosophical and scientific advocates of the global neuronal workspace model of consciousness, who have repeatedly tried to rebut it by arguing that it rests on some cognitive illusion or other. This dispute is the focus of my paper. One reason I focus on this dispute is that Putnam has recently acknowledged the impact of Block's disputed claim on his own thinking.[5] I start by a succinct reconstruction of Putnam's own approach to some of the metaphysical and epistemological puzzles raised by the phenomenology of visual experiences. As it will turn out, acceptance of Block's thesis has surprising consequences on introspective self-knowledge about one's own visual experiences.

1. Putnam on naïve realism and qualia

[4] Dennett (2001, p. 225) reprinted as chapter 6 of Dennett (2005).
[5] Putnam (2016a) says of two papers by Block (2007a, 2007b) that they 'have had an impact on my thinking about the phenomenology of perception comparable to the impact on my later philosophy of mathematics that reading Quine's (1948) "On What There Is" and (1951) "Two Dogmas of Empiricism" in my twenties turned out to have.'

Over the years of his incredibly prolific career, Putnam has been both praised and criticized for changing his mind on several fundamental issues;[6] but he has been incredibly generous in crediting other philosophers for their role in causing these changes. As I see it, Putnam's published work reflects two related features that are quite unique in contemporary philosophy: one is Putnam's unrivalled ability to display through his writing the extent to which the quality of a philosopher's arguments depends on the challenges of other philosophers. The other is his unrivalled ability to display in his writing the extent to which addressing philosophical issues is entangled with the intellectual experience of feeling "torn" between conflicting positions.[7]

One philosophical issue about which Putnam's writings over the past three decades exemplify these two features is how to best think of *qualia* (sensations or *sense-data*), i.e., mental representations with a qualitative or phenomenal character. Do *qualia* exist? Can the individual who experiences a *quale* know it? Is the individual's knowledge about it infallible? Are *qualia* expressible by ordinary natural language predicates? Can *qualia* be known by science? Putnam (1994, pp. 445-446) himself has warned us against what he describes as 'the pattern of "recoil" that causes philosophy to leap from frying pan to fire, from fire to a different frying pan, from different frying pan to a different fire, and so on, apparently without end," and he has urged us to *overcome* it. My account of Putnam's philosophical struggle over the nature of *qualia* starts with his (1977) paper, "Models and Reality," in which he criticized what he described as Austin's and Skinner's respective attempts at *driving sense-data out of existence* — a train of thought Dennett (1988) later called *quining qualia*.[8] As my succinct

[6] As Putnam (1988, p. xi) ironically once put it, "the fact that I change my mind in philosophy has been viewed as a character defect. When I am lighthearted, I retort that it might be that I change my mind so often because I make mistakes, and that other philosophers don't change *their* minds because *they* simply never make mistakes".
[7] As Putnam himself non-ironically put it in a Note to his (1978) paper "There is at least one a priori truth," 'we philosophers are frequently torn between opposing considerations, but we rarely show it in print. What we do is let our selves be torn in private until we finally "plonk" for one alternative or the other; then the published paper only shows what we plonked for, and not the being-torn. For once, the preceding paper-plus-potentially-infinite series-of-notes will show the "being torn"', reprinted in Putnam (1983, p. 111-112).
[8] In "Models and Reality", reprinted in Putnam (1983, p. 15).

account in this section will hopefully reveal, I do not think of Putnam's writings on the topic of *qualia* as embodying a "pattern of recoil" *to be overcome*. I think of them instead as making manifest — to himself and others — the insuperable intellectual experience of feeling torn between contrasting metaphysical positions. I take it that the genuine experience of feeling torn can be put to rest temporarily, but the appeal to overcome the feeling once and for all sounds utopian.[9]

In his 1981 book, *Reason, Truth and History*, while Putnam endorsed *internal* realism, his way of tackling the nature of *qualia* lay at the intersection of two preoccupations: the mind/body problem and the rejection of metaphysical realism. On the one hand, he used an intra-subjective version of the inverted spectrum scenario to argue that one might not have introspective knowledge of the quality of one's visual experiences, which might lie beyond what one might report using ordinary color terms (pp. 80-81). He further argued that the quality of visual sensations defeats functionalism, construed as "the right *naturalistic* description of the mind/body relation" (p. 79). But far from either repudiating *qualia* or opting for ontological dualism, he instead endorsed a rather traditional version of the physicalist identity thesis for *qualia* and urged that they should be identified with some physical property of the brain. On the other hand, he was also concerned to criticize the metaphysical realist interpretation of the physicalist identity thesis applied to *qualia*. Thus, he argued that, in light of the many alternative ways of specifying the correlations between sensations and physical states of the brain, *knowing* which mind-brain identity (or correlation) theory is the correct one is likely to be an unsolvable epistemological problem. In a nutshell, Putnam (1981) did not rule *qualia* out of existence since he was willing to identify them with physical properties of the brain. But he argued that both objective scientific and non-scientific introspective *knowledge* of *qualia* were deeply problematic.

[9] In fact, Putnam (1994, p. 445) did not merely urge us to overcome the pattern of recoil mentioned above, but also to "attempt to understand and, *to the extent that it may be humanly possible*, to overcome" it [my italics].

In his (1994) Dewey Lectures, reprinted as part of his (1999) book, *The Threefold Cord: Mind*, *Body and World,* Putnam shifted from *internal* realism to a metaphysical position, which, following William James, he called "the natural realism of the common man" (or *natural realism* for short). There, he came closest to yielding to the temptation of ruling *qualia* out of existence, as he took "the idea that there has to be an interface between our cognitive powers and the external world" to be a *disaster*.[10] His rejection of the mind-world *interface* itself laid at the intersection of two converging epistemological strands of thought, one of which was his acceptance of John Austin's (1962) attack on *sense-data* in *Sense and Sensibilia*, and the other of which was his endorsement of John McDowell's (1994) rejection of visual *qualia* construed as unconceptualized human visual experiences, on the grounds that they are bound to fail to constitute genuine *epistemic justifications* (or *reasons*) for perceptual beliefs. As a result, Putnam embraced a version of the *disjunctivist* position according to which only veridical perception enables us to "reach all the way to the objects themselves" or "things out there" (Putnam, 1994, p. 454).[11] In addition to ruling *qualia* out of existence, acceptance of disjunctivism also involves the rejection of the assumption that veridical and non-veridical perception may share a *common* representational and qualitative (or phenomenal) *factor*.[12] Furthermore, to the extent that disjunctivism construes veridical perception as a relation to *mind-independent* objects it may be hard to reconcile *natural* realism with *internal* realism.

Recently, as a result of his acquaintance with two papers by Ned Block, both published in 2007, Putnam (2016a) has backtracked from endorsing disjunctivism and ruling

[10] Putnam (1994), p. 453

[11] For a critical appraisal of Putnam's rejection of the mind-world interface in the philosophy of perception and a defense of a representationalist framework, cf. Fodor (2000).

[12] Putnam (2008, p. 111) wrote: 'it was not until I became aware of the "disjunctivist school" in the philosophy of perception, of which John McDowell is today one of the most distinguished representatives, that I came to see it is possible to defend what James called "the natural realism of the common man"'.

qualia out of existence.[13] In his "Wittgenstein and *qualia*," Block spells out what he calls the *dangerous* version of the inverted spectrum hypothesis, acceptance of which shows that ordinary language color predicates like 'looks red' are bound to fail to capture what it is like to have visual color experiences. Block's (2007b) thesis that *qualia* are not expressible in ordinary color terms seems to fit Putnam's (1981) earlier skeptical view that both scientific knowledge and non-scientific introspective knowledge of *qualia* face insuperable epistemological difficulties. However, far from accepting the view that there might be no fact of the matter with respect to "sameness of qualitative character," Block (2007a, 2007b) instead has argued for the joint claims that the neuroscientific evidence supports the physicalist identification between *qualia* and states of the early visual primary cortex and that scientific knowledge of the identity cannot be based on the contents of subjects' introspective reports. Furthermore, Block (2007b, p. 78) argues that one can *attend* to the content of one's own perceptual experience. Putnam's (2016a) acceptance of Block's (2007a) scientific research program into the neural basis of visual experience has caused him to express reservations towards the idea of the introspective *transparency* (or *diaphanousness*) of experience.[14] Consequently, he is thereby led to re-evaluate his earlier (1994, 1999) diagnosis that the idea of the mind-world interface is a disaster.

2. A tripartite distinction

While in his (1995) paper, Block stressed the *binary* contrast between phenomenal and access consciousness, Block's (2007a) approach to visual *qualia* really involves a *tripartite* distinction between the *phenomenology* of visual perception (i.e., *visual phenomenology*), *cognitive accessibility* and *report*. As I already mentioned, Block (1995, 2005, 2007a)

[13] Cf. Putnam's quote in footnote 4. In his latest paper, in which he refers to Block's recent work as eye-opening, Putnam calls 'naïve' the version of realism, which he now favors, by reference to the doctrine that Russell (1912) rejected.

[14] For some further remarks, see section 7 of the present paper.

endorses the global neuronal workspace model of consciousness as a good model of what he calls *access* consciousness, not phenomenal consciousness. Advocates of the global workspace model assume a *modular* view of brain activity, in which independent neuronal networks process *unconscious* information in parallel (cf. Dehaene and Naccache, 2001). They further assume that the modularity of unconscious information processing is overcome by the establishment of *long distance* neural connections between neuronal activity taking place in e.g., early primary visual occipital cortical areas and neuronal activity taking place in e.g., parietal and frontal cortical areas. Such long distance neuronal connections are taken to underlie the broadcasting of information that creates the global availability of information that is said in turn to give rise to conscious experience and to make it reportable (Dehaene and Changeux, 2004, 2008). On this model, the content of conscious visual experience is equated with reportable content.

Typical evidence for the global workspace model of the reportability of visual stimuli involves use of the *attentional blink* paradigm: participants are shown two visual stimuli T1 and T2 in rapid succession; when T2 is presented between 100 and 500 ms after T1, the ability to report it drops, as if participants' attention had *blinked*. Dehaene, Sergent and Changeux (2003) report that when T2 occurred 180 ms after T1, participants' ability to report T2 could be as low as if T2 were absent, but nonetheless the occipito-temporal event-related potentials evoked by a non-reported T2 could be as large as the event related potentials in the same brain areas when participants could report T2. Thus, intense occipito-temporal activation (of purely visual areas) can be accompanied by a complete lack of conscious report (cf. Dehaene et al. 2006). Furthermore, advocates of the global workspace model of consciousness are prone to express the verificationist worry that insofar as visual experience is entirely divorced from what participants are able to report, it runs the risk of lying beyond the scope of scientific investigation (cf. Dehaene and Changeux, 2004; Kouider et al., 2007).

Nonetheless, there is room for a tripartite distinction between visual phenomenology, cognitive accessibility and report along the following lines. Visual experiences are the output of perceptual processes and they have a distinctive kind of phenomenology (i.e., *visual* phenomenology). Phenomenology is a property of perceptual experience; and the phenomenology of visual experiences depends on the processing of visual attributes of objects in areas of the visual cortex.[15] Thus, visual phenomenology is a property of the experiences generated by *modular* perceptual (visual) processes. Reporting a visual experience is not experiencing, nor is it part of a perceptual process. Reporting (whether verbally or not) is performing an intentional action of a special sort, i.e., a communicative action, or responding to an action performed by a communicative agent. If asked to report what one saw, one expresses (verbally or otherwise) the content of one's belief about what one saw. What one reports is the content of the *belief* formed on the basis of one's visual experience, not one's visual *experience* itself.[16] Whether or not the phenomenology of one's visual experience (visual phenomenology) is necessary for turning the content of one's visual experience into the content of one's reported perceptual belief, it is not sufficient. It is not sufficient because only that part of the content of one's visual experience that is being *accessed* by inter alia attention and working memory can be part of the content of perceptual belief, and hence is reportable. According to the global workspace model, only by overcoming the modularity of purely visual processing is visual information made accessible to attention and working memory, and hence made reportable.

Furthermore, even if the intentional communicative dimension of report can be, to some extent, overcome (e.g., with non-human primates), as the framework of signal detection

[15] By being linked to the modular activity of the visual cortex and the processing of visual attributes of objects, the phenomenology of visual experiences differs from the phenomenology of experiences in different sensory modalities, which depend on the activity of different brain areas and the detection of different attributes of objects. I assume that while the processing of visual attributes is internal to the visual cortex, visual attributes of objects are properties of objects lying in the environment, not properties of the brain. Thus, I assume that the phenomenology of visual experiences does not rest entirely on properties of the brain.

[16] A point stressed by Dretske (1993, 1994), but disputed by Dennett (1991, 1994).

theory makes clear, a subject's report depends not only on the content of her perceptual belief, but on her motivation as well (cf. Block, 2005, p. 48). Subjects may shift their criterion for what counts as a sufficient experiential (or evidential) basis for engaging in report as a function of their representation of either reward for success or punishment for failure.[17] Another way to put the point based on signal detection is to stress that although accessibility is a necessary condition for reportability, it cannot be sufficient. The fact that a piece of information is accessible to an individual's attention and working memory does not guarantee that the individual will find the reward for success or the punishment for failure sufficient to comply with a request to report it.

3. The Overflow interpretation of Sperling's findings

Block's (2007a) thesis that phenomenology overflows cognitive access primarily rests on classical experimental findings reported long ago by Sperling (1960) and more recent converging findings by Lamme and colleagues (cf. Landman et al., 2003). As the opening of his paper makes clear, Sperling was acutely aware of the gap between visual experience and report. As he paradoxically put it (p. 1), 'the apparently simple question: "What did you see?" requires the observer to report both what he remembers and what he has forgotten.' As the meaning of the verb 'to report' implies, what one reports at $t + 1$ is the content of one's visual experience, i.e., an event that occurred at t. Thus, reporting one's experience requires remembering at $t + 1$ an event that occurred at t. But since plausibly more information is available in perceptual experience than in memory, perceptual experience could only be fully reported if one could remember what must have been forgotten.

[17] Some philosophers and cognitive scientists subscribe to the so-called *enactivist* view that perception is an action (cf. O'Regan and Noë, 2001; Noë, 2004). While it is not contentious that 'perception' can both name a process and its resulting output (the perceptual experience), I don't accept the view that perception is an action if that means that there is no distinction to be drawn between perceiving (i.e., detecting objects, properties or facts), and either acting, which consists in turning a possible into an actual state of affairs, or intending, which consists in representing a possible, non-actual state of affairs. Cf. Jacob (2006, 2008).

In a first experiment, Sperling (1960) showed participants an array of 4 x 3 letters presented for 15-500 ms and asked them to report as many letters from the array as possible after stimulus off-set. Sperling found that participants could report 3-4 from the whole array of 12. In a second experiment, participants were asked to report as many letters as possible from a particular row cued by a tone emitted 300 ms after stimulus off-set. Participants heard one of three distinct tones: a high, a medium or a low tone, each of which cued respectively the top, the medium or the low row. In this condition, Sperling (1960) found that participants were able to report 3-4 items from any cued row of four letters. He called this remarkable effect the *partial report superiority*. He further took it to show that information about 9-12 letters is stored in what he called *iconic memory*, i.e., a memory buffer with a large storage capacity, but in which the information is stored in a format inappropriate to support report. Finally, he conjectured that report requires the recoding of the information from an iconic into a digitalized format and its transfer from the iconic memory buffer to *working memory* with a smaller storage capacity but longer persistence than the iconic memory system. One cannot visually perceive a letter unless one perceives a token in either a lower or upper case, in a specific font, with a determinate size and color. But the name by means of which the letter is reported fails to specify the above perceptual features. Segregating information about the identity of a letter from information about the properties of the token that are visually perceived is what digitalization achieves.[18]

Sperling's (1960) evidence for the partial report superiority effect seems to support the dissociation between the mechanisms underlying respectively the phenomenology of visual experience and selective attention, along the following two conspiring lines of thought. First of all, what is controversial is whether letters contained in an array are being consciously represented or encoded in representations with visual phenomenology when participants *fail*

[18] Cf. Dretske (1981, ch. 6) for a sophisticated exploration of the contrast between the iconic (or analog) and digital encoding of information.

to report them. But the following assumption is beyond controversy: the 3-4 letters which *are* being reported 300 ms after stimulus off-set and which belong to a cued row are consciously represented and have been encoded in representations with visual phenomenology. In the context of Sperling's paradigm, visual phenomenology is not sufficient for report, but report is evidence of visual phenomenology. No report of specific letters unless the specific letters were encoded within representations with visual phenomenology. Call this assumption the *visual phenomenological basis of report*.

A second plausible assumption is that the phenomenology of participants' representation of the letters in the cued row after stimulus off-set derives from participants' visual perception of the letters during stimulus presentation. Call the second assumption the *perceptual basis of visual phenomenology*. The alternative to this latter assumption is that the visual phenomenology of letter-representations is somehow retrospectively generated by the attentional cue, which follows the perceptual processing of the visual array of letters. From the *visual phenomenological basis of report* and the *perceptual basis of visual phenomenology*, it follows that the perceptual experience of the reported letters prior to the attentional cue (the tone) must have had visual phenomenology.

Furthermore, it seems hard to reject the following *counterfactual parity* assumption: if any other row *had* been cued by a different tone instead of the actual one, participants *would* have reported 3-4 letters in that hypothetical other row as well. To deny the counterfactual parity assumption is to introduce a mysterious asymmetry among the rows of a Sperling array. Why posit an arbitrary (undemocratic) processing bias favoring the visual information conveyed by one row to the detriment of the other two at any instant? Acceptance of the counterfactual parity assumption entails that the participants' representation of the *hypothetically* 3-4 reported letters from any *hypothetically* cued row *would* have visual phenomenology, just as much as their representations of the actually reported letters from the

actually cued row. From the *visual phenomenological basis of report* and the *counterfactual parity* assumptions, it follows both that participants must have represented 9-12 letters and that their representations of 9-12 letters all had visual phenomenology.[19] As Block (2011, p. 567) puts it, "all or almost all of the 12 items are consciously represented, perhaps fragmentarily but well enough to distinguish among the 26 letters of the alphabet. However, only 3-4 of these items can be cognitively accessed, indicating a larger capacity in conscious phenomenology that in cognitive access."

The goal of opponents of the Overflow interpretation of the Sperling findings is to dispense with Block's notion of phenomenal consciousness as distinct from access consciousness. For example, Kouider et al. (2010) propose a threefold framework involving the notion of *partial* awareness, the availability of a hierarchy of representational levels and the possibility of separate access to different levels. In a nutshell, they propose to replace "dissociable forms of consciousness" by "dissociable levels of access." What the Sperling-like findings show, on Kouider et al.'s (2010, p. 303) view, is not that phenomenology overflows cognitive access but instead that "access overflows verbal report."[20] In support of their contention, they have argued that "dissociative theories are inherently unfalsifiable and beyond the scope of science because inaccessible conscious states are intrinsically off-limits to investigation" (Cohen and Dennett, 2011, p. 358).[21] Kouider et al. (2010, p. 304) ascribe to Block's version of the Overflow thesis what they call "the neural purity" assumption, "according to which phenomenological consciousness can be probed regardless of reportability." They further argue that "this strategy leads to circularity because validation of the neural index necessarily requires reliance on access mechanisms."

---

[19] I am aware that the counterfactual parity assumption has been questioned by Philipps (2011), on the grounds that initial display experience in partial report trials is subject to postdictive effects due to the subsequent cue. I will indirectly address Philipps' retrodictive account when I examine the possibility that visual phenomenology might be retrospectively generated by the presence of an attentional cue.

[20] In Sperling's paradigm, participants refer to letters by names.

[21] This verificationist worry has been repeatedly expressed by advocates of the global neuronal workspace model of consciousness (cf. Dehaene and Changeux, 2004; Dehaene et al., 2006; Kouider et al., 2007).

But on the one hand, as I argued above, while the fact that some information is accessible to attention and working memory is necessary for its being reportable, it is not sufficient for report. On the other hand, as Block (2011, p. 567) has pointed out, the verificationist criticism fails to hit its target because it rests on the further confusion between two distinguishable claims: the stronger claim that *some* items in a given array are such that their representations have *inaccessible* phenomenology and the weaker claim that *not all* items in a given array can be *accessed*, let alone at a single time. The Overflow interpretation of the Sperling findings only requires the weaker, not the stronger, claim. It merely asserts that, at any given time, some items which are being represented with visual phenomenology are not being accessed by attention and working memory. It does not assert that at any given time, there are items whose representations have visual phenomenology and are *inaccessible* to attention and working memory. The fact that some representation with visual phenomenology (or part thereof) remains unaccessed by attention and working memory at a given time, in a given task, does not make it inaccessible in principle, at all times, in every possible task.

4. Change blindness and the Refrigerator Light illusion a

Arguably, the most radical rejection of the Overflow interpretation of the Sperling findings straightforwardly follows from O'Regan and Noë's (2001, p. 960) *enactivist* assertion that experiences are not states, but "ways of acting. They are things we do. There is no introspectibly available property determining the character of one's experiential states, for there are no such states. Hence, there are, in this sense at least, no (visual) *qualia*. *Qualia* are an illusion, and the explanatory gap is no real gap at all." If there are no visual experiences (no *qualia*), then clearly the phenomenology of visual experience cannot exceed what people can report in virtue of the fact that it overflows what is cognitively accessible.

Less radical criticisms of the Overflow interpretation of the Sperling findings appeal to results of *change blindness* experiments as evidence for the existence of a *cognitive illusion of seeing*, whereby humans would spontaneously tend to *overestimate* the richness and fine-grainedness of their visual experience (Dehaene et al., 2006). A particularly attractive version of the cognitive illusion of seeing is the so-called *Refrigerator Light* illusion, i.e., the illusion that the refrigerator light is always on because it is always on when one looks (cf. Block, 2001; Block, 2007a; O'Regan and Noë, 2001).

In typical change blindness experiments, participants have been demonstrated to fail to notice quite large differences between two visual stimuli presented in alternation. The differences are generated by a change concealed by means of skillful experimental techniques, such as the flicker paradigm or the use of mud splashes. In the flicker paradigm, a visual scene and its altered version are switched back and forth at a high rate that makes it difficult for observers to detect the difference. The mud splashes paradigm consists in masking a change between two images by scattering high contrast small shapes over an image while the change takes place. In so-called "forced-choice detection paradigm" experiments, participants are shown two distinct images for a short equal amount of time and asked to detect the difference (cf. Simons and Rensink, 2005 for review).[22]

As the following instance of a common sense version of the change blindness phenomenon that fits the forced-choice detection paradigm illustrates, the failure to detect the relevant difference between two visual stimuli demonstrated by change blindness experiments is open to two rival interpretations, which have been called by Block (2011) respectively the inattentional *blindness* and the inattentional *inaccessibility* interpretations. One may fail to notice the difference between two occurrences of Sam's face, one in which Sam has a moustache at t and the other in which Sam's moustache has been shaved at $t + 1$.

[22] The psychologist Kevin O'Regan (who co-authored with Alva Noë the 2001 paper entitled "A sensorimotor approach to vision and visual consciousness") has done groundbreaking work on change blindness.

Furthermore, each interpretation can make use of a distinctive tripartite distinction in order to account for the phenomenon.

According to inattentional blindness, the reason one fails to notice the difference between Sam's face with and without a moustache is simply that one *failed to see* the difference between Sam's face on both occasions, namely the presence of his moustache at *t* and its absence at *t* + 1 (cf. O'Regan and Noë, 2001). Among the extreme versions of inattentional blindness are the *anti-representationalist* slogans that we should "use the world as its own model" (Brook, 1991), that "the world is an outside memory" (O'Regan, 1992) or that "there's no need to re-present the world on one's own internal memory drive. Off-loading internal processing onto the world simplifies our cognitive lives and makes good engineering and evolutionary sense" (Noë, 2004, p. 50). On this *anti-representationalist* approach to visual perception, we experience the detail of a visual scene as *virtual* or *accessible*. Our experience of the details of a visual scene does not arise from a detailed visual internal representation. The detail belongs so to speak to the outside world; it is not intrinsic to our visual representation of it. We implicitly know that the detail is available for us to retrieve if we need to, and how to retrieve it.[23]

Advocates of the inattentional *blindness* interpretation of change blindness phenomena (who endorse the global workspace model of consciousness) have drawn a tripartite distinction between three kinds of processing of visual information, which they call respectively *subliminal*, *preconscious* and *conscious*, as a function of two main parameters: the *bottom up strength* of a visual stimulus and the *top down amplification* of attention. When the bottom up activity of purely visual occipito-temporal cortical areas is strong and there is top down attentional amplification, the processing is conscious and the stimulus is *seen*.

[23] Noë (2004, pp. 21-22) fluctuates between the radical denial of the view that "vision is a process whereby the brain produces an internal representation of the world" and the more moderate denial of the view that "vision is a matter of generating a detailed internal representation of the visual world on the basis of information *available at the retina alone*" [my emphasis].

When the bottom up activity of purely visual occipito-temporal cortical areas is too weak to reach the global workspace, the processing is subliminal and the stimulus is *invisible*. When the bottom up activity of purely visual occipito-temporal cortical areas is strong, but attention is focused on some other competing stimulus, the processing is preconscious and the stimulus is *visible*, but *not seen* (cf. Dehaene et al., 2006 and Kouider et al., 2010 for detailed versions of this approach). Thus, on the inattentional blindness interpretation of the change blindness phenomenon, one fails to notice the item or property that constitutes the difference between two stimuli (e.g., Sam's moustache), because although it may have been visible, it was unattended and therefore unseen: one failed to see it. The inattentional blindness interpretation supports the claim that humans are prey to the cognitive (Refrigerator light) illusion of seeing more than they do.

But on the inattentional *inaccessibility* interpretation, evidence that one fails to notice the difference is not *ipso facto* evidence that one was blind to, or failed to see (visually experience), the item or property that constitutes the difference.[24] As Dretske (2004, 2007) has emphasized, one may fail to notice the difference between two occurrences of Sam's face, one in which Sam has a moustache and the other in which Sam's moustache has been shaved. Failure to notice the difference does not mean that Sam's face looked to one the same on both occasions: on one occasion Sam's moustache occluded parts of his upper lip; on the other it did not. One saw (or visually experienced) Sam's moustache on the first occasion just as one saw Sam's eyebrows: they both occluded some relevant portion of Sam's skin. One's visual experience of Sam's face must have been different when Sam had shaved his moustache since the shaved moustache failed to occlude Sam's upper lip. Failure to notice the difference arises from failure to categorize or conceptualize the difference, which is required for making a comparison, i.e., coming to judge or believe that Sam's face did not look the same with and

[24] The difference in meaning between the two ordinary English verbs 'to see' and 'to notice' reflects the inattentional inaccessibility interpretation.

without the moustache.

In Dretske's (2004, 2007) terminology, while the change between two stimuli is an event, it causes a difference.[25] The difference between two stimuli can consist of either one object present in one stimulus and absent from the other or in a property being instantiated in one stimulus and not in the other. But of course the difference caused by a change can be visually concealed. If Sam wore a mask on his face when he had a moustache at *t* and one saw his face at that time, then his moustache would have been invisible and one would have failed to see it. Although the presence of Sam's moustache at *t* and its absence at *t* + 1 constitutes the difference between Sam's face at *t* and Sam's face at *t* + 1, one could not notice the difference if one failed to see Sam's moustache at *t* because it was occluded by Sam's mask.

However, failure to *notice* the *fact* (or *believe*) that two stimuli are different is compatible with seeing (or experiencing) the object or property that constitutes the difference. Conversely, seeing the object or property that constitutes the difference is necessary, but not sufficient to notice the difference or to form the belief that the two stimuli are different. Thus, seeing Sam's moustache a *t* and not seeing it at *t* + 1 is not sufficient for noticing the difference between Sam's face at t and Sam's face at *t* + 1. Far from supporting the claim that humans are prey to the cognitive (Refrigerator light) illusion of seeing more than they do, on the inattentional inaccessibility interpretation, results of change blindness experiments are fully consistent with the Overflow interpretation of the Sperling findings: the phenomenology of visual experience outstrips participants' ability to report because only a small portion of visual phenomenology is cognitively accessible (or available) to attention and working memory.

5. The generic phenomenology illusion

[25] As Dretske (2004, p. 3) has emphasized, one of the reasons why participants fail to notice the difference between two stimuli is precisely that experimental psychologists are experts at concealing the change.

In reaction to Dretske's (2007) example of a face with and without a moustache, Tye (2010, pp. 413-414) concedes that it is necessary that an item be visible (e.g., a moth on a tree trunk), that it reflects photons onto an individual's retina and that it occludes other items (e.g., part of the surface of the tree trunk), for it to be consciously seen. But it is not sufficient that "light reflected from the moth carries information that reaches the eyes" for one's representation of the moth to have visual phenomenology. As Tye (2010, p. 420) puts it, "what is represented need not be consciously represented." In a similar vein, Kouider et al. (2010, p. 304) have argued that "a supposed neural index of inaccessible consciousness might… simply reflect an unconscious form of processing." Cohen and Dennett (2011, p. 359) have also attempted to rebut Block's (2007a, 2011) Overflow interpretation of the Sperling partial report superiority effect by arguing that representations of letters may be "stored unconsciously" in the iconic memory buffer "until the cue brings them to the focus of attention." On this view, uncued letters may be non-consciously represented: unattended letters in uncued rows may be represented. But if so, then letter-representations lack visual phenomenology.[26]

Cohen and Dennett's view that letter-representations stored in the iconic memory buffer might lack visual phenomenology amounts to the view that the attentional cue (i.e., the tone) confers visual phenomenology onto the representation of the seen letters. But for three related reasons, this seems odd. First, it sounds odd that a tone that is heard might confer visual phenomenology onto the representations of letters that are visually perceived. Secondly, what participants perceive and presumably store in the iconic memory buffer are not just letters of the alphabet, but alphanumeric characters in either lower or upper case, in a particular font, at a particular distance from one another, with a particular size and color.

[26] This view seems in accordance with Dehaene et al.'s (2006) tripartition (discussed in section 4) between invisible, visible and seen stimuli, where a subliminal stimulus is an instance of an invisible one and the difference between a visible and a seen stimulus is a function of bottom up strength and top down attentional amplification.

When, however, they report what they saw, they use the *names* of the letters of the alphabet thereby abstracting away from the case, font, relative distance, size and color of the perceived symbols. Thirdly, only by denying that the phenomenology of participants' representations of the letters in the cued row after stimulus off-set derives from participants' visual perception of the letters during stimulus presentation (*the perceptual basis of visual phenomenology*) could Cohen and Dennett endorse the idea that the attentional cue could *retrospectively* confer visual phenomenology to the representation of letters.[27]

To support their denial of the *perceptual basis of visual phenomenology*, Kouider et al. (2010) and Cohen and Dennett (2011) appeal to findings by de Gardelle and colleagues (2009), which are interpreted as showing that in a task of so-called *free subjective report*, based on a modified version of Sperling's paradigm, participants fall prey to the cognitive illusion of construing pseudo-letters as letters as a result of their expectation that a visual array contains either letters or else non-letters.

In the modified version of Sperling's paradigm, participants were first exposed for 500 ms to an array containing mostly letters, some non-letters and some rotated letters. After the brief presentation of a masking pattern (for 50 ms), participants heard an auditory cue instructing them to report as many letters as possible from the cued row (that contained only letters). Only non-cued rows (i.e., the unattended part of the visual array) contained either non-letters ('wingdings') or rotated letters. Finally, some trials involved a free subjective report task in which participants were asked to move a cursor over a set of symbols, some of which were rotated letters and non-letters from the non-cued rows: the task was to click each time a symbol belonged to the unattended part of the original array. What de Gardelle et al. (2009) found was that participants did not click on non-letters (wingdings) that had been present in the non-attended part of the array, but that they construed what they call 'pseudo-

[27] This seems to be part of Philipps' (2011) retrospective account.

letters' from the initial array as if they were letters. They interpret the latter finding as evidence for the *illusory* construal of pseudo-letters as letters.

However, what de Gardelle et al. (2009) call 'pseudo-letters' are just letters *in a non-canonical orientation* (e.g., a rotated 'R'). While it is true that evidence of a genuine cognitive illusion whereby non-letters (wingdings) from non-cued rows would be interpreted as letters could support the rejection of the Overflow interpretation of the Sperling finding, it is misleading to say that participants are under an illusion when they recall a letter in a non-canonical orientation as a letter. It is misleading because a letter in a non-canonical orientation is a letter, not a wingding. Thus, perceiving a rotated letter as a letter should not count as an illusion.

In the same vein as Cohen and Dennett (2011), several commentators of Block's (2007a) paper have argued that in Sperling's partial report superiority effect, participants confuse what can be called the *generic phenomenology* of the representation of the instantiation of the property *letterness* (or *letterhood*) with the visual phenomenology of *specific* letter representations (e.g., 'A'). Thus, Grush (2007, p. 505) has argued that "we seem to have a full, richly detailed, phenomenal representation of the visual scene, though in fact what we have is, albeit full and clear…, only generic." Papineau (2007, p. 521) argues that Sperling's partial report superiority effect can be accounted for by crediting participants with representations displaying merely "scene phenomenology," not "item phenomenology," where to have the former is to be phenomenally conscious of a 3 x 4 array of letters, and to have the latter is to be phenomenally conscious of individual characters. As Grush (2007, *Ibid*.) argues, the representation of letters stored in iconic memory has *generic* phenomenology. It *affords* answers to detailed queries. Only after the queries have been solved does the representation of letters acquire *detailed* visual phenomenology. But participants in Sperling experiments confuse the former and the latter. This is what I call the

generic phenomenology illusion (which Block, 2007a, p. 533 calls the *affordance* illusion).

If generic phenomenology is taken to be a feature of the representation of the property letterness (or of the property letterness being instantiated), then the question arises: to what extent could the representation of the property letterness have *visual* phenomenology? Presumably, unlike the visual perception of some particular alphanumeric character, the representation of letterness does not require making a choice between upper and lower case. Not does it mandate the representation of a specific font, size and color. Thus, it is not clear whether the generic representation (or belief) that some subset of the set of 26 letters of the alphabet is being instantiated can have *visual* phenomenology or not. Arguably, on the basis of the representation of some particular alphanumeric character with visual phenomenology, one can form the judgment or belief that the property letterness is being instantiated. But if so, then the content of the latter judgment or belief could only have *cognitive* phenomenology (in the sense of Strawson, 1994), *not visual* phenomenology. Thus, the thesis of the *generic phenomenology illusion* seems to face two related problems.

On the one hand, it seems committed to the unparsimonious assumption that the explanation of Sperling's partial report superiority requires positing two kinds of phenomenology: non-cognitive or visual phenomenology and cognitive or non-visual phenomenology. On the other hand, while it is unmysterious how one can move from the representation of specific letters with visual phenomenology to the judgment that letterness is being instantiated, it is mysterious how one could conversely move from the judgment that letterness is instantiated to the representation of a specific letter with visual phenomenology.[28]

6. On the persistence of visual phenomenology

[28] For a critical discussion of Strawson's (1994) view of cognitive phenomenology, see Jacob (1998). For present purposes, I need not assess the possibility of cognitive phenomenology. It is sufficient to point out that advocates of the generic illusion interpretation of the Sperling findings are forced to posit the existence of two kinds of phenomenology.

Block's (2007a) thesis that phenomenology overflows cognitive access has been further corroborated by findings reported by Landman et al. (2003) and Lamme (2003), based on an experimental paradigm that combines change blindness and Sperling's paradigm. Participants saw an *initial* array of eight rectangles of equal area for 500 ms, followed by a blank screen for a period that can vary from 200 to 1,500 ms, followed by a novel instance of the same array of eight rectangles (the *probe* array), in which the *orientation* of one rectangle, which is being cued by a yellow line, has either changed or not. Landman et al. (2003) compared three conditions. In the first condition, the appearance of the cue coincides with that of the probe array. In the second condition, the cue appears with the initial array. In the third condition, the cue appears with the blank screen, when no array is visible. In each condition, participants are asked to tell whether the orientation of the cued rectangle is the same or different in the initial and the probe arrays. Landman et al. (2003) report that while participants' ability to detect the difference in the second condition is almost 100%, if falls to 60% in the first condition. Interestingly, participants are almost as good in the third condition as they are in the second condition.

As Philipps (2011, p. 405) has noted, there are a couple of important differences between Sperling's paradigm and the paradigm used by Landman and colleagues. In Landman et al.'s paradigm, participants are presented with a two-alternative forced-choice: change or no change of orientation. By contrast, in Sperling's paradigm, participants are asked to identify letters, which involves selecting one among 26 possible complex items for each of the 3-4 choices made. Nonetheless, the results of Landman et al.'s (2003) experiment are clearly open to the Overflow interpretation. If the relevant rectangle is cued before its orientation changes (as in the second and third experimental conditions), then information about the first orientation of the rectangle, which is first encoded in a visual iconic format with phenomenology and stored in iconic memory, can be recoded in a non-iconic or digital

format and transferred into working memory. Only information stored in a digital format in working memory is appropriate for comparison with information about the second orientation of the rectangle, which is necessary for answering the forced-choice question. If the relevant rectangle is only cued after its orientation has changed, then information about the first orientation, which has been encoded in a visual iconic format and stored in iconic memory, has become unavailable for recoding in a digital format and for transfer into working memory.

The Overflow interpretation of the findings by Landman et al. (2003) is open to the same rebuttal as the interpretation of Sperling's findings, by applying to the orientation of a rectangle the distinction drawn by Byrne et al. (2007, pp. 501-502) between *visible persistence* and *informational persistence*, which is reminiscent of the earlier distinction between *specific* phenomenology and *generic* phenomenology. On this view, when the relevant rectangle is uncued (as in the initial array in the first experimental condition), it is represented as having *some orientation or other* (horizontal *or* vertical) and this representation has generic phenomenology. Only after the rectangle has been cued is its *determinate* orientation encoded in a representation with visual phenomenology. Thus, the cue changes the phenomenology of the representation of the rectangle's orientation from generic (or cognitive) phenomenology to specific (or visual) phenomenology: it confers visual (i.e. specific) phenomenology to the representation of the rectangle's orientation by making it more determinate (or less determinable).

The view that the representation of the orientation of an uncued rectangle stored in the iconic memory buffer has generic (i.e. cognitive) phenomenology and lacks visual phenomenology until it is cued gives rise to the same puzzles as the view that participants store a representation of letterness with generic (cognitive) phenomenology in the iconic memory buffer. On the one hand, it is unclear what it could be like to store in iconic memory

a representation of a rectangle as having *some orientation or other*. If one rectangle in an array of eight rectangles is being represented as having a *determinable* orientation, is it also represented as having a determinate *size* and a determinate *color* or a determinable *size* and a determinable *color*?

On the other hand, there is some consensus that making information suitable for report involves recoding it from a more iconic to a more digital (or less iconic) format, and transferring it from iconic memory (with larger storage capacity and shorter persistence) to working memory (with smaller storage capacity and longer persistence). Presumably, what digitalization can do is *segregate* information about the *orientation* of a rectangle from information about its size and color, in the way that a linguistic report, unlike a picture, can specify orientation (e.g., *horizontal*), while being mute about size and color.[29] But digitalization of information could not make the represented property (i.e., horizontality) more determinate (and less determinable).

Further recent work by Victor Lamme and his colleagues has extended Landman et al.'s (2003) paradigm by increasing the number and complexity of items to be stored in the iconic memory buffer and also by presenting attentional cues at three distinct temporal stages: an *iconic* cue introduced at 10-40 ms after off-set of the initial array of rectangles, a *retro-cue* introduced at 1000-4000 ms after off-set of the initial array, and a *post-change cue* introduced with the probe array, which has been demonstrated to overwrite the initial array. On this basis Sligte et al. (2008) propose the following tripartition of short-term memory. They found that the pure iconic memory buffer, whose persistence is within the limits of the iconic cue, 10-40 ms after the off-set of the initial (memory) array, supports a change detection capacity of arrays of almost 32 items. They found that what they call *fragile visual short-term memory*, whose persistence lies within the limits of the retro-cue, 1000-4000 ms after the off-set of the

[29] This is Dretske's (1980) account of the distinction.

initial array, supports a detection change capacity of 10 items. Finally, they found that working memory tested by means of the post-cue has a storage capacity of 4 items (from the initial array). As Block (2011, p. 573) puts it, these findings show that what was construed on the basis of Sperling's findings as a single iconic memory system is in fact comprised of two distinguishable systems: 'a rod-based "pure iconic" memory lasting at most a few hundred milliseconds and a much longer-lived "fragile visual short-term memory" that is based higher up in the visual system and lasts up to 4-5 seconds.'

In a follow up change detection task conducted by Sligte et al. (2010), participants saw pairs of initial and probe arrays composed of eight complex colored objects, one of which could be changed from one array to the other. The relevant object was cued at three different stages: 10 ms after off-set of the initial array (iconic cue), 1000 ms after off-set of the initial array (retro-cue), or during the probe array (post-change cue). In addition, in trials involving a change of item, participants were also asked to identify the item present in the initial array but not in the probe array and to select it from three distractors not present in the initial array. The results of both the change detection task and the identification task show that both pure iconic memory and fragile visual short-term memory have a far greater storage capacity than working memory.

In short, the extension of the paradigm first used by Landman et al. (2003) corroborates the Overflow thesis that visual phenomenology outstrips cognitive access necessary for report. Not only does it vindicate the assumption of the perceptual basis of visual phenomenology, but it also offers empirical support for a graded view of the persistence of visual phenomenology by introducing a novel kind of fragile visual short-term memory, whose storage capacity is slightly smaller than that of pure iconic memory but vastly greater than that of working memory. This graded view stands in contrast with the winner-take-all competitive view of conscious phenomena.

7. Whose experience?

On my view, recent experimental findings support Block's Overflow thesis that visual phenomenology outstrips cognitive access. But if so, then the following puzzle arises: one could fail to be introspectively aware of some of one's own phenomenal experiences. This puzzle has been explicitly addressed by Block (2008, pp. 290-292), when he considers the case of patient GK who exhibits visuo-spatial extinction for stimuli in his left hemi-field caused by a lesion in his right inferior parietal lobe. When such neglect patients see a single object on either side, they can identify it, but if there are objects on both sides, then they can only identify the object on their right and report no visual experience on their left. Stimuli on the right are said to *extinguish* stimuli on the left.[30] Recent fMRI studies (discussed by Block) show that when a face presented in patient GK's left hemi-field is being extinguished by a competing stimulus on his right, GK reports not seeing the face. But nonetheless the experimenters report that activity in the right portion of his fusiform gyrus (or face area) is almost as strong as when GK reports seeing a face.

As Block (2008, p. 291) notes, this observation gives rise to the following dilemma, one horn of which is that activity in the right fusiform face area is not sufficient for supporting conscious visual experiences of faces located on one's left. Patients with hemi-neglect are known to be impaired in their ability to attend to stimuli on their left when there is a competing stimulus on their right. GK's inability to report extinguished stimuli on his left is evidence that his capacity to attend to stimuli on his left is impaired by the lesion. His attentional impairment might in turn be taken to support the contribution of attention to the phenomenology of visual experience, in accordance with the global workspace model of consciousness. On the other horn of the dilemma, we might take the result of the fMRI study

[30] Cf. Driver and Vuilleumier (2001) for review.

as evidence that GK has visual conscious experiences of faces on his left, but he cannot report it because his ability to attend to his left has been damaged. This alternative interpretation is consistent with the Overflow thesis that the phenomenology of visual experience outstrips cognitive access (i.e., the availability of what is experienced to attention and working memory).

This dilemma is reminiscent of another one generated by research on split-brain patients (conducted by Sperry), who have been shown to be able to verbally report words seen on their right and processed by their left hemisphere, where their language processing system is located. But they have been shown to be unable to verbally report words seen on their left and processed by their right hemisphere, when the latter has been disconnected from the left hemisphere (after severing of the corpus callosum). However, these patients have also been shown to be able to select by touch with their left hand the object referred to by the word seen on their left from a set of distractors. Does the fact that these patients cannot verbally report words seen on their left show that they are not conscious of these words? Does the fact that tactile experience with their left hand enables them to choose the right object show that they are conscious of words on their left, in spite of their inability to make a verbal report?[31]

In accordance with the latter horn of the dilemma, Block (2008, p. 291) has argued that the hypothesis that patient GK has unreported (and perhaps unreportable) visual experience of a face on his left is perfectly intelligible. Now if patient GK has a visual experience of a face on his left that he does not (and cannot) know about, what makes it *his experience*? As a way of answering this question, Block (2008) appeals to the possibility of taking an *objective* optical measure of an individual's *visual field*: in healthy people, the shape of the visual field is oval, elongated to the right and left, and slightly larger on the bottom. On

[31] Cf. Gazzaniga (2005). While Putnam (1981, pp. 85-92) argued from this dilemma that knowledge of the identity between *qualia* and brain states might remain epistemologically inaccessible, Dretske (2006) takes the view congenial to Block that both split-brain patients and neglect patients may have unreportable visual experiences.

this criterion, if an individual's visual experience is caused by items that happen to be located within the objective boundaries of her visual field, then the experience is hers. But if the question of *ownership* is about an individual's *introspective* access to the content of his or her own experience, then it is far from clear that an objective optical measure of the individual's visual field can be the whole answer.

As I see it, there is a basic bifurcation between two main philosophical approaches to introspective awareness of one's own sensory experiences. One is the traditional empiricist "inner sense" approach, well exemplified by Russell (1912). On Russell's threefold epistemological picture, one does *not* really perceive (one is *not* directly acquainted with) ordinary physical objects. Instead, one *perceives* (one is *directly acquainted* with) one's own *sense data*. Finally, knowledge about ordinary physical objects is indirect (propositional) knowledge of truths about ordinary physical objects inferred from one's direct acquaintance with one's own sense data.[32] The alternative approach squarely rejects the *observational* model of introspection and denies that one can have a *perceptual experience of one's own perceptual experience*. As Shoemaker (1994), one of the sharpest critic of the observational model, has famously argued, none of the conditions that should be met for introspection to be construed as a kind of sensory perception does in fact prevail.

Some philosophers reject the observational model of one's awareness of one's own perceptual experiences because they reject the basic distinction between perceptual experience and belief or judgment. This rejection underlies Dennett's (1991, p. 132) endorsement of *first-person operationalism*, which "denies the possibility in principle of consciousness of a stimulus in the absence of the subject's belief in that consciousness." On this view, an individual's introspective belief that she is aware of something in the world, e.g., a red cube, is a necessary condition for her being aware of a red cube. Rosenthal's (1986,

[32] Cf. Jacob (2004) for further discussion.

2005) *higher-order thought* theory of conscious states offers another alternative to the observational model of introspection. On Rosenthal's approach, the phenomenology of one's perceptual experience is constituted by one's higher-order thought about one's perceptual experience. In other words, what confers visual phenomenology to one's perceptual experience is some higher-order conceptual thought that refers to it. In Rosenthal's (1986, 2005) useful terminology, what makes a creature's mental state *intransitively* conscious is that the creature whose state it is, is *transitively* conscious of it.

First-person operationalism is not acceptable to an advocate of the Overflow interpretation of Sperling's findings, because the latter is predicated on the distinction between belief or judgment and perceptual experience, which is rejected by first-person operationalism. Nor is the higher-order thought theory of conscious experiences acceptable to an advocate of the Overflow thesis, because by making the phenomenology of perceptual experiences depend on the ability to form higher-order thoughts about perceptual experiences, it denies the principle of the perceptual basis of visual phenomenology, which is crucial to the Overflow interpretation of Sperling's findings.

There is, however, an alternative to both first-person operationalism and Rosenthal's higher-order thought theory of conscious states, which is both consistent with the distinction between belief (or judgment) and perceptual experience and inconsistent with the observational model of one's own awareness of one's perceptual experiences. On Rosenthal's higher-order thought theory, the phenomenology of a creature's perceptual experience derives from the creature's being transitively conscious of it in virtue of forming a higher-order thought about it. An alternative option is to give primacy to *intransitive state* consciousness over transitive and intransitive creature consciousness. On this alternative option (espoused by Dretske, 1997 as I interpret it), not only is the phenomenology of a creature's perceptual experience *independent* from the creature's transitive higher-order consciousness of it, but the

creature's transitive consciousness of objects and properties in the world depends on the phenomenology of her intransitive visual experiences, not vice-versa.

Not all neurophysiological stages and bodily events (e.g., head- and eye-movements, retinal stimulation), which are part of the perceptual process whereby a creature becomes visually conscious of a red cube, are conscious experiences. Nor is the red cube itself (i.e., the distal stimulus) conscious. The red cube, however, is one of the necessary conditions of the creature's perceptual awareness of it.[33] According to the thesis of the primacy of intransitive state consciousness over creature consciousness of objects and properties in the environment, a creature cannot be transitively conscious of the red cube unless a particular state of her visual cortex qualifies as an intransitively conscious visual experience.[34] While attention and working memory are necessary conditions for report, advocates of the Overflow thesis deny that attention and memory are necessary for visual phenomenology. Nor is higher-order thought necessary for visual phenomenology. Instead, in line with Block's (2007a) approach, advocates of Overflow take it that it is an open empirical question which computation in the visual cortex confers visual phenomenology to a representation of a distal stimulus (e.g., a red cube). Some neuroscientists take it that *binding* (or integration) of different visual attributes (e.g., size, shape, color, location, etc.) processed in distinct areas of the visual cortex is crucial. Others take the contrast between *feed-forward* and *recurrent cortical processing* as crucial to the distinction between conscious and unconscious processing of a visual stimulus.[35]

Advocates of the Overflow thesis who reject the observational model of introspective self-knowledge cannot endorse first-person operationalism. Since they further reject the

[33] This is why the phenomenology of visual experiences cannot supervene on properties of the brain only.
[34] Plausibly, transitive creature consciousness entails intransitive creature consciousness: if a creature is conscious of something or other, then the creature is intransitively conscious. If so, then by entailing transitive creature consciousness, intransitive state consciousness also entails intransitive creature consciousness (cf. Dretske, 1993).
[35] Cf. Lamme (2006, 2010).

higher-order thought theory of phenomenal experiences, they should, I think, endorse the thesis of the primacy of intransitive state consciousness over transitive and intransitive creature consciousness. In a nutshell, they should embrace Dretske's (1993, 1997) view that what makes a visual experience a conscious experience is that it makes the creature whose experience it is conscious of objects and properties (e.g., a red cube) in her environment. To reject the observational model of introspective awareness is to accept that the only way one can be aware of one's own visual experience of a red cube is by forming a belief about the reportable content of one's own experience. To be introspectively aware that one saw a red cube is to *believe* that one saw a red cube. One cannot in turn believe that one saw a red cube unless one possesses and deploys not only the color and shape concepts RED and CUBE (that apply to the red cube) but also the psychological concept SEE and applies it, not to the red cube, but to oneself.[36]

In a nutshell, according to Overflow, one can have a visual experience of a red cube that fails to be reportable because its phenomenology is not (entirely or at all) accessible to attention and working memory. On this approach, it is not necessary to have and deploy the concepts RED and CUBE for having a visual experience of a red cube with a distinctive phenomenology. Nor is it necessary to be introspectively aware of one's own experience for having an experience with a distinctive phenomenology. But it is necessary to have and deploy the concepts RED and CUBE for believing of a perceived object that it is a red cube. Furthermore, the only way one can become introspectively aware of seeing a red cube is by believing that one is. One could not believe that one is seeing a red cube unless one believed of the perceived object that it is a red cube and also one applied the concept SEE to oneself. Since one's introspective belief could be false, it follows that one could believe that one sees a red cube even if there is no red cube to be seen or if the background conditions (distance

[36] It is an open question currently very much discussed to what extent the same processes underlie first-person and third-person mindreading, i.e. attribution of the belief that one is seeing a red cube to self and others.

from the perceived object and/or lighting conditions) are such that what one takes to be a red cube is in fact not a red cube.

On Block's (2008, p. 292) view, when a competing stimulus on his right extinguishes his perception of a face on his left, patient GK should be diagnosed as having a visual experience that "he does not and cannot know about." In this neuropsychological case, the patient can be said to have lost first-person epistemic authority on both the content and very existence of some of his own visual experiences. If so, then the patient can learn from a third-person fMRI report whether he had a visual experience and what it was like. fMRI evidence of activity in his right fusiform gyrus (face area) can tell the patient that he saw a face. If the patient is given a chance to see the same face again while he is aware that he sees it, then he can learn what his extinguished experience of the face was like.

Acceptance of the Overflow thesis and rejection of the observational model of introspection opens the path for the possibility that healthy human adults' introspective knowledge of their own experiences is neither authoritative nor infallible. Other people may be better positioned than I am or they may have more adequate conceptual resources than I do for capturing the content of my own experiences.[37] If one saw a numbat that one mistook for a fox, then one may mistakenly believe that one saw a fox: one could thus be corrected by someone else. One may also see things to which one can only apply concepts such as THING, OBJECT or SUBSTANCE. Someone else might supply concepts that better capture the content of one's own experience. It also opens the path for the more radical possibility that if an individual remains introspectively unaware of some of her own experience, then the individual is unlikely to *claim it as her own*. But nobody else either is likely to claim it as his or her own. If so, then the question arises how we can and how we should make sense of the possibility of *disowned* experiences, i.e., experiences hanging in mid air, which belong to no

[37] Since use of fMRI studies is not restricted to neuropsychological patients, healthy human adults too can learn from them that they had visual experiences of which they are not introspectively aware.

one and which nobody can claim as his or her own.

Conclusions

In this paper, I have examined some of the experimental evidence for Block's disputed thesis that visual phenomenology overflows cognitive access. I have tried to rebut the claim made by opponents of the Overflow thesis that it rests on some cognitive illusion. I have embraced Block's thesis, but in the last section of the paper, I have tried to spell out the consequences of accepting Block's thesis for introspective self-awareness of one's own visual experiences. One puzzling corollary of Block's thesis is that one may have conscious visual experiences of which one might be unaware. Since nobody else might be aware of them, one may have conscious visual experiences of which no one is aware. If so, then such experiences will not make any difference to anyone: their phenomenology will not matter to anyone.[38] The deepest puzzle is not that one may not be either authoritative or infallible and may learn from others about the contents of one's experiences. The deepest puzzle instead arises from the possibility of experiences that may lie beyond anybody's *sense of ownership*. Can there be something it is like to have that nobody can recognize as one's own?[39]

References

Austin, J.L. (1962) *Sense and Sensibilia*, New York: Oxford University Press.

Baars, B. J. (1988) *A Cognitive Theory of Consciousness*. Cambridge University Press.

Block, N. (1995) On a confusion about a function of consciousness. *Behavioral and Brain Sciences* 18, 2, 227–287.

[38] Kouider et al. (2010, p. 303-304) argue that the very fact that this puzzle arises is a reason to reject the Overflow thesis, because it entails that one may have a visual experience with a distinctive phenomenology, which might make no difference to anyone since the person whose experience it is may be unaware of it.

[39] I am grateful to Ned Block, Pierre Buser, Fred Dretske, Sid Kouider and Hilary Putnam for their comments on this paper.

Block, N. (2001) Paradox and cross purposes in recent work on consciousness. *Cognition*, 79, 197-219.

Block, N. (2005) Two neural correlates of consciousness. *Trends in Cognitive Sciences*, 9, 46–52.

Block, N. (2007a) Consciousness, accessibility, and the mesh between psychology and neuroscience. *Behavioral and Brain Sciences*, 30, 481–548.

Block, N. (2007b) Wittgenstein and qualia. *Philosophical Perspectives*, 21, *Philosophy of Mind*, 73-115.

Block, N. (2008) Consciousness and cognitive access. *Proceedings of the Aristotelian Society*, 108, 289–317.

Block, N. (2011) Perceptual consciousness overflows cognitive access. *Trends in Cognitive Sciences*, 15, 12, 567-575.

Brooks, R. (1991) Intelligence without representation. *Artificial Intelligence*, 47, 139–59.

Byrne, A., Hilbert, D. R. and Siegel, S. (2007) Do we see more than we can access? *Behavioral and Brain Sciences*, 30, 501–2.

Cohen, M.A. and Dennett, D. (2011) Consciousness cannot be separated from function. *Trends in Cognitive Sciences*, 15, 358–364.

de Gardelle, V., Sackur, J. and Kouider, S. (2009) Perceptual illusions in brief visual presentations. *Consciousness and Cognition*, 18, 569–577.

Dehaene, S. and Changeux, J. P. (2004) Neural mechanisms for access to consciousness. In: *The Cognitive Neurosciences* III, M. Gazzaniga (ed.), pp. 1145–58, Cambridge, Mass.: MIT Press.

Dehaene, S. and Naccache, L. (2001) Towards a cognitive neuroscience of consciousness: basic evidence and a workspace framework. *Cognition*, 79, 1-37.

Dehaene, S., Sergent, C. and Changeux, J-P. (2003) A neuronal network model linking subjective reports and objective physiological data during conscious perception. *Proceedings of the National Academy of Sciences USA*, 100, 14, 8520–8525.

Dehaene, S., Changeux, J-P., Naccache, L., Sackur, J. and Sergent, C. (2006) Conscious, preconscious, and subliminal processing: a testable taxonomy. *Trends in Cognitive Sciences*, 10, 5, 204–11.

Dennett, D. (1988) Quining Qualia. In: Marcel, A. & Bisiach, E. (eds.) *Consciousness in Contemporary Science*, pp. 381–414, Oxford: Oxford University Press.

Dennett, D. (1991) *Consciousness Explained*, Boston: Little & Brown Co.

Dennett, D. (1994) Get real. *Philosophical Topics*, 22, 1 & 2, 505-568.

Dennett, D. (2001) Are we explaining consciousness yet? *Cognition*, 79, 221–37.

Dennett, D. (2005) *Sweet Dreams, Philosophical Obstacles to a Science of Consciousness*, Cambridge, Mass.: MIT Press.

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

Dretske, F. (1993) Conscious experience. In Dretske, F. (2000) *Perception, Knowledge and Belief*, Cambridge: Cambridge University Press, pp. 113-137.

Dretske, F. (1994) Differences that make no difference. In Dretske, F. (2000) *Perception, Knowledge and Belief*, Cambridge: Cambridge University Press, pp. 138-157.

Dretske, F. (1997) What good is consciousness? In Dretske, F. (2000) *Perception, Knowledge and Belief*, Cambridge: Cambridge University Press, pp. 178-192.

Dretske, F. (2000) *Perception, Knowledge and Belief*, Cambridge: Cambridge University Press.

Dretske, F. (2004) Change blindness. *Philosophical Studies*, 120: 1–18.

Dretske, F. (2006) Perception without awareness. In Gendler, T. and Hawthorne, J. (eds.) *Perceptual Experience*, Oxford: Oxford University Press, pp. 147-180.

Dretske, F. (2007) What change blindness teaches us. *Philosophical Perspectives*, 21, 1, 215-230.

Driver, J., and Vuilleumier, P. (2001) Perceptual awareness and its loss in unilateral neglect and extinction'. *Cognition*, 79,1–2, 39–88.

Fodor, J.A. (2000) A science of Tuesdays. *London Review of Books*, 22, 14, 21-22.

Gazzaniga, M.S. (2005) Forty-five years of split-brain research and still going strong. *Nature Reviews Neuroscience,* 6, 653–659.

Grush, R. (2007) A plug for generic phenomenology. *Behavioral and Brain Sciences*, 30, 504-505.

Jacob, P. (1998) What is the phenomenology of thought? *Philosophy and Phenomenological Research,* 58, 2, 443-448.

Jacob, P. (2004) Do we know how we know our own minds yet? In Schantz, R. (ed.) *The Externalist Challenge: New Studies on Cognition and Intentionality*, Berlin: Walter de Gruyter, pp. 401-418.

Jacob, P. (2006) Why visual experience is likely to resist being enacted. *Psyche*, 12, 1, 1-11, http://www.psyche.Cs.monach.edu.au/

Jacob, P. (2008) The scope and limits of enactive approaches to visual experience. *Perception*, 37, 3, 446–461.

Kouider, S., Dehaene, S., Jobert, A. and Le Bihan, D. (2007) Cerebral Bases of Subliminal and Supraliminal Priming during Reading. *Cerebral Cortex*, 17, 2019–2029.

Kouider, S., de Gardelle, V., Sackur, J. & Dupoux, E. (2010) How rich is consciousness? The partial awareness hypothesis. *Trends in Cognitive Sciences*, 14, 301-307.

Lamme, V. (2003) Why visual attention and awareness are different. *Trends in Cognitive Sciences*, 7, 12–18.

Lamme, V. (2006) Towards a true neural stance on consciousness. *Trends in Cognitive Sciences*, 10, 11, 494–501.

Lamme, V. (2010) How neuroscience will change our view on consciousness. *Cognitive Neuroscience*, 1, 3, 204–240.

Landman, R., Spekreijse, H. & Lamme, V. A. F. (2003) Large capacity storage of integrated objects before change blindness. *Vision Research*, 43, 2, 149–64.

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

Nagel, T. (1974) What is it like to be a bat? *The Philosophical Review*, 83, 4, 435–50.

Noë, A. (2004) *Action in Perception*, Cambridge, Mass.: MIT Press.

O'Regan, K. (1992) Solving the "real" mysteries of visual perception: The world as an outside memory. *Canadian Journal of Psychology*, 46, 3,461–88.

O'Regan, K. and Noë, A. (2001) A sensorimotor approach to vision and visual consciousness. *Behavioral and Brain Sciences*, 24, 883–975.

Papineau, D. (2007) Reuniting (scene) phenomenology with (scene) access. *Behavioral and Brain Sciences,* 30, 521.

Putnam, H. (1975) *Mind, Language and Reality*, *Philosophical Papers*, vol. 2, Cambridge: Cambridge University Press.

Putnam, H. (1977) Models and reality. In Putnam, H. (1983) *Realism and Reason, Philosophical Papers*, vol. 3, Cambridge: Cambridge University Press, pp. 1-25.

Putnam, H. (1978) There is at least one *a priori* truth. In Putnam, H. (1983) *Realism and Reason, Philosophical Papers*, vol. 3, Cambridge: Cambridge University Press, pp. 98-114.

Putnam, H. (1981) *Reason, Truth and History*, Cambridge: Cambridge University Press.

Putnam, H. (1983) *Realism and Reason, Philosophical Papers*, vol. 3, Cambridge: Cambridge University Press.

Putnam, H. (1988) *Representation and Reality*, Cambridge, Mass.: MIT Press.

Putnam, H. (1994) Sense, nonsense, and the senses: an inquiry into the powers of the human mind. *The Journal of Philosophy*, 91, 9, 445-517.

Putnam, H. (1999) *The Threefold Cord: Mind, Body and World*, New York: Columbia University Press.

Putnam, H. (2008) 12 Philosophers—and their influence on me. *Proceedings and Addresses of the American Philosophical Association*, 82, 2,102-115.

Putnam, H. (2016a) Naïve realism and qualia. In Putnam (2016b) edited by de Caro, M. (2016) *Naturalism, Realism, and Normativity*, Cambridge, Mass.: Harvard University Press, pp. 169-196.

Rosenthal, D.M. (1986) Two concepts of consciousness. *Philosophical Studies*, 49, 3, 329–359.

Rosenthal, D.M. (2005) *Consciousness and Mind*, Oxford: Oxford University Press.

Russell, B. (1912) *The Problems of Philosophy,* Oxford: Oxford University Press.

Shoemaker, S. (1994) Self-knowledge and "inner sense": Lecture III: the phenomenal character of experience. *Philosophy and Phenomenological Research*, 54, 249–314. In Shoemaker, S. (1996) *The First-Person Perspective and Other Essays*, Cambridge: Cambridge University Press.

Shoemaker, S. (1996) *The First-Person Perspective and Other Essays*, Cambridge: Cambridge University Press.

Simons, D. J. and Rensink, R. (2005) Change blindness: past, present, and future. *Trends in Cognitive Sciences*, 9, 16–20.

Sligte, I.G., Scholte, H.S. and Lamme, V.A.F. (2008) Are There Multiple Visual Short-Term Memory Stores? *PLoS ONE*, 3, e1699.

Sligte, I.G., Vandenbroucke, A.R.E., Scholte, H.S. and Lamme, V.A.F. (2010) Detailed sensory memory, sloppy working memory. *Frontiers in Psychology*, 1, 1–10.

Sperling, G. (1960) The information available in brief visual presentations. *Psychological Monographs: General and Applied*, 74 (11, Whole No. 498), 1–29.

Strawson, G. (1994) *Mental Reality*, Cambridge, Mass.: MIT Press.

Tye, M. (2010) Attention, seeing and change blindness. *Philosophical Issues*, 20, 1, 410-437.