Cognitive Neuroscience

Stage 3 and what we see

Gideon Paul Caplovitz[a]; Michael J. Arcaro[a]; Sabine Kastner[a]

[a] Department of Psychology, Princeton University, Princeton, NJ, USA

Online publication date: 25 August 2010

PLEASE SCROLL DOWN FOR ARTICLE

Ψ Psychology Press
Taylor & Francis Group

Discussion Paper

How neuroscience will change our view on consciousness

Victor A. F. Lamme

University of Amsterdam, Amsterdam, The Netherlands

Is there consciousness in machines? Or in animals? What happens to consciousness when we are asleep, or in vegetative state? These are just a few examples of the many questions about consciousness that are troubling scientists and laypersons alike. Moreover, these questions share a striking feature: They seem to have been around forever, yet neither science nor philosophy has been able to provide an answer. Why is that? In my view, the main reason is that the study of consciousness is dominated by what we know from introspection and behavior. This has fooled us into thinking that we know what we are conscious of. The scientific equivalent of this is Global Workspace theory. But in fact we don't know what we are conscious of, as I will explain from a simple experiment in visual perception. Once we acknowledge that, it is clear that we need other evidence about the presence or absence of a conscious sensation than introspection or behavior. Assuming the brain has something to do with it, I will demonstrate how arguments from neuroscience, together with theoretical and ontological arguments, can help us resolve what the exact nature of our conscious sensation is. It turns out that we see much more than we think, and that Global Workspace theory is all about access but not about seeing. The exercise is an example of how neuroscience will move us away from psychological intuitions about consciousness, and hence depict a notion of consciousness that may go against our deepest conviction: "My consciousness is mine, and mine alone." It's not.

Keywords: Neuroscience; Consciousness; Qualia; Global workspace; Visual perception; Re-entrant.

A STALEMATE BETWEEN INTROSPECTION AND BEHAVIOR

This paper is about seeing. So look at the scene of Figure 1, just briefly. What did you see? You are probably aware that there were colored objects, arranged in a circle. You may remember some colors, red and green, some shapes or even the identity of some objects, like the bread or the motorcycle. While you are reminiscing, you will undoubtedly also get the impression that you are forgetting. Somehow, the visual experience seems to

fade away, from the rich and detailed representation you had at the moment you looked, to the impoverished, almost verbal trace that you are pondering over now. So what were you really seeing, what was in your conscious mind *at the moment* you looked? When you look again, the idea will no doubt settle in your mind that during the looking itself your visual sensation is in fact pretty rich: you see the whole picture. Sort of.

What is evident from this small exercise is that introspection is a poor guide to conscious visual sensation. To "know" what you are seeing, you need
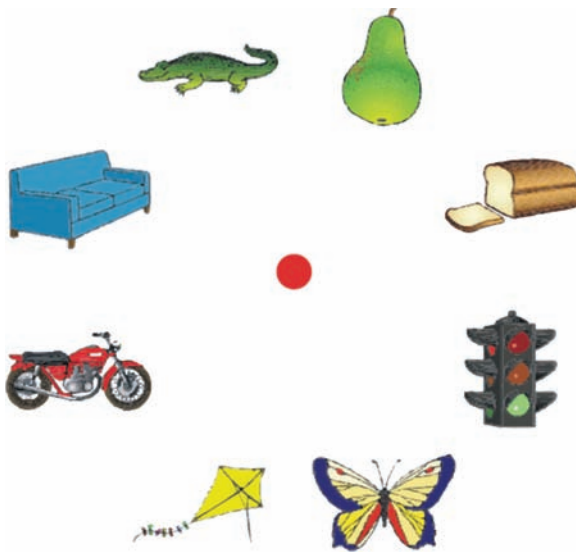
Figure 1. Look at this image for a second, and try to figure out what you saw.

to resort to cognitive functions such as attention, memory, and verbalization. At the same time, you notice that these functions impose a limit on the transfer from visual sensation to cognition. They do not seem to capture your full conscious experience. So which is it? Do I see what I see, or do I see what I know I am seeing?

Psychology to the rescue. Can't we design a task in order to really know how much we are seeing? A classic way to do such a thing would be the *whole report* paradigm. Subjects see an array of items, and the task is to name as many as possible immediately after the array has disappeared (Sperling, 1960). What is found in such experiments is that the capacity to store and retrieve even simple items, such as letters or numbers, is disappointingly low. The average subject will typically only reliably report about four objects. A somewhat more modern way of gauging the capacity of conscious vision is the *change detection* paradigm (Simons & Rensink, 2005). Here an array of items is shown as well, but the task of the subject is to compare that array to a second array that is presented after the first one, with a brief interruption in between (see Figure 2a). Typically, only one item of the array has changed between the first and second display in 50% of the trials, while in the other 50% there was no change at all. The task of the subject is to indicate whether a change was present or not, or even more simply, to indicate whether the object that is cued in the second array has changed or not. Surprisingly, subjects perform very poorly on this task (Landman, Spekreijse, & Lamme, 2003a, 2004b; Sligte, Scholte,

& Lamme, 2008), typically only moderately above chance level (~70%). Percentage correct, in combination with the number of items that has been used, can be converted to the number of items that subjects can simultaneously monitor for change; in other words, the capacity of consciously accessible information that is extracted from the scene. Change detection paradigms come in many flavors (Simons & Rensink, 2005), but in all cases the capacity of conscious access that comes out of these experiments is rarely higher than four objects (Figure 2f; Sligte et al., 2008), and it may even drop to (much) lower values depending on the exact paradigm, or the complexity of the objects (Figure 2c).

The picture of conscious vision that emerges from these findings is that of limitation. Apparently, consciousness is fairly sparse, and much more limited than our introspection would suggest. We may think we see a full and detailed image of what is in front of our eyes, but we take in only a small subset of the information (Dehaene & Naccache, 2001; O'Regan & Noe, 2001). This is all the more evident when we look at other paradigms that have been used to gauge the capacity of conscious access, such as the *inattentional blindness* experiments, where the appearance of large and fully visible objects is missed when attention is focused on something else (Mack & Rock, 1998), or *attentional blink* experiments where items are missed when another item is detected briefly before (Sergent, Baillet, & Dehaene, 2005; Shapiro, 2009; Shapiro, Raymond, & Arnell, 1994). Introspectively, consciousness seems rich in content. We see colors, shapes, objects, and seemingly everything that is in front of us. From the third-person perspective of the behavioral scientist, however, consciousness is rather miserable. Which is the truth (Block, 2005)?

A RICHER REPRESENTATION?

One might argue that these paradigms measure not so much the capacity of conscious vision, but rather the capacity of working memory, attention or any other function necessary for access. It could be that the limit sits in one of these functions, not so much in the visual sensation itself. They are necessary to complete the behavioral task, and if they have limited capacity they will impose a limit on a potentially much richer visual sensation. There is indeed evidence that this might be the case. Many decades ago, it was shown that the whole report paradigm can be modified slightly to yield a completely different result. Instead of asking subjects what they saw, Sperling (1960) used a cue to point at the location of items that had to
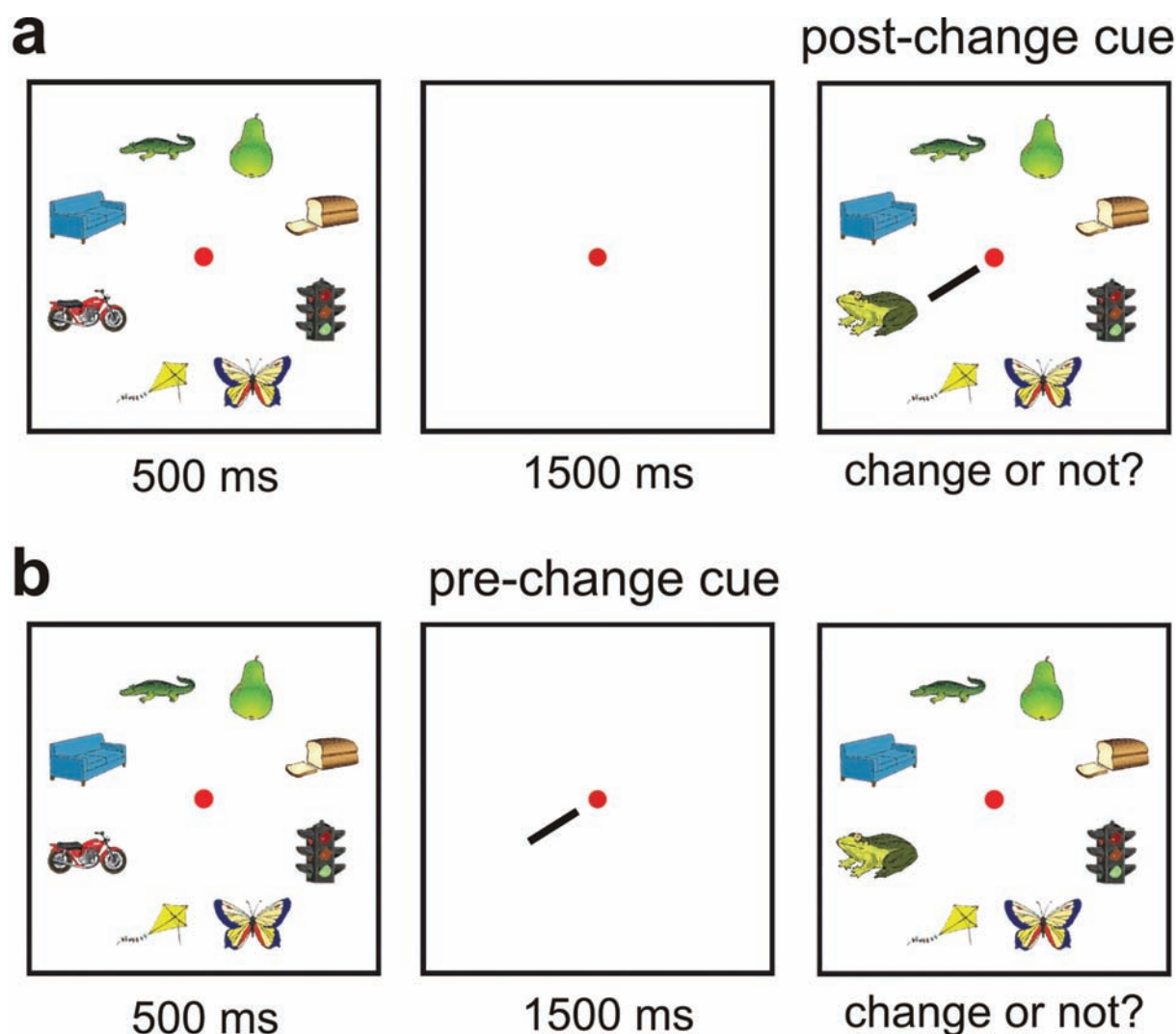
Figure 2 (a and b).   Change detection paradigms, revealing different capacity representations. (a) The traditional change blindness paradigm, where an array of objects is presented for 0.5 s or so, followed by a blank screen for several seconds, followed by a second array. The task of the subject is to indicate whether the cued object has changed or not (in this case it changes, but in 50% of trials it doesn't). Subjects perform poorly in this task, and the percentage correct can be converted to a capacity, which in this case is about two objects that can be monitored for change (see result in panel c). (b) The same paradigm, but now with a cue somewhere during the interval. This yields a much higher percentage correct, and hence a higher capacity (see results in panel c).

be reported. He presented three rows of four letters, and after disappearance of the array a tone indicated to subjects whether they had to report the letters of the top, middle, or bottom row. Again, only three or so letters were reported. But it didn't matter what row was cued. That implies that at the moment of the cue (up to a second after presentation) the subject must have had a representation in his mind that was much richer than the three items that were reported—in fact three times (the number of rows) as large, i.e. about nine letters. This richer representation was named *iconic memory*, and many subsequent *partial report*

experiments revealed that it consisted of an almost-carbon copy of what is presented to the subject (Coltheart, 1980).

We (and others; Becker, Pashler, & Anstis, 2000)) combined the change blindness and partial report paradigms in a single experiment to compare the different representations directly (Figures 2a and 2b). An array of objects is presented, followed by a blank interval of up to several seconds, again followed by the array. One of the objects might have changed, and a cue points to that object. Largely different results are obtained depending on when exactly the cue is presented
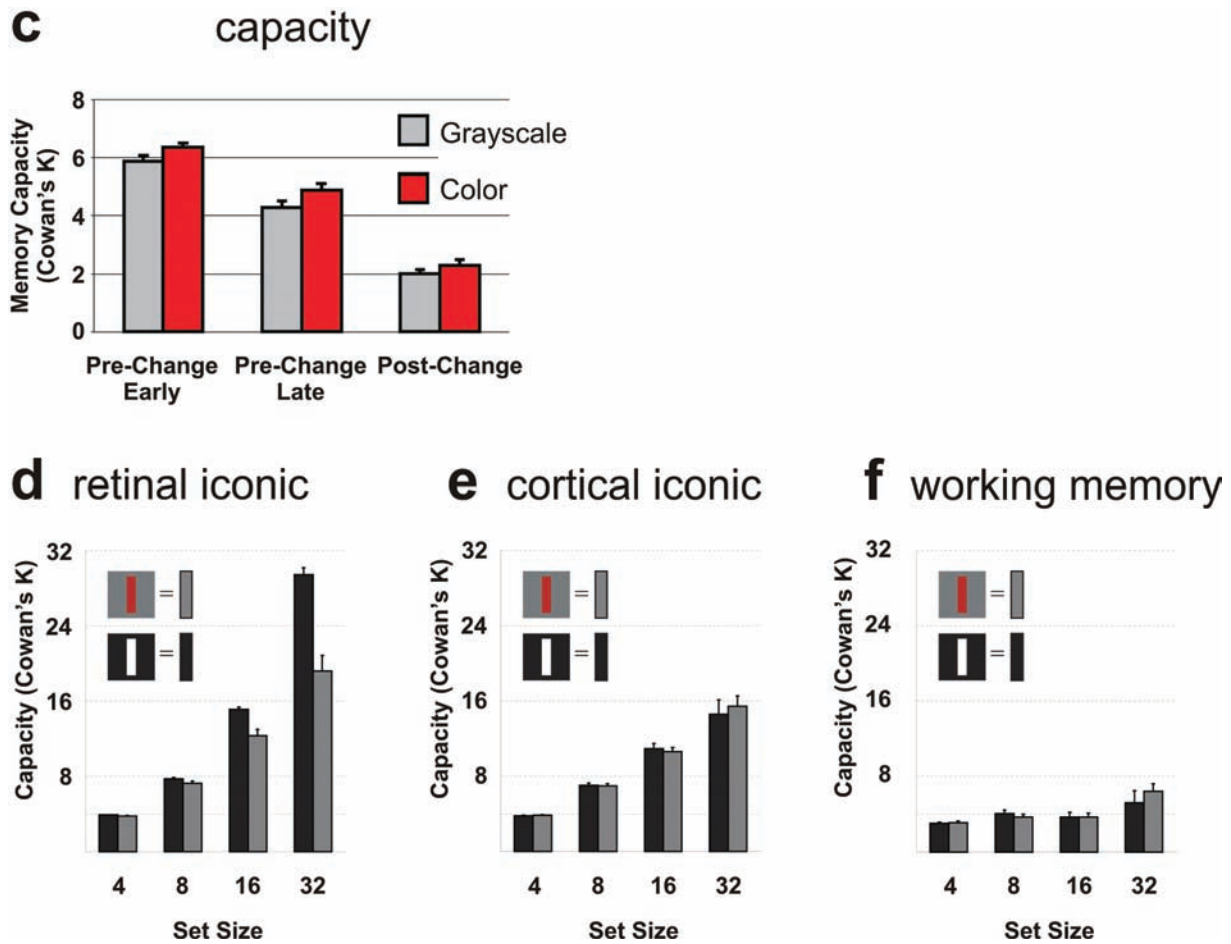
Figure 2 (c, d, e and f).   (c) Results for the change blindness experiments shown in panels a and b. Capacity is high for the pre-change cues (i.e., paradigm b), and low for the post-change cues (paradigm a). Using b/w or colored objects makes no difference. (d) Results for change detection using a pre-change cue immediately after an array of randomly oriented rectangles has disappeared. Capacity increases with set size, i.e., the number of objects used. Black bars are for white rectangles on a black background, gray bars for isoluminant red on gray. These results probably reflect a retinal afterimage. (e) Results from the same paradigm as in panel d, but with a cue 1 s after disappearance of the first array. Results probably reflect a cortical "after-image". (f) Results from the same paradigm as in panel d, but with a post-change cue. Results reflect the (fixed) capacity of working memory.

(Figure 2c). Presenting the cue together with the first array makes the task fairly trivial, as this directs attention to the items that might change—a change that is not difficult to detect in itself. A cue together with the second array yields the low performance and limited capacity of change detection results. However, a cue presented during the interval, even as long as 1 or 2 s after disappearance of the first array, yields a capacity that is much larger, almost as large as when the cue is shown with the first array still there (Landman et al., 2003a, 2004b; Sligte et al., 2008). This confirms and extends the partial report/iconic memory results. Large capacity representations are found even seconds after disappearance of the first array, provided the second array is not allowed to overwrite these.

Research using this combined paradigm so far has shown the large capacity/iconic memory representa-

tion to be of fairly long duration (up to 4 s), and to be quasi-linearly dependent on the number of items that are presented; the more items are shown, the more are stored, but probably with a plateau (Sligte et al., 2008; see Figure 2e). Capacity also depends on the complexity of objects. Presenting a new scene erases (or interferes with) this form of iconic memory, and this has been confirmed using other intervals than the blank screen shown in Figure 2. Critical for this interference is the presentation of objects at the same location. For example, a homogeneous texture of oriented line segments has no effect, but when these line segments compose figures at the same location as the objects of the memory array, performance drops to the low capacity of working memory. However, objects presented at other locations pose few problems. Even when an object is shown at fixation during the interval

there is little interference, suggesting that the large capacity is not erased by the capturing of attention (Landman et al., 2004b). This is confirmed by an experiment where the cue during the interval was followed in some trials by a second cue later in that interval. The capacity of the representation addressed by this second cue is equal to the capacity addressed by the first cue, i.e., the focusing of attention on the item in iconic memory does not make the iconic memory representation 'collapse' (Landman et al., 2003a).

Some controversy exists as to whether the large capacity representation that is measured using the cued change detection task is identical to iconic memory as it was classically defined, or maybe is better seen as some fragile form of working memory. Indeed different forms exist. Immediately after presentation of the array, a cue will reveal an almost exact carbon-copy of what is on the retina, which is erased by any light or colored screen (Sligte et al., 2008). This is probably a retinal after-image (Figure 2d). However, the large capacity store that is obtained with a cue several seconds after disappearance (which is what is discussed above) is not that easily erased, and there are sufficient arguments against it being a mere retinal after-image. It is probably best viewed as a "cortical after-image", i.e., the neural activity that remains in the visual brain after a visual stimulus has been removed. A recent neuroimaging study showed that the neural correlate of an item being part of the large capacity (iconic) representation is neural activity in visual area V4 (Sligte et al., 2009).

WHAT ARE WE SEEING?

Whether this type of iconic memory is better viewed as a fragile form of working memory is not relevant for the discussion here. The key question is this: These results force us to acknowledge that different neural representations of a scene exist (or at least, different representations remain immediately after that scene has disappeared). There is a stable representation, linked to working memory and attention, that allows access and report, and can be maintained across views, yet has fairly limited capacity. And there is a much larger capacity representation—perhaps a virtual copy of the outside world—that is however very fragile, fading away in a few seconds and overwritten as new visual information hits the eyes (Figure 3a). Which is the conscious one? Which is better evidence of what we are—or rather were[1]—seeing when we look at a scene (Block, 2007)?

Face value arguments don't really resolve the issue. Introspectively, it is almost impossible to know

whether you really see all the objects of Figure 1, or just the few you focus attention on. As has been argued, it could be that the impression of seeing all objects is an illusion, created by the fact that every time you focus on one of the objects it is there (often referred to as the refrigerator light illusion: When you open the door it's always on, but in reality it is not; Dehaene et al., 2006). Behaviorally, we can only argue for the presence of two representations, a limited and a large capacity one. Whether there is phenomenality (i.e., a conscious sensation) in either of the two is not directly addressed by the experiment.

Deciding whether there is phenomenality in a mental representation implies putting a boundary—drawing a line—between different types of representations (regardless of whether that is a sharp or a fuzzy boundary). Where to draw that line cannot be decided on the basis of this experiment alone. We have to start from the intuition that consciousness (in the phenomenal sense) exists, and is a mental function in its own right. That intuition immediately implies that there is also *unconscious* information processing. These intuitions come from the extreme ends: cases where the presence or absence of conscious sensations is undisputed. For example, when I show a picture of a face to someone and he replies by confirming to see it, verbally sketching an accurate outline, giving a description of the texture of the face, the emotional expression, and every other feature of it, there is little reason to doubt the presence of a conscious sensation of that face. Likewise, when I show that picture for only 5 ms, followed by a strong mask, and then the subject is incapable of telling whether an image was presented at all, incapable of making a higher than chance forced choice guess whether it was a face or a house, black or white, or bearing any other feature, let alone giving a description of the identity or expression of that face, there is little reason to doubt the absence of a conscious sensation. Especially when the subject is fully focusing his attention on the location where the image is shown, has no neurological disorders, or any other condition that would logically prevent a conscious sensation—if present—to be reported, the only logical conclusion would be to infer the absence of such a sensation.

From that starting point, we infer phenomenality in other situations, or not. A conservative approach would be to infer phenomenality only when the subject confirms having it. This runs into the problem of what

---

[1] Please note that neither iconic nor working memory representations have any visual phenomenality themselves. They are just memory traces, of course. But they are the closest evidence of what you were seeing when the image was still there, and as such give us a window on the representations that are present *during* the seeing.

Which representation is the conscious one?
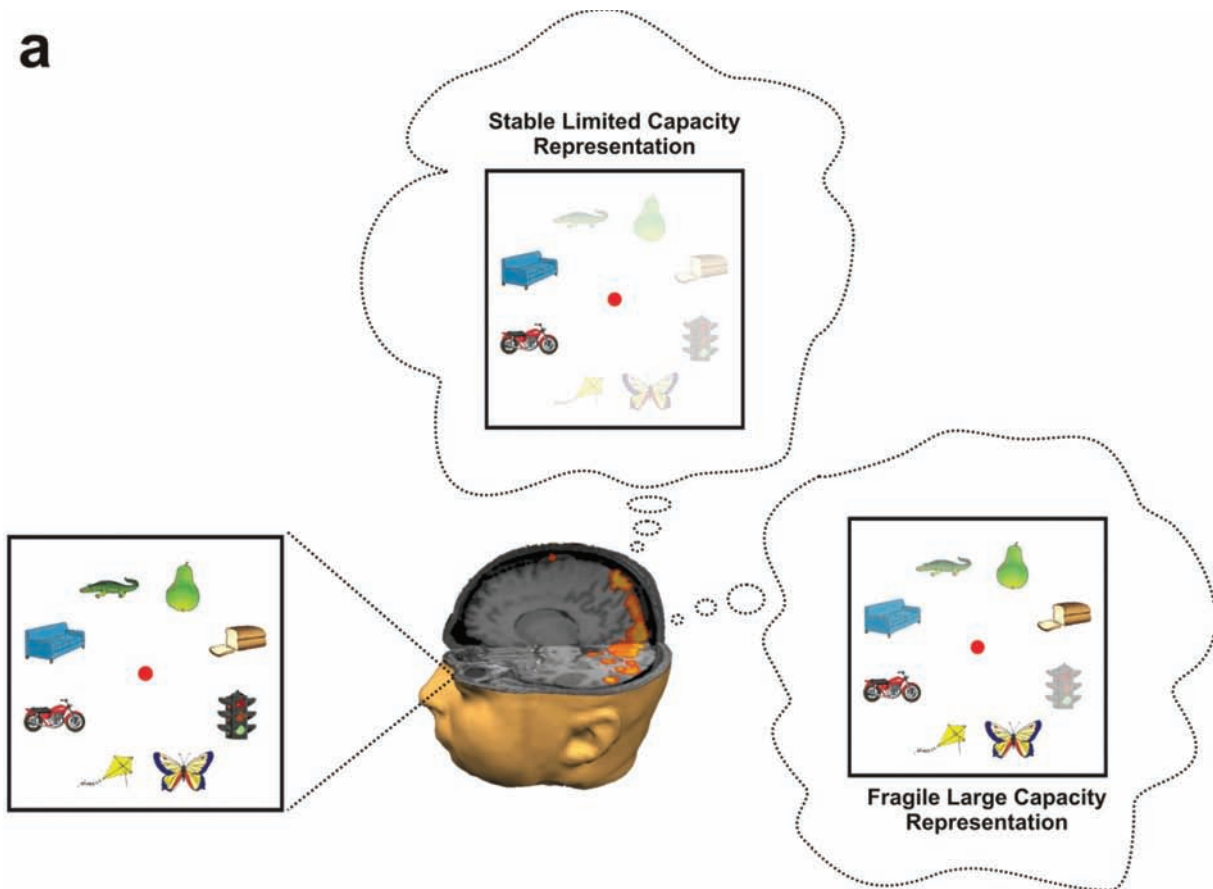
b

	Iconic Memory	Working Memory
Capacity	Large: scales with number of objects in scene, up to 16 documented	Limited: typically 4 or less, depending on complexity
Duration	Short: Up to 4 seconds	Medium: Minutes to hours
Stability	Fragile: Overwritten by any new scene containing objects at approx. the same location. Probably also erased by eye movements.	Stable: Resistant to new visual information, eye movements. Interference from other working memory load
Frame of reference	Mostly retinotopic	Mostly spatiotopic
Quality	Some feature binding, high visual accuracy. Capacity drops with complexity of objects	Feature binding, 'object files', etc. capacity drops with complexity of objects
Cognitive value	Limited	Extensive

Figure 3.   (a) Two representations are generated by our brains when we see a scene, one fragile with large capacity, one stable with limited capacity. (b) Properties of the two representations.

should be considered proper "confirmation." If every behavioral measurement—verbal report, yes–no detection, forced choice discrimination, pointing, priming, etc.—indicates its presence, there is little problem. But what if some behaviors indicate "present" while others do not? This happens all the time when we study consciousness, either by manipulating it (masking, change blindness, attentional blink, rivalry, etc.) or when it has been altered by neurological conditions (blindsight, split-brain, neglect, extinction, etc.). Then some—essentially—arbitrary choice has to be made as to what type of behavior constitutes proper evidence for the presence of conscious sensation and what does not (Lamme, 2006; Seth, Dienes, Cleeremans, Overgaard, & Pessoa, 2008). An alternative would be to take a more liberal approach, and to assign phenomenality whenever there is no proof of absence, i.e., when any behavioral measurement gives a green light. In that case, however, one immediately runs into the problem of drawing a line with the unconscious, as *some* response, such as an altered skin conductance or pupil dilatation, is always present (Stoerig, 1996).

In the situation we focus on here, the two representations discussed above, we face a similar problem. There is little reason to doubt that there is (was) phenomenality in the limited capacity working memory representation. The issue hinges on whether there is any such thing in the iconic memory representation. Should we group that with the conscious, or rather with the unconscious?

FUNCTIONAL ARGUMENTS

A way to answer this question may come from functional arguments. Soon after the discovery of the high capacity representation of iconic memory, it was discarded as having no functional use (Haber, 1983). Seemingly, any cognitive manipulation of visual input requires it to be accessed by working memory or attention. And that imposes the limits discussed above. Also the instability of iconic memory—it being overwritten as soon as we move our eyes to a new fixation—argues against it being of any cognitive use (Figure 3b). So if the *raison d'être* of consciousness is cognitive access to the information, and the ability to cognitively manipulate that information, and combine it with information stored in working or long-term memory, or with inputs from other senses, then the most sensible solution would be not to view the iconic representation as part of consciousness. A theoretical framework that supports this idea is the *Global Workspace* model of consciousness (Baars, 2005).

There are, however, two arguments against this. First, by linking consciousness so much with cognition, there is some "throwing away of the baby with the bathwater," because cognition and access do very little to explain the key feature of consciousness that we consider here, which is phenomenality. Why would combining visual input with working memory make it "visible" to the mind's eye—in other words, produce qualia? That makes me smell a homunculus, in the sense that visual information seemingly needs to *go somewhere* to achieve phenomenality. Second, recent experiments force us to acknowledge that also in the unconscious there is a lot of cognition going on, such as multisensory integration (de Gelder, Pourtois, & Weiskrantz, 2002), interaction with long- and short-term memory (Schacter, Chiu, & Ochsner, 1993; Watanabe, Nanez, & Sasaki, 2001), cognitive control (Lau & Passingham, 2007; van Gaal, Ridderinkhof, Fahrenfort, Scholte, & Lamme, 2008; van Gaal, Ridderinkhof, van den Wildenberg, & Lamme, 2009), attention and selection (Kentridge, Heywood, & Weiskrantz, 1999; Koch & Tsuchiya, 2007), and even reasoning and thinking (Bechara, Damasio, Tranel, & Damasio, 1997; Dijksterhuis, Bos, Nordgren, & van Baaren, 2006). That makes these functional arguments intrinsically ill-suited to put a sharp divide between conscious and unconscious processing.[2]

PHENOMENAL QUALITIES

A better approach to answer the question would be to study the phenomenal qualities of the two representations. Much is known about such qualities in the case of accessible and reportable percepts. For example, conscious percepts typically show the integration of features; in other words, grouping and binding. In a conscious percept, the loose set of elements that make up a scene are typically grouped into coherent surfaces and objects (Nakayama & Shimojo, 1992; Serences & Yantis, 2006; Singer, 1999). Moreover, different features, such as color, shape, motion, and size, are linked to each other. For example, when consciously seeing (and attending) the motorcycle of Figure 1, you instantaneously also see that it is red, and is

---

[2] Functions, whether cognitive or not, are of course also seen as irrelevant to consciousness in the original formulation of the so called *hard problem* of consciousness (Chalmers, 1995). I am not implying that that line of reasoning should be followed fully, as that way of posing the problem makes phenomenality—or qualia—almost impossible to study. For example, it renders invalid the very intuitions on which the conscious–unconscious divide is based (see above). But I do agree that many functions—cognitive functions in particular—do very little towards explaining qualia.

TABLE 1
Which key aspects of phenomenality (feature binding, inference, competition) are present in unattended or unreported visual representations?

	Iconic memory/Change blindness	Inattentional blindness	Attentional blink	Neglect extinction	Split brain, nondominant hemisphere
Feature binding or fig–ground segregation	Landman et al., 2003a, 2003b, 2004b	Scholte et al., 2006			
Object recognition		Mack & Rock, 1998	Marois et al., 2004		Sperry, 1984
Inference (illusions)				Vuilleumier et al., 2001	Corballis, 2003
Competition (rivalry)		Lee et al., 2007			O'Shea & Corballis, 2003

Notes: The table lists reported evidence for the presence of these features in various conditions that are characterized by the absence of attention and report. References are not meant to be exhaustive. Blank fields indicate the absence of knowledge, i.e., a fruitful area of further research. Object recognition is added to the list for completeness, but it may be questioned whether this is a feature of phenomenality, as categorical discrimination is also present in clearly unconscious states, such as masking or blindsight.

not wearing the blue of the couch next to it. We know that grouping and binding exist in working memory, and there is a strong tendency to believe they depend on attention (Luck & Vogel, 1997; Treisman, 1996). However, it has been shown that also in iconic memory there is figure–ground segregation of oriented line segments (Landman, Spekreijse, & Lamme, 2004a; Landman et al., 2004b), and binding of features such as size and orientation (Landman et al., 2003a).

Another property of conscious percepts is that they often are the end result of a competition between several possible groupings and bindings. This is most prominent in bi-stable phenomena such as perceptual or binocular rivalry (Blake & Logothetis, 2002). In addition, perception often is nonveridical, in that inferences are drawn that move away from the physical input. This is evident from the many visual illusions that exist, where we are tricked into seeing things that are not "really" there (e.g., Churchland & Churchland, 2002). It is unclear whether these phenomena—perceptual competition and inference—also exist when we do not attend to the visual input, such as in the case of the iconic memory representation, or during inattentional blindness or change blindness. It is known that both phenomena are largely independent of attention: Rivalry switches cannot be fully suppressed at will (Meng & Tong, 2004), and visual illusions are the prime example of the cognitive impenetrability of visual perception (Pylyshyn, 1999). Even when you know that the two double arrows of the Muller-Lyer illusion are of equal length, you still see them as different (Bruno & Franz, 2009). It is interesting to note that in visual neglect or extinction, visual illusions coming from the neglected hemifield still "work" to influence the percepts that are reported by the patient from the intact hemifield (Vuilleumier,

Valenza, & Landis, 2001). That is a first example that phenomenal qualities exist in a representation that is inaccessible by the patient.

Studying the phenomenal qualities of the two different representations would be a good research agenda to answer the question of whether there is phenomenality in iconic memory. If it turns out that the iconic memory representation shares almost all phenomenal qualities with attended/working memory representations—except cognitive access—it would make the most scientific sense to acknowledge phenomenality to this large capacity representation (and conversely to conclude that our phenomenal experience is widespread rather than limited). However, at present this is unclear. Similarly, it would be useful to study precisely the phenomenal qualities of other states where attention or report is absent, such as inattentional blindness, change blindness, attentional blink, neglect, or split-brain. In some cases, evidence for phenomenal qualities in those conditions has already been reported (see Table 1). What if we could fill the whole table? Shouldn't we conclude that the existence of phenomenality without report is the more parsimonious conclusion?

NEURAL ARGUMENTS

Would it help to find the neural correlates of the two representations (Crick & Koch, 1998a; Crick & Koch, 2003; Rees, Kreiman, & Koch, 2002)? Seemingly not. If we know that the large capacity iconic memory representations sits in visual cortex, while the limited working memory representation depends on the fronto-parietal network, how could that ever answer whether there is phenomenality to either of them? The

explanatory gap only seems larger between neuroscience and phenomenality. Neuroscience can, however, help to solve the seemingly impossible, and it can do so in different ways (Haynes, 2009; Rees, 2007; Tononi & Koch, 2008). My approach is to go beyond neural correlates, and turn these into neural arguments (Lamme, 2006; Seth, 2008; Seth et al., 2008; Seth, Izhikevich, Reeke, & Edelman, 2006).

The approach would be more or less equivalent to what is described above for the study of phenomenal qualities of the different representations. What if we could show that the neural correlate of the iconic memory representation shares all its essential qualities with the working memory representation—except the neural qualities that enable access and report? This would require finding the neural correlates of both types of representations—which is far less difficult than finding *the* neural correlate of consciousness—and finding out which are the essential qualities of these correlates that matter for phenomenality, which is almost as difficult as finding the neural correlate of consciousness but is a much more theoretical exercise (Seth et al., 2006; Tononi, 2004, 2008).

What would be the neural correlates of the two representations gauged via the paradigm of Figure 1? It has been shown (and explained at greater length in Lamme, 2004; Lamme & Roelfsema, 2000; Lamme, Super, & Spekreijse, 1998a) that each time we lay our eyes on a scene (by making an eye movement, or when it is flashed in an experiment), cortical visual processing goes through a succession of stages. First, information flows from visual to motor areas in what is called the fast feedforward sweep (FFS; Lamme & Roelfsema, 2000). Within 100 to 120 ms (in monkeys; probably about 200 ms in humans), activity spreads from V1 to the extrastriate and dorsal and ventral stream areas, all the way up to motor cortex, and prefrontal regions involved in controlling and executing movement. In some respects we could call this a cortical reflex arch, mediated by the feedforward connections. During the FFS, early visual areas extract features of the image such as orientation, shape, color, or motion (Bullier, 2001; Lamme & Roelfsema, 2000). But high level features are also detected. Cells in inferotemporal cortex distinguish between face and non-face stimuli with their first spikes (Oram & Perrett, 1992; Rolls & Tovee, 1994). The FFS thus enables a very rapid categorization of visual stimuli into all sorts of (probably) behaviorally relevant categories. Potentially related motor responses are initiated when the FFS reaches motor regions (Dehaene et al., 1998), and control centers are activated in prefrontal cortex (Lau & Passingham, 2007; van Gaal et al., 2008, 2009).

Not all stimuli travel all the way up. If multiple stimuli are presented, many can be represented at the early stages. However, in successively higher areas, competition between multiple stimuli arises. Attentional selection (in one way or another; Egeth & Yantis, 1997) may resolve this competition (Desimone, 1998; Desimone & Duncan, 1995). In the end, only a few stimuli reach the highest levels, such as the areas involved in planning and executing behavior. And because only attended stimuli are selected for deep processing, only these influence behavior, can be reported, are stored in working memory, etc. Unattended ones "die out" in the early stages of processing (Sergent et al., 2005). From a neural perspective, attention (or the consequence of attention) can thus be straightforwardly defined as the depth of processing that a stimulus reaches (Dehaene, Changeux, Naccache, Sackur, & Sergent, 2006; Lamme, 2003, 2004).

As soon as the FFS has reached a particular area, horizontal connections start to connect distant cells within that area, and feedback connections start sending information from higher level areas back to lower levels, even all the way down to V1 (Bullier, 2001; Salin & Bullier, 1995). Together, these connections provide what is called recurrent processing (RP) (Edelman, 1992; Lamme & Roelfsema, 2000), which is fundamentally different from the FFS. RP allows for dynamic interactions between areas that can grow ever more widespread as time after stimulus onset evolves. RP may thus form the basis of dynamic processes such as perceptual organization, where different aspects of objects and scenes are integrated into a coherent percept (Lamme & Spekreijse, 2000; Lamme et al., 1998a; Lamme, Vandijk, & Spekreijse, 1993; Sporns, Tononi, & Edelman, 1991). Also in RP, attentional selection makes a difference. Attentional amplification may turn RP into the widespread coactivation of visual and fronto-parietal areas, including parts of the brain that mediate action or control, so as to produce a coordinated and planned response to selected visual information (Dehaene et al., 2006).

The FFS is more or less automatically followed by RP. It is in fact very difficult to prevent FFS activation from being followed by RP. The only way to do this is by forcing a second FFS to follow the first one before RP related to the first one can fully ignite. This is what happens when a visual stimulus is masked: The FFS activation of the mask prevents RP for the target stimulus, which then becomes invisible (Enns & Di Lollo, 2000; Fahrenfort, Scholte, & Lamme, 2007; Lamme, Zipser, & Spekreijse, 2002).

FOUR STAGES OF NEURAL PROCESSING

Because depth of processing (attention) and the FFS/RP distinction are orthogonal, a visual stimulus can reach any of four stages of processing, illustrated in Figure 4.

- *Stage 1:* Superficial processing during the FFS. This would happen when a stimulus is not attended and masked. Unattended and masked words, for example, do not activate word-form selective areas, only visual areas (Dehaene et al., 2006), so do not even penetrate deeply into the ventral stream hierarchy.
- *Stage 2:* Deep processing during the FFS; for example, a stimulus that is attended, yet masked (and hence invisible). This stimulus does travel through the whole hierarchy of sensory to motor and prefrontal areas, and may influence behavior, as in unconscious priming (Dehaene et al., 1998; Eimer & Schlaghecken, 2003; Thompson & Schall, 1999).
- *Stage 3:* Superficial processing of a recurrent/re-entrant nature (RP); for example, a visual stimulus that is given sufficient time to evoke RP (i.e., is not masked within ~50 ms) yet is not attended or is neglected, as in neglect (Driver & Mattingley, 1998), inattentional blindness (Scholte, Witteveen, Spekreijse, & Lamme, 2006), change blindness (Landman, Spekreijse, & Lamme, 2003b; Landman et al., 2004a; Schankin & Wascher, 2007), or the attentional blink (Marois, Yi, & Chun, 2004).
- *Stage 4:* Deep (or a better word may be "widespread") RP. This is the case when RP spans the whole hierarchy from low level sensory to high level executive areas. This occurs when a stimulus is given sufficient time to engage in RP and is attended. Others have equated this to the situation that a stimulus has entered global workspace (Baars, 2005; Dehaene & Naccache, 2001).

How would this apply to glancing at the image of Figure 1? As soon as we look at the image, the FFS will be activated. Depending on where your attention happens to be focused, the activation from some objects will travel all the way up to motor and prefrontal areas, while others activate only some visual areas. Given that the image is not masked, this will then be followed by RP for all objects in the image. Attentional selection determines the extent of RP for each object. Objects that were attended during the FFS will have an advantage, but attentional selection can also switch to items that proved to be more salient during the FFS (i.e., penetrated more deeply, despite attention being focused elsewhere). Subsequently, many objects will evoke RP that is limited to a few visual areas (Stage 3) (Landman et al., 2003b), while only some evoke widespread RP (Stage 4), like in global workspace activation (Dehaene et al., 2006). Finally, the stimulus is removed. What then remains is the traces of the pattern of RP that was present at the moment the stimulus was switched off. Stage 3 turns into iconic memory, while Stage 4 turns into working memory (Figure 5).

The question we try to answer in this paper can now be propped up by neural arguments. We know that the large capacity representation of iconic memory corresponds to the remains of Stage 3 processing, while the limited capacity working memory representation corresponds to what remains of Stage 4. What would be the neural argument to grant phenomenality to Stage 3? Assuming there is phenomenality in Stage 4, and not in Stage 1 (remember the starting intuitions), we have to decide which are the essential qualities of Stage 4 that would "produce" the phenomenality, and see whether these neural qualities are also present in Stage 3. In Stage 4, we have recurrent processing and activation of fronto-parietal areas, which we don't have in Stage 1. Either of the two—or their combination—therefore is a candidate neural ingredient for phenomenality.

Let us first turn to the involvement of the frontoparietal network, a key ingredient of the neural equivalent of Global Workspace theory (Dehaene & Naccache, 2001). It is argued that activation of prefrontal areas is a prerequisite for consciousness (Rees, 2007) because it has the long range connections that enable the integration of information from otherwise widely separated regions of the brain (Del Cul, Dehaene, Reyes, Bravo, & Slachevsky, 2009). However, a convergence of information towards prefrontal cortex in itself does not appear to be sufficient for conscious sensations to arise. For example, there are now several studies showing activation of regions such as the frontal eye fields, anterior cingulate, pre-supplementary motor area, inferior frontal gyrus, and anterior insula by masked stimuli or other unconscious events (Klein et al., 2007; Lau & Passingham, 2007; Thompson & Schall, 1999; van Gaal et al., 2008, 2009). The type of masking used in these studies leaves little doubt about the absence of any visual sensation, as even forced choice detection and other rigorous behavioral measurements are at chance. Yet despite being unconscious, the prefrontal activation is functional, in the sense that it evokes effects of control on subsequent visible stimuli, such as response inhibition or strategic switching. Apparently, a feedforward convergence of information towards the
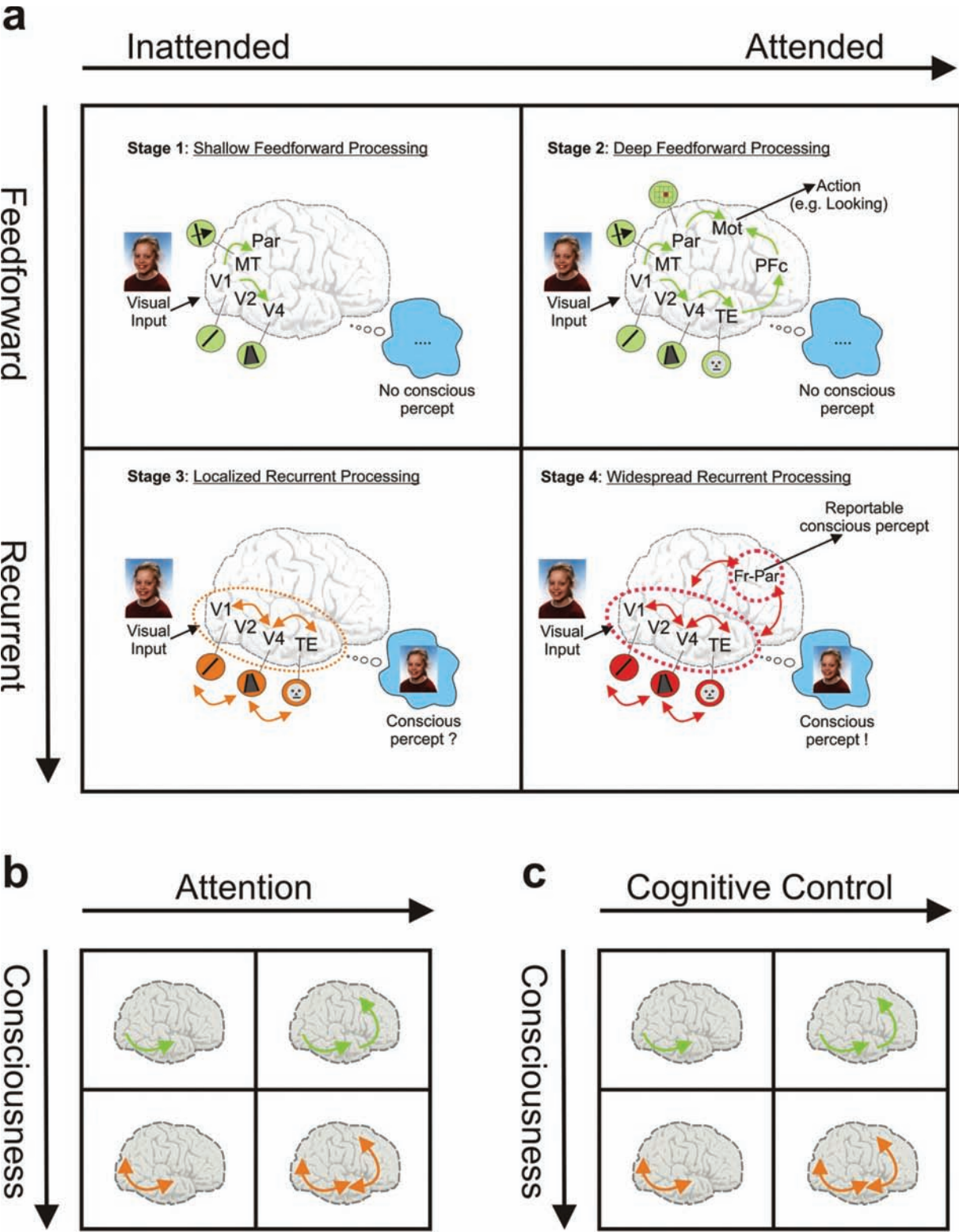
Figure 4.   (a) Four stages of cortical processing (see text for explanation). (b, c) When the difference between feedforward and recurrent processing (vertical axis, panel a) is identified to the difference between unconscious and conscious processing (vertical axis, panels b and c), how consciousness is orthogonal and independent of attention (b) and cognitive control (c) is readily explained.
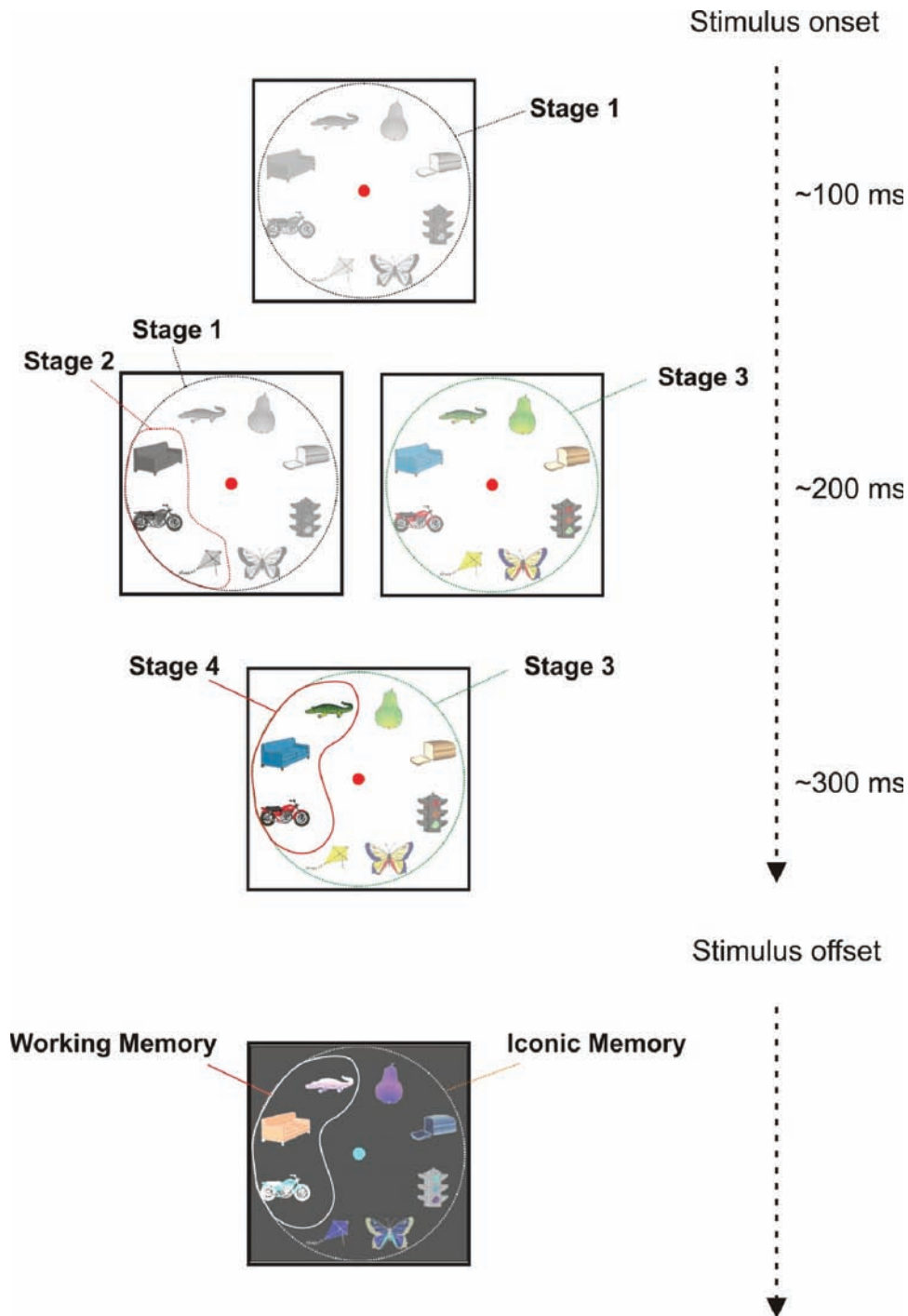
Figure 5. (a) A graphical depiction of the temporal evolution of processing stages that each of the objects in Figure 1 may reach at successive moments in time after presentation of the stimulus array, and how these processing stages change into iconic and working memory representations once the stimulus is removed. Initially, all objects are processed by low level areas in a feedforward fashion, so that basic features are extracted (Stage 1: faint gray shades). Some objects are processed more deeply (Stage 2: higher contrast gray shades), depending on top down and bottom up attentional selection. Meanwhile, recurrent processing in early visual areas emerges (Stage 3; faint colors) for all or most of the objects. Later still, recurrent processing grows more widespread (Stage 4, vivid colors) for those objects that are selected by attention (potentially slightly different ones than those that were favored initially, as attentional selection is influenced by previous processing). After stimulus removal, Stage 3 processing turns into iconic memory, while Stage 4 processing turns into working memory (inverted colors). In change detection paradigms (Figure 2), the sequence is repeated once the second array appears, and all stages are reset or overwritten, except for the Stage 4/working memory representation.

prefrontal cortex in itself is not yielding any conscious sensation, even when this information activates processes such as control and inhibition. In other words, both Stage 1 and Stage 2 processing are unconscious, no matter what areas are reached.

The remaining difference between Stage 4 and Stages 1 and 2 is that in the latter there is only feedforward processing, while in Stage 4 (and Stage 3) there is recurrent processing. Could that be the essential ingredient that gives phenomenality to Stage 4? And if so, what is so special about recurrent processing compared to feedforward processing that it yields conscious sensation? That recurrent processing is necessary for visual awareness is now fairly well established, and supported by numerous experiments (Boehler, Schoenfeld, Heinze, & Hopf, 2008; Camprodon, Zohary, Brodbeck, & Pascual-Leone, in press; Haynes, Driver, & Rees, 2005; Lamme, Super, Landman, Roelfsema, & Spekreijse, 2000; Lamme, Zipser, & Spekreijse, 1998b; Lamme et al., 2002; Pascual-Leone & Walsh, 2001; Silvanto, Cowey, Lavie, & Walsh, 2005; Super, Spekreijse, & Lamme, 2001). The key issue here is to what extent recurrent processing can be considered the ingredient of Stage 4 that *explains* phenomenality, and whether it does so better than the other main ingredient, which is the involvement of the fronto-parietal network.

KEY PROPERTIES OF PERCEPTION EXPLAINED BY RECURRENT PROCESSING

It is well established that many computational problems in vision can be solved better using recurrent rather than feedforward processing architectures. Among the many examples are feature grouping and segregation, figure–ground assignment, depth sorting of surfaces, and occlusion (Grossberg & Pessoa, 1998; Jehee, Lamme, & Roelfsema, 2007a; Roelfsema, Lamme, Spekreijse, & Bosch, 2002; Thielscher & Neumann, 2008). These processes are the first steps in going from the features that make up the image towards a description of the surfaces and their relative layout in depth (Nakayama, He, & Shimojo, 1995). What emerges from that is perceptual organization: the grouping and sorting of the elements that make up the image. That recurrent processing is mediating perceptual organization is also supported by many experimental findings (Lamme, 1995; Lamme & Spekreijse, 2000; Singer 1999; Zipser, Lamme, & Schiller, 1996).

Perceptual organization is essential for understanding visual perception as it appears to us, i.e., phenomenality. An example can clarify this. Consider seeing a face.

We know that when we see a face, neurons in the inferotemporal cortex are selectively activated. This occurs with extremely short latency, and is probably mediated by feedforward connections (Oram & Perrett, 1992). It is also known, however, that these neurons often signal the presence of a face in a feature-invariant way, that is to say, regardless of its size, position, color, details, or identity (Leopold, Bondar, & Giese, 2006; Rolls, 2000). In other words, these neurons mainly categorize the stimulus as being a face, as opposed to, say, a house or a box. That is not how we see the face. In the phenomenal experience of seeing a face, the face-ness (in a categorical sense) is joined by all the other features that constitute the face, such as its color, shape, emotional expression . . . In other words, a key feature of the phenomenality of seeing the face is that features and categories are integrated. This aspect of seeing the face is not captured by the feedforward activation of the face selective cell, but may be perfectly accounted for by recurrent interactions between face selective neurons and neurons that encode shape, color, etc. (Haxby et al., 2001).

The result of recurrent interactions between neurons is twofold. Neurons in lower regions modify their spiking activity so as to reflect the higher level properties. For example, a V1 neuron receiving feedback signals will fire more strongly when it is responding to features that are part of an object (Albright & Stoner, 2002; Lamme, 1995; Zipser et al., 1996). Conversely, higher level neurons start to reflect lower level properties, which enables them to further categorize the stimulus (Jehee, Roelfsema, Deco, Murre, & Lamme, 2007b). For example, face selective cells that initially only categorize stimuli as face vs. nonface will now become selective for the expression or identity of that face (Sugase, Yamane, Ueno, & Kawano, 1999). Recurrent processing thus makes it possible for neurons to signal shared information.

Sharing of information and feature integration are probably even better supported when neurons engage in recurrent interactions that result in synchronous activity. The role of neural synchrony, whether of an oscillatory nature or not, is still controversial, but much evidence supports the idea that synchronous firing is important for feature binding and conscious perception (Engel, Fries, & Singer, 2001; Engel & Singer, 2001; Singer, 1999; Uhlhaas et al., 2009).

Whether recurrent processing also explains other features of conscious percepts, such as inference or competition, is unclear. Models of visual illusions often use a recurrent architecture (Finkel & Edelman, 1989), but feedforward models exist as well. Cells in V2 respond with very short latency to illusory contours, indicating that at least some inferences are computed

in a feedforward fashion in the brain (Peterhans & Vonderheydt, 1989). Likewise, the jury is still out on the exact neural mechanism of perceptual or binocular rivalry (Leopold & Logothetis, 1999; Sterzer, Kleinschmidt, & Rees, 2009). Mutual inhibitory interactions between competing neural assemblies obviously are important, but these may operate at various levels of the visual system (Tong, Meng, & Blake, 2006). An important characteristic of rivalry is that different perceptual levels often switch simultaneously, as do separate regions with similar features. This is best explained by viewing rivalry as a competition between two recurrent assemblies, each representing a single potential percept (Grossberg, Yazdanbakhsh, Cao, & Swaminathan, 2008). Complete proof of the point would be obtained if it could be shown that a dominating percept "kills" the recurrency of the suppressed one.

In sum, recurrent processing has high explanatory power in accounting for important features of conscious percepts, as there is a strong homology between the integrated structure of perception and the structure of recurrent processing.

THE Φ ARGUMENT

Theoretical arguments further stress the importance of recurrent processing in explaining consciousness. Tononi argues that we should understand consciousness as the integration of information (Tononi, 2004, 2008). He uses a measure, Φ, to denote the amount of integrated information that is generated by a system when it goes from one state of processing to the next. This is determined by two factors: First, effective information must be generated, in the sense that the current state of the network—given its way of information processing—rules out a (large) number of previous states. Second, this information must be integrated, so that the amount of information that is generated by the system as a whole is larger than that of the sum of its parts. Tononi uses a powerful metaphor to explain this: Consider a digital camera. Each of the pixels of its sensor carries a bit of information, and so the camera can enter a huge number of different states. The camera is not conscious, however, because it would make no difference when the sensor is divided into individual pixels that work independently. The thalamocortical system consists of elements (neurons or maybe clusters of neurons) that each carry independent bits of information, while at the same time these elements are highly interconnected. That gives it the propensity to generate integrated information, and hence consciousness.

There is a fine balance between the requirements for independent information carried by each element and the elements being interconnected. Too low a connectivity lowers Φ because of the lack of shared information. But too high a connectivity also lowers it, because the elements lose specificity, resulting in less effective information being generated. This aspect of the theory has high explanatory power, in that it explains why the cortico-thalamic network has the capacity to generate high levels of Φ, while other brain structures, such as the cerebellum or basal ganglia, do not (Tononi, 2004). It also explains why we lose consciousness in sleep or epileptic seizures, even while synchronous activity and recurrent processing remain present: Connectivity becomes too high and aspecific (Alkire, Hudetz, & Tononi, 2008; Tononi & Massimini, 2008). Recurrent interactions that produce consciousness therefore should not be too strong (adding to arguments already made by Crick & Koch, 1998b, and countering those by Macknik & Martinez-Conde, 2009). A V1 cell that signals the orientation of a line segment should remain selective for that feature, regardless of whether that feature is part of an object or of the background. But it should alter its response somewhat, reflecting the different context. That is precisely what V1 neurons do in awake animals that report seeing the object (Zipser et al., 1996), and what they don't do during anesthesia (Lamme et al., 1998b), when a percept is absent (Super et al., 2001), or when recurrent interactions are disrupted (Lamme et al., 2002).

Tononi tested different architectures, and feedforward networks typically generate low Φ (Tononi 2004), which explains the absence of consciousness in Stage 1 and Stage 2 processing. In those cases, the brain works more or less like in the digital camera metaphor. Stage 4, on the other hand, probably has high Φ, given the simultaneous specificity and interdependence that is mediated by the recurrent interactions. The presence of consciousness in Stage 4 is therefore readily explained by this theory. Φ is difficult to measure in real networks such as the brain, but related measures are available, such as *causal density* (Seth et al., 2006, 2008). This method uses Granger causality, i.e., the way in which past activity at one point x of a network accounts for the activity at another point y above and beyond past activity of y itself. Like Φ, causal density increases when elements are independently influencing each other. Gaillard et al. (2009) used masking to compare causal density between (roughly) Stage 1/2 and Stage 4 processing, and indeed showed that there is a large difference between the two, confirming these theoretical notions.

The key question here is whether there is a sufficient level of Φ in Stage 3 to grant it phenomenality

as well. No direct experimental evidence is available. Theoretically, there probably is a much larger jump in Φ or causal density going from Stage 2 to Stage 3 than going from Stage 3 to Stage 4 (Seth, 2009; Seth et al., 2008). The Stage 2 to Stage 3 transition is characterized by the involvement of horizontal and feedback connections that introduce precisely the interactions that are necessary for high Φ. From Stage 3 to Stage 4, these interactions only grow more widespread. Importantly, Tononi's proposal allows for multiple complexes to coexist at the same time, each supporting its own conscious experience (Tononi, 2004). Therefore, the presence of a Stage 4 complex (the attended objects in Figure 5) does not preclude the presence of a Stage 3 complex (the unattended ones in Figure 5) that is conscious as well.

A FUNDAMENTAL NEURAL DIFFERENCE

It is also relevant to consider whether there are fundamental differences between Stages 1/2 and Stages 3/4—i.e., between feedforward and recurrent processing—other than the ones already mentioned. One such difference could be the extent to which these stages evoke synaptic plasticity. In recurrent processing, in particular when it involves synchronous firing, large numbers of neurons are simultaneously active, satisfying the Hebb rule (Singer, 1995). This is an ideal situation for the massive activation of NMDA receptors and ensuing synaptic plasticity (Dudai, 2002). NMDA receptor activation will obviously also occur during feedforward activation, particularly in the form of spike-time dependent plasticity (Dan & Poo, 2004), but it has been argued that this type of plasticity is all the more effective and specific in the case of oscillatory and synchronous activity in the gamma range, which depends on re-entrant connections. Moreover, for stimulus-specific learning, attention, and consciousness, large scale resonance in thalamocortical circuits combined with spike time dependent plasticity seems essential (Grossberg & Versace, 2008).

These theoretical considerations would imply that learning occurs mainly when neurons engage in recurrent interactions. Consequently, there might be a large jump going from Stage 2 to Stage 3 processing, in the extent to which the different types of processing evoke changes to the brain. Stages 1 and 2 do not evoke much learning, while Stages 3 and 4 do (Lamme, 2006). Indeed, masked stimuli (Stage 2) have only very brief effects on consequent behavior (Eimer & Schlaghecken, 2003), while unattended ones (Stage 3) can have much longer effects, which

are often equal to, and sometimes even larger than those of attended stimuli (Stage 4) (Bornstein, 1989).

Whether indeed recurrent processing is more tightly linked to NMDA receptor activation and learning than feedforward processing still awaits direct experimental evidence (but see Dudkin, Kruchinin, & Chueva, 2001). We have recently obtained evidence that in monkey visual cortex, blockade of NMDA receptors reduces recurrent signals, while blockade of AMPA receptors has its main effect on feedforward activity (unpublished data). Further support comes from studies of anesthesia. It has been noted that many anesthetic agents have as their final common pathway the blockade of NMDA receptor activation (Flohr, Glade, & Motzko, 1998), while at the same time it has been shown that anesthesia abolishes recurrent processing in the visual cortex (Lamme et al., 1998b), leaving feedforward signals relatively untouched.

How would this constitute an argument to grant phenomenality to Stage 3? First of all, it would show once more that there is no fundamental difference between Stages 3 and 4, but only between Stages 3 and 2. However, the argument comes from a neuroscience perspective entirely (Lamme, 2006), which makes it different from the previous ones, which are mixes of phenomenological, functional, and theoretical arguments. Second, the reasoning would improve the ontological status of consciousness, as it can be associated with a basic neural mechanism. This provides a metaphysical argument to grant phenomenality to Stage 3: If we can improve our science of consciousness by granting phenomenality to Stage 3, we—as proper scientists—are forced to do so (Lamme, 2006).

ONTOLOGICAL ISSUES

That brings another player to the stage, which is the science of consciousness. As is readily evident from Figure 4b, by laying the unconscious and conscious processing divide between Stages 1/2 and 3/4 respectively, attention and consciousness (in the sense of phenomenality) become orthogonal and hence independent properties (Lamme, 2003, 2004). There is currently much debate about whether these two functions are better considered independent, and a variety of experiments support this idea (Koch & Tsuchiya, 2007).

In fact, in the same stroke we can make consciousness orthogonal to other functions, such as cognitive control (Figure 4c). For example, both in the Go-NoGo and in the Stop-Signal paradigms, a reaction to a stimulus is required from the subject, unless this stimulus is preceded or followed by a NoGo or Stop

stimulus. Such inhibition of planned motor responses typically requires the activation of a prefrontal inhibition network (Eagle et al., 2008; Ridderinkhof, Ullsperger, Crone, & Nieuwenhuiss, 2004). Until recently, it was assumed that this network is only activated by conscious NoGo or Stop signals. It was shown that the very same network is also activated by masked and hence invisible NoGo or Stop signals. This activation furthermore is functional, in that it slows down the response to the visible stimuli (van Gaal et al., 2008, 2009). The finding is readily explained by the schema of Figure 4c. Masked stimuli activate the prefrontal inhibition network via the feedforward sweep (Stage 2), yet remain unconscious because of the absence of recurrent processing. In Stage 4, this activation would have been consciously reportable. But in both cases the function of cognitive control (response inhibition in this case) is executed. Stage 2 yields unconscious cognitive control; Stage 4 yields conscious cognitive control. Of course there are more differences than the phenomenal experience of the inhibitory stimulus between Stages 2 and 4. But these relate mostly to what could be called "motor" phenomenality, such as the sense and strength of control (Rowe, Friston, Frackowiak, & Passingham, 2002).

The schema of Figure 4 not only dissociates consciousness from attention and cognitive control, but may also explain why we can both consciously and unconsciously categorize stimuli (Reder, Park, & Kieffaber, 2009), or even perform unconscious reasoning and inference (Bechara et al., 1997; Dijksterhuis et al., 2006). Which cognitive function is executed depends on *which* area—which cognitive module—is activated. Whether this produces a conscious sensation or not is determined by whether the areas involved engage in recurrent interactions.

The schema of Figure 4, where the conscious–unconscious divide is between Stages 1/2 and Stages 3/4 respectively (Lamme, 2003, 2006), gives a much better ontological status to consciousness (or phenomenality) than Global Neural Workspace (GNW) theory, where that divide lies between Stages 1/2/3 and 4 respectively (Dehaene et al., 2006). In the latter situation, there is no orthogonality between attention and control on one hand and consciousness on the other. In fact, the two are highly confounded. That is another metaphysical argument to grant consciousness to Stage 3. If we only grant phenomenality to Stage 4, we force ourselves into a science of consciousness that is critically flawed from an ontological point of view. The first thing in science always is to demarcate concepts as clearly as possible.

WHY NOT BOTH?

At this point, it will be clear that there are several scientific arguments to conclude that the reason why we have conscious *visual* sensations in Stage 4 is the recurrent interactions between visual areas. These sensations can be reported and manipulated because Stage 4 also includes the prefrontal and motor areas. But if they are not, as in Stage 3, the visual sensation should still be there. All the neural ingredients that seem to matter for visual phenomenality are present in Stage 3, as they are in Stage 4.

From the neural description, we understand why there is access in Stage 4, and not in Stage 3. Stage 4 allows for the widespread integration of visual information with other sensory modalities, motor programs, executive control, and report, simply because here the visual activity is linked with cortical structures that enable such functions (Dehaene et al., 2006). RP in Stage 3 is limited to visual areas, and hence cannot directly influence motor control and other functions necessary for direct report. Several experiments, indeed, have shown that this is what happens in conditions such neglect (Driver & Mattingley, 1998), inattentional blindness (Scholte et al., 2006), or change blindness (Landman et al., 2003b, 2004a; Schankin & Wascher, 2007). In other words, it is perfectly understandable why we have reportable conscious visual sensations in Stage 4, or cognitive access to visual information. There are however no reasons whatsoever to assume that taking away the modules that enable access and report (Stage 3) also takes away the visual phenomenality.

In fact, linking visual phenomenality to access and report gives the whole notion of consciousness a poor ontological status. Stating that consciousness requires both recurrent processing *and* the inclusion of frontal areas (as in GNW theory) seems justified from a behavioral or maybe even an introspective point of view (although the latter is in fact neutral). A closer inspection shows that holding on to this idea impedes progress in our science of consciousness. It disregards the neuroscience argument (Block, 2007; Lamme, 2006). GNW theory is great to explain access, not to explain seeing.

VISION IS RICH EVEN WHEN YOU DON'T KNOW IT

Our conclusion is that Stage 3 processing is just as visually conscious as Stage 4. Change blindness is not blindness, it is the overwriting of one rich conscious visual sensation with another one. You might not know it, but you see all the objects in Figure 1. Likewise, in inattentional blindness, you don't remember having seen an

object, but that does not imply you had no conscious visual sensation at the moment it passed by. The event was just not stored in working memory. That is not blindness; it is forgetfulness (Wolfe, 1999). This "seeing without knowing" may sound strange, but in attentional blink paradigms this can be observed directly. When a T1 target is detected, this precludes detection of T2 a few hundred milliseconds later. You really don't know T2 was there. But this only works when the sensation of T2 is wiped out by a following stimulus. A brief presentation of T2, not followed by a mask, gets rid of the attentional blink. Likewise, we should be inclined to conclude that in neglect or extinction, or in split brain patients (Gazzaniga, 2005; Sperry, 1984) the problem is with access and report, but not with seeing.

That is the point of view of a science that goes beyond neural correlates of things we believe to exist introspectively or behaviorally. In this account, neuroscience is used to produce *explanatory correlates* (Seth, 2009) to arrive at a framework with maximal explanatory power regarding consciousness and its relation to other cognitive functions. The approach can also be compared to a factor analysis where behavioral and neural data are simultaneously reduced to underlying principal components, or basic constructs. In psychology, "raw" behavioral data are traditionally boiled down to arrive at underlying concepts such as attention or control. In much cognitive neuroscience, it is then tried to link these concepts to neural structures or mechanisms (Kosslyn, 1999). Here, the behavioral and neural data are taken *together* to arrive at concepts that are better than the ones that can be arrived at by either psychology or neuroscience independently (Lamme, 2004, 2006; Seth, 2008; Seth et al., 2008). In doing so, we automatically move away from behavioral or introspective starting points. If we didn't, neuroscience would not add anything.

This approach is particularly discomforting for consciousness. How can it be that we are mistaken about the identity of a phenomenon that derives its existence from introspection? Is this not a return to behaviorism, with its deep mistrust of anything mental? Or is it a form of eliminative materialism, where mental phenomena are entirely replaced by neural mechanisms, up to the point where talking about the mental would no longer make scientific sense? I think not. I accept the introspective intuition that there is something real and scientifically tractable about conscious experience. In fact it is the starting point of the approach to grant different mental states to Stages 1 and 4—almost entirely from introspection—and from there to extrapolate to Stages 2 and 3. The approach thus is obviously different from behaviorism, maybe even more so than the Global Workspace account, where behavior is taken as the primary evidence for the presence or absence of conscious experience. Neither is there a hidden agenda of eliminative materialism, because I embrace rather than deny the existence of qualia (Dennett, 1988), albeit not in their strictest form. My main objection is against a form of cognitive psychology where mental constructs are taken as undeniable truths to which neuroscience has to be fitted. I would argue that in the study of consciousness, there are no undeniable truths.

That is the standard approach in science. Intuition told us the sun revolves around the earth, while in fact it is the other way around. Intuition dictated creation, where evolution is the counterintuitive scientific answer. To make scientific headway in our science of consciousness, we need to acknowledge that our intuitions may be wrong and need to be set aside. The upshot is that—finally—we may start solving the questions that have been bothering us for the ages.

Commentaries

Stage 3 and what we see

Gideon Paul Caplovitz, Michael J. Arcaro, and Sabine Kastner

Department of Psychology, Princeton University, Princeton, NJ, USA
E-mail: gcaplovi@princeton.edu

Abstract: In his article, Lamme provides a neurotheoretical argument that recurrent processing (RP) produces the phenomenological sensations that form the contents of our conscious experiences. Importantly, he argues that this processing includes local intra-areal (i.e., horizontal connections) as well as local inter-areal feedback (i.e., from higher level sensory areas to lower level ones) interactions that

occur within the sensory cortices. This has direct implications for what the contents of these experiences may be and the role that neuroscience can play in identifying them.

Lamme argues that we subjectively experience Stage 3 processing that includes local intra-areal (i.e., horizontal connections) as well as local inter-areal feedback (i.e., from higher level sensory areas to lower level ones) interactions that occur within the sensory cortices. Having attributed phenomenology to stage 3, a question remains of how much information it can represent, and how much it thereby contributes to our phenomenological experiences.

Referencing the large capacity representation revealed through studies of iconic memory, Lamme suggests that Stage 3 processing represents a virtual copy of the physical world, and concludes that when examining his Figure 1, one in fact experienced all of the objects in the circular array. This implies Stage 3 processing and that our phenomenological experiences have an unlimited capacity to represent the world around us. However, the capacity of Stage 3 processing (in fact all processing of visual information) is limited by the optics of the eye, the receptive field properties of individual neurons, and their impacts on the dynamics of the circuits to which they belong. Starting at the earliest stages of processing, these physiological limitations lead to numerous ambiguities (e.g., the aperture problem: Adelson & Movshon, 1982) such that an infinite number of physical stimuli can produce the same neuronal activity. It could be argued that a fundamental goal of visual processing in general is to resolve such ambiguities.

As demonstrated by visual illusions and the principles of perceptual grouping, it is clear that we do not experience a direct representation of the physical world, but rather a limited "best guess" of what it might be (Wertheimer, 1924/1950). Certain aspects of how this best guess is constructed, such as the local interactions that support perceptual grouping (i.e., colinear facilitation: Polat & Sagi, 1993), texture segmentation (De Weerd, Sprague, Vandenbussche, & Orban, 1994), as well as competitive interactions that can weaken the representations of co-occurring stimuli (Beck & Kastner, 2009; Desimone & Duncan, 1995), are likely embodied at least partially if not entirely in Stage 3 processing. As such, Stage 3 processing does not allow us to experience a virtual copy of the physical world nor even a direct representation of the retinal image. You likely did not, as Lamme suggests, really see all the objects in his Figure 1, but instead saw the limited information about those objects (as well as the entire visual field) that the visual system was capable of representing.

This leads to an intriguing argument that the contents of our phenomenological experiences should not be considered in terms of the physical stimulus alone, but rather in terms of how information is represented in visual cortex, which, depending on the circumstance, may only loosely correspond to the stimulus itself or even the stimulus-driven inputs. This view provides a direct link between the neuroscience of consciousness and the systems neuroscience of vision. Namely, if we can understand the nature of the information being represented in the visual system, we may begin to understand the contents of our conscious experiences.

Finally, the various parallel pathways (both spatially specific and feature-specific) of visual cortex represent a dynamic system in which activity (both feedforward and recurrent) is ongoing, ever-changing, and dependent not only on its own internal dynamics (e.g., adaptation) but also on numerous inputs that are both stimulus and nonstimulus driven. As such, our phenomenological experiences (and their underlying neural mechanisms) are likely not isolated constructs that appear when a stimulus is present and then disappear when it is removed (as tacitly implied by Lamme's four-stage model), but rather are ever-present and simply change in response to changes that occur within the underlying neural mechanisms. When a stimulus is taken away, there is not a *loss* of consciousness, but a *change* in consciousness. Such changes need not be stimulus driven *per se*, but could arise, for example, from strictly top-down influences of spatial attention (Kastner, Pinsk, De Weerd, Desimone, & Ungerleider, 1999), prior experience (e.g., Bar, 2009), or as a result of internal dynamics as is the case in phenomena such as Troxler filling-in (De Weerd, Gattass, Desimone, & Ungerleider, 1995; Troxler, 1804).

We suggest that in order to understand the contents of our visual experiences, the focus may be better placed on the neural circuits that underlie them (e.g., those that support RP throughout the visual system) and frame questions in terms of how perturbations in the ongoing activity of these circuits arise. Thus, the question is reframed from "How is the conscious experience of a visual stimulus *created*?" to "How is the ongoing stream of conscious experience (and the underlying neural activity) *changed* and what has it *changed to*?" This viewpoint naturally allows the past states of the underlying network to be taken into account when considering how specific inputs will influence the information that will be currently represented.

Furthermore, it naturally dissociates the phenomenological experience of a given stimulus from the stimulus itself.

* * *

Consciousness minus retrospective mental time travel

Álvaro Machado Dias[1] and Luiz R. G. Britto[2]

[1]*University of São Paulo, Institute of Psychiatry, Neuroimage Lab Rua Dr. Ovidio Pires Campos s/n, São Paulo, Brazil*
[2]*Biomedical Institute, University of São Paulo*
E-mail: alvaromd@usp.br

Abstract: In this paper we apply the concept of mental time travel to introduce the basic features of full-blown conscious experiences (encapsulation in mental models and recollection). We discuss the perspective that Lamme's 'Level 3' experiences can be considered as part of the scope of phenomenological consciousness, in relation to which we emphasize the necessity to consider the different degrees of consciousness and how a particular situation compares to the conscious experiences present in resting states of wakefulness.

The understanding of consciousness is intrinsically bound to the notion of time in the mind, much like a DNA double helix. Much as we can gain insights about time by assuming some features about consciousness, the inverse path can aid our understanding of the deep and complex discussion conducted by Lamme.

"Phenomenological consciousness" can serve as a strategic tool in establishing the epistemological basis of a monist philosophical system (as in Merleau-Ponty, 1962, where the anti-dichotomous character is at the heart of the concept), or it can be used simply to express the way we relate to mental content and the outside world. In the latter case, conscious experiences can be thought of as: (1) what (and no more than what) is manifested from instant t to the very next instant (which cannot be defined precisely, for obvious reasons); (2) what (and no less than what) can be framed in mental models and presented to oneself ("access") or to someone else ("reportability").

All levels of reportability relate to past experiences, which are sometimes so close that we do not even realize that we are recollecting. Things are not so clear in relation to conscious access, but the generally accepted idea that full-blown conscious experiences (Lamme's "Level 4") are structurally built (from the bottom up) and then "accessed" (top-down) within specific cognitive domains suggests that access also relates to some sort of past.

The basic process involved in reportability and access (in an even narrower time frame) is the capacity to switch from a modus operandi where construction is in the core (bottom-up) to another that is driven by recollection. In both cases the mechanism may be conceived as a specific type of retrospective mental time travel (MTT) in narrow time frames (Arzy, Adi-Japha, & Blanke, 2009), which encapsulated appearance; conversely, the whole discussion about phenomenality and its relation to particularities of cognitive architecture (e.g., recurrent processing) can be assumed to be the effort to determine whether any level of retrospective MTT is needed for one to consider that some conscious phenomenon has taken place. The idea that someone suffering from a neglect syndrome might have conscious experiences of events that cannot be remembered instants later (see Lamme) can be thought of in terms of an inability to activate MTT within the narrowest intervals.

With this perspective in mind, the boldest and most interesting aspect of Lamme's thesis is his persuasive defense of the standpoint that neglect and other conditions where retrospective MTT is precluded should be considered within the scope of phenomenological consciousness, since the basic biological dimensions of conscious experience are present. However, it is attractive to consider that when we take account of conscious experience in the absence of retrospective MTT, we immediately assume that what is being approached is a manifestation that should be attributed to alterations in parameters usually characterized by low encapsulation and reportability. *Ceteris paribus*, changes in bodily expressions that lead to reportability in the appropriate biocomputational environment (leading to retrospective MTT) can be assumed as indices of conscious experience in Lamme's "Level 3" conditions, since these are equivalent to the former minus encapsulation and activation of the recollection mode.

In effect, this means that it is possible to leave aside the delicate issue of cognitive architecture (recurrent networks vs. synchronicity) and still reach

the same endpoint in the debate. For example, orientation reflex (OR) is a fast response that tends to occur before any possibility of conscious access and reportability, which is associated to evoked potentials and changes in skin conductance response (SCR) (Barry & Rushby, 2006); considering that the latter is an index of reportable stress and arousal (Critchley, Elliott, Mathias, & Dolan, 2000), it can be said that their manifestation in, e.g., neglect syndrome for aversive stimuli raises "Level 3" conditions to the status of part of the scope of phenomenological consciousness.

What is most interesting in this example is that the discussion surrounding OR is much more prosaic than the one that relates to neglect syndromes, providing us with an opportunity to consider that the real issue is not whether OR is conscious or not, but the extent to which it is—a question that we can literally think about from minute to minute.

This perspective discloses the importance of integrating the categorical "stage system" to a gradual frame of conscious experience (probably defined by fuzzy parameters). There is no definitive objection to assuming that there is an angle from which all "Level 3" phenomena can be taken as part of the scope of consciousness (regardless of what happens inside the brain). Nevertheless, when we consider the nature and extent of what a subject may be conscious of in these situations where no encapsulation/recollection is provided, we are forced to accept the conclusion that conscious experience in these cases should not be considered to be much different from conscious experience in resting states of wakefulness, which is a practical way of saying that, in effect, it is conscious of nothing except the passage of time from one moment to another.

ACKNOWLEDGMENTS

* * *

Explaining seeing? Disentangling qualia from perceptual organization

Agustin Ibáñez[1] and Tristan Bekinschtein[2]

[1]*Institute of Cognitive Neurology (INECO), Buenos Aires, Argentina, and; Career of the National Scientific and Technical Research Council (CONICET), Buenos Aires; Favaloro University, Buenos Aires; and Universidad Diego Portales, Santiago, Chile*
[2]*Medical Research Council, Cambridge, UK, and INECO, Buenos Aires, Argentina*
E-mail: neurologiacognitiva.org

Abstract: Visual perception and integration seem to play an essential role in our conscious phenomenology. Relatively local neural processing of reentrant nature may explain several visual integration processes (feature binding or figure–ground segregation, object recognition, inference, competition), even without attention or cognitive control. Based on the above statements, should the neural signatures of visual integration (via reentrant process) be non-reportable phenomenological qualia? We argue that qualia are not required to understand this perceptual organization.

The main point in Lamme's paper is that "we need other evidence about the presence or absence of a conscious sensation than introspection or behavior." (p. 240). We could not be more supportive of this proposal since, in fact, we have been developing measures to address it using electrophysiology (Bekinschtein et al., 2009a) and electromyography (Bekinschtein et al., 2009b). Lamme advocates that seeing is rich but reporting what you see is poor (because the transfer from visual sensation to cognition is limited). Moreover, seeing without access (and report) should have an independent phenomenological quality.

By considering that consciousness is not about access (as proposed by the global neuronal workspace (GNW) theory) but is about phenomenal qualities, Lamme is steering towards visual integration of features as a key component of a conscious percept, and considers qualia a necessary component of it. The GNW does not take the phenomenology of the stimuli into account, and in this sense the criticism of the GNW that working memory is needed to produce qualia loses validity. GNW does not "need" qualia for conscious access (Dehaene & Naccache, 2001).

Furthermore, we consider that even in Lamme's proposal, no qualia are explained but some properties of visual integration (e.g., feature binding or figure–ground segregation, object recognition, inference, competition) are. Those properties may already occur when reentrant process is restricted to a few visual areas (Stage 3). Consequently, Lamme assumes that

Stage 3 is the neural ingredient of phenomenology. Nevertheless, no one of those properties (or their combination) constitutes a *quale* (at least in the phenomenological sense; Block, 1990).

Lamme attributes a phenomenological quality to visual neural processing. It might as well be that there is something that it *is like* to retain briefly a certain number of objects (Stage 3 in Lamme). However, to constitute qualia requires being introspectively accessible, and this is not the case, since Lamme's Stage 3 does not reach the status of being reportable.

There is no need to invoke a phenomenological property to explain visual processing. As Lamme advocates, there are many computational models that yield feature grouping and segregation, figure–ground assignment, depth sorting of surfaces, and occlusion (e.g., Grossberg & Pessoa, 1998; Thielscher & Neumann, 2008), or robots with simulated brain-based recurrent process that learn object recognition (Edelman, 2007). Should we grant qualia properties to those computational networks in order to understand perceptual organization? We believe that *quale* as a property is certainly not needed to understand that kind of perceptual organization (Ibáñez, 2007).

Perceptual organization is part of our experience of qualia (and that is an introspective judgment). But perceptual organization at the same time can be reproduced by natural and artificial neural networks that do not exhibit or require phenomenological properties. Why does Lamme grant phenomenological properties to perceptual organization? Does he implicitly assert an introspective judgment when he classifies a non-necessary phenomenological process (e.g., a reentrant neural process of visual integration) as having a phenomenological property (qualia)? In opposition to Lamme's proposal, this argument does not go beyond introspection.

The neural arguments of Lamme resemble an oldfashioned and frequent categorical error in inter-level explanations of mind (Ryle, 1949). Nevertheless, a fruitful but paradoxical consequence of Lamme's proposal is that introspection allows us to think of qualia *as* perceptual organization. After implicitly making this analogy, we can propose that the neural correlates of qualia (e.g. local reentrant processing) may also have the status of a phenomenological property. It is important to note that this is a metonymic explanation: the part—*e.g. perceptual grouping*—for the whole—*qualia*); probably valid, but not causal. Despite Lamme's assertions, at this stage, his model does not go beyond the neural correlates of a consciousness agenda.

*   *   *

How consciousness will change our view on neuroscience

Morten Overgaard

CRNU, Hammel Neurorehabilitation and Research Center, Aarhus University, Denmark
E-mail: mortover@rm.dk

Abstract: Victor Lamme proposed that the study of consciousness should not be based on introspection. Nevertheless, Lamme understands consciousness as a subjective phenomenon, and introspection as the way in which we acquire knowledge about consciousness. This makes the task to find introspective-free methods to study consciousness difficult. Lamme attempts to make progress by introducing "neural arguments," but fails to show how such arguments are independent of introspective methods which seem necessary in order to decide how any neural process relates to mental phenomena. This commentary paper thus aims to show that our understanding of neural correlates is shaped by introspection.

Lamme has proposed that our scientific investigations of consciousness should not be based on introspection. The argument states that our intuitive feeling of knowing what we are conscious of may be challenged under certain conditions. When perceiving a visual scene like the one in Sperling's (1960) experiments, we may immediately believe we have consciously seen all the presented stimuli, but after a closer introspective examination we may doubt the validity of this belief.

Before returning to this interpretation of Sperling's experiments, let me however introduce my take on the central concepts involved. Consciousness, however hard to define, seems in most recent publications to be defined as subjective experience. Although Lamme does not present a more formal definition, this seems to be his way of using the term as well. Lamme's conception of introspection seems to be a directing of attention towards the contents of consciousness, and thus different from directing it towards external objects. This definition of introspection resembles

those of other researchers (e.g., Jack & Roepstorff, 2002; Overgaard & Sørensen, 2004) and is in opposition to others (e.g., Dretske, 1995). According to Lamme, consciousness does not depend on or necessarily lead to introspection, as we are able to consciously perceive and report about scientific findings with different ("better") validity than we can do based on introspection.

This conception, however, does not come for free, but carries certain necessary consequences. For one thing, introspection becomes the *sine qua non* for knowledge about consciousness. We may think and report about conscious contents only by way of introspection—i.e., attending to consciousness. Being a "disbeliever" in the validity of introspection, Lamme distances himself from other researchers endorsing the same conceptions (e.g. Jack & Roepstorff, 2002; Overgaard, 2006), and he places himself in a rather odd position in trying to study consciousness with little trust in the method he must accept as crucial.

Lamme attempts to get out of this dilemma by introducing "neural arguments." Neural arguments, it seems, differ from neural correlates in such a way that they may be used to make conclusions about conscious experiences. Lamme (p. 213) argues that "we have to decide which are the essential qualities . . . that would 'produce' the phenomenality" and then look for conditions where these essential qualities are present. This, then, would be our introspection-free method to decide whether a subject is conscious without having to ask any direct question about it.

Lamme poses "neural arguments" to decide whether "superficial recurrent processing" should be associated with conscious experience, as Lamme thinks "widespread recurrent processing" should. Were we now to believe that "recurrent processing" is so strongly associated with consciousness that the latter never would appear without the former, we would still not have found such an introspection-free method. To arrive at this association, one would have to conduct several experiments correlating recurrent processes with consciousness—using introspecting experimental participants. Consequently, this method would not be independent of introspection but would carry the strengths, weaknesses, and limitations of introspection. Hence, the "neural argument" method can be no stronger than "neural correlates of introspective reports."

The issue is ironically characterized by the introductory example from the Sperling experiment. The method by which Lamme rejects introspective evidence is . . . introspective evidence. It is only on "closer introspective examination" that doubts may be raised about the validity of the initial, introspectively based belief.

Sadly, the attempt to disregard introspection and find oneself fully dependent on has often been seen before. In fact, it seems a necessary logical consequence of any method suggesting an independent objective measure of consciousness.

According to the view presented here, contrary to the title of Lamme's paper, neuroscience on its own should never change anyone's view on consciousness. However, insights into consciousness and the methods of its study (e.g., Sandberg, Timmermans, Overgaard & Cleeremans, in press) would be of great value to neuroscience, and might indeed change our view of it. By introspection, we form the very categories we put to use when analyzing brain activations as correlations of "colour perception" or "happiness." Thus, neural correlates to subjective states are shaped by introspection in the scientific process.

ACKNOWLEDGMENTS

* * *

Electrophysiological evidence for phenomenal consciousness

Antti Revonsuo[1] and Mika Koivisto[2]

[1]*Centre for Cognitive Neuroscience, University of Turku, Turku, Finland, and University of Skövde, Sweden*
[2]*University of Turku, Turku, Finland*
E-mail: revonsuo@utu.fi

Abstract: Recent evidence from event-related brain potentials (ERPs) lends support to two central theses in Lamme's theory. The earliest ERP correlate of visual consciousness appears over posterior visual cortex around 100–200 ms after stimulus onset. Its scalp topography and time window are consistent with recurrent processing in the visual cortex. This electrophysiological correlate of visual consciousness is mostly independent of later ERPs reflecting selective attention and working memory functions. Overall, the ERP evidence supports the view that phenomenal consciousness of a visual stimulus emerges earlier than access consciousness, and that attention and awareness are served by distinct neural processes.

Event-related brain potentials (ERPs) track stimulus processing with high temporal resolution. Different ERP waves can be distinguished from each other by their typical onset and peak latency, and their scalp topography. Thus, by exploring the ERP correlates of visual consciousness, it may be possible to answer the following three sets of questions.

1. How many distinct or empirically dissociable ERP waveforms correlate with visual consciousness? More specifically do ERPs reveal only *one* correlate of consciousness or *two separable* correlates of consciousness (one for phenomenal, the other for access consciousness)?

2. In what time window do the ERP correlate(s) of consciousness appear and what kind of scalp topography do they show? More specifically, is there any ERP correlate to be found with a temporal evolution and scalp topography consistent with localized recurrent processing in the visual cortex and thus likely to reflect purely phenomenal consciousness? Or are there only ERP correlates of consciousness that are dependent on attention and working memory, and whose latency and scalp topography suggest widespread "global workspace" processing and frontoparietal involvement? If the latter is true, then the ERP evidence would not support the existence of purely phenomenal consciousness, only the access type of consciousness.

3. Are the ERP correlates of consciousness dissociable from the ERP correlates of attention? More specifically, does the ERP evidence support Lamme's proposal that awareness and attention are served by different neural mechanisms?

To answer questions such as the above, we have recently published a series of ERP experiments of visual awareness and attention. We have also reviewed the relevant wider ERP literature on visual awareness (Koivisto & Revonsuo, 2010). From the published evidence, a coherent pattern emerges that promises to answer the above questions. The consistent pattern of results (and our conclusions below) are based on converging evidence from a number of different kinds of experiments using different experimental paradigms (e.g., masking, change blindness, attentional blink). In all the relevant experiments, the idea has been to contrast a condition where a visual stimulus enters consciousness with a condition where it does not, and to study the differences between these two conditions in the event-related responses of the brain.

Overall, three different candidates for ERP correlates of visual consciousness have emerged: an early positive enhancement around 100 ms from stimulus onset (P1) (Pins & ffytche, 2003), a negative difference wave, visual awareness negativity (VAN; Koivisto & Revonsuo, 2003), typically occurring 150–250 ms from stimulus onset, and a late positive wave (LP) (Niedeggen, Wichmann, & Stoerig, 2001), occurring after 300 ms from stimulus onset. Of these, the P1 enhancement has gained the least support as a genuine correlate of consciousness. It has been only occasionally observed, and probably reflects an attentional effect to near-threshold stimuli that are very difficult to distinguish. By contrast, VAN is the most consistently observed ERP correlate of visual consciousness. Its time window (onset invariably after 100 ms, peaking usually between 200 and 300 ms) and its occipito-temporal, posterior scalp topography are perfectly consistent with the localized recurrent processing in the visual cortex suggested in Lamme's model of consciousness. The LP is similar to other ERP waves in the P3 family of ERPs, with a central and widespread scalp topography. P3 is generally thought to reflect the updating of working memory and other higher cognitive functions that involve frontoparietal attentional networks. Thus, the LP is most naturally interpreted as a correlate of access consciousness.

In experiments where visual consciousness and attention have been separately manipulated it has been possible to test whether phenomenal consciousness (as reflected by VAN) is independent of top-down attention. These experiments have revealed that VAN indeed is independent of selective attention and of the scope of attention (global/local) (Koivisto & Revonsuo, 2008; Koivisto, Revonsuo, & Lehtonen, 2006; Koivisto, Revonsuo, & Salminen, 2005): VAN emerges even for nonselected and nontarget stimuli. The effects of attentional selection affect at most the latter part of VAN.

However, spatial attention appears to be a special case. In a recent experiment (Koivisto, Kainulainen, & Revonsuo, 2009) where spatial attention was strongly manipulated, the stimuli that did not receive any spatial attention also did not elicit any VAN. Therefore, according to these results, spatial attention, but not other forms of attention, is necessary for visual consciousness. On the basis of these results it can be suggested that patients suffering from neglect have no phenomenal consciousness of the neglected stimuli, at least insofar as the stimuli are neglected because of absent spatial attention to them. Here our interpretation is in conflict with Lamme's suggestion that neglect patients are phenomenally conscious of the neglected stimuli.

In conclusion, ERPs are a useful tool to test hypotheses concerning visual awareness and atten-

tion. The evidence so far mostly supports the main ideas of Lamme's model: that phenomenal consciousness is independent of higher cognition and attention, but with the important exception of spatial attention.

ACKNOWLEDGMENTS

* * *

Neural theories need to account for, not discount, introspection and behavior

Anil K. Seth and Adam B. Barrett

University of Sussex, Brighton, UK
E-mail: a.k.seth@sussex.ac.uk

Abstract: A satisfying neuroscience of consciousness must account for phenomenological properties in terms of neural properties. While pursuing this project may challenge our intuitions about what we are conscious of, evidence from behavior and introspection should not be discounted. All three lines of evidence need to be integrated in order to naturalize phenomenal experience.

The science of consciousness is rapidly maturing, thanks in large part to highly informative work in cognitive neuroscience, exemplified by Lamme and colleagues. In his target article, Lamme argues that consciousness science is best served by allowing neural evidence to trump intuitions and evidence derived from introspection and behavior. It does seem right to allow that phenomenal consciousness could in principle exist in the absence of introspective or behavioral report (assuming one does not *define* consciousness in terms of reportability). The challenge then is epistemological: How can one infer the presence or absence of consciousness without the validation provided by subjective report? Lamme suggests, and we agree, that a useful approach is to consider which properties of neural processing *account for* rather than merely *correlate with* basic features of phenomenal consciousness (i.e., "explanatory correlates"; Seth, 2009). But we are not convinced that such neural evidence should necessarily trump evidence from other sources.

A neural process may be said to account for an aspect of consciousness when there are identifiable homologies among neural processes, phenomenal properties, and their cognitive and behavioral accompaniments. Lamme focuses on *recurrent processing* (RP), arguing that it can account for a variety of properties of conscious experience, including figure–ground segregation, feature grouping, occlusion, and the like. So does the existence of RP in Lamme's "stage 3" allow the inference of phenomenal consciousness in the absence of subjective report?

A first challenge is that RP is prevalent within the brain, occurring even in states that are generally considered unconscious including general anesthesia, dreamless sleep, and epileptic absence (Arthuis et al., 2009). RP may therefore be necessary but is unlikely to be sufficient for consciousness. Secondly, some of the properties identified by Lamme, plausibly underpinned by RP, may also characterize unconscious processing. For example, studies of hemispatial neglect have shown that preattentive feature grouping can induce illusory contours in the absence of reportable awareness (Vuilleumier, Valenza, & Landis, 2001). Correct interpretation of these results is, however, difficult. On one hand the absence of reportable awareness in neglect does not necessarily exclude phenomenal consciousness (moreover, the inference to illusory contours still relied on a non-introspective form of behavioral report, line bisection). On the other hand, there seems little *a priori* reason to assume that feature grouping and illusory contour induction are uniquely properties of conscious experience. Thus, Vuilleumier's results could indicate (1) non-reportable phenomenal consciousness, assuming that illusory contours uniquely characterize conscious as opposed to nonconscious contents; (2) reportable phenomenal consciousness, assuming a dissociation between introspective and non-introspective behavioral report; or (3) unconscious processing, challenging the association of phenomenal consciousness with grouping processes of the sort needed for illusory contour induction. We have dwelt on this example because it highlights the difficulty of determining via experiment the fact of the matter about phenomenal consciousness.

To do better, we need explanatory correlates that have stronger *a priori* connections with phenomenal consciousness. One example is that every conscious scene is different from every other possible conscious scene (differentiation), yet is experienced as a unified whole (integration). This property (let's call it "dynamical complexity") is central to recent theoretical approaches including Edelman and Tononi's "dynamic core hypothesis" (Edelman, 2003) and Tononi's "information integration theory" (Tononi, 2008). According to these theories, and *contra* Lamme, we consciously see a face as a face not only

virtue of the integration of face-specific perceptual features but also because these integrated features are discriminated from a vast repertoire of alternative possibilities: A face is a face to the extent that it is not a house, a car, the smell of a lemon, an explicit belief, etc.

As Lamme recognizes, dynamical complexity can be operationalized via measures such as neural complexity, Φ (phi), and causal density (Seth, Izhikevich, Reeke, & Edelman, 2006). These measures offer an advance over RP because they are more plausibly associated with conscious as opposed to nonconscious contents (illustrated by the face example above), and because they explain why unconscious conditions (seizures, anesthesia, sleep) can nonetheless show high levels of RP. Unfortunately, existing measures are extremely hard to apply in practice. For example, the current Φ is well defined only for discrete memory-less dynamical systems (Tononi, 2008) and causal density, while measurable from time series data, depends on assumptions of statistical stationarity. More importantly, the plausibility of dynamical complexity as an explanatory correlate is *not* derived purely from neural evidence, but also from introspection (i.e., what are the key invariant phenomenal features of consciousness?) buttressed by inferences relating to behavior and cognition (i.e., that the function of consciousness, with respect to dynamical complexity, is to provide informative discriminations within a vast possibility space of potential conscious scenes).

In conclusion, a mature neuroscience of consciousness will indeed show how neural processes can account for phenomenal properties. In the limit, such processes may shed light on dissociations between phenomenal consciousness and subjective report. However, reaching this limit will require not only more thoroughly worked out explanatory correlates, but also an improved understanding of the mechanisms underpinning report itself. And even then, a mature theory will need to explain why our introspection appears as it does, and which behavioral and cognitive functions are subserved by consciousness, whether accompanied by report or not.

* * *

Localized phenomenology: A recurrent debate

Murray Shanahan

Department of Computing, Imperial College London, London, UK
E-mail: m.shanahan@imperial.ac.uk

Abstract: The neuroscience carried out by Lamme and colleagues is fascinating and important. But his case for localised phenomenology rests on a flawed understanding of rival theories and a misguided view of introspective report.

Lamme's paper is the lastest attempt, following in Block's footsteps, to argue for "phenomenology without access." The topic is fraught with difficulty since the relevant data, deriving from experiments such as Sperling's as well as those of Lamme and colleagues, are hard to interpret without taking a stand on philosophical issues. For Lamme, localized recurrent processing (Stage 3) is sufficient for "phenomenology," and the widespread cortical processing that is the hallmark of "access" (Stage 4) is not necessary. Lamme is surely to be applauded for orienting the debate towards the findings of neuroscience. But his arguments are problematic in a number of ways.

In his criticisms of global workspace theory (GWT), Lamme assumes a neural version according to which prefrontal activation precedes broadcast and is therefore a prerequisite for the conscious condition. But this is a faulty assumption. According to the basic tenets of the theory, global broadcast goes hand in hand with the conscious condition. But the idea that a single brain area is the locus of broadcast does not follow from this. A more plausible view is that broadcast is effected by a brain-wide communications infrastructure realized by the cerebral white matter (Shanahan, 2008, 2010). In other words, it's the network that enables access, not some "module" as in Lamme's caricature of the theory. Or more precisely (because "access" is a vague notion), it's thanks to this network that a localized brain process, or coalition of brain processes, can exercise systemic influence.

So Lamme is wrong to ascribe to GWT the view that "taking away the modules that enable access and report . . . also takes away the visual phenomenality." (p. 219). Recent imaging studies have shown that white matter connectivity presents a hierarchically modular small-world topology with a pronounced connective core comprising multiple hub nodes (Hagmann et al., 2008). Such a topology is ideally suited to the global dissemination of influence and information, and is robust to damage. No lesion of an individual hub node in such a network is sufficient to disable the communications infrastructure that is hypothesized to underpin broadcast.

Lamme's misinterpretation of GWT is also apparent when he notes sceptically that "visual information

seemingly needs to *go somewhere* to achieve phenomenality" and claims to "smell a homunculus" in the theory (p. 210). Perhaps the "somewhere" Lamme has in mind is the prefrontal cortex. But GWT does not claim that information has to go "somewhere" for the conscious condition to arise. On the contrary, the very essence of the theory is its claim that, to give rise to the conscious condition, information has to go *everywhere*.

The same misunderstanding guides one of Lamme's key arguments for ascribing "phenomenality" to Stage 3 processing. He asks us to consider what distinguishes (indisputably unconscious) Stage 1 and Stage 2 processing from (indisputably conscious) Stage 4 processing. Having dismissed prefrontal processing alone as a candidate (by implication dismissing his misrepresented version of GWT), Lamme alights on the fact that "the remaining difference between Stage 4 and Stages 1 and 2 is that in the latter there is only feedforward processing, while in Stage 4 (and stage 3) there is recurrent processing." (p. 216). The conclusion that recurrent processing is the vital ingredient naturally follows, along with the ascription of "phenomenality" to Stage 3. But *widespread* recurrent processing, the signature of Stage 4, is more than just recurrent *prefrontal* processing. And according to GWT, properly construed, it is widespread activation, not prefrontal activation, that counts. So Lamme is wrong to claim that recurrent processing is the only candidate for what distinguishes Stages 1 and 2 from Stage 4. Wide-spread (recurrent) processing remains a candidate, and GWT as a consequence is still in the frame.

In short, Lamme's claim that recurrent processing alone is sufficient for the conscious condition, even without widespread activation and the resultant capacity for introspective report, is unconvincing. Moreover, there are profoundly important reasons for taking introspective report as a valid indicator of both consciousness and its absence. Suppose a company develops a drug to relieve pain. It works by alleviating the effects of pain only at Lamme's Stage 4, leaving activation at Stage 3 unaffected. (Of course, Lamme's discussion concerns vision, not pain, but his claim must generalize if it is to be taken seriously.) In clinical trials, patients who take the drug report relief from their pain.

However, suppose the authorities refuse to license the drug (under the influence of Lamme's paper, perhaps). Their justification states: "Despite the patients' introspective reports to the contrary, we must assume that the phenomenology of pain is still present, because recurrent neural processing at Stage 3 is unaffected by the drug. The patients only think

they are not in pain. Thanks to neuroscience, we know better." The point of the story is obvious. The neuroscience itself is not in dispute. What matters is how we characterize its findings in ordinary human terms.

* * *

Experiencing more complexity than we can tell

Bert Timmermans[1], Bert Windey[2], and Axel Cleeremans[2]
[1]*Department of Psychiatry, Neuroimaging Group, University of Cologne, Germany*
[2]*Université Libre de Bruxelles, Brussels, Belgium*
E-mail: bert.timmermans@ulb.ac.be

Abstract: The notion of unreportable conscious contents is misguidedly premised on the idea that access necessarily follows phenomenal representation. We suggest instead that conscious experience should be viewed as a constructive, dynamical process that involves representational redescription: The brain continuously and unconsciously performs signal detection on its own representations, so developing an understanding of itself that subtends conscious experience. Cases where phenomenality seems to overflow access are thus illusory and depend on interactions between task instructions and stimulus complexity. We support this perspective through recent evidence suggesting that properly graded, qualitative subjective reports appear to be exhaustive in revealing conscious knowledge.

We agree with two intuitions put forward in the target paper. The first is that we see more than we can tell—phenomenal experience does indeed seem to overflow our ability to report on its contents. The second is that consciousness is a graded rather than a dichotomous phenomenon. We note that such gradedness is not incompatible with nonlinearity (as proposed by the global workspace theory, GWT), of which the sigmoid function is a perfect example.

We disagree, however, that consciousness should be defined purely in neural terms. It simply does not make sense to us to speak of unreportable phenomenal contents. Thus, awareness must be distinguished from mere sensitivity. I can be sensitive to some stimulus yet remain unaware of it. In such a case, it makes little sense to think of this sensitivity as involving *any* sort of phenomenal content. It simply reflects the

fact that I can react to the stimulus in certain ways, just as an obviously unconscious motion detector can be properly said to be sensitive to movement without this sensitivity implying awareness in any form. Conscious sensitivity, however, crucially involves phenomenal content. It is this specific difference between sensitivity and awareness that one should be focused on.

Both the GWT and the recurrent processing hypothesis (RPH) defend the idea that recurrent processing is essential for conscious experience. However, only GWT assumes that parieto-frontal activity is necessary to amplify initial posterior activity, causing "ignition." In contrast, the RPH proposes that recurrent processing in visual regions is sufficient for conscious perception.

It strikes us, however, that there is a crucial difference in the characteristics of the stimuli used to support either GWT or RPH. Indeed, studies reporting anterior correlates of consciousness generally use complex stimuli (e.g., Del Cul, Baillet, & Dehaene, 2007), whereas studies reporting posterior correlates of consciousness mostly use very low-level stimuli (e.g., Fahrenfort, Scholte, & Lamme, 2008). Whether or not one is a aware of a given stimulus could thus depend on the region that is involved in processing it, so integrating GWT and RPH by letting the emergence of consciousness in Lamme's Stage 3 or Stage 4 depend on the complexity of the stimulus.

In this respect, recent evidence (Sandberg, Timmermans, Overgaard, & Cleeremans, 2010) suggests that for simple stimuli, introspection is in fact rich, graded, and fairly accurate when properly probed by qualitative graded scales referring to the stimulus (from "No experience" to "A clear experience") rather than through dichotomous (yes/no) reports. Strikingly, Overgaard, Fehl, Mouridsen, Bergholt, and Cleeremans (2008) found that, using this method, blindsight patients reveal (severely degraded) awareness of stimuli presented in their blind field. Thus, such graded reports correlate better with behavior and indicate that above-chance identification is always associated with some awareness, at least for simple stimuli.

Whether a subjective report is exhaustive could thus depend on the extent to which feature integration is necessary to respond appropriately to a stimulus (Timmermans, Sandberg, Cleeremans, & Overgaard, 2010). Conversely, stimulus complexity could lie at the heart of the impression of unreportable phenomenal overflow. Kouider, de Gardelle, Sackur, and Dupoux (2010) recently proposed the partial awareness hypothesis, which holds that rich phenomenality is a "perceptual illusion" brought about by partial bottom-up

information that is accessed at *some* but not *all* representational levels, in combination with prior top-down information at the accessed level. Thus, phenomenal awareness never overflows access in this framework. In this sense, becoming aware of a stimulus does not merely involve filtering and selective amplification of a (overflowing) phenomenal field through attention, but rather the active construction of content based on fragmentary input of complex material, biased by priors.

Our own perspective on these issues begins with the notion that the brain *learns to be conscious* by continuously and unconsciously redescribing its own activity to itself (see Cleeremans, 2008). For such redescriptions to be possible at all, the target first-order representations need to be strong, stable, and distinctive—a condition that is itself only possible through recurrent processing. On this view, thus, phenomenal experience depends on the interaction between sufficiently strong first-order representations and the existence of learned redescriptions (metarepresentations) that reflect the manner in which the target first-order representations are known at some level (i.e., their meaning). There may be many levels of such metarepresentations in the brain. Which end up being active during some information-processing episode will depend on both stimulus complexity and task instructions.

*   *   *

Is recurrent processing necessary and/or sufficient for consciousness?

Naotsugu Tsuchiya[1] and Jeroen J. A. van Boxtel[2]

[1]*Division of Humanities and Social Sciences, California Institute of Technology, Pasadena, CA, USA; and Tamagawa University, Tamagawa, Japan*
[2]*Division of Biology, California Institute of Technology, Pasadena, CA, USA*
E-mail: naotsu@klab.caltech.edu

Abstract: While we agree with Lamme's general framework, we are not so convinced by his mapping between psychological concepts with their underlying neuronal mechanisms. Specifically, we doubt if recurrent processing is either necessary or sufficient for consciousness. A gist of a scene may be consciously perceived by purely

feedforward, without recurrent, processing. Neurophysiological studies of perceptual suppression show recurrent processing in visual cortex for consciously invisible objects. While the neuronal correlates of attention and consciousness remain to be clarified, we agree with Lamme that these two processes are independent, evinced by our recent demonstration of opposing effects of attention and consciousness.

Lamme's hypothesis powerfully explains several phenomena such as masking and iconic memory. However, other lines of research suggest that recurrent processing (RP) is neither necessary nor sufficient for conscious perception.

First, RP seems unnecessary for conscious gist perception of visual scenes. A gist of an unfamiliar and unexpected natural scene can be extracted with a very brief exposure and rapidly reported (Kirchner & Thorpe, 2006). Even when the iconic trace is masked (and presumably RP is terminated), it can be consciously perceived in the near absence of attention (Li, VanRullen, Koch, & Perona, 2002). This happens before the details of the scene become available. These properties of gist perception suggest that its computation can be performed in a purely feedforward manner.

Second, RP seems insufficient for conscious perception. For example, in perceptual suppression phenomena (Leopold & Logothetis, 1996; Maier et al., 2008), objects evoke the same sustained neuronal firing regardless of their conscious visibility. Assuming that sustained firing is a reflection of RP, as Lamme does in other papers, these findings suggest an insufficiency of RP for conscious visibility.

Perhaps, varying amounts of RP are correlated with different kinds of qualia (e.g., no RP for a gist); feedforward activation in an area that has widespread connectivity with the rest of the brain may be sufficient to produce consciousness because it has a higher possibility to produce larger phi. On the other hand, RP in an area that is not connected with frontal areas, like V1, may not give rise to consciousness.

While consciousness may not be tightly correlated with RP, some forms of attention are, and they promote long-distance coherent activity (Womelsdorf & Fries, 2007). In the cases of sustained invisibility mentioned above, some visual aftereffects are enhanced by attention (Kanai, Tsuchiya, & Verstraten, 2006). It would be interesting to study whether attention enhances processing of objects with sustained invisibility via enhanced RP and/or widespread activation.

Although RP and depth of widespread activation may not map onto consciousness and attention, respectively, we do believe that consciousness and attention are supported by distinctive neuronal mechanisms (Tsuchiya & Koch, 2008) based on two lines of evidence: (1) By classifying percepts based on their relation with attention and consciousness, we find examples of attention without consciousness and consciousness without attention, the latter including gist perception and iconic memory. (2) By independently manipulating attention and consciousness, one can demonstrate the opposite effects of attention and consciousness.

As to the second point, the perception of afterimages is modulated in opposite ways by attention and consciousness. By manipulating the visibility (a proxy for the content of consciousness) of an afterimage inducer, *perceptual invisibility* of afterimage inducers is shown to *reduce* afterimage duration. On the other hand, *attending* to afterimage inducers *reduces* afterimage duration. Recently, we directly demonstrated the opposite effects with a 2×2 factorial design, removing any stimulus or task confound (van Boxtel, Tsuchiya, & Koch, 2010). We explain the opposite effects by assuming that attention and consciousness enhance luminance and contrast adaptation to different degrees (Brascamp, van Boxtel, Knapen, & Blake, 2010). It would be difficult to explain the opposite effects of attention and consciousness using the line of reasoning based on RP and extent of activation.

While we agree with Lamme's point that neuroscience should go beyond introspection and that attention and consciousness are independent, evidence from gist perception, perceptual suppression, and afterimages suggests that local RP may not explain consciousness.

* * *

Reply to Commentaries

What introspection has to offer, and where its limits lie

Victor A. F. Lamme

University of Amsterdam, Amsterdam, The Netherlands

A proper science of consciousness combines all the available evidence – either coming from introspection, behavior, neuroscience or theory – in such a way that a 'best of all worlds' perspective is attained. Introspection shows us that qualia are all about perceptual organization. Neuroscience can then tell us where and when perceptual organization occurs, and whether this is independent of attention, access or report. Access, no matter in what guise it comes, remains ill-suited to explain where, when and how qualia emerge.

INTROSPECTION AND NEUROSCIENCE AS EQUAL PARTNERS

To begin with, I feel some clarification of my position is needed. Some commentaries—Overgaard; Timmermans, Windey, and Cleeremans; Ibáñez and Bekinschtein; Seth and Barrett—see me as a neuroscience chauvinist disregarding introspection (and behavior) altogether. But fully disregarding introspection in favor of neuroscience is not what I propose. On the contrary, my research agenda starts from the notion that qualia—in some form—exist. I follow the intuition that in deep sleep or strong masking there are no conscious sensations, whereas when someone gives a detailed report about what he sees, there are. There are simply no good reasons to doubt these basic intuitions. Without these, the whole point of looking for a better definition of conscious sensation seems moot.

At some point, however, my embracing of intuition stops. While we can all agree on the extreme ends of consciousness, I note that there is sufficient doubt on where exactly the boundary between the conscious and unconscious should be laid (also in the commentaries). I have previously noted that behaviorally, laying such a boundary is impossible (Lamme, 2006). Here I argue that it is equally difficult to do so introspectively. To me that gives sufficient foundation for a scientific debate on that boundary, and neuroscience should be one of the debaters, alongside with psychology, introspection, theoretical reasoning, etc. These contestants should however be on equal footing. That introspection came first doesn't allow it to set the agenda.

WHY PERCEPTUAL ORGANIZATION AND QUALIA ARE LINKED

What I *do* take from introspection is that qualia are better explained by perceptual organization than by access. This is a choice that apparently not everyone agrees on. Ibáñez and Bekinschtein argue that perceptual organization in itself does not constitute a quale, as perceptual organization is just a function—one that can equally well be implemented in a presumably unconscious robot. This argument obviously leads towards the denial of any phenomenality (*à la* Dennett or Churchland; e.g., Churchland, 1985[1]; Dennett, 1988), and indeed they argue that also global workspace theory (GWT) does not need qualia to explain whatever GWT is explaining. If all functionality is taken out of qualia—as in the original formulation of the hard problem—of course all scientific argumentation ends. My stance on qualia is "soft" in the sense that I do link them to some function, in this case perceptual organization.

This has a long empirical background. Gestalt psychologists have always emphasized the importance of perceptual organization for understanding conscious vision. It is hard to imagine any visual illusion occurring unconsciously. Direct evidence for the link

---

[1] Dennett is opposed to the notion of 'pure' qualia, i.e. phenomenality without any consequence. This is probably what Ibanez and Beckinschtein mean by perceptual organization not "requiring" qualia. Qualia as defined by me, strongly linked to perceptual organization, will probably meet much less opposition from Dennett. Churchland, the "father" of eliminative materialism, is in fact more strongly opposed to the propositional attitudes coming from folk psychology (wanting, believing, etc.) than to qualia.

between visual illusions and consciousness came from the work of Goodale and Milner (1992), who showed that only conscious percepts, and not unconscious reflexes, suffer from visual illusions.

This touches on the difference between what Timmermans, Windey, and Cleeremans call sensation and perception. At some point, the information from the outside world that is registered by our senses is transformed into our interpretation of that world. I have always found the phenomenon of color constancy a nice example of that transformation. When we look at an apple on our table, to us that apple will have the same shade of green, regardless of the time of day. The composition of wavelengths that radiate from the apple may however differ greatly depending on whether it is illuminated by morning or afternoon sunlight. Our photoreceptors pick up different wavelengths; we *see* a constant color. Perceptual organization is at the core of explaining this phenomenon: It is by combining the wavelengths of surrounding objects with that of the apple that our brain distills the perceived color. This link between color constancy and consciousness finds direct support in the empirical observation that unconscious priming goes according to wavelength, whereas conscious priming goes according to perceived color (Breitmeyer, Ro, Ogmen, & Todd, 2007).

As Caplovitz, Arcaro, and Kastner eloquently put forward, this going from the physics of the world towards our interpretation of it is what conscious vision is all about. It is the brain that imposes structure and meaning on the incoming information, and conscious sensations are all about how the structure of perception changes from one moment to the next. That is why perceptual organization is the key to understanding qualia. And that is why the boundary between neural representations that go from unconscious to conscious should be laid at the point where neural representations go from, say, wavelengths to colors. Only then does this boundary inform us about the essence of conscious sensations. Putting the boundary at the location where representations go from colors to the access to colors (or maybe their names, or emotional associations) doesn't explain anything about seeing colors instead of wavelengths.

Seth and Barrett do not endorse this *a priori* link between qualia and perceptual organization, feature grouping, or illusions. They argue that even if these were all present in a patient with neglect, it would still be unknown whether the subject had a phenomenal sensation. They argue to look for qualities of neural representations that may have a stronger *a priori* link with qualia, and suggest that these may be found in the intuition that conscious representations seem to be both unified and differentiating: We consciously see a

face not only "by virtue of the integration of face-specific perceptual features but also because these integrated features are discriminated from a vast repertoire of alternative possibilities: A face is a face to the extent that it is not a house, a car, the smell of a lemon, . . . etc." (p. 228).

First, in putting so much emphasis on differentiation they lean somewhat towards the unconscious. Face-selective cells, for example, discriminate between a "vast repertoire of alternative" stimuli even when a conscious sensation is undoubtedly absent, as in masking. It is the *combination* of differentiation and integration that makes representations conscious. To me that seems almost synonymous with what perceptual organization or recurrent processing in the visual cortex does: When face-selective cells engage in recurrent interactions with cells that encode lower level features such as color, the face-selective cells start to express both face and color sensitivity, and thus become capable of discriminating not just a face from a house, but also the face of Person A from that of Person B (Sugase, Yamane, Ueno and Kawano, 1999; Jehee, Roelfsema, Deco, Murre and Lamme, 2007b). Simultaneously, the low-level cells start to discriminate between the color red coming from a face and the color red coming from a house. Together, the ensemble of cells constitutes exactly what I would consider a representation that is integrating and differentiating—or, in my words, where perceptual organization has occurred. Therefore, I think that the *a priori* link between integration–differentiation and qualia that Seth and Barrett propose leads to the same conclusion as mine: that neglect patients in fact do have conscious sensations of their unattended/neglected stimuli.

BITING THE ORTHOGONALITY BULLET

This brings another player to the stage, which is attention. It is telling that in most cases where one would argue for the presence of unreportable conscious sensations, attention has been withdrawn from the stimuli at hand (neglect, attentional blink, change blindness, inattentional blindness). The issue boils down to the question of whether taking away attention just removes the ability to access and report these stimuli, or takes away all phenomenality associated with them. By definition, this is a question that cannot be answered introspectively. What does seem open to scientific investigation is the extent to which neural correlates of attention and phenomenal sensations are independent.

Revonsuo and Koivisto provide convincing ERP data to argue for three stages of processing, roughly corresponding to what I call Stages 1/2, 3, and 4. Stage 1/2 processing, reflected in the P1 potential, has no strong correlation with visibility or awareness. The visual awareness negativity (VAN), corresponding to Stage 3 processing, is invariably correlated with visibility. Moreover, this VAN is largely independent of (nonspatial) attention. Stage 4 is reflected in a P3 like positivity (LP) that is reflecting access and attention. These results therefore strongly argue for the independence of attention and consciousness: or of attention and conscious visibility, to be more precise. Tsuchiya and Van Boxtel take this one step further by showing opposite effects of attention and visibility on the duration of after-images.

There thus seems to be increasing experimental support for the independence of attention and consciousness. Together with the ontological argument put forward in my target paper, this makes one wonder why one should not simply bite the bullet and let consciousness and attention be fully orthogonal functions. In the light of many recent results, the arguments seem to swing in favor of doing so.

The notion of attention shares many components with that of access. Both attention and access evoke the amplification of signals, to the extent that they become more widely available. Attention and access have the same behavioral consequences: A target is detected, reported, or put in working memory. Attention and access depend on virtually identical neural structures: the fronto-parietal network. The experimental evidence and ontological arguments in favor of seeing attention as independent of consciousness would therefore also call for making consciousness orthogonal to access. The burden of proof for not doing so is in the hands of those who want to stick to the conflation of consciousness and access.

RECURRENT PROCESSING IN DIFFERENT GUISES

Shanahan posits a somewhat different stance on access, aligned with his interpretation of GWT, where information is broadcast via multiple network hubs via the white matter structure of the cerebral cortex. As soon as localized processing modules are linked through this network to become globally available, access is realized and consciousness ensues. It may be considered a neuroanatomical equivalent of recurrent processing theory or dynamical complexity theory discussed above. And just like in those theories, the key question then becomes how widespread the dissemina-

tion of information has to be before consciousness is produced. Shanahan thinks that the dissemination has to be widespread, but why that is required is unclear. How widespread does it have to be? If these white matter tracts and hubs link visual information to, say, working memory and report modules, but not to auditory or language modules, are we then conscious or not? In the end, also for this position it becomes necessary to answer the question of what modules should be included for a representation to be called conscious.

Dias and Britto consider the amount of "mental time travel" (MTT) associated with a particular representation as critically relevant to the (graded) amount of consciousness that goes along with it. Their idea adds a separate dimension to the discussion: It is not just about which modules are involved (and whether or not these are recurrently activated), but also whether the representation can be thought of to invoke some sort of MTT. The example they give for a representation without MTT—the orienting reflex—is intriguing, because I would consider that to be a strictly feedforward process, i.e., Stage 2. I would say that as soon as any recurrency is involved (i.e., Stages 3 and 4), some sort of MTT is activated along with it, in the sense that there is a meeting of bottom-up construction with top-down expectations. Another argument to link MTT with recurrent processing lies in the relation between recurrent processing and memory formation, as pointed out in the target paper. I thus see the MTT idea as confirmation of the essential dichotomy lying between Stages 1/2 and Stages 3/4.

MORE LEVELS OF CONSCIOUSNESS

Tsuchiya and Van Boxtel empirically question whether recurrent processing is required for conscious experience. Scene gist, for example, can be reported with only very brief presentations of natural scenes, suggesting that they become available through feedforward processing alone. Representations seem to become available in reverse hierarchy, with global properties reaching awareness before details of a scene. Somewhat related is the observation made by Timmermans, Windey, and Cleeremans that access may be operating at different representational levels, from simple to more complex, but is typically accurate for the level of description that is selected.

Indeed, scene gist can be computed in a feedforward sense. The activation of a set of high-level neurons selective for complex and ecologically relevant features (faces, bodies, scene layout, etc.) will give you exactly that. For this representation to become

conscious, recurrent interactions between these representations would be sufficient. This is probably what happens when a complex natural scene is partially masked. The masking precludes the interaction with lower level features, and hence the subject is not aware of the scene in all its detail. So partial consciousness of the gist of a complex scene is not incompatible with the recurrent processing theory. The theory predicts that as soon as you have modules representing different aspects of a scene engage in recurrent interactions, you will have a core of conscious sensation linked to whatever is represented by these modules. Adding more modules—of a lower or higher level—makes the conscious sensation of a richer content, but not more or less conscious.

The thought experiment put forward by Shanahan (about the drug that alleviates pain only at Level 4 and not at Level 3, yet is refused a license because neuroscience shows that patients—despite their denial— still "feel" pain) deep down is also about the levels of consciousness. What can we take away from pain and still genuinely call it pain? Obviously, if we take away a subject's capability to report pain by cutting off his tongue, or making him aphasic, the license should indeed not be granted; pain will still be felt. The pain may lose much of its feel if we take away the fear that goes along with it ("is it cancer?"), or the memories or emotions ("this is the worst pain I ever had"). But

wouldn't it still be pain? I even question whether it is critical to attribute the pain to one's own body. Empathic pain may feel just as awful as pain we suffer ourselves. At some point, by taking away sufficient reactive dispositions, we enter the realm of reducing the pain to the level that we should no longer call it pain. When enough has been stripped away, pain may turn into something categorically different, like an itch. This will, however, invariably go along with a change in the Level 3 representation as well. In fact, it is unclear to what extent the feelings of itching and pain are related and may be carried by the same nerve fibers (Ikoma, Steinhoff, Ständer, Yosipovitch, & Schmelz, 2006). Some theories suggest that the main difference between pain and an itch is in the balance or pattern of activation of different afferent fibers (McMahon and Koltzenburg, 1992). So the difference between feeling a pain or an itch may be very similar to the difference between seeing light of 600 nm as either orange or red. It's all about context, about perceptual organization, about combining information.

It is the same with consciousness. From introspection and psychology consciousness looks very different than from neuroscience or theory. Only by putting all these perspectives in their proper context, and by integrating all their information into a unified framework, will we really see what consciousness is about.

References from the Discussion Paper, the Commentaries, and the Reply

Adelson, E. H., & Movshon, J. A. (1982). Phenomenal coherence of moving visual patterns. *Nature*, *30*, 523–525.

Albright, T. D., & Stoner, G. R. (2002). Contextual influences on visual processing. *Annual Review of Neuroscience*, *25*, 339–379.

Alkire, M. T., Hudetz, A. G., & Tononi, G. (2008). Consciousness and anesthesia. *Science*, *322*, 876–880.

Arthuis, M., Valton, L., Régis, J., Chauvel, P., Wendling, F., Naccache, L., et al. (2009). Impaired consciousness during temporal lobe seizures is related to increased long-distance cortical–subcortical synchronization. *Brain*, *132*(8), 2091–2101.

Arzy, S., Adi-Japha, E., & Blanke, O. (2009). The mental time line: An analogue of the mental number line in the mapping of life events. *Consciousness and Cognition*, *18*(3), 781–785.

Baars, B. J. (2005). Global workspace theory of consciousness: Toward a cognitive neuroscience of human experience. In S. Laureys (Ed.), *Boundaries of consciousness:*

Neurobiology and neuropathology (Vol. 150, pp. 45–53). Amsterdam: Elsevier.

Bar, M. (2009). The proactive brain: Memory for predictions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*, 1235–1243.

Barry, R., & Rushby, J. (2006). An orienting reflex perspective on anteriorization of the P3 of the event-related potential. *Experimental Brain Research*, *173*(3), 539–545.

Bechara, A., Damasio, H., Tranel, D. & Damasio, A. R. (1997). Deciding advantageously before knowing the advantageous strategy. *Science*, *275*, 1293–1295.

Beck, D. M., & Kastner, S. (2009). Top-down and bottom-up mechanisms in biasing competition in the human brain. *Vision Research*, *49*, 1154–1165.

Becker, M. W., Pashler, H., & Anstis, S. M. (2000). The role of iconic memory in change-detection tasks. *Perception*, *29*, 273–286.

Bekinschtein, T. A., Dehaene, S., Rohaut, B., Tadel, F., Cohen, L., & Naccache, L. (2009a). Neural signature of

the conscious processing of auditory regularities. *Proceedings of the National Academy of Sciences of te United States of America*, *3*, 1672–1677.

Bekinschtein, T. A., Shalom, D. E., Forcato, C., Herrera, M., Coleman, M. R., Manes, F. F., et al. (2009b). Classical conditioning in the vegetative and minimally conscious state. *Nature Neuroscience*, *12*, 1343–1349.

Blake, R., & Logothetis, N. K. (2002). Visual competition. *Nature Reviews Neuroscience*, *3*, 13–23.

Block, N. (1990). Inverted earth. *Philosophical Perspectives*, *4*, 53–79.

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

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

Boehler, C. N., Schoenfeld, M. A., Heinze, H. J., & Hopf, J. M. (2008). Rapid recurrent processing gates awareness in primary visual cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *105*, 8742–8747.

Bornstein, R. F. (1989). Exposure and affect – Overview and meta-analysis of research, 1968–1987. *Psychological Bulletin*, *106*, 265–289.

Brascamp, J. W., van Boxtel, J. J., Knapen, T., & Blake, R. (2010). A dissociation of attention and awareness in phase-sensitive but not phase-insensitive visual channels. *Journal of Cognitive Neuroscience*, 22, 2326–2344. doi:10.1162/jocn.2009.21397

Breitmeyer, B. G., Ro, T., Ogmen, H., & Todd, S. (2007). Unconscious, stimulus-dependent priming and conscious, percept-dependent priming with chromatic stimuli. *Perception & Psychophysics*, *69*, 550–557.

Bruno, N., & Franz, V. H. (2009). When is grasping affected by the Muller-Lyer illusion? A quantitative review. *Neuropsychologia*, *47*, 1421–1433.

Bullier, J. (2001). Integrated model of visual processing. *Brain Research Reviews*, *36*, 96–107.

Camprodon, J. A., Zohary, E., Brodbeck, V., & Pascual-Leone, A. (in press). Two phases of V1 activity for visual recognition of natural images. *Journal of Cognitive Neuroscience*. Advance online publication. Retrieved September 31, 2009. doi:10.1162/jocn. 2009.21253

Chalmers, D. J. (1995). Facing up to the problem of consciousness. *Journal of Consciousness Studies*, *2*, 200–219.

Churchland, P. M. (1985). Reduction, qualia and the direct inspection of brain states. *Journal of Philosophy*, *82*, 8–28.

Churchland, P. S., & Churchland, P. M. (2002). Neural worlds and real worlds. *Nature Reviews Neuroscience*, *3*, 903–907.

Cleeremans, A. (2008). Consciousness: The radical plasticity thesis. *Progress in Brain Research*, *168*, 19–33.

Coltheart, M. (1980). Iconic memory and visible persistence. *Perception & Psychophysics*, *27*, 183–228.

Corballis, P. M. (2003). Visuospatial processing and the right-hemisphere interpreter. *Brain and Cognition*, *53*, 171–176.

Crick, F., & Koch, C. (1998a). Consciousness and neuroscience. *Cerebral Cortex*, *8*, 97–107.

Crick, F., & Koch, C. (1998b). Constraints on cortical and thalamic projections: The no-strong-loops hypothesis. *Nature*, *391*, 245–250.

Crick, F., & Koch, C. (2003). A framework for consciousness. *Nature Neuroscience*, *6*, 119–126.

Critchley, H. D., Elliott, R., Mathias, C. J., & Dolan, R. J. (2000). Neural activity relating to generation and representation of galvanic skin conductance responses: A functional magnetic resonance imaging study. *Journal of Neuroscience*, *20*(8), 3033–3040.

Dan, Y., & Poo, M. M. (2004). Spike timing-dependent plasticity of neural circuits. *Neuron*, *44*, 23–30.

de Gelder, B., Pourtois, G., & Weiskrantz, L. (2002). Fear recognition in the voice is modulated by unconsciously recognized facial expressions but not by unconsciously recognized affective pictures. *Proceedings of the National Academy of Sciences of the United States of America*, *99*, 4121–4126.

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

Dehaene, S., Changeux, J. P., Naccache, L., Sackur, J., & Sergent, C. (2006). Conscious, preconscious, and subliminal processing: A testable taxonomy. *Trends in Cognitive Sciences*, *10*, 204–211.

Dehaene, S., Naccache, L., Le Clec'H, G., Koechlin, E., Mueller, M., Dehaene-Lambertz, G., et al. (1998). Imaging unconscious semantic priming. *Nature*, *395*, 597–600.

De Weerd, P., Gattass, R., Desimone, R., & Ungerleider, L. G. (1995). Responses of cells in monkey visual cortex during perceptual filling-in of an artificial scotoma. *Nature*, *377*(6551), 731–734.

De Weerd, P., Sprague, J. M., Vandenbussche, E., & Orban, G. A. (1994). Two stages in visual texture segregation: A lesion study in the cat. *Journal of Neuroscience*, 14, 929–948.

Del Cul, A., Baillet, S., & Dehaene, S. (2007). Brain dynamics underlying the nonlinear threshold for access to consciousness. *PLoS Biology*, *5*(10), e260.

Del Cul, A., Dehaene, S., Reyes, P., Bravo, E., & Slachevsky, A. (2009). Causal role of prefrontal cortex in the threshold for access to consciousness. *Brain*, *132*, 2531–2540.

Dennett, D. (1988) Quining qualia. In A. Marcel and E. Bisiach (Eds.), *Consciousness in modern science*. Oxford, UK: Oxford University Press.

Desimone, R. (1998). Visual attention mediated by biased competition in extrastriate visual cortex. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences*, *353*, 1245–1255.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual-attention. *Annual Review of Neuroscience*, *18*, 193–222.

Dijksterhuis, A., Bos, M. W., Nordgren, L. F., & van Baaren, R. B. (2006). On making the right choice: The deliberation-without-attention effect. *Science*, *311*, 1005–1007.

Dretske, F. (1995). *Naturalizing the mind*. Cambridge, MA: MIT Press.

Driver, J., & Mattingley, J. B. (1998). Parietal neglect and visual awareness. *Nature Neuroscience*, *1*, 17–22.

Dudai, Y. (2002). Molecular bases of long-term memories: A question of persistence. *Current Opinion in Neurobiology*, *12*, 211–216.

Dudkin, K. N., Kruchinin, V. K., & Chueva, I. V. (2001). Neurophysiological correlates of delayed differentiation tasks in monkeys: The effects of the site of intracortical blockade of NMDA receptors. *Neuroscience and Behavioral Physiology*, *31*, 207–218.

Eagle, D. M., Baunez, C., Hutcheson, D. M., Lehmann, O., Shah, A. P., & Robbins, T. W. (2008). Stop-signal reaction-time task performance: Role of prefrontal cortex and subthalamic nucleus. *Cerebral Cortex*, *18*, 178–188.

Edelman, G. M. (1992). *Bright air, brilliant fire: On the matter of the mind*. New York: Basic Books.

Edelman, G. M. (2003). Naturalizing consciousness: A theoretical framework. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(9), 5520–5524.

Edelman, G. M. (2007). Learning in and from brain-based devices. *Science*, *16*(318), 1103–1105.

Egeth, H. E., & Yantis, S. (1997). Visual attention: Control, representation, and time course. *Annual Review of Psychology*, *48*, 269–297.

Eimer, M., & Schlaghecken, F. (2003). Response facilitation and inhibition in subliminal priming. *Biological Psychology*, *64*, 7–26.

Engel, A. K., Fries, P., & Singer, W. (2001). Dynamic predictions: Oscillations and synchrony in top-down processing. *Nature Reviews Neuroscience*, *2*, 704–716.

Engel, A. K., & Singer, W. (2001). Temporal binding and the neural correlates of sensory awareness. *Trends in Cognitive Sciences*, *5*, 16–25.

Enns, J. T., & Di Lollo, V. (2000). What's new in visual masking? *Trends in Cognitive Sciences*, *4*, 345–352.

Fahrenfort, J. J., Scholte, H. S., & Lamme, V. A. F. (2007). Masking disrupts reentrant processing in human visual cortex. *Journal of Cognitive Neuroscience*, *19*, 1488–1497.

Fahrenfort, J. J., Scholte, H. S., & Lamme, V. A. F. (2008). The spatiotemporal profile of cortical processing leading up to visual perception. *Journal of Vision*, *8*(1), 1–12.

Finkel, L. H., & Edelman, G. M. (1989). Integration of distributed cortical systems by reentry: A computer-simulation of interactive functionally segregated visual areas. *Journal of Neuroscience*, *9*, 3188–3208.

Flohr, H., Glade, U., & Motzko, D. (1998). The role of the NMDA synapse in general anesthesia. *Toxicology Letters*, *101*, 23–29.

Gaillard, R., Dehaene, S., Adam, C., Clemenceau, S., Hasboun, D., Baulac, M., et al. (2009). Converging intracranial markers of conscious access. *PLOS Biology*, *7*, 472–492.

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

Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in Neuroscience*, *15*, 20–25.

Grossberg, S., & Pessoa, L. (1998). Texture segregation, surface representation and figure–ground separation. *Vision Research*, *38*, 2675–2684.

Grossberg, S. & Versace, M. (2008). Spikes, synchrony, and attentive learning by laminar thalamocortical circuits. *Brain Research*, *1218*, 278–312.

Grossberg, S., Yazdanbakhsh, A., Cao, Y. Q., & Swaminathan, G. (2008). How does binocular rivalry emerge from cortical mechanisms of 3-D vision? *Vision Research*, *48*, 2232–2250.

Haber, R. N. (1983). The impending demise of the icon: A critique of the concept of iconic storage in visual information-processing. *Behavioral and Brain Sciences*, *6*, 1–11.

Hagmann, P., Cammoun, L., Gigandet, X., Meuli, R., Honey, C. J., Wedeen, C. J., et al. (2008). Mapping the structural core of human cerebral cortex. *PLoS Biology*, *6*(7), e159.

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, *293*, 2425–2430.

Haynes, J. D. (2009). Decoding visual consciousness from human brain signals. *Trends in Cognitive Sciences*, *13*, 194–202.

Haynes, J. D., Driver, J., & Rees, G. (2005). Visibility reflects dynamic changes of effective connectivity between V1 and fusiform cortex. *Neuron*, *46*, 811–821.

Ibáñez A. (2007). The neurodynamic core of consciousness and neural Darwinism. *Revista de Neurologia*, *45*, 547–555.

Ikoma, A., Steinhoff, M., Ständer, S., Yosipovitch, G., & Schmelz, M. (2006). The neurobiology of itch. *Nature Reviews Neuroscience*, *7*, 535–547.

Jack, A., & Roepstorff, A. (2002). Retrospection and cognitive brain mapping: From stimulus-response to script-report. *Trends in Cognitive Sciences*, *6*, 333–339.

Jehee, J. F. M., Lamme, V. A. F., & Roelfsema, P. R. (2007a). Boundary assignment in a recurrent network architecture. *Vision Research*, *47*, 1153–1165.

Jehee, J. F. M., Roelfsema, P. R., Deco, G., Murre, J. M. J., & Lamme, V. A. F. (2007b). Interactions between higher and lower visual areas improve shape selectivity of higher level neurons: Explaining crowding phenomena. *Brain Research*, *1157*, 167–176.

Kanai, R., Tsuchiya, N., & Verstraten, F. A. (2006). The scope and limits of top-down attention in unconscious visual processing. *Current Biology*, *16*(23), 2332–2336.

Kastner, S., Pinsk, M. A., De Weerd, P., Desimone R., & Ungerleider, L. G. (1999). Increased activity in human visual cortex during directed attention in the absence of visual stimulation. *Neuron*, *22*, 751–761.

Kentridge, R. W., Heywood, C. A., & Weiskrantz, L. (1999). Attention without awareness in blindsight. *Proceedings of the Royal Society of London, Series B: Biological Sciences*, *266*, 1805–1811.

Kirchner, H., & Thorpe, S. J. (2006). Ultra-rapid object detection with saccadic eye movements: Visual processing speed revisited. *Vision Research*, *46*(11), 1762–1776.

Klein, T. A., Endrass, T., Kathmann, N., Neumann, J., von Cramon, D. Y., & Ullsperger, M. (2007). Neural correlates of error awareness. *NeuroImage*, *34*, 1774–1781.

Koch, C., & Tsuchiya, N. (2007). Attention and consciousness: Two distinct brain processes. *Trends in Cognitive Sciences*, *11*, 16–22.

Koivisto, M., Kainulainen, P., & Revonsuo, A (2009). The relationship between awareness and attention: evidence from ERP responses. *Neuropsychologia*, *47*, 2891–2899.

Koivisto, M., & Revonsuo, A. (2003). An ERP study of change detection, change blindness and visual awareness. Psychophysiology, 40, 423–429.

Koivisto, M., & Revonsuo, A. (2008). The role of selective attention in visual awareness of stimulus features: Electrophysiological studies. *Cognitive, Affective, & Behavioral Neuroscience*, *8*, 195–210.

Koivisto, M., & Revonsuo, A. (2010). Event-related brain potential correlates of visual awareness. *Neuroscience and Biobehavioral Reviews*, *34*, 922–934.

Koivisto, M., Revonsuo, A., & Lehtonen, M. (2006). Independence of visual awareness from the scope of attention: An electrophysiological study. *Cerebral Cortex*, *16*, 415–424.

Koivisto, M., Revonsuo, A., & Salminen, N. (2005). Independence of visual awareness from attention at early processing stages. *NeuroReport*, *16*, 817–821.

Kosslyn, S. M. (1999). If neuroimaging is the answer, what is the question? *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences*, *354*, 1283–1294.

Kouider, S., de Gardelle, V., Sackur, J., & Dupoux, E. (in press). How rich is consciousness? The partial awareness hypothesis. *Trends in Cognitive Sciences*. Advance online publication. doi:10.1016/j.tics.2010.04.006

Lamme, V. A. F. (1995). The neurophysiology of figure ground segregation in primary visual-cortex. *Journal of Neuroscience*, *15*, 1605–1615.

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

Lamme, V. A. F. (2004). Separate neural definitions of visual consciousness and visual attention: A case for phenomenal awareness. *Neural Networks*, *17*, 861–872.

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

Lamme, V. A. F., & Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends in Neurosciences*, *23*, 571–579.

Lamme, V. A. F., & Spekreijse, H. (2000). Modulations of primary visual cortex activity representing attentive and conscious scene perception. *Frontiers in Bioscience*, *5*, D232–D243.

Lamme, V. A. F., Super, H., Landman, R., Roelfsema, P. R. & Spekreijse, H. (2000). The role of primary visual cortex (V1) in visual awareness. *Vision Research*, *40*, 1507–1521.

Lamme, V. A. F., Super, H. & Spekreijse, H. (1998a). Feedforward, horizontal, and feedback processing in the visual cortex. *Current Opinion in Neurobiology*, *8*, 529–535.

Lamme, V. A. F., Vandijk, B. W., & Spekreijse, H. (1993). Contour from motion processing occurs in primary visual-cortex. *Nature*, *363*, 541–543.

Lamme, V. A. F., Zipser, K., & Spekreijse, H. (1998b). Figure–ground activity in primary visual cortex is suppressed by anesthesia. *Proceedings of the National Academy of Sciences of the United States of America*, *95*, 3263–3268.

Lamme, V. A. F., Zipser, K., & Spekreijse, H. (2002). Masking interrupts figure–ground signals in V1. *Journal of Cognitive Neuroscience*, *14*, 1044–1053.

Landman, R., Spekreijse, H., & Lamme, V. A. F. (2003a). Large capacity storage of integrated objects before change blindness. *Vision Research*, *43*, 149–164.

Landman, R., Spekreijse, H., & Lamme, V. A. F. (2003b). Set size effects in the macaque striate cortex. *Journal of Cognitive Neuroscience*, *15*, 873–882.

Landman, R., Spekreijse, H. & Lamme, V. A. F. (2004a). Relationship between change detection and post-change activity in visual area VI. *NeuroReport*, *15*, 2211–2214.

Landman, R., Spekreijse, H., & Lamme, V. A. F. (2004b). The role of figure–ground segregation in change blindness. *Psychonomic Bulletin & Review*, *11*, 254–261.

Lau, H. C. & Passingham, R. E. (2007). Unconscious activation of the cognitive control system in the human prefrontal cortex. *Journal of Neuroscience*, *27*, 5805–5811.

Lee, S. H., Blake, R., & Heeger, D. J. (2007). Hierarchy of cortical responses underlying binocular rivalry. *Nature Neuroscience*, *10*, 1048–1054.

Leopold, D. A., Bondar, I. V., & Giese, M. A. (2006) Norm-based face encoding by single neurons in the monkey inferotemporal cortex. *Nature*, *442*, 572–575.

Leopold, D. A., & Logothetis, N. K. (1996). Activity changes in early visual cortex reflect monkeys' percepts during binocular rivalry. *Nature*, *379*(6565), 549–553.

Leopold, D. A., & Logothetis, N. K. (1999). Multistable phenomena: Changing views in perception. *Trends in Cognitive Sciences*, *3*, 254–264.

Li, F. F., VanRullen, R., Koch, C., & Perona, P. (2002). Rapid natural scene categorization in the near absence of attention. *Proceedings of the National Academy of Sciences of the United Sstates of America*, *99*(14), 9596–9601.

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*, 279–281.

Mack, A. & Rock, I. (1998). *Inattentional blindness*. Cambridge, MA: MIT Press.

Macknik, S. L., & Martinez-Conde, S. (2009). The role of feedback in visual attention and awareness. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 1165–1179). Cambridge, MA: MIT Press.

Maier, A., Wilke, M., Aura, C., Zhu, C., Ye, F. Q., & Leopold, D. A. (2008). Divergence of fMRI and neural signals in V1 during perceptual suppression in the awake monkey. *Nature Neuroscience*, *11*(10), 1193–1200.

Marois, R., Yi, D. J., & Chun, M. M. (2004). The neural fate of consciously perceived and missed events in the aftentional blink. *Neuron*, *41*, 465–472.

McMahon, S. B. and Koltzenburg, M. (1992) Itching for an explanation. *Trends in Neuroscience*, *15*, 497–501.

Meng, M., & Tong, F. (2004). Can attention selectively bias bistable perception? Differences between binocular rivalry and ambiguous figures. *Journal of Vision*, *4*, 539–551.

Merleau-Ponty, M. (1962). *Phenomenology of perception*. London: Routledge & Kegan Paul.

Nakayama, K., He, Z. J., & Shimojo, S. (1995). Visual surface representation: A critical link between lower-level and higher level vision. In S. M. Kosslyn & D. N. Osherson (Eds.), *An invitation to cognitive science: Visual cognition* (pp. 1–70). Cambridge, MA: MIT Press.

Nakayama, K., & Shimojo, S. (1992). Experiencing and perceiving visual surfaces. *Science*, *257*, 1357–1363.

Niedeggen, M., Wichmann, P., & Stoerig, P. (2001). Change blindness and time to consciousness. *European Journal of Neuroscience*, *14*, 1719–1726.

O'Regan, J. K., & Noe, A. (2001). A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences*, *24*, 939–973.

O'Shea, R. P., & Corballis, P. M. (2003). Binocular rivalry in split-brain observers. *Journal of Vision*, *3*, 610–615.

Oram, M. W., & Perrett, D. I. (1992). Time course of neural responses discriminating different views of the face and head. *Journal of Neurophysiology*, *68*, 70–84.

Overgaard, M. (2006). Introspection in science. *Consciousness and Cognition*, *15*, 629–633.

Overgaard, M., Fehl, K., Mouridsen, K., Bergholt, B., & Cleeremans, A. (2008). Seeing without seeing? Degraded conscious vision in a blindsight patient. *PLoS ONE*, *3*(8), e3028.

Overgaard, M., & Sørensen, T. A. (2004). Introspection distinct from first order experiences. *Journal of Consciousness Studies*, *11*(7–8), 77–95.

Pascual-Leone, A., & Walsh, V. (2001). Fast backprojections from the motion to the primary visual area necessary for visual awareness. *Science*, *292*, 510–512.

Peterhans, E., & Vonderheydt, R. (1989). Mechanisms of contour perception in monkey visual-cortex. 2. Contours bridging gaps. *Journal of Neuroscience*, *9*, 1749–1763.

Pins, D., & ffytche, D. (2003). The neural correlates of conscious vision. *Cerebral Cortex*, *13*, 461–44.

Polat U., & Sagi D. (1993). Lateral interactions between spatial channels: Suppression and facilitation revealed by lateral masking experiments. *Vision Research*, *33*, 993–999.

Pylyshyn, Z. (1999). Is vision continuous with cognition? The case for cognitive impenetrability of visual perception. *Behavioral and Brain Sciences*, 22, 341–365.

Reder, L. M., Park, H., & Kieffaber, P. D. (2009). Memory systems do not divide on consciousness: Reinterpreting memory in terms of activation and binding. *Psychological Bulletin*, *135*, 23–49.

Rees, G. (2007). Neural correlates of the contents of visual awareness in humans. *Philosophical Transactions of the Royal Society, B: Biological Sciences*, *362*, 877–886.

Rees, G., Kreiman, G., & Koch, C. (2002). Neural correlates of consciousness in humans. *Nature Reviews Neuroscience*, *3*, 261–270.

Ridderinkhof, K. R., Ullsperger, M., Crone, E. A., & Nieuwenhuiss, S. (2004). The role of the medial frontal cortex in cognitive control. *Science*, *306*, 443–447.

Roelfsema, P. R., Lamme, V. A. F., Spekreijse, H., & Bosch, H. (2002). Figure–ground segregation in a recurrent network architecture. *Journal of Cognitive Neuroscience*, *14*, 525–537.

Rolls, E. T. (2000). Functions of the primate temporal lobe cortical visual areas in invariant visual object and face recognition. *Neuron*, *27*, 205–218.

Rolls, E. T., & Tovee, M. J. (1994). Processing speed in the cerebral-cortex and the neurophysiology of visual masking. *Proceedings of the Royal Society of London, Series B: Biological Sciences*, *257*, 9–15.

Rowe, J., Friston, K., Frackowiak, R., & Passingham, R. (2002). Attention to action: Specific modulation of corticocortical interactions in humans. *NeuroImage*, *17*, 988–998.

Ryle, G. (1949). *The concept of mind.* Chicago: New University of Chicago Press.

Salin, P. A., & Bullier, J. (1995). Corticocortical connections in the visual-system: Structure and function. *Physiological Reviews*, *75*, 107–154.

Sandberg, K., Timmermans, B., Overgaard, M., & Cleeremans, A. (in press). Measuring consciousness: Is one measure better than the other? *Consciousness and Cognition*. Advance online publication. doi:10.1016/j.concog.2009.12.013

Schacter, D. L., Chiu, C. Y. P., & Ochsner, K. N. (1993). Implicit memory – A selective review. *Annual Review of Neuroscience*, *16*, 159–182.

Schankin, A., & Wascher, E. (2007). Electrophysiological correlates of stimulus processing in change blindness. *Experimental Brain Research*, *183*, 95–105.

Scholte, H. S., Witteveen, S. C., Spekreijse, H., & Lamme, V. A. F. (2006). The influence of inattention on the neural correlates of scene segmentation. *Brain Research*, *1076*, 106–115.

Serences, J. T., & Yantis, S. (2006). Selective visual attention and perceptual coherence. *Trends in Cognitive Sciences*, *10*, 38–45.

Sergent, C., Baillet, S., & Dehaene, S. (2005). Timing of the brain events underlying access to consciousness during the attentional blink. *Nature Neuroscience*, *8*, 1391–1400.

Seth, A. K. (2008). Theories and measures of consciousness develop together. *Consciousness and Cognition*, *17*, 986–988.

Seth, A. K. (2009). Explanatory correlates of consciousness: Theoretical and computational challenges. *Cognitive Computation*, *1(1)*, 50–63.

Seth, A. K., Dienes, Z., Cleeremans, A., Overgaard, M., & Pessoa, L. (2008). Measuring consciousness: Relating behavioural and neurophysiological approaches. *Trends in Cognitive Sciences*, *12*, 314–321.

Seth, A. K., Izhikevich, E., Reeke, G. N., & Edelman, G. M. (2006). Theories and measures of consciousness: An extended framework. *Proceedings of the National Academy of Sciences of the United States of America*, *103(28)*, 10799–10804.

Shanahan, M. P. (2008). A spiking neuron model of cortical broadcast and competition. *Consciousness and Cognition*, *17*, 288–303.

Shanahan, M. P. (2010). *Embodiment and the inner life: Cognition and consciousness in the space of possible minds*. Oxford, UK: Oxford University Press.

Shapiro, K. L. (2009). The functional architecture of divided visual attention. *Progress in Brain Research*, *176*, 101–121.

Shapiro, K. L., Raymond, J. E., & Arnell, K. M. (1994). Attention to visual-pattern information produces the attentional blink in rapid serial visual presentation. *Journal of Experimental Psychology: Human Perception and Performance*, *20*, 357–371.

Silvanto, J., Cowey, A., Lavie, N., & Walsh, V. (2005). Striate cortex (V1) activity gates awareness of motion. *Nature Neuroscience*, *8*, 143–144.

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

Singer, W. (1995). Development and plasticity of cortical processing architectures. *Science*, *270*, 758–764.

Singer, W. (1999). Neuronal synchrony: A versatile code for the definition of relations? *Neuron*, *24*, 49–65.

Sligte, I. G., Scholte, H. S., & Lamme, V. A. F. (2008). Are there multiple visual short-term memory stores? *PLOS One*, *3*, e1699.

Sligte, I. G., Scholte, H. S., & Lamme, V. A. F. (2009). V4 activity predicts the strength of visual short-term memory representations. *Journal of Neuroscience*, *29*, 7432–7438.

Sperling, G. (1960). The information available in brief visual presentations, *Psychological Monographs*, *74*, 1–29.

Sperry, R. (1984). Consciousness, personal identity and the divided brain. *Neuropsychologia*, *22*, 661–673.

Sporns, O., Tononi, G., & Edelman, G. M. (1991). Modeling perceptual grouping and figure ground segregation by means of active reentrant connections. *Proceedings of the National Academy of Sciences of the United States of America*, *88*, 129–133.

Sterzer, P., Kleinschmidt, A., & Rees, G. (2009). The neural bases of multistable perception. *Trends in Cognitive Sciences*, *13*, 310–318.

Stoerig, P. (1996). Varieties of vision: From blind responses to conscious recognition. *Trends in Neurosciences*, *19*, 401–406.

Sugase, Y., Yamane, S., Ueno, S., & Kawano, K. (1999). Global and fine information coded by single neurons in the temporal visual cortex. *Nature*, *400*, 869–873.

Super, H., Spekreijse, H., & Lamme, V. A. F. (2001). Two distinct modes of sensory processing observed in monkey primary visual cortex (V1). *Nature Neuroscience*, *4*, 304–310.

Thielscher, A., & Neumann, H. (2008). Globally consistent depth sorting of overlapping 2D surfaces in a model using local recurrent interactions. *Biological Cybernetics*, *98*, 305–337.

Thompson, K. G., & Schall, J. D. (1999). The detection of visual signals by macaque frontal eye field during masking. *Nature Neuroscience*, *2*, 283–288.

Timmermans, B., Sandberg, K., Cleeremans, A., & Overgaard, M. (in press). Partial awareness distinguishes between measuring conscious perception and conscious content. *Consciousness and Cognition*. Advance online publication. doi:10.1016/j.concog.2010.05.006

Tong, F., Meng, M., & Blake, R. (2006). Neural bases of binocular rivalry. *Trends in Cognitive Sciences*, *10*, 502–511.

Tononi, G. (2004). An information integration theory of consciousness. *BMC Neuroscience*, 5, 42.

Tononi, G. (2008). Consciousness as integrated information: A provisional manifesto. *Biological Bulletin*, *215*, 216–242.

Tononi, G., & Koch, C. (2008). The neural correlates of consciousness: An update. In *Year in Cognitive Neuroscience 2008* (vol. 1124, pp. 239–261). Boston: Blackwell.

Tononi, G., & Massimini, M. (2008). Why does consciousness fade in early sleep? In D. W. Pfaff and B. L. Kieffer (Eds.), *Molecular and biophysical mechanisms of arousal, alertness, and attention* (Vol. 1129, pp. 330–334). New York: New York Academy of Sciences.

Treisman, A. (1996). The binding problem. *Current Opinion in Neurobiology*, *6*, 171–178.

Troxler, D. (1804). Uber das Verschwinden gegebener Gegenstände innerhalb unseres Gesichtskreises [On the disappearance of given objects in visual space]. In K. Himly & J. A. Schmidt (Eds.), *Ophthalmologische bibliothek II* (pp. 51–53). Jena, Germany: Fromman.

Tsuchiya, N., & Koch, C. (2008). The relationship between consciousness and attention. In S. Laureys & G. Tononi (Eds.), *The neurology of consciousness: Cognitive neuroscience and neuropathology* (pp. 63–78). New York: Academic Press.

Uhlhaas, P. J., Pipa, G., Melloni, L., Neuenschwander, S., Nikolic, D. & Singer, W. (2009). Neural synchrony in cortical networks: History, concept and current status. *Frontiers in Integrative Neuroscience*, *3*, 17.

van Boxtel, J. J., Tsuchiya, N., & Koch, C. (2010). Opposing effects of attention and consciousness on afterimages. *Proceedings of the National Academy of Sciences of the United Sstates of America*, *107*(19), 8883–8888.

van Gaal, S., Ridderinkhof, K. R., Fahrenfort, J. J., Scholte, H. S., & Lamme, V. A. F. (2008). Frontal cortex mediates unconsciously triggered inhibitory control. *Journal of Neuroscience*, *28*, 8053–8062.

van Gaal, S., Ridderinkhof, K. R., van den Wildenberg, W. P. M., & Lamme, V. A. F. (2009). Dissociating consciousness from inhibitory control: Evidence for unconsciously triggered response inhibition in the stop-signal task. *Journal of Experimental Psychology: Human Perception and Performance*, *35*, 1129–1139.

Vuilleumier, P., Valenza, N., & Landis, T. (2001). Explicit and implicit perception of illusory contours in unilateral spatial neglect: Behavioural and anatomical correlates of preattentive grouping mechanisms. *Neuropsychologia*, *39*(6), 597–610.

Watanabe, T., Nanez, J. E., & Sasaki, Y. (2001). Perceptual learning without perception. *Nature*, *413*, 844–848.

Wertheimer M. (1924/1950). Gestalt theory. In W. D. Ellis (Eds.), *A sourcebook of Gestalt psychology* (pp. 1–10). New York: Humanities Press.

Wolfe, J. M. (1999). Inattentional amnesia. In V. Coltheart (Ed.), *Fleeting memories*. Cambridge, MA: MIT Press.

Womelsdorf, T., & Fries, P. (2007). The role of neuronal synchronization in selective attention. *Current Opinion in Neurobiology*, *17*(2), 154–160.

Zipser, K., Lamme, V. A. F., & Schiller, P. H. (1996). Contextual modulation in primary visual cortex. *Journal of Neuroscience*, *16*, 7376–7389.