Is Attention Necessary and Sufficient for Consciousness?[1]

Jesse J. Prinz
jesse@subcortex.com

Elsewhere, I have argued that consciousness arises at a particular stage of sensory processing (2000; 2005; forthcoming). Following Jackendoff (1987), I claimed that consciousness occurs at an intermediate level of representation that lies between a low level of processing, which represents local features of a stimulus in a disunified way, and a high level, which abstracts away from vantage point and surface details in the service of object recognition. The low level might be compared to a pixel map, and the high level might be compared to the structural descriptions used by some computer aided design programs, whereas the intedermediate level is more like a 3D movie: it represents whole objects, which rich surface details, located in depth, and presented from a particular point of view. If the intermediate-level hypothesis is right, it is a major boon in the search to find the neural correlates of consciousness (NCCs). When we identify the neural mechanisms underlying intermediate-level perceptual processing, we have, in effect, located consciousness in the brain. One might think this is enough. If our goal is to find the NCCs, then we need look no further. Intermediate-level processing areas are well known in the brain, and we have overwhelming reason to think consciousness resides there.

Unfortunately, it is premature to pop open the celebratory champagne. For while there is good reason to think activation of intermediate-level representations and their neural realizers is necessary for conscious experience, there is equally good reason for thinking that such activations are not sufficient. The reason is simple. We sometimes perceive things in the absence of conscious experience. When we do so, we are presumably engaging the entire perceptual hierarchy, from low level to high; otherwise, we would not recognize objects that we unconsciously perceive. That means there is activation of intermediate-level representations in cases of unconscious perception, and, therefore, mere activation at this level is not sufficient for conscious experience (see also Kanwisher, 2001).

The point can be put as follows: The intermediate-level hypothesis gives us an account of *what* we are conscious of. The representations at that level correspond to the contents of experience. But it doesn't tell us *how* these representations become conscious. We need a theory of what goes on when perceptual states come to be consciously experienced. Without such a theory, we'll have an incomplete story about the psychological conditions that are necessary and sufficient for consciousness. We need such a story if we are to locate the neural correlates. Put differently, the intermediate-level hypothesis does tell us where to find our NCCs, but this is mere cartography. Once we locate the right brain regions we still need to figure out what kinds of processes in those regions correspond to conscious experience. Answering that question is the goal of this chapter. And, to dispel any unpleasant suspense, I will reveal

[1] I am grateful to Chris Mole, Declan Smithies, and Wayne Wu for extensive comments, corrections, and critique.

the answer now: consciousness arises when and only when we attend. The neural processes underlying attention are the physical mechanisms by which the neural correlates of intermediate-level representations become conscious. How do we become conscious? We attend.

1. Evidence for the Necessity and Sufficiency of Attention

1.1 Unconscious Perception

In 1957, James Vicary instigated a consumer panic when he reported that he had dramatically increased soft drink and popcorn sales at a New Jersey movie theater by inserting subliminal messages. Evidence later suggested that these results were fabricated, and Vicary himself admitted that he did not collect enough data to consider the results reliable. Since then, however, there have been numerous studies establishing that stimuli can be subliminally perceived.

In studies of subliminal perception, the subliminal stimulus is either presented very briefly, in a degraded form (e.g., with low intensity), or in competition with another stimulus or task demand. Afterwards, researchers must measure two things: consciousness and perception. Consciousness is measured either objectively or subjectively (Szczepanowski and Pessoa, 2007). On objective measures, subjects are asked to make a forced choice guess about what they have seen; errors suggest that the stimulus was not consciously perceived. On subjective measures, subjects are asked to report on whether they saw anything, and, sometimes, on what they saw; if the subject reports not having seen anything, that is taken as evidence that stimulus was unconscious. Sometimes subjective and objective measures are combined. For example, Kunimoto et al. (2001) say that a stimulus is unconscious if confidence judgments about having seen it fail to predict accuracy in recall.

There are many ways to test for stimulus perception, but they generally involve some form of priming: the pre-activation of representation that influence performance on a subsequent task. Most priming studies present a meaningful stimulus very quickly, followed by a mask (a screen of meaningless visual noise), which prevents an iconic memory from forming. In some cases, the stimulus is so brief that subjects are unaware that anything has been presented at all, but there is still evidence, under such conditions, for semantic processing. For example, Naccache and Dehaene (2001) presented masked numbers and then asked subjects whether a consciously presented target number was higher or lower than 5. If the masked number was in the same direction from five (higher or lower) as the test number, reaction times improved. In the study, the masked stimuli were presented for 43 milliseconds, which is below subjective thresholds, meaning subjects report seeing nothing. Naccache et al. (2005) have also demonstrated priming below objective thresholds: they founds that emotionally significant words generated activity in emotion centers at the brain even when presented at 29 milliseconds, at which point subjects ability to guess whether a stimulus was presented is at chance. Such unconscious priming can even influence consumer choices, adding credible support to Vicary's dubious movie theater study. Winkielman et al. (2005) showed pictures of faces to subjects that were angry, happy, or neutral for a mere 16 milliseconds, followed by a mask. Afterwards subjects were given a soft drink and asked to rate it, pour as much as they desired into a glass, and say how much they were willing to pay for it. All of these measures were influenced by the valence of the prime.

Unconscious perception can be found in all sensory modalities. For example, Hillyard et al. (1971) had subjects listen to auditory noise and try to determine whether a brief tone had been played in the background. Even when subjects failed to detect the tone, electrical activity measured on their scalps indicated that the tone had been unconsciously perceived. Pagano and Turvey (1998) report that people can determine the length of a wielded object even under conditions of anesthesia or neuropathology that prevents conscious experience of touch. Schnall et al. (2008) show that unconsciously smelling fart spray can lead people to make harsher moral judgments than they would otherwise make.

In short, researchers have found evidence for unconscious perception using a wide range of experimental methods across all modalities that have been examined. In each case, it is clear that stimuli are being semantically processed, and hence represented up to the highest levels of perceptual processing, even in the absence of conscious experience. And this leaves us with a question. What makes the difference between perceiving consciously and perceiving unconsciously?

1.2 Attention Is Necessary and Sufficient for Consciousness

One way to answer the preceding question is to consider pathological cases. One can look for brain injuries that lead to subliminal perception under ordinary viewing conditions. If such cases can be found, one could identify the locus of the injury to generate a hypothesis about the mechanisms that matter for consciousness. In taking this approach, one might immediately think of blindsight, which is one of the most celebrated neurological disorders (Weiskranz, 1986). People with blindsight have injuries in their primary visual cortices, which prevent them from consciously seeing things presented in the visual field corresponding to the injury, yet they can correctly guess the location of objects presented in these blind fields. Blindsight is certainly intriguing, but it's not exactly what we are looking for, because people with the disorder cannot recognize objects in their blind fields, even on implicit measures. This suggests that they are not perceiving those objects in the sense under consideration here (representing those objects as such), and they are not using the full extent of their visual processing hierarchies in response to those objects. Instead, the residual capacity probably involves subcortical structures and, perhaps, a select subset of spatially sensitive cortical visual areas. We need a condition in which objects are in fact recognized in the absence of consciousness.

The best example of this is unilateral neglect. Neglect is a disorder typically caused by injuries to the right inferior parietal cortex (Driver and Mattingly, 1998). People with this condition seem to have no conscious experience in the left visual fields or, sometimes, of the left sides of objects. They often lack experience of the left sides of their bodies as well. Phenomenalogically, they seem to be blind on the left. But there is good evidence that many people with neglect retain a capacity for unconscious perception. For example, Marshall and Halligan (1988) presented a neglect patient with two vertically aligned pictures of houses that were exactly the same except one of them had flames shooting out on the left. The patient insisted that the houses were the same, but when asked which one she would rather live in, she chose the one without flames on

9 out of 11 trials. Some neglect patience do not show this pattern of performance on the house task (Bisiach and Rusconi, 1990), but the result has been replicated. Doricchi and Galati (2000) tested a neglect patient who showed a preference for the intact house on 17 of 19 trials, despite seeing the houses as the same. They found similar results on a wide range of items, with the patient prefering the intact object in pair of pictures that she perceived as identical.

Further evidence for unconscious perception in neglect comes from brain imaging studies. Rees et al. (2000) found that objects presented on the left caused brain activation in the right visual pathway of a patient with neglect. Important for our purposes, the activations included intermediate-level visual areas, confirming that mere activity here is not sufficient for conscious experience.

Patients with neglect offer just what we are looking for: unconscious visual object recognition and unconscious activity in cortical visual areas. The question we must ask is what is preventing conscious experience in neglect? The answer is that neglect is an attention deficit. The inferior parietal brain areas that usually cause the disorder are known to play a role in allocating attention. Neglect is also sometimes associated with injuries to the frontal eye fields, which are frontal cortex structures associated with attentional orienting (Hussain & Kennard, 1996; see also Ruff et al., 2008). Patients with neglect cannot consciously perceive things on the left, because they can't attend to them. The reason why we rarely see neglect for the right visual field is that attention mechanisms in the right hemisphere, but not the left, seem to be capable of allocating attention to either side, so damage to the left hemisphere, as opposed to the right, leaves attention comparatively intact (Posner and Raichle, 1994).

Research on unilateral neglect gives us a candidate mechanism for consciousness: attention. When attention mechanisms are damaged, consciousness is lost, even though perception remains. But there is always some risk in inferring normal mechanisms from pathological cases. There is always a chance that the injuries in neglect compromise something other than attention. To be test the hypothesis that attention is responsible, it is important to test people with intact brains. It has been shown that one can induce neglect symptoms in healthy people by delivering transcranial magnetic stimulation to attention areas in parietal cortex (Meister et al., 2006), but like real lesions, the exact locus of these TMS lesions is difficult to determine. To increase confidence that attention is the culprit, we should see what happens to consciousness in healthy people when attention is withdrawn without doing anything that directly interferes with brain processing. Two phenomena are especially relevant.

First, consider the phenomenon called "the attentional blink." This occurs when a subject is asked to detect two target stimuli in a series of rapidly displayed stimuli (typically presented at a rate of 10 per second). Under such conditions, the first stimulus captures attention, and the second stimulus is not consciously perceived if it appears soon afterwards (typically within 200 to 500 milliseconds of the first stimulus). Electrical recordings from the scalp suggest that the second stimulus is perceived unconsciously, but it does not reach consciousness because attention is consumed by the first (Luck et al., 2000). Research on the attentional blink has led to the discovery of a related phenomenon called the emotional blink. Arnell et al. (2007) gave people a rapidly presented series of words, including some color terms. After a color term was presented, subjects were asked to name the color. In some conditions, they included an emotionally charged word (e.g., "orgasm") shortly before the color term. When this occurred, subjects tended to miss the color term. Emotionally charged words attract attention, leading to disruptions in conscious perception.

Second consider inattentional blindness (Mack and Rock, 1998). People often fail to consciously experience an unexpected stimulus if it is presented while they are engaged in an attention-demanding task. For example, Mack and Rock instructed subjects to judge which of two intersecting lines in a crosshair was longer, and while engaged in this task, they flashed a word or shape. Many of their subjects failed to detect the unexpected stimulus, and had no recollection of seeing anything other than the crosshair when probed afterwards. These stimuli exhibited priming effects, but seem to have had no impact on conscious experience. Strikingly, inattentional blindness can even occur for stimuli that are presented for a relatively long time. Most et al. (2000) gave subjects a task in which they needed to look at a field of moving black and white shapes and count how frequently one class of letters (black or white) bounced against the side of the screen; during this display a red object that was different from all the other objects in shape, luminance, and pattern of motion scrolled across the center of the screen for five seconds. 28% of the subjects failed to perceive it consciously. When explicitly asked, they said they had not seen anything other than the black and white shapes that had been present on all the other trials. Given the duration and strikingly different visual features of the unexpected stimulus, it is overwhelmingly unlikely that these people consciously perceived that stimulus and forgot about it (*pace* Wolfe, 1999; see also Mack & Rock, 1998: chapter 9).

Together with research on unilateral neglect, inattentional blindness and attentional blink provide powerful evidence for the claim that attention is needed for consciousness. When attention is withdrawn, due to brain injury, bottom-up capture, or top-down allocation to a demanding task, stimuli that are presented in clear view become invisible. In each case, the unconscious stimuli show priming effects, suggesting that they are represented at all levels of the visual hierarchy. Mere activation at the intermediate-level is not enough. Attention is necessary.

There is also evidence that attention is *sufficient* for making intermediate-level perceptual states conscious. Consider the phenomenon of visual pop-out. If subjects are asked to find a target stimulus in a group of contrasting distracters, the target stimulus often seems to pop-out. Imagine a blue ball, for example, in a grid of red balls. Pop-out is believed to occur when a target stimulus captures attention (Treisman & Gelade, 1980), and the stimulus that pops-out is consciously experienced.

Similarly, conscious perception can be improved by the presence of a cue that indicates where a stimulus is going to occur. In a method developed by Posner (1980), subjects see an arrow that either accurately or inaccurately predicts where a target will appear. When the arrow is accurate, conscious detection is improved.

Pop-out and Posner cases are usually considered low-level or "early selection" phenomena. In both cases, visual forms capture or direct attention without much semantic processing. But attention can also be captured by meanings; there can be "late selection." We all know the cocktail party effect, in which you can hear your own name being mentioned by someone across a crowded noisy room. A moment earlier the surrounding conversations were all an unintelligible din, but, when you hear your name, it is crystal clear. That suggests the surrounding conversations are being processed, to some degree, and filtered for words that might be especially relevant. When such a word is found, attention is captured. What could be more relevant to you than your own name? Mack and Rock (1998) discovered a visual analogue of the cocktail party effect. In their cross-hairs studies, they found that inattentional blindness did not occur when the unexpected stimulus was the name of the subject in the experiment. The name would capture attention and become visible as a result.

Attention-capture can also facilitate conscious perception in individuals with brain damage. For example, individuals with blindsight can have experiences in their blindfield when presented with very high contrast stimuli, and when this occurs, areas of the brain associated with attention become active (Sahraie et al., 1997). For some neglect patients, the locus of blindness can shift during attentional object tracking. Behrmann and Tipper (1994) showed a neglect patient a barbell picture and then rotated it 180 degrees as they watched it; initially, he was blind to the circle on the left, but after rotation, he was blind do the circle on the right. The blindsight case involves bottom-up attention, and the neglect case involves top-down attention. In both cases, attention helps to bring about conscious experiences in a visual region that is usually blind.

1.3 Attention and Subliminal Perception

I have been arguing that attention is necessary and sufficient for consciousness. This hypothesis was first inferred from research on unilateral neglect and then supported by appeal to studies in healthy people. But I have not yet shown that this hypothesis can explain the cases of subliminal perception with which this chapter began. Perception in neglect and inattentional blindness qualifies as subliminal in informal parlance, but the term is usually used to refer to cases when a stimulus is masked, extremely fast, or degraded. If it cannot explain those cases, then it cannot offer the needed account of how to draw the conscious/unconscious divide.

To explain core cases of subliminal perception, we need to begin with the simple fact that attention takes time. Once a stimulus is presented, it captures attention, either because it pops out or because we are looking for it. In either case, the initial visual response to the stimulus and the attention direction to it, are two distinct processes, and the latter happens after the former. We might attend to a region of space before a stimulus is presented there, but, even in this case, the stimulus presentation must cause a stimulus representation to be formed before that representation can itself become an object of attention. Or, put more accurately, we attend to an external stimulus only by attentionally modulating a representation of that stimulus, not by merely attending to the location in which it is presented.

This fact has a simple consequence. The representation that is caused by a stimulus can be modulated by attention only if it endures for a temporal interval that is long enough for attention to do its work. It is known that representations of stimuli do endure in perceptual systems after a stimulus is removed. This is called iconic memory. A visual stimulus, such as a color, can get to intermediate visual areas in about 110 milliseconds (Plendl et al. 1993). Attention to color has been recorded as fast as 125 milliseconds after stimulus onset (Connor et al. 2004). Other features might generate faster latencies for both visual response and attention, but attention always seems to lag behind (e.g., Maunsell and Gibson, 1988, find V1 responses within 20 milliseconds; and

Schoenfeld et al., 2007, find attention at 90 milliseconds). When stimuli are presented faster than the time-course of attention, they can still usually be consciously see, because they produce iconic images, which can last for about 300-500 milliseconds if the stimuli are sufficiently intense. But fleeting stimuli do not always produce iconic images that are available for attention. Consider three cases. First, if a stimulus is followed by a mask, the representation of the mask will quickly replace the representation of the stimulus, preventing an iconic memory trace from arising. Second, if the stimulus is low in intensity or contrast it may produce a perceptual representation that is correspondingly weak, and decay time may increase. Third, decay time is affected by stimulus presentation time, so stimuli that are presented very briefly may leave a weak and short-lived trace. These cases correspond to the conditions that are used in subliminal perception studies. My conjecture is that perception is unconscious in these cases because the stimulus conditions do not generate perceptual representations that are strong or long enough to be modulated by attention.

This interpretation is supported by recent work on visual masking by Enns and Di Lollo (2000). In traditional masked priming, the mask is placed in the same location as the stimulus: the mask in effect replaces the stimulus. But Enns and Di Lollo have shown that masking can be achieved by a mask that simply surrounds the stimulus without covering it. A letter followed by four dots placed in the space around where the letter was located can be masked by the four dots. Enns and Di Lollo explain the effect by proposing a new model of masking according to which attention is centrally involved. The dots attract attention before the representation of the stimulus can be attentionally modulated. This model, which they use to explain more traditional masking results, suggesting that standard cases of subliminal perception may be attentional effects. In standard cases of masking, attention does not have time to set in, and in cases where the mask does not overlap with the stimulus, and thus fails to wipe out the iconic memory trace, it may serve to distract attention away from that trace.

The upshot is that research on subliminal perception is consistent with the suggestion that attention is the mechanism by which consciousness is attained. Indeed, current models of masked priming explicitly hypothesize that attention makes the crucial difference. Enns and Di Lollo explicitly compare visual masking to inattentional blindness and point towards a unified account.

1.4 The AIR Theory of Consciousness

The evidence just reviewed suggests that attention makes the difference between conscious and unconscious perception. When we attend, perceptual states become conscious, and when attention is unavailable, consciousness does not arise. Attention, in other words, is necessary and sufficient for consciousness.

To avoid misunderstanding, let me underscore that attention is an answer to the How question: How do mental states become conscious? It is not an answer to the What question: What are the contents of conscious experience? The answer to the What question is that we are conscious of representations at an intermediate-level of representation in perceptual systems. These representations become conscious when we attend. Attention is necessary and sufficient for making intermediate-level representations conscious, not sufficient for making any mental state conscious.

Putting the What and the How together, we get the following theory of consciousness:

> *The AIR Theory of Consciousness*
> Consciousness arises when and only when intermediate-level representations are modulated by attention.

"AIR" stands for Attended Intermediate-level Representation. Conscious states are AIRs. I will have more to say about what it means for attention to modulate an intermediate-level representation in the next section. For now, the basic idea is that, when we attend, there is a change in the way intermediate-level representations are processed. That change is what makes the difference between these representations being conscious and not.

The AIR theory is a two-part theory of consciousness, because it has an account of the contents of consciousness and an account of the mechanisms by which we become conscious. There are other two part theories. For example, defenders of higher-order theories of consciousness, distinguish the target mental states, that are conscious and the representations of those states, that render them conscious (Rosenthal, 1997; Lycan, 2001; for critique, Prinz, forthcoming). Unlike these, the AIR theory is not metacognitive; attention does not work by re-representing the attended states. To see this, I need to say more about what attention is.

2. How Does Attention Give Rise To Experience?

2.1 What Is Attention?

To some ears, the claim that attention gives rise to consciousness sounds utterly uninformative because they think "attention" and "consciousness" are synonyms. Such semantic intuitions reveal a close link between attention and consciousness, but the two constructs can be defined independently. By consciousness I mean to refer to the property of having phenomenal qualities. Mental states are conscious if they feel like something, or, in Nagel's (1994) phrase, if there is something it is like to have them. Attention can be defined without reference to phenomenal qualities.

I treat "attention" as a natural kind term. It is not something that has an essence that can be discovered by conceptual analysis. Pre-theoretically, we grasp the concept of attention by appeal to a range of different activities and phenomena. A couple of those have already been mentioned. There is the phenomenon of pop-out, when a stimulus seems to stand out from things around it. Pop-out is passive, but attention can also be effortful. There is the phenomenon of search, as when you are looking for a specific object in a complex scene. Attention can also involve monitoring, as when we retain perceptual contact with something; tracking, as when we watch an object move through space; or vigilance, as when we remain alert and responsive to anything that might come before our senses. Attention sometimes involves selection, as when we focus in on a feature of an object. But it can also be diffuse, as when we survey our surroundings. Meditation can put one in a state of being hyper-attentive without attending to any specific thing. Put differently, attention can be thought of as a process that, in principle, could be applied to everything in the visual field at once, even if, in practice, it is usually selective.

I don't think any of these phenomena constitutes a definition of attention. Rather they are all cases in which we say that attention is taking place. From a pre-theoretical point of view, it is possible that these phenomena do not involve any overlapping mechanisms. They may be fundamentally different. But it is also possible that there is some shared mechanism running across all these cases or many of them. There may be a common denominator that can be empirically discovered. If such a common mechanism were found, we might say that "attention" refers to that mechanism. If these phenomena share nothing in common, then we might say "attention" should be dropped as a term from scientific psychology. We might become eliminativists.

To look for a common denominator, we might begin with one paradigm case, and then see whether the underlying, empirically discovered mechanisms are also operative in other cases. Consider pop-out. One thing that happens in cases of pop-out is stimulus competition. Various objects in an array compete with each other, and one (or more) wins the competition. Now it would be a mistake to say that attention is the processes of competing (e.g., Reynolds et al. 1999). Competition can occur outside of attention. Indeed, prior to the moment when a stimulus pops out, it is competing with other stimuli, but it hasn't yet captured our attention. So "attention" doesn't refer to competition, as such, but rather to process that occurs when a competition is won. But what is this process? What happens when one stimulus wins out over others while "vying for attention"? The best empirically informed answer to this question is that victory involves becoming available for certain kinds of further processing. But what kind? The losers, in a bout for attention, can also be processed further. For example, they can passively activate a network of semantically associated representation. This is what happens in priming. The victor does more than that. The victor becomes available for processes that are controlled and deliberative. For example, we can *report* the stimulus that pops out, we can reason about it, we can keep it in our minds for a while, and we can willfully choose to examine it further.

Psychologists postulate the existence of a capacity that plays all of these roles associated with victory in pop out. It's called "working memory." Working memory is a short term storage capacity, but one that allows for "executive control" (Baddely, 2007; D'Espisito & Postle, 1999). Once something is encoded in working memory, it becomes available to language systems for reporting, and with systems that allow effortful serial processing. Working memory can play a role in guiding effortful attention (e.g., Cowan, 1995), but it is also where attended perceptual states get temporarily stored (Knudsen, 2007). It is widely recognized that attention is a "gatekeeper" to working memory (Awh et al., 2006). Attention determines what information gets in.

Evidence for this view of attention comes from many sources. For example, consider a study by Rock and Gutman (1981) in which subjects had to attend to one of two overlapping shapes. Then, on a subsequent memory test, the attended shape was recalled and the unattended shape was not, even though both were presented for the same duration of time in plain view. There is also evidence that working memory capacity limits the allocation of attention. When working memory is full, it is harder to attend. In fact, studies have shown that inattentional blindness increases when people have to keep many items in their mind (Klein and Acevedo, 2002; Fougnie and Marois, 2007).

Such interactions between attention and working memory suggest an intimate relationship. The simplest explanation for this relationship is an identify claim: attention can be identified with the processes that allow information to be encoded in working memory. When a stimulus is attended, it becomes available to working memory, and if it is unattended, it is unavailable.

In the case of pop-out, this hypothesis amounts to the following. Pop-out occurs when the representation of one stimulus competes with the representations of surrounding stimuli and wins. When it wins, it is processed in a way that makes it available to working memory. This process is the psychological correlate of attention. I turns out that very same process—a process that makes perception available to working memory—may be implicated in all the phenomena that we called attention above. When you visually search for an item, like a can of Coke in crowded room, you use a template of the sought object as a filter on the representation of the room; when a match is found, it becomes available to working memory. When you monitor something or track it, a representation of it becomes available to working memory. When you are vigilant, you are in a state that disposes any new incoming representation to become available to working memory. When you selectively focus in on some part of an object, that part becomes available.

Working memory access can be determined top-down or bottom-up. Rather than seeing these as two different kinds of attention, we can see them as different control structures that make use of the same resource. There is some process that makes information available to working memory and that process can be passively triggered by stimulus features (bottom up) or actively recruited (top-down). But it's the same process in both cases. In the case of control, working memory is both the cause and the effect of attention. A representation in working memory is used to guide a search process and a successful match makes an input representation available for working memory encoding. Attention can be controlled by different kinds of representations. We can search for an object (the Coke can) or for a location. We can attend to color, shape, or shadow, to vision or sound. In each case, it is plausible that the same processes are taking place, even thought the control structure used to guide the process differs. By analogy, there are many ways to apply paint to a canvas: by hand, brush, or spray can; in strokes, small dabs, or expansive washes. But in all these cases, the end result in physically analogous; we have paint on a canvas. If we equate attention with the process in virtue of which perceptual information becomes available to working memory, then we can say that all examples of attention involve that process and differ only in what allows that process to take place.

Thus, there exists a uniform, empirically motivated account of what attention is. This is a satisfying result, because disparate cases seem to converge on one process. This explains why folk psychology has used the same term to cover all these cases, despite obvious and dramatic differences between them. The construct of working memory is not part of folk psychology, but there may be an implicit recognition of the fact that some of the items we perceive become available for reporting, deliberation, and so on. This idea of availability underlies all the phenomena we call attention. Psychological research reveals that reporting and deliberation are underwritten by a common short-term storage capacity, working memory. So the folk psychological insight implicit in the range of phenomena that we call "attention," can map onto the empirical construct of availability to working memory. We need not eliminate the folk construct; we have found a functional analysis.

Against this proposal, it might be objected that attention has an influence on perceptual processing in brain areas that never become available to working memory. It is know that attention leads to increased activity throughout the visual stream, for example, including primary visual areas whose contents are neither conscious, on my view, nor capable of being stored. So it seems to be a mistake to say that all attentional modulation renders perceptual activity available to working memory. In response to this objection, there are two lines of defense. First, it may be that the increased brain activity in low-level visual areas is simply an effect of increased activity in intermediate-level areas, which are potentially available to working memory. Second, the low-level increases may occur in the service of bringing about processing changes in areas that are potentially available to working memory. On the first alternative, attention operates directly only on brain areas that can become conscious, but back projections from these areas result in more widespread effects. On the second alternative, attention acts directly on early visual areas, but it does so in the service of making representations in subsequent processing areas available to working memory. Here, attention enhances some cells in V1, so that the cells to which they are connected in extrastriate cortex can become available. Either of these proposals can explain the widespread effects of attention while preserving the proposal that attention is process that allows information to flow from perception into structures that allow for temporary storage.

This analysis resolves the circularity worry. At first glance, it might have seemed unhelpful to propose that consciousness occurs when we attend. One might have thought that attention and consciousness are synonyms. But this is not the case. Consciousness is phenomenal character and attention is a process by which perceptual representations become available to working memory (compare Kirk, 1994, and Tye, 1995, who argue for a similar conclusion without appeal to attention or the psychological construct of working memory). The AIR theory can be unpacked accordingly:

> *The AIR Theory of Consciousness (Unpacked)*
> Consciousness arises when and only when intermediate-level representations undergo changes that allow them to become available to working memory.

This revised formation resolves the circularity, but it does leave one important question unanswered: what are the processes that allow for availability? I think those processes are likely to be specifiable only in neural, rather that psychological terms. Thus, a complete theory of consciousness might unpack the above formulation even further with a specification of the neural correlates of availability. Elsewhere, I speculate about those neural correlates, and I can only sketch that story here (Prinz, forthcoming). Roughly, current evidence points to the conclusion that attention involves an increase in the activity of inhibitory interneurons, which result in phase locked oscillations in stimulus encoding pyramidal cells; once synchronized, these cells can propagate forward to structures that can maintain neural activity during delay periods (working memory structures). The basic idea is that synchrony allows a population of neurons to produce a signal that can be picked up by other brain areas despite the noisy neuronal environment.

I speculatively submit that such synchrony is absent in cases of masked priming and inattentional blindness, but present whenever we consciously perveive.

This neural mechanism directly explains availability, and it also helps address a nagging worry that I have so far ignores. Availability is dispositional term, and one might think that consciousness cannot be defined by appeal to a disposition. For one thing, consciousness is an occurrent state and dispositions are mere potentials, and, for another thing, the disposition in question seems to easy to satisfy many unconscious stimuli could are available to working memory, in that we would encode them if conditions were right. But these worries evaporate once we see that availability is not a mere potentiality, but rather a physical process that takes place, changing functional connectivity in the brain. To be available is not a matter of what would happen if things were different in the brain, but what is happening in virtue of a change in how neurons are firing.

This sketch of the neural correlates will suffice for current purposes. My goal here has been to argue that the AIR theory is true and non-circular, and those arguments do not hinge on the precise neural details. There is, however, one issue at the psychological level of description that must be clarified.

2.3 Accessed or Accessible

The AIR theory says consciousness arises when intermediate-level representations become available to working memory. This raises a question. Why say that consciousness involved *availability* as opposed to actual encoding in working memory? Metaphorically, why think that consciousness involves broadcasting, rather than receiving? This question is pressing, because there is a popular approach to consciousness according to which receiving is necessary. This is the Global Workspace model, proposed by Baars (1988) and defended in neurobiological terms by Dehaene (2001). These authors say consciousness arises when information from the senses is brought into a global workspace, where it can be used to guide deliberation, reporting, and intentional behavior. The global workspace can be equated with working memory, insofar as working memory brings information in one sense modality to a functional space where it can play these disparate roles. In a similar spirit, Crick and Koch (1990) have claimed that consciousness depends on encoding in frontal cortex, where executive processes are realized. But I suggested that consciousness involves availability to the working memory, rather than encoding in working memory. Conscious states are accessible to working memory, but not necessarily accessed. Thus, consciousness is local (that is, located in sensory pathways) not global (that is, dependent on the involvement of "central" processes associated with higher cognition). I think there is considerable evidence favoring the local view over the global view.

The first problem for the view that consciousness involved working memory encoding is that working memory encodes the wrong things. Working memory encodes high-level perceptual representations—representations that abstract away from details on the stimuli that we consciously experience. Evidence for this comes from the sizable discrepancy that exists between discrimination and recall. Hasley and Chapanis (1951) demonstrated that people can discriminate about one million colors. When actually looking at pairs of colors, there are a million different colors that we can tell apart. But color recall is extremely limited. If we are presented with a color and then shown several similar color chips to choose from, moments later, we tend to do very badly. There are only about eleven to sixteen colors we can recognize. Since working memory is a storage system, and color storage is bad, it follows that working memory is not storing representations that are anywhere near as fine grained as the representations that underwrite conscious experience.

Second, there are many cases where we experience something that is too complex to readily encode in working memory, but not too complex to experience. Consider the displays used in visual search experiments (Treismann & Gelade, 1980). A subject might be present with a group of letters at various orientations, including one T and a dozen Ls. It's quite laborious to find a T in a sea of Ls. We might see such a display for well over one second and not notice whether there is a T. We may also see such a display without having any idea afterwards how many letters there were. Beyond four or five, numerosity is hard to encode in working memory without serious effort. So a typical visual search display may have features that are not encoded in working memory. But we nevertheless experience these features. If there were 13 big clear letters presented you for a few seconds, then it's overwhelmingly likely that you experienced all of them. And, if there was a T among them, you experienced that T even if you can't report that you did afterwards. The fact that such features go unreported, suggests that they are not encoded in working memory, even though they are experienced.

A third line of evidence comes from change blindness. "Change blindness" refers to the widely publicized fact that people often fail to notice when something directly before their eyes undergoes a change. For example, Rensink et al. (1997) showed subjects pairs of altered photographs with a moment of visual noise in between. In each case, a central object might change color, or size, or disappear. People in the photos switched hats, parrots changed from red to green, buildings vanished, and yet subjects failed to notice. Simons and Levin (1997) found that people failed to notice when a stranger on the street who stopped to ask them for directions was surreptitiously switched for another person wearing somewhat different clothing. Ballard (1994) found that people failed to notice when a stack of colored blocks changed colors during a video game in which they had to match those blocks on the other side of the screen. Change blindness is sometimes confused with inattentional blindness, but there is a crucial difference, and this is what I want to focus on here.

In inattentional blindness, subjects don't seem to have any experience corresponding to the unexpected stimulus. They are confident they have seen nothing aside from the items they are focusing on. In change blindness, however, subjects usually experience the whole image that undergoes the change. In some cases, the stare at these images for very long periods of time, even indefinitely long. They scan the whole image strategically, hoping they might pick up on the change. Every millimeter is examined and experienced. What subjects fail to notice is that some of the features that they experience change from one moment to the next. Thus, change blindness is essentially a deficit of memory, not of experience. People do not store what they are seeing. And, the failure of storage means they miss out on changes. They cannot match the vivid experience they have at second, with the vivid experience they have only a moment later. For example, in the Ballard study, people stare at the blocks they are trying to copy, and then those blocks change colors when they saccade away, even if just for a few milliseconds. Then, when subjects look back at the blocks, they don't notice the change. The reason is simple: they haven't stored the colors in working memory. There are too many colors to keep track of, and color usually isn't that important for object identification. Now here's the rub: the colors were experienced, we can presume, but not encoded. That means that experience can occur without working memory encoding.

A fourth reason for thinking encoding is not necessary derives from research on subliminal perception. In these studies, there are sometimes three different outcomes, depending on the stimulus conditions. In some cases, subjects experience nothing. If you ask them whether a stimulus was presented, they are at chance in guessing. In other cases, they are confident that there was a stimulus, but they do not know what it was. And in still other cases, they know they saw a stimulus and can readily report it (for these three outcomes, see, for example Kuider et al., 2007). It is obvious to anyone in these studies that the second and third cases are *both* supraliminal. Subjects experience something in these cases, even though they can only say what they experienced in the third case. The difference between case two and three involves working memory. The stimulus is encoded in working memory in the third case, but not in the second. This strongly suggests that encoding is not necessary for consciousness.

Finally, consider a study by Hasson et al. (2004) in which subjects watched movies in an fMRI scanner. The authors show that the brain responses in perceptual pathways are highly and predictably responsive to the film. For example, while watching Sergio Leone's spaghetti Western, *The Good, The Bad, and The Ugly*, visual areas associated with face processing became active when an actor's face was clearly visible on the screen. Hasson et al. report great consistency across subjects in their neural responses. For our purposes, the most important finding is that subjects showed no significant response in frontal areas of the brain. Passive movie watching is a perceptual affair. Working memory centers, and other brain areas associated with thought and executive control seem to go into a rest mode (presumably this wouldn't happen during films that are more challenging). So, there is reason to conclude that we are not encoding what we see on the screen in working memory. But movies are clearly experienced consciously. We do not become zombies when we enter a theater. So conscious experience does not depend on working memory encoding.

Collectively, these findings provide evidence for the AIR theory as I presented it above: consciousness depends of availability to working memory, not encoding. It depends on accessibility, not access. This way of putting it raises a further question, however. Why say that consciousness even involved availability? Block (1995) draws a distinction between what he calls access-consciousness, which arises when a perceptual state is poised for reporting and deliberation, and phenomenal-consciousness, which occurs whenever there are phenomenal experiences. He suggests that phenomenal-consciousness can occur without access-consciousness. As evidence, he cites a classic study by Sperling (1960), in which subjects are presented with 3 x 3 arrays of letters, and have to report what they have seen (see Philips, this volume). Subjects can typically report about three or four letters, but no more. But, which letters they report, can be determined by a cue that comes *after* the stimulus is presented. If, after the stimulus is removed, subjects are instructed to report the top row, they can succeed, but they cannot report the rest, and likewise for any other row. Subjects in this experiment report having an experience of the whole array, and the fact that they can report on any row they are asked about supports this. But Block says they have access to only one row. The other rows are phenomenally conscious, but not access conscious. On his interpretation of the result, these rows are not even available to working memory; they are not poised for reporting or deliberation.

This interpretation is difficult to defend. Sperling's study firmly establishes that the unreported rows *could* have been reported. Thus, Block is wrong to say they are not poised for reporting. In the terminology I have been using, the unreported rows are *available* to working memory, they just have not been encoded. They are accessible, but not accessed. So, in one sense, I agree with Block. A perceptual state can be phenomenally conscious without having been accessed by centers of higher cognition. Global theories of consciousness, like the Global Workspace theory, are wrong if they are intended as theories of phenomenal consciousness. But, Block is wrong to think that phenomenal consciousness does not involve information access. It is only in virtue of being accessible that perceptual states are experienced.

Block would reject this claim and push his case for phenomenal consciousness further back. He would say there can be phenomenal states that are not even accessible to working memory. That, I think, is implausible. To see why, consider again the contrast between change blindness and inattentional blindness. In change blindness, the changes are not noticed, but the items that change are experienced. The items could be reported at any moment, even if the changes go unnoticed. If, just after looking at the pile of colored blocks, Ballard asked a subject to report on one, the subject would probably supply a correct answer. In inattentional blindness, the surprise stimulus is not experienced at all, and not reportable. So we have a striking contrast. In both cases, there is no working memory encoding, but in one case there is experience and the other there is not. This difference in experience correlated with a difference in accessibility. In change blindness the item could be reported, if subjects were probed immediately after viewing. In innattional blindness, there is no reportability. That suggests that availability to working memory is necessary for experience.

Further support for this conclusion comes from subliminal perception studies. Consider the three cases mentioned above: no experience, experience without identification, and experience with identification. It seems utterly reasonable to say that in the first case, the stimulus is perceived but unavailable to working memory. Thus, phenomenology seems to disappear with availability. In the second case, the stimulus is available to some degree. The identity of the stimulus is not available, but, when probed, subjects recall that they saw *something*. In some cases, they can even recall when given a forced choice test just after the display. Here again, experience seems to come with availability.

In short, Block has one distinction too few. He says there can be phenomenal-consciousness without access, but "access" is ambiguous between accessibility and being accessed (see also Chalmers, 1997). Because he has not drawn this distinction, he erroneously concludes that phenomenal consciousness can arise without access of any kind—without even availability to working memory. But that claim makes it impossible to account for the distinction between change blindness and inattentional blindness, or between the first and second cases in subliminal perception studies. The Sperling cases that he uses to support phenomenal consciousness without access, actually provide powerful support for the conclusion that consciousness comes with availability to working memory, even though it doesn't require encoding.

In conclusion, I think the preponderance of empirical evidence favors the AIR theory as I presented it above. Consciousness arises when we attend, and attention makes information available to working memory. Consciousness does not depend on storage in working memory, and, indeed, the states we are conscious of cannot be adequately stored.

3. Objections

3.1 Alleged Evidence Against the Sufficiency of Attention

I have been explicating the AIR theory of consciousness, according to which attention is necessary and sufficient for rendering intermediate-level perceptual conscious. To defend the necessity claim, I argued that, when attention is absent, there is no consciousness. To defend the sufficiency claim, I argued that when we attend, we are conscious. Both of these claims have recently been challenged. Some argue that consciousness can arise is the absence of attention—hence challenging necessity. Others have argued that attention can arise in the absence of consciousness, challenging sufficiency. I will consider both of these challenges, beginning with the latter.

Consider an experiment that Kentridge et al. (2004) conducted with GY, the most studied individual with blindsight. They presented GY with an arrow in the center of a screen, followed by either a vertical or horizontal line in one of two locations in his blind visual field. The arrows were visible to him, but the oriented lines were not. At a tone, GY had to guess the orientation of the line he could not see. Kentridge et al. found that his accuracy increased if the line was located in the direction that the arrow was pointing. They concluded that the arrow leads GY to direct attention within his blind field, and it is that attention that facilitates performance. Thus, attention seems to be possible in the absence of consciousness.

It is possible that the behavior can be explained without supposing that GY attends to the unconscious stimulus. To see why, it is necessary to mention two physiological processes that typically co-occur with attention. First, attention usually co-occurs with eye movements (saccades). Overt saccades can be suppressed by asking a subject to stare at a fixation point, but even then fast microsaccades occur. This is significant because saccades shift the position of the fovea, allowing finer resolution processing at the point of gaze. If GY microsacceds in the direction of the cue, he can get a sharper visual representation of stimulus located there. This is likely to happen because saccades remain intact even after complete movement of V1. Second, attention normally co-occurs with a shrinking of the receptive fields in the attended location in retinotopic neural areas such as V4 and V5. That means more cells respond to the stimulus, and the resulting representation has higher resolution. Such receptive field shifts could explain GY's enhanced stimulus detection without assuming that GY also attends to the stimulus.

In response to the latter suggestion, one might object that shrinking receptive fields are part of the neural correlate of attention. If so, this reply would unwittingly concede that there can be attention without consciousness. There are three ways to reply.

First, I think receptive field shrinking is a contingent concomitant of attention, not a constituent part. Receptive field shrinking seems to be a separate and perhaps more primitive mechanism. Receptive fields shrink when eyes move (Tolias et al., 2001), and when small lights are flashed on the retinae of anesthetized cats (Wörgötter et al. 1998); they also shrink in and also in simpler creatures, such as fish (Umino and Ushio, 1998). Arguably, the mechanisms that drive these changes are dissociable from, and more ancient than the mechanisms that allow information to become available to working memory. To the extent that I have motivated the latter analysis of attention, it would be a mistake to equate attention with changes in receptive fields even if such changes often occur in conjunction with attentional shifts. Second, even if receptive field changes are a component of attention, they are not the full story. Without availability, which is clearly absent in blindsight, GY cannot be instantiating all the processes necessary for attention if the arguments for the availability analysis are compelling. Finally, GY's receptive field changes may reflect a shift in *spatial* attention rather than object-based attention. Recall that the cue occurs before the stimulus is presented. That could lead GY to attend to the cued region of space, but, because there is no stimulus there yet, it doesn't follow that he has an attended intermediate-level object representation—an AIR. Indeed, people with blindsight have grave difficulties representing objects; they show little neural activation in intermediate-level perception centers, and their residual abilities are thought to driven by subcortical mechanism that bypass the usual visual pathways (Stoerig and Cowey, 1997). If GY's performance stems from a shift in perceptual fields, that shift may occur in parts of the visual system that are primarily involved in spatial processing. The shift might heighten sensitivity to primitive perceptual features in the cued location (such as orientation) by lowering detection thresholds, but we need not suppose that there is a further attentional modulation of the stimulus representation once it is presented. In summary, the shrinking of receptive fields is either not a neural correlate of attention, only a contingent part of attention, or a correlate of spacital rather than object-based attention. On any of these alternatives, the discovery of shirking fields outside of consciousness would not refute the AIR theory.

The Kentridge study has two limitations when used as evidence against AIR. First, GY may suffer from a general deficit in his ability to form object representations, so his success may not reflect the presence of AIRs, and second, the attentional cue precedes the stimulus, so it is hard to confirm that an object representation has been modulated by attention, rather than a region of space. Both limitations are overcome in a clever study by Jiang et al. (2006). They used a paradigm called interocular suppression to generate an unconscious stimulus. Interocular suppression works by presented different images to each eye; if one image is dominant (e.g., brighter) it suppresses the other, making the other invisible. As their invisible stimulus, Jiang et al. used a figure that showed a naked body on one side, and a scrambled version of that body on the other. Naked bodies influence attention, but, in this case, the attraction was unconscious, because subjects were unaware of the stimulus. Then, in a test phase of the experiment, subjects were presented with a visible target that was either on the side where the nude had been or on the other side. The results were fascinating. For example, when presented with a female nude, heterosexual men showed an improvement in their reaction times, but gay men and both gay and heterosexual women tended to show a decrement in performance. For present purposes, the important finding is that the nudes clearly increased attention in some conditions, despite the fact that they were invisible. And for this to happen, the nudes need to have been recognized as such, hence processed throughout the visual hierarchy. Moreover, attention was not attracted by a cue that occurs prior to the stimulus. The nude itself attracted attention, suggesting that there was an attended intermediate-level perceptual representation in the absence of consciousness.

Fortunately for the AIR theory, there is an alternative interpretation, and it appeals to one of the same resources already discussed in response to the Kentridge study. Perhaps the nude is not modulated by attention, but rather induces microsacces. Nudes normally capture gaze in a way that is difficult to suppress even when we try not to look. The instruction to stare at fixation may not be strong enough to resist fast surreptitious glances towards the unconsciously perceived nude. Of course, nudes usually capture attention as well, but, in the case of interocular suppression, that may be impossible. The competing stimulus may prevent subjects from attending to the nude. If so, there is no attended intermediate-level representation. If this interpretation is right, the improved performance comes from gaze, not attention.

This interpretation may seem ad hoc, but it is actually supported by other research on interocular suppression. Fang and He (2005) measured neuronal response during this paradigm and found enhancement in the dorsal part of the visual stream, but not in the ventral part. The dorsal stream has been implicated in saccading as well as attention, but, when attention is involved there are detectable ventral increases as well (Corbetta et al. 1990). In interocular suppression, we don't seem to find such ventral enhancement. Thus, the Jiang study does not establish the existence of unconscious AIRs. An intermediate-level representation of the nude is probably generated, and it causes a shift in gaze that improves subsequent performance, but that representation is never itself the object of attentional enhancement. If it were, we should see the kind of ventral stream increases, which are believed to be absent interocular suppression.

There are a number of other studies attempting to establish attention in the absence of consciousness, but these are two of the best. The response strategies presented here can be readily extended to other experiments reported in the literature. I conclude that existing research fails to establish a clear case of unconscious attention.

3.2 Alleged Evidence Against the Necessity of Attention

Attention may be sufficient for consciousness, but is it necessary? I read research on inattentional blindness and inattentional blind as suggesting a resounding yes. When attention is withdrawn, consciousness seems to go as well. But some researchers are unconvinced. They think there are clear cases of consciousness in the absence of attention. This would mean that the AIR theory is mistaken.

Consider, first, an argument from Cristof Koch (personal communication). Imagine you are looking at an equally luminous, equally saturated wall of color (a "ganzfeld"). Your entire visual field is taken up by the color, and there is no variation in it, and no objects to focus on. In the case, there in no need to allocate attention, because attention is a selective capacity, and there is nothing to select. So, Koch reasons, under such conditions, attention is not engaged. Yet, it is obvious that we would experience the color.

The problem with this argument is that Koch is wrong to assume we would not allocate attention while looking at a ganzfeld. First of all, we might adopt a visual search strategy of scanning different parts of the field to different degrees at different times. We might let direction of gaze dictate focus of attention, in this case. Second of all, attention does not need to be focal or selective. We can attend diffusely to a whole field, just as we can attend to an object or a region of space.

In presenting his objection, Koch is assuming that attention is used only when we need to select between competing stimuli, but this is not the case. We can attend when selection problems are trivially easy, and this is just a limiting case of that. Attention is primarily a capacity for bringing perceptual information into working memory, and it just so happens that selection is often necessary for that. In this case, there is no competition, but there is certainly access to working memory, and, if my analysis of attention is right, this proves that attention is at work. One might even put this to an empirical test by first seeing what neural mechanisms underlie attention in paradigm cases of selection, and then looking to see whether those mechanisms are operative when we look at a ganzfeld. If the mechanisms of attention are interneuron inhibition and gamma synchrnony, then I would predict that these would be observed if we tested for them while staring at a wall of color.

Another piece of evidence for consciousness without attention comes from Reddy et al. (2006). They devised an interesting task that combines divided attention and masking. Subjects were presented with a cluster of letters (Ts and Ls) in the center of the screen and asked to determine whether they were all the same or different. Then, at the same time, a photograph of a celebrity was flashed in a corner of the screen followed by a mask. Subjects could identify them, even though the central task was extremely demanding on attention. Subjects do not report having clear experiences of the faces, but they do seem to experience *something* when the faces are flashed. This looks like conscious perception without attention.

The problem with this study is that the authors do not establish that the central task consumes all available attention. In fact, they describe it as a case of perception in the "near absence" of attention. That means, some attention was available, and that might account for why the faces, were consciously perceived. In fact, the experimental set up is similar in crucial ways to the Mack and Rock (1998) studies of inattentional blindness, in that both tasks show a stimulus while subjects are engaged in tasks that demand attention. We know from that work, that removal of attention can completely eliminate consciousness. Why then do Reddy et al. get different results? The answer may already be found in Mack and Rock. In one of their studies, Mack and Rock used smiley faces as the surprise stimulus, and they found that these faced poped-out and were consciously experienced. Faces are very significant stimuli, and familiar faces of people we admire are all the more so. In addition, Reddy et al. do not control for facial expressions, and the image the reproduce in their paper is Tom Cruise smiling broadly. Faces like this may capture attention and enter into conscious experience as a result. The study has one more limitation, worth reporting.

There is one more experiment that deserves consideration. Lamme (2003) has developed a paradigm that combines Sperling (1960) method mentioned above with change blindness. Subjects are presented with a ring of eight rectangles, each of which is oriented either horizontally or vertically. Afterwards, there is a blank screen and then the circle of rectangle returns. Subjects also see an arrow pointing to one rectangle in the circle and they have to say whether that arrow has changed its orientation. The key manipulation is when the arrow appears. On some trials the arrow appears with the original ring of rectangles, so subjects know which rectangle to monitor; on some trials, it appears only when the ring is presented for the second time; and on some, the arrow appears on the blank screen just after the ring is removed. Unsurprisingly, subjects make few errors when the arrow is presented at the start, because they know which rectangle to monitor, and they make many errors when the arrow is presented at the end, because they did not know which rectangle to monitor, and they could not keep track of all eight. The noteworthy finding is that subjects make few errors when the arrow is presented on the blank screen. The original ring is gone at this point, but there may be a trace in iconic memory that subjects can attend to, and from which they can recover information about the orientation of the rectangle that was in the location indicated by the rectangle a moment earlier. This replicates Sperling's results, and it also seems to suggest the possibility of consciousness in the absence of attention. Prior to the presentation of the rectangle on the blank screen, it seems implausible that any subjects were attending, except perhaps by chance, to the rectangle that was in the location to which the rectangle points. Yet the fact that subjects can recover information about the orientation of this rectangle, and use it to accurately and explicitly judge that the rectangle changed orientation suggests that the rectangle was consciously experienced in the initial presentation of the circle. Thus, in the original display, there is consciousness of all the rectangles, but not attention to all the rectangles, and this suggests that attention is not necessary for consciousness.

I am not convinced. Lamme's interpretation of his study echoes Block's interpretation of the Sperling studies. Following Block, he says his study shows that there can be phenomenal consciousness without access. But this is to conflate being accessed with being accessible. Clearly, all the rectangles are accessible. The fact that we can report on any, if cued, shows that. It is radically unlike inattentional blindness, where people are utterly unaware of the stimulus, and would presumably not be able to recover information about it if a cue were presented after the stimulus is taken away. The Lamme study also contrasts with inattention blindness in another way: there is no attention-demanding task. Thus, attention is available to scan and monitor the display. As with the ganzfeld, one might surmise that subjects attend diffusely to the whole screen, making the total array of rectangles available to working memory. Lamme is right that subjects are not attending selectively to each specific rectangle, but he has no grounds for saying subjects are not attending to the full assembly. And consequently, he has no grounds for saying there is consciousness without attention. At best he can say that the orientation of each rectangle is not encoded in working memory, but that conclusion, far from refuting the AIR theory, simply adds further support to my earlier conjectures that mere availability is sufficient for consciousness.

I have also argued that there is no compelling example of consciousness in the absence of attention. I have not surveyed every bit of counter-evidence, but I hope that the replies here can be extrapolated to other studies alleging to show consciousness without attention. The evidence that attention is necessary is much stronger than any extant evidence to the contrary.

4. Conclusion

In this chapter, I have argued that attention is necessary and sufficient for making out perceptual states conscious. I also argued that attention is a process by which information becomes available to working memory, and can arise without that information actually getting encoded. Finally, I addressed a number of empirical studies designed to show that attention and consciousness are dissociable. I argued that none of these provides evidence powerful enough to overturn the empirical case for the claim that attention and consciousness come and go together. There are other studies in the literature, some of which I review elsewhere, but they suffer from similar limitations (see de Brigard and Prinz, 2010; Prinz, 2010; Prinz, forthcoming). I conclude that current evidence offers supports the conjecture that attention is the mechanism by which perceptual representations become conscious.

References

Arnell, K.M., Killman, K.V., & Fijavz, D. (2007). Blinded by emotion: Target misses follow attention capture by arousing distractors in RSVP. *Emotion*, 7: 465-477.

Awh, E., Vogel, E.K. & Oh, S.H. (2006). Interactions between attention and working memory. *Neuroscience*, 139: 201–208.

Baars, B. (1988). *A cognitive theory of consciousness*. Cambridge, MA: Cambridge University Press.

Baddeley, A. D. (2007). *Working memory, thought and action.* Oxford: Oxford University Press.

Ballard, D.H., Hayhoe, M.M., & Pelz, J.B. (1994). Memory representations in natural tasks. *Journal of Cognitive Neuroscience*, 7: 66-80.

Behrmann, M., & Tipper, S.P. (1994). Object-based attentional mechanisms: Evidence from patients with unilateral neglect. In C. Umilta & M. Moscovitch (Eds.), *Attention and performance: Vol. 15. Conscious andnonconscious processing and cognitive functioning* (pp. 351–375). Cambridge, MA: MIT Press

Bisiach, E., & Rusconi, M.L. (1990). Break-down of perceptual awareness in unilateral neglect. *Cortex*, 26: 643-649.

Block, N. (1995). On a Confusion about the Function of Consciousness. *Behavioral and Brain. Sciences,* 18: 227-287.

Chalmers, D. (1997). Availability: the cognitive basis of experience? *Behavioral and Brain Sciences*, 20: 148-149.

Connor, C.E., Egeth, H.E., & Yantis, S. (2004). Visual attention: Bottom-up versus top-down. *Current Biology*, 14, R850-R852.

Cowan, N. (1995). *Attention and memory: An integrated framework*. New York, NY: Oxford University Press.

Corbetta, M., Miezin, F.M., Dobmeyer, S., Shulman, G.L., & Petersen, S.E. (1990). Attentional modulation of neural processing of shape, color, and velocity in humans. *Science*, 248: 1556-1559.

Crick, F., & Koch, C. (1990). Towards a neurobiological theory of consciousness. *Seminars in Neuroscience*, 2: 263-275.

D'Esposito, M. & Postle, B.R. (1999). The dependence of the mnemonic components of working memory on prefrontal cortex. *Neuropsychologia*, 37: 89-101.

De Brigard, F., & Prinz, J.J. (2010). Attention and consciousness. *Wiley Interdisciplinary Reviews: Cognitive Science*, 1.

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

Doricchi, F., & Galati, G. (2000). Implicit semantic evaluation of object symmetry and contralesional visual denial ina case of left unilateral neglect with damage of the dorsal paraventricular white matter. *Cortex*, 36: 337-350.

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

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

Fang, F. & He, S. (2005). Cortical responses to invisible objects in the human dorsal and ventral pathways. *Nature Neuroscience,* 8: 1380-1385.

Fougnie, D., & Marois, R. (2007). Executive load in working memory induces inattentional blindness. *Psychonomic Bulletin & Review,* 14: 142-147.

Halsey, R., & Chapanis, A. (1951) On the number of absolutely identifiable spectral hues. *Journal of the Optical Society of America,* 41: 1057-1058.

Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., & Malach, R. (2004). Intersubject synchronization of cortical activity during natural vision. *Science*, 303: 1634-1640.

Hillyard, S.A., Squires, K, C., Bauer, J.W, & Lindsav, P.H. (1971). Evoked potential correlates of auditory signal detection. *Science*, 172: 1357-1360.

Husain, M. & Kennard, C. (1996). Visual neglect associated with frontal lobe infarction. *Journal of Neurology*, 243, 652-657.

Jackendoff, R. (1987). *Consciousness and the computational mind*. Cambridge, MA: MIT Press.

Jiang, Y., Costello, P., Fang, F., Huang, M. & He, S. (2006). A gender- and sexual orientation-dependent spatial attentional effect of invisible images. *Proceedings of The National Academy of Science*, 103: 17048-17052.

Kanwisher, N. (2001). Neural events and perceptual awareness. *Cognition,* 79: 89-113.

Kentridge, R.W., Nijboer, T.C.W., & Heywood, C.A. (2008). Attended but unseen: Visual attention is not sufficient for visual awareness. *Neuropsychologia*, 46: 864-69.

Kirk, R. (1994). *Raw feeling: A philosophical account of the essence of consciousness*. New York, NY: Oxoford University Press.

Klein, K. & Acevedo, C. (2002). Working memory and inattentional blindness. . Poster presented at the Annual meeting of the North Carolina Cognition Group, Chapel Hill, NC.

Knudsen, E.I. (2007). Fundamental components of attention. *Annual Review of. Neuroscience*, 30, 57-78.

Kouider, S., Dehaene, S., Jobert, A. & Le Bihan, D. (2007). Cerebral bases of subliminal and supraliminal priming during reading. *Cerebral Cortex*, 17: 2019-2029.

Kunimoto, C., Miller, J., & Pashler, H. (2001). Confidence and accuracy of near-threshold discrimination responses. *Consciousness and Cognition*, 10: 294–340.

Lamme, V.A.F. (2003). Why visual attention and awareness are different. *Trends in Cognitive Sciences*, 7: 12-18.
Luck, S.J., Woodman, G.F., & Vogel, E.K. (2000). Event-related potential studies of attention. *Trends in Cognitive Sciences*, 4: 432-440.
Lycan, W.G. (2001). A Simple Argument for a Higher-Order Representation Theory of Consciousness. *Analysis,* 61: 3-4.
Mack, A., & Rock, I. (1998). *Inattentional blindness*. Cambridge, MA: MIT Press.
Marshall, J.C., & Halligan, P.W. (1988). Blindsight and insight in visiospatial neglect. *Nature*, 336: 766 – 767.
Maunsell, J.H.R., Gibson, J.R. (1992) Visual response latencies in striate cortex of the macaque monkey. *Journal of Neurophysiology*, 68:1332-1344.
Meister, I.G., Wienemann, M., Buelte, D., Grünewald, C., Sparing, R., & Dambeck, N. (2006). Hemiextinction induced by transcranial magnetic stimulation over the right temporo-parietal junction. *Neuroscience*, 142: 119–123.
Most, S.B., Scholl, B.J., Clifford, E., & Simons, D.J. (2005). What you see is what you set: Sustained inattentional blindness and the capture of awareness. *Psychological Review,* 112: 217-242.
Naccache, L. & Dehaene, S. (2001). Unconscious semantic priming extends to novel unseen stimuli. *Cognition*, 80: 215-229.
Naccache, L., Gaillard, R., Adam, C., Hasboun, D., Clémenceau, S., Baulac, M., Dehaene, S., & Cohen, L. (2005). A direct intracranial record of emotions evoked by subliminal words. *Proceedings of the National Academy of Science*, 102: 7713-7717.
Nagel, T. (1974). What is it like to be a bat? *Philosophical Review*, 83: 435– 450.
Pagano, C., & Turvey, M. T. (1998). Eigenvectors of the inertia tensor and perceiving the orientations of limbs and objects. *Journal of Applied Biomechanics*, 14: 331-359.
Plendl, H., Paulus, W., Roberts, I. G., Gotzel, K., Towell, A., Pitman, J. R., Scherg, M., & Halliday, A. M. (1993). The time course and location of cerebral evoked activity associated with the processing of color stimuli in man. *Neuroscience Letters, 150*: 9-12.
Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology,* 32: 3-25.
Posner, M. I. & Raichle, M. E. (1994). *Images of mind.* New York, NY: Scientific American Library.
Prinz, J.J. (2000). A neurofunctional theory of visual consciousness. *Consciousness and Cognition,* 9: 243–259.
Prinz, J.J. (2005). A neurofunctional theory of consciousness. In *Cognition and the brain: Philosophy and neuroscience movement,* ed. Andy Brook and Kathleen Akins, 381-396. Cambridge: Cambridge University Press.
Prinz, J.J. (2010). When is perception conscious? In B. Nanay (ed.), *Perceiving the world: New essays on perception*. New York: Oxford University Press.
Prinz, J.J. (forthcoming). *The conscious brain*. New York, NY: Oxford University Press.
Reddy, L., Reddy, L., & Koch, C. (2006). Face identification in the near-absence of focal attention. *Vision Research*, 46: 2336-2343
Rees, G., Wojciulik, E., Clarke, K., Husain, M., Frith, C.D., Driver, J. (2000).

Unconscious activation of visual cortex in the damaged right hemisphere of a parietal patient with extinction. *Brain*, 123: 1624-1633.
Rensink, R.A., O'Regan, J.K.. & Clark, J.J. (1997). To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science*, 8: 368–373,
Reynolds, J.H., Chelazzi, L. & Desimone, R. (1999). Competitive mechanisms subserve attention in macaque areas V2 and V4. *Journal of Neuroscience*, 19: 1736-1753.
Rock, I., & Gutman, D. (1981). The effect of inattention on form perception. *Journal of Experimental Psychology: Human Perception & Performance,* 7: 275-285.
Rosenthal, D. M. (1997). A theory of consciousness. In *The nature of consciousness: Philosophical debates*, ed. Ned Block, Owen Flanagan, and Guven Güzeldere, 729-753. Cambridge, MA: MIT Press.
Ruff, C. C., Bestmann, S., Blankenburg, F., Bjoertomt, O., Josephs, O., Weiskopf, N., Deichmann, R., & Driver, J. (2008). Distinct causal influences of parietal versus frontal areas on visual cortex:.evidence from concurrent TMS-fMRI. *Cerebral Cortex*, 18: 817-827.
Schnall, S., Haidt, J., Clore, G. L., & Jordan, A. H. (2008). Disgust as embodied moral judgment. *Personality and Social Psychology Bulletin*, 34: 1096-1109.
Schoenfeld, M.A., Hopf, Martínez, A, , Mai, H.M., Sattler, C., Gasde, A., Heinze, H. -J. & Hillyard, S.A. (2007). Spatio-temporal analysis of feature-based attention. *Cerebral Cortex*, 17: 2468-2477.
Simons, D.J., & Levin, D.T. (1997). Change blindness. *Trends in Cognitive Science,* 1: 261-267.
Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs,* 74: 1-29.
Stoerig, P. & Cowey, A. (1997). Blindsight in man and monkey. *Brain*, 120: 535-59.
Szczepanowski, R., & Pessoa, L. (2007). Fear perception: Can objective and subjective awareness measures be dissociated? *Journal of Vision*, 7:10, 1-17.
Tolias, A.S., Moore, T., Smirnakis, S.M., Tehovnik, E.J., Siapas, A.G., & Schiller, P.H. (2001). Eye movements modulate visual receptive fields of V4 neurons. *Neuron*, 29: 757-767.
Treisman, A., & Gelade, G. (1980). A feature integration theory of attention. *Cognitive Psychology*, 12: 97-136.
Tye, M. (1995). *Ten problems of consciousness*. Cambridge, MA: MIT Press.
Umino, O. & Ushio, T. (1998). Spatio-temporal receptive fields in carp retinal horizontal cells. *Journal of Physiology*, 508: 223—236.
Weiskrantz, L. (1986). *Blindsight: A case study and implications*. Oxford: Oxford University Press.
Winkielman, P., Berridge, K. C., & Wilbarger, J. L. (2005). Unconscious affective reactions to masked happy versus angry faces influence consumption behavior and judgments of value. *Personality and Social Psychology Bulletin*, 1: 121-135.
Wolfe, J.M. (1999). Inattentional amnesia. In V. Coltheart (ed.), *Fleeting Memories* (pp. 71-94). Cambridge, MA: MIT Press.
Wörgötter, F., Suder, K., Zhao, Y., Kerscher, N., Eysel U.T., & Funke, K. (1998). State-dependent receptive-field restructuring in the visual cortex. *Nature,* 396: 165-168.