When is Perception Conscious?[1]

Jesse J. Prinz

Once upon a time, people thought that all perception was conscious. Indeed, it was widely believed that all mental states are conscious, so the problem of explaining consciousness collapses into the problem of explaining mentality. But things have changed. Most people now believe that a lot goes on unconsciously. Indeed, some people believe that mental states that are not perceptual in nature are never conscious. That's a matter of controversy. Less controversial is the claim that perceptual states are conscious some of the time, but not all of the time. This raises a question. When are perceptual states conscious? A theory of consciousness is, in large part, an answer to that question. In this chapter, I will offer a few critical remarks on one answer that has been popular in philosophy, and then I will offer a defense of another answer that has emerged out of cognitive science. To avoid undue suspense, the answer that I favor is that perceptual states become conscious when and only when the perceiver is attending.

1. The Conscious/Unconscious Divide

1.1 Unconscious Perception

In case there is any doubt, it will help to briefly review some reasons for thinking that perception can occur unconsciously. Evidence comes in various forms. I will focus here and throughout on the visual modality, because vision has been most thoroughly studied. One line of evidence comes from brain injury. Consider blindsight. People with lesions in the primary visual cortex (V1) are said to be cortically blind: they claim not to see objects presented in the affected portion of the visual field. In cases of global V1 damage, these individuals cannot spontaneously detect any changes made available to the eyes, even if they are quite dramatic. In one recent case report, for example, a subject with global V1 damage could not tell when the bright lights in a room were turned on and off (Hamm et al., 2003). But, in some cases of cortical blindness, there is residual visual ability: when asked to guess where an object is located, these individuals do quite well, even though they think they do not see the objects they are pointing at (Weiskrantz, 1986).

Blindsight, as it is called, is a case of visual perception without conscious experience of what is being perceived. It counts as perception because people with blindsight are receiving information through a sensory transducer (the ideas) and responding to it. Some people with blindsight even retain some ability to recognize objects in the blind field. De Gelder et al. (2005) presented a blindsight subject with emotional facial expressions and these influenced subsequent emotion tasks, even

[1] In writing this chapter, I had helpful discussions with Felipe De Brigard, Dave Chalmers, Anya Farennikova (who also corrected my English), Bill Lycan, and Bence Nanay (without whom this chapter would not exist). I also benefitted considerably from audience feedback at the 2008 SPAWN conference at Syracuse University.

though he showed no sign of having consciously perceived the faces. Hamm et al. (2003) were able to condition a person with blindsight to have an aversive response to pictures of airplanes even though he reported no conscious awareness of the airplane pictures or control pictures when they were being presented.

Unconscious perception can also be demonstrated in individuals with intact brains. The most widely practiced method for doing this is backwards masking. Subjects are briefly presented with a visual stimulus which is followed by a second stimulus (the "mask") that prevents them from seeing the first. If the first stimulus is presented for a long enough duration (say, 200 milliseconds) subjects can see it, identify it, and spontaneously report on their experience. At shorter durations (say, 50 milliseconds), the stimulus cannot be identified, but subjects are confident that they saw *something* flash before their eyes. At even shorter durations (say, 16 milliseconds) subjects do not report seeing a first stimulus, prior to the mask. Indeed, if they are given a sequence of trials in which some show a stimulus before the mask and others don't, they are at chance in guessing which kind of trial they are on. They have no idea that there has been a first stimulus.

Masking can be achieved in various ways. In the most typical designs the mask is a stimulus that occupies the same region of space that was occupied by the first stimulus. In other cases, the mask does not occupy the same region. In "metacontrast masking", subjects typically see a mask that occupies the area surrounding the space in which the first stimulus had been presented. For example, they see a colored disk followed by a colored ring that would surround the disk had the two stimuli been presented simultaneously. A study using metacontrast masking will be discussed bellow.

For present purposes, the crucial thing to note is that masked stimuli are perceived. They influence information processing in measurable ways. This is shown using a family of methods called priming. A masked word can influence answers given on a word completion task or facilitate memory access to associated words. A masked color can facilitate detection of the same color. In some cases, the influence is dramatic and wide ranging. For example, Winkielman et al. (2005) presented subjects with masked facial expressions, which were either angry or happy, and then asked them to evaluate a soft drink. As after unconsciously seeing happy faces, as compared to angry faces, subjects said the soft drink was more delicious, they were willing to pay more for it, and they poured more of it into their cups. Clearly they were recognizing the emotional valence of the face, but they were doing so unconsciously.

1.1 Are Conscious Perceptual Representations Special?

A good strategy for finding a theory of consciousness is to compare conscious and unconscious perceptual states and see if there are any physical or functional differences between them. We want to know, what is distinctive about the perceptual states that we experience consciously? One possibility that might look promising at first is that there are two distinct kinds of percepts. Some are conscious and others are not. The idea would be to show that perceptual systems use a variety of different kinds of representations, of which only a special subset are

conscious. I think this suggestion won't ultimately succeed, but it's not obviously wrong. Indeed, there is a kernel of truth to the idea that will play an important role in developing an adequate theory of consciousness.

The hypothesis that each perceptual systems use different kinds of representations is true, and, it turns out, some of these may be privileged when it comes to consciousness. The representations vary along two dimensions, what they represent, and how abstractly. In vision, there are cell populations that respond to color, others to shape, and still others to motion (Zeki, 1993). These can be regarded as different types of representation. A psychological-level theory would treat these as separate symbols systems with their own primitive representations and rules of combination. But it can't be the case any one of these systems is privileged with respect to consciousness: we consciously experience shape, color, and motion.

More interesting, in this context, is the dimension of abstraction. Like all sensory systems, the visual system is hierarchical (Marr, 1982). It begins with cell populations that operate relatively independently of each other and respond to very discrete features of a stimulus, such as a small edge or bit of color. At this level (associated with V1), the visual system is not very interested in how these edges and colors hang together, or how foreground differs from background. This is low-level vision. The connections between edges arise at an intermediate level of visual processing (found in extrastriate visual areas, V2-V5). Here, edges become bounded contours, figure separates from ground, and binocular information is more thoroughly integrated to reveal depth. We also see global context effects at the intermediate level: color sensitive cells fire in a way that is sensitive to the color in other areas, spanning over distances that are larger than the receptive fields of cells in early vision. Illusory contours are also registered at the intermediate level, completing shapes that are only partially present in the stimulus. The intermediate level is not the last stage however. Representations here are very specific to vantage point. Objects are represented form a specific point of view and they occupy a specific location and size in the visual field. This level of specificity is not ideal for object recognition. To recognize objects, it is useful to abstract away from specific vantage points and discern the underlying structure of objects. Cells in high-level visual areas (in inferior temporal cortex) do just that. They abstract away from location, visual size, and, to a considerable degree, viewing angle. In some sense, cells at these different stages are representing different things (a vantage-point specific representation encodes different features than a vantage-point invariant representation), but the crucial difference is in the degree of abstraction (the same kinds of stimulus dimensions are represented in ways that are less and less like the pattern of stimulation on the retina).

It was against the background of this picture, that Ray Jackendoff (1987) put forward his Intermediate-Level Theory of Consciousness. Of the three levels just described, it is the intermediate level that best corresponds to what people report in conscious experience. If I see a walrus in the zoo, I experience it as a bounded whole, separate from its environment, and from a point of view. It occupies a certain region of my conscious visual field, and I could, were it still enough and I skilled enough, draw its portrait from the particular vantage point at which I'm

standing. My portrait would not be a Picasso monstrosity, simultaneously representing multiple perspectives, as one might expect if I were copying the contents of high-level vision, and nor would it be like one of Seurat's pointillist paintings viewed from an inch away, obscuring the boundaries and decomposing the surfaces.

I have argued elsewhere, on neurobiological grounds, that Jackendoff's conjecture is right (Prinz, 2000; see also Koch and Braun, 1996). Visual consciousness arises at an intermediate level of processing, and other visual representations are always unconscious. This suggests an answer to the central question of this chapter: conscious arises when, and only when, intermediate-level perceptual representations are activated.

It's a tempting conjecture, but, alas, it won't do. The difficulty is that mere activation of the intermediate-level or any other level of perceptual processing is not sufficient for consciousness. Indeed, we have already seen why. There is such a thing as unconscious perception. And unconscious perception often involves the unconscious recognition of objects. Priming studies show that people can respond in content-specific ways to unconsciously perceived faces, spiders, airplanes, words, numbers, and everyday objects, such as desk lamps. This suggests that cases of unconscious vision involve processing through the visual hierarchy (see, e.g., Bar, 2000). For many objects recognition is achieved only in high-level visual centers, and this, in turn, requires prior activation in intermediate- and low-levels of processing. This conclusion, confirmed by neuroimaging, shows that mere activation of intermediate-level visual representations is not sufficient for consciousness.

Thus, it looks like we won't find any special class of representations in the visual system that are the conscious representations. There very same neurally realized representations can be activated consciously or unconsciously. We have reason to believe that some visual representations can never be conscious, while others—those at the intermediate level—are conscious some of the time. But when? We need a theory of how intermediate-level representations become conscious. What is the difference between conditions under which these representations are conscious and when they are not? Let's turn now, to two proposals.

2. From HOR to AIR

2.1 Higher-Order Theories

Philosophers sometimes point out that theories of consciousness need two parts: an account of qualitative character of conscious states (what kinds of things are we conscious of) and an account of how states become conscious. The intermediate-level hypothesis is a contribution to a theory of the first part. But what about the second part? Philosophers who are careful about distinguishing these two parts have come up with various theories of the second, but one class of theories is particularly prevalent. According to the prevailing approach, a mental state becomes conscious when there is some other mental state that represents it. A

mental representation that represents another is a higher-order representation, and so these are known as HOR theories of consciousness. In this section I want to indicate why I think HOR theories won't succeed. The concerns I will raise provide some of the philosophical motivation for the alternative proposal defended in the next section.

HOR theories come in several varieties, but the differences need not concern us here. The question I want to ask is why think that any HOR theory is true? In answering this question, HOR theorists of different stripes often converge on variants of the same *a priori* argument. It goes like this (compare Rosenthal, 1997; Lycan, 2001):

P1. A conscious state is a state of which a subject is conscious
P2. To be conscious *of* a state a subject must represent it
C . Therefore a conscious state only if it is represented by a subject

I think this argument is dubious. In fact, there is reason to doubt it from the outset. We are trying to figure out what distinguishes conscious and unconscious states. It is hard to imagine why that could possibly be determined *a priori*. It's not a conceptual question. Analysis of the concept *consciousness* is not what we are after. We may not even have a firm concept of consciousness, and concepts, notoriously, encode information in the form of stereotypes and folk theories that can be profoundly mistaken (some people conceptualize the world as flat, gorillas as ferocious, dolphins as docile, various ethnic groups as having traits they lack, and so on). An account of conscious should explain why some states feel like something and others don't, and it should ultimately also explain various empirically discovered facts about the functional role of the states that feel like something. There is no reason to think an *a priori* argument can prove a substantive theory of consciousness or even identify an interesting necessary condition that points us towards such a theory.

The argument also has dubious premises. A conscious state is a state that feels like something. Does that mean there is someone who is conscious of it? Not necessarily. It could work out that feeling like something in an intrinsic property of certain states. For all we know in advance, it could turn out that the brain states that are conscious would feel like something if they were extracted and put in a Petri dish. Of course, when conscious states are in full brains, there is a subject who has them. There is then, at least a nominal sense, in which conscious states are conscious to their bearers. But it does not follow from this that the bearer is conscious of the state. "Conscious of" is often used to imply information access. When you look at a complex scene, it seems plausible that many details feel like something to you, even though you are not conscious of them. We'd need substantive arguments to prove otherwise. Short of that, P1 can be defended as a stipulative definition of some technical sense of "conscious state," but, if so, there is no reason to think that conscious states so defined are the ones of interest—the ones that feel like something rather than nothing.

P2 fares no better. It is perfectly possible that one can be conscious of a mental state without representing it. First, "consciousness of" mental states could just be a

grammatical construction that doesn't reveal the real way we gain access to our mental states. Just as the German languages forces uses to postulate a subject when they say "Es gibt" (there is), we may have a syntactic transformation that goes from "state S is conscious" to "someone is conscious of S." There is no reason to expect this pattern in every language. If so, there is no reason to think that the locution "someone is conscious of S" is using the "of" in an intentional sense. Second, even if the "of" is intentional, it may not be representational. Representations are used by the mind to keep track of these in the external world, but, for things inside the mind, representations may be unnecessary. We can keep track of a mental state by having the state. On this view, when a subject is conscious of a mental state, that is in virtue of direct experience the states itself. It is an experience of the state, but not a representation of the state; the state itself is directly experienced. I don't mean to suggest that the locution "conscious of" is never used representationally. We have a doxastic use that is a near synonym for "know that," as in "I am conscious of the fact that aardvarks are nocturnal." But *this* use has little if anything to do with the kind of consciousness that is of interest here: the property of feeling like something as opposed to not feeling like anything.

Thus far I have been arguing that a core argument for HOR theories is deeply problematic. At this point, the HOR theorist might move away from the armchair and try to provide empirical support for the view. The problem is that such a defense has not been seriously undertaken and probably won't succeed. There is no study, to my knowledge, testing a HOR theory, nor any experimentally established method for manipulating or measuring HORs of the kind postulated in these theories. There is no known case of selective damage to HORs leading to deficits in consciousness, and deficits in the use of mental concepts (which are required on some HOR theories) are not correlated with diminished or distorted experience.

We might try to rectify these problems by actively looking for HORs in the brain during conscious episodes. This is probably a fool's errand. Extensive research has been done on what the cells in our sensory systems respond to, and none of that research points to cells that represent what other cells are doing. No significant model of sensory circuits postulates metarepesentations, as far as I know.

The HOR theorist might move outside of sensory pathways to find HORs that are active during conscious perception, but not built into perceptual systems. Such a search is also likely to fail. The closest thing to higher-order representations recognized in sensory neuroscience are working memory encodings that allow us to store visual states during brief temporal intervals. But it would be a mistake to think of working memory encodings as HORs. First, working memory encodings do not seem to re-represent perceptual states. Rather, they maintain activation in perceptual pathways of the brain. Second, even if working memory contained representations corresponding to perceptual states, they wouldn't qualify as representations of those states; rather they would be stored copies of those states, and like the states they copy, they would represent features of the world. For example, if you hold a phone number in your head while walking to the phone, your working memory represents the number, not your perception of the number. Third, working memory seems to encode perceptual information in a way that is much more coarsely grained than intermediate-level perceptual states. For example, if

you try to store a specific shade of blue in working memory, you won't succeed. If shown one blue patch followed by three others that include the first as well as two that are similar, you won't know which one was the original. So, if working memory represents perceptual states, it represents them using a code that is coarser in grain than the representations that we consciously experience.

In summary, I think HOR theories lack adequate *a priori* support and adequate empirical support. I think we need a theory of consciousness that is driven by empirical data. I turn to such a theory now.

2.2 The AIR Theory

One way to empirical identify the factor that is responsible for conscious experience is to see what factor leads to an absence of consciousness when it is taken away. We have already seen one such factor: time. If a stimulus is presented too quickly, it won't be experienced. But time itself probably isn't the key. There are, as we will see in a moment, conditions under which stimuli presented for reasonably long durations do not get consciously experienced. It is more likely that time plays an indirect role. Stimuli that are presented for long enough durations become candidates for other processes to act on them, and one of these other processes is responsible for rendering perceptual states conscious. But which one? What process leads to blindness when it is prevented from occurring? Recent research points to one very plausible answer: attention.

It's a commonplace that we can fail to see things when we're not paying attention. But platitudes are often false. To put this one to the test, vision veterans Mack and Rock (1998) developed an experimental paradigm in which subjects are presented with shapes or words while performing a concurrent task that demands a lot of attention. I particular, subjects are asked to determine which of two intersecting lines is longer—hard work if line lengths are close. While doing this, an unexpected polygon, face, or word flashes in the center of the visual field for 200 milliseconds, which is normally well above the conscious threshold. In these studies, about 25% of observers fail to notice the unexpected stimulus. It's not just that they see something and can't identify it. They seem to see nothing at all. Immediately after the critical trial, subjects are a series of increasingly leading questions to see if they experienced anything other than the intersecting lines. In many cases, the answer is no; they subjects simply don't see an object presented in clear view. This is called inattentional blindness, and it has now been replicated many times, often showing even higher rates of detection.

Inattentional blindness suggests that attention is necessary for consciousness. This conjecture gains further support from other experimental paradigms. For example, Macdonald and Lavie (2008) have demonstrated what they call "load-induced blindness." Subjects are asked to search for a target letter in a group of letters, and, while performing this task, a meaningless shape is flashed. Subjects know what the shape will look like in advance but only 37% detect it when they are given six letters for the search task, which introduces a large attentional load. In addition, there is a phenomenon called the attentional blink, in which subjects fail to see the second of two targets in a rapid series letters or numerals

(Raymond et al., 1992). The first, target captures attention briefly, and the second goes unseen if presented shortly thereafter. Researchers have recently discovered a related phenomenon called the emotional blink (Arnell et al., 2007). If subjects are asked to look for the name of a color in a series of words, they will fail to see it if it is displayed shortly after an emotionally charged word that captures attention, such as "orgasm."

These behavioral experiments on healthy subjects actually re-confirm something that has been well known in neurology for a long time: damage to attention centers of the brain disrupts consciousness. The most familiar case of this is visual neglect. People who sustain injuries to attention centers in right inferior parietal cortex seem to be blind in the contralesional region. They are oblivious to stimuli presented on the left (or the left side of object), despite the fact that their visual systems are intact and responsive to those stimuli. When asked to compare two objects that differ only on the left, subjects with neglect report that the objects are identical, even though there is sometimes evidence for unconscious processing of the invisible features (Marshall and Halligan, 1988). Neglect can also affect perception in other sense modalities. Some patients become oblivious to the left sides of their bodies, even insisting that their left limbs are not their own. In auditory neglect, words heard in the left go unnoticed during dichotic listening tasks.

The evidence strongly suggests that attention is necessary for consciousness. Attention may also be sufficient. This is evidence from many of the studies just mentioned. In inattentional blindness studies, subjects attend to the intersecting lines, and they consciously see them. In attentional blindness, the first target is consciously seen, because subjects are looking for it. These are cases of top down attention: we experience what we chose to pay attention to. In other cases, attention is bottom up. Features of a stimulus capture our attention. And, when they do, consciousness results. A circle in a sea of squares will "pop out," meaning it will capture attention and come into awareness. Pop out sometimes depends on unconscious filters that respond to significant stimuli. At a cocktail party, we can hear our own names if they are mentioned from across the room, even if the conversation there had been inaudible. Mack and Rock (1998) found a visual analogue for the cocktail party effect in their inattentional blindness studies. Subjects always consciously see the surprise stimulus when it is their own name. Subjects also experience pop-out effects for smiling faces and for emotionally changed words.

In sum, it seems that attention is both necessary and sufficient for making intermediate-level representations conscious. Conscious states are attended intermediate-level representations, or AIRs (Prinz, 2000; 2005). Unlike HOR theories, this claim is empirically motivated, and it does not require metarepresentation. The mind does not need to represent perceptual states in order for them to be experienced.

2.3 What Is Attention?

I have been claiming that consciousness arises when, and only when, we attend. But, as stated, this may seem unsatisfying because I have not yet said what attention is. By way of elucidation, let me consider four objections that will lead to a substantive theory of attention, and more precise statement of the AIR theory.

First, one might object that the theory is circular. I say perception becomes conscious when we attend. But, one might think that attention should be defined in terms of consciousness. An attended stimulus, on this view, is by definition one when are conscious of.

I do not think the conceptual link between consciousness and attention is so tight. Indeed, we will see that some researchers claim that they are dissociable. In any case, I think attention can be defined without reference to consciousness. Empirically, there is a close link between attention and working memory, the mechanisms that allow us to temporarily store information for use in controlled cognitive processes. In a word, attention makes information *available* to working memory. Attention is a change in the way perceptual representations are processed that allows them to send signals to working memory. This is not a bit conceptual analysis, but rather an *a posteriori* identity, supported by empirical evidence. For example, Rock and Gutman (1981) showed that, when attending to one of two overlapping shapes, subjects can remember the attended one, but not the unattended one. When attention is spread thin, fewer things can get into working memory (Sperling, 1960). And, when working memory is occupied, attention is limited. Fougnie and Marois (2007) have shown that inattentional blindness can be induced by giving people a heavy working memory load. Neuroimaging studies show that when compared to unconsciously perceived stimuli, conscious perception is associated with co-activation of parietal attention structures and dorsolateral prefrontal working memory structures (Rees et al., 2002).

In response to this *a posteriori* definition of attention, critics might advance a second objection. Surely, they'll say, attention cannot be identified with any one thing. It is very tempting to say attention is not a natural kind. After all, we use the word to refer to a wide range of different mental phenomena. These include vigilance ("Pay attention!"), monitoring ("Maintain attention on this spot here"), tracking ("Attend to the ball!"), pop-out ("The stimulus captured my attention"), focus ("Attend to the fine details"), and selection ("I attended to Xs and ignored Ys"). I already noted that attention can be top down or bottom up. Why think that all these phenomena have a common essence? Attention seems to be a mongrel category. If so, the claim that consciousness arises when we attend is unpromising, because attention is not one thing, but many.

Here again, I think the working memory account can do some work. It is true that there are many ways to control attention. One can monitor a whole scene, track an object, or select one object over another. Attention can be controlled by a perceiver's plans or by features of the stimuli perceived. But the variability of control should not dupe us into thinking attention itself is varied. In each of these cases, attention coincides with availability. If you monitor an object in the scene in front of you, its features and changes become available to working memory, and if you allow things to pop out, those things will become available to working memory.

It seems the impact of attention is more or less the same regardless of the source. This supports the conjecture that attention is a single process.

One important consequence of the single process view is that attention can retain its identity as a process even when applied very differently. Attention can be highly focused, as when we study discrete features of an object in our field of view, or quite diffuse, as when we scan a scene or passively take in a vista. Stage light metaphors are useful here. Attention can act like a floodlight or a spotlight, and there can even be an attentive spotlight (point of focus) within a diffusely lit scene. Attention also seems to have a limited capacity; some things go unlit. As focus increased, there may be less attention left for diffuse monitoring.

Talk of diffuse attention raises a third objection. Working memory has a highly limited capacity. It seems we can only store about four items at once (Cowan, 1999). And what we store seems quite coarse grained. For example, recognition of color patches after short temporal delays is very limited. So it looks like working memory stored small numbers of relatively abstract chunks. But attention, I have just said, can be diffuse, covering large areas, and it can also be highly detailed. We can attend to specific colors and shapes, which have no hope of getting stored in working memory. The hypothesis that attention is a mechanism that makes information available to working memory seems hopeless when these facts are considered.

The difficulty can be addressed if we distinguish availability and encoding. A perceptual stimulus is available to working memory when it is processed in a way that would allow for working memory encoding, but many things are available and never get encoded. If you glance at a scene, you could store information about many parts of it, but you don't. Most things are forgotten instantly, because they are not stored in working memory. That is why change blindness is so pervasive (Simons and Levin, 1997). When a change is detected it is usually because the changing feature was, by luck, encoded. The feature could have been encoded prior to detection. It was available, but not stored.

Working memory encoding is not simple copying. When a perceived item is stored in working memory, working memory does not encode a duplicate of the item. Indeed, the duplication metaphor is false on two counts. First, as already seen, working memory stores information in a coarser code than what we experience in perception. Second, working memory may store information in form of procedural knowledge, rather than be re-representing features of perception. More specifically, working memory encodings can be thought of as commands for maintaining activity in perceptual systems rather than reproducing copies of perceptual states. Putting these two points together, we might imagine the following picture. When an intermediate-level perceptual representation is encoded, that means a high-level perceptual representation that captures the gist of the intermediate-level representation sends a signal to working memory systems, which, in turn, uses this representation as a set of instructions for maintaining perceptual states during a temporal delay. The high-level representation can be used to generate a mental image of the stored item, by projecting back into intermediate-level areas, but, because it is course grained, the resulting image will be indeterminate and unstable.

If this picture is right, then consciousness does not require working memory encoding. It just requires availability for encoding. Available representations can be coarse-coded and stored. The picture in borne out by various psychological studies including Sperling's (1960) widely discussed experiments with letter arrays. When subjects are briefly presented with a three by three array, they can recall three or four letters, but not more. Nevertheless, they seem to experience the entire array, and, absolutely any of the letters could be stored if it is cued after the stimulus is removed. Thus, each letter is available for encoding and consciously experienced, but only a small handful are encoded. Availability is what matters.

This raises one final objection. Availability is a dispositional property. Consciousness, on the other hand, cannot be dispositional. Having an experiential quality is an occurrent property of a perceptual state, not merely something that the state *could* do. So, if attention is availability, attention cannot be what makes perception conscious.

The slogan, attention is availability, is a bit of loose talk. It would be more accurate to say, attention is the categorical basis of availability. This is what I implied above when I introduced the proposal. Attention is a process, I said, that makes information available to working memory. It's an actual change that takes place in perceptual representations in virtue of which they become candidates for encoding.

But what is this process? What is the categorical basis for availability? Here the science is still underway, but it is possible to advance an empirically driven guess (see Prinz, forthcoming, for more discussion). To see how perceptual representations become available for working memory encoding, it is necessary to descend to the level of computational neuroscience. The mechanisms of availability cannot be adequately described using psychological vocabulary. What accounts for availability at the neuronal levels seems to be something like this. The various structures that control levels of attention seem to work by increasing activity in interneurons, which are inhibitory cells that modulate activity in pyramidal cells, which respond to the features of perceptual stimuli. By inhibiting pyramidal cells, interneurons cause them to synchronize their activity (Sohal and Huguenard, 2005). Synchrony can be measured in the axon potentials and in the local field potentials around dendrites, though it's not fully clear which, if either, form of synchrony is more important. They key thing is that neural synchrony seems to be a very good candidate for availability. Neurons that are in sync, speak as one voice and can be heard above the din (Salinas and Sejnowski, 2001). This may turn out to be mistaken, but it gives some idea of what a theory of availability should look like, and it currently enjoys some empirical support.

In this section, I've considered four objections that have lead, in turn, to helpful insights about the nature of attention. Attention is the categorical basis of availability. Attention can be controlled by different sources and can therefore be diffuse or focused, top-down or bottom-up. Attention can occur without encoding, and has a finer grain. And, at the neurocomputational level, attention may be identified with processes such as interneuron inhibition and pyramidal synchronization, which make neural signals available for downstream propagation.

The AIR theory of consciousness states that consciousness arises when and only when we have attended intermediate-level representations. This can now be understood more precisely as the view that consciousness arises when intermediate level representations become available for working memory encoding as a result of neural synchronization brought on by inhibitory interneurons that operate under multiple sources of influence. The theory is driven by empirical evidence, but, as we will now see, it is also vulnerable to empirical objections.

3. Objections

3.1 Attention is Not Necessary

I have claimed that attention is necessary and sufficient for making intermediate-level perceptual states conscious, but this claim can be challenged. Within cognitive neuroscience, it has become popular to argue that consciousness and attention are dissociable. Some of this evidence is reviewed by Koch and Tsuchiya (2006), and other evidence has emerged since that publication. Koch used to think attention and consciousness were closely linked (Crick and Koch, 1990), but not thinks they are independent mechanisms. My goal here is not to assess each line of evidence that has been given for this conclusion, but rather to consider a few representative examples. In showing how that AIR theory can handle these counter-examples, I am hoping that readers will be able to extrapolate or devise replies to others (see also Prinz, forthcoming; and De Brigard and Prinz, forthcoming). I will begin with two studies that purport to show that attention is not necessary for consciousness.

First consider a study by Li et al. (2002). Here subjects are presented with the attention-demanding task of finding a rotated T in a group of rotated Ls. While performing the task, an image is briefly presented in the periphery and subjects are asked to make a judgment about the image, such as whether it contained an animal or whether it contained a vehicle. For some of these discriminations, subjects perform exceptionally well. This leads the Koch and Tsuchiya (2006) to describe the study as evidence for conscious perception in the near absence of attention.

This counter-example is problematic in various ways. First, "near absence" is not absence. Subjects may be allocating some attention to the periphery. In fact, if they were not we have antecedent reason to think they would not be able to freely report on what they saw, because this is precisely what change blindness studies establish. Second, the flashed stimuli, which are complex, richly colored, and high contrast, may capture attention, unlike the small shapes used in inattentional blindness studies. Performance was poor for meaningless color patterns or letters, and excellent for meaningful objects, which may be especially effective as attention lures. Third, subjects had 10 hours of training, going through 12,000 trials before being tested, which may have reduced the attentional demands of the central task. Forth, we don't even know for sure that subjects are conscious of the peripheral stimuli. Success at forced choice guessing after extensive training is not necessarily a measure of conscious awareness. In sum, the research by Li et al. (2002) does not provide strong evidence for consciousness without attention

Turn now to another study purporting to show that there can be consciousness without attention. Landman et al. (2003) devised a clever

experiment that combines change blindness with Sperling's array paradigm. Subjects see an array of rectangles, which are either horizontal or vertical, followed by a gray screen, and then a second array. In the second array, one rectangle has changed its orientation, but subjects are usually incapable of detecting the change. However, on some trials, subjects see a cue during the gray screen pointing to where a particular rectangle has been located in the original array. When this is done, subjects can reliably report whether than rectangle changed when the second array is presented. Thus subjects have the potential to detect every change if attention is directed. The crucial finding is that the attention cue can come after the stimulus. The authors interpret this as showing that the stimulus has been consciously perceived when the display was originally presented, but unattended (see also, Lamme, 2003; Block, 2007). This interpretation rests on two conditional assumptions. If the stimulus had been attended, it would have been reportable. If it had been unconscious, the presentation of a post-display cue shouldn't be effective.

Both of these conditional assumptions can be questioned. First, consider the claim that unconscious stimuli cannot be cued after they are removed. Sperling's original study proved that visual stimuli produce iconic memories: rapidly fading traces in the visual system. For all we know from this research, iconic memories can be produced by unconsciously presented stimuli. If so, a cue presented during the period in which an unconsciously induced iconic memory is fading may serve to bring that iconic trace into consciousness. On the AIR theory, this would be readily explained. The cue brings attention to a visual trace, and the trace becomes conscious thereby. In the Landman et al. study, this may be exactly what takes place. The rectangles used in their displays are depicted with a noise gradient against a noisy background, and it's far from obvious that every rectangle is consciously perceived.

But suppose these rectangles in the original display are consciously perceived. There is still a possibility that the second conditional assumption is false. To argue that this study illustrates consciousness without attention, Landman et al. must say that, if a rectangle had been attended, it would have been reportable. But I reject that assumption. Reportability arises when attended stimuli are encoded in working memory, but I have already argued that attention outstrips working memory. So Landman et al. do not provide evidence for thinking their stimuli are not attended. It's quite plausible that subject try to attend to the whole display, since they know they will be tested on it. And the allocation of diffuse attention may bring each rectangle into consciousness. If so, the study established only that consciousness can arise without encoding, which is consistent with the AIR theory. In summary, Landman et al. do not establish consciousness without attention because they neither establish that the stimuli in question are conscious nor that they are unattended.

These studies are illustrative of the best recent attempts to show that consciousness can arise without attention, and they have been emphasized by leading defenders of that dissociation. But the studies fail to establish any such thing.

3.2 Attention is Not Sufficient

Granting that attention is necessary consciousness, critics may still object that it is not sufficient. Recent experiments have been constructed to establish that attention can occur in the absence of consciousness. These experiments pose a threat to the AIR theory. Fortunately for AIR, they don't provide compelling evidence for what they seek to establish.

Let's begin with a study by Kentridge et al. (2008), which extends earlier work by the first author. The study combines metacontrast masking with attentional cueing. Subjects see colored disks followed by colored rings, and the rings mask the disks, resulting in unconscious perception. Their task is to detect the ring as quickly as possible. On some trials the disk is preceded by an arrow that serves a lure for attention. The arrow does not bring the disk into consciousness, but it does have a significant affect: if the disk is the same color as the ring, it facilitates ring detection, but this happen only when the disk is preceded by an arrow. Thus the arrow seems to give this unconscious stimulus the power to exert priming influence. The authors interpret this as showing that the unconscious stimulus has been enhanced by attention. Thus, it seems to be a case of attention without consciousness.

As compelling as this may seem, it is not a counter-example to the AIR theory. The AIR theory says that consciousness arises when there is an attended intermediate-level perceptual representation, but the experience does not establish any such thing. Alternative interpretations are available. The arrow may have a number of different effects that are sufficient for explaining task performance. I will just mention one. The arrow may result in saccadic eye movements to the region in which the disk will be presented, and that could result in a more accurate representation of the disk. Foveal vision has a higher concentration of color receptive cells, and results in more saturated representations. If the cue leads to a more saturated representation, that could explain why color-based priming is found in the cued condition.

This interpretation is not *ad hoc*, because it appeals to a known process, but Kentridge et al. may still object that I have not given evidence against their interpretation. Showing that other processes can explain the effect does not rule out unconscious attention. But I do think there is a reason to think attention is not responsible. The disk is presented very briefly and followed by a ring in the surrounding area. According to the leading interpretation of metacontrast masking, the ring is able to mask the disk precisely because it draws attention away from the region of space in which the disk is located (Enns and Dilollo, 2000). Attention does not have time to enhance the representation of disk, and it cannot enhance the iconic trace of the disk because the ring draws attention away. Thus, even if my saccading explanation turns out not to be responsible for the effect, I seriously doubt that attention is doing the work.

Turn now to a second study that attempts to establish attention in the absence of consciousness. Jiang et al. (2006) use a technique called interoccular suppression in which a low-contrast stimulus presented to one eye is masked by a higher contrast stimulus presented to the other eye. Cleverly, Jiang et al. used nude photographs as the masked stimuli assuming that these would attract attention,

despite the fact that they could not be consciously perceived. Sure enough, when the interoccular displays were taken away, subjects were asked to detect a target and target detection was superior in the location that had been occupied by the nude. The effect worked only when the nude was a member of the subject's preferred sex. The study has two features that make it a very powerful response to the AIR theory. First, the method allows the stimulus to be presented for an extended duration, which means attention has enough time to act on the stimulus representation. Second, the stimulus itself serves as the attention cue, so it seems especially plausible that the representation of the stimulus is modulated by attention. This looks like a case of an unconscious AIR.

But, once again, an alternative interpretation is available. Perhaps the nude attracts saccades and not attention. Or, it may even attract spatial attention, without attracting attentional modulation of the represented object. Spatial attention is attentional enhancement of a region of space, rather than the objects that occupy that space. Either or both of these effects could facilitate target detection when the target is displayed in the region that was occupied by the nude.

As above, it is one thing to present an alternative explanation, and another to refute the explanation offered by the authors. What reason do I have to reject the interpretation given by Jiang et al., according to which attention is being allocated to the unconscious stimulus? The answer is simple. In interoccular suppression the rival stimulus is an attention lure. Suppression may result from the fact that the high contrast stimulus is able to attract attention away from the low contrast stimulus. Some authors argue that binocular rivalry, of which interoccular suppression is a special case, involves the same mechanisms as selective attention (Mitchell et al., 2004). Moreover, if the nude were being attentionally enhanced, we might expect to see increased activation in the ventral stream, where object representations are processed. Jiang et al. did not measure brain response, but fMRI studies of interoccular suppression suggest that suppressed stimuli are not in fact associated with increased ventral processing (Fang and He 2005). The increases are observed in the dorsal stream, which plays a role in saccadic eye movements and spatial perception. Thus, the existing evidence offers better support for my interpretation than for the interpretation given by Jiang et al.

The two studies I have been discussing fail to establish that there can be unconscious AIRs. I think these studies are the best efforts to establish that attention is insufficient for consciousness, and, if I am right, they do not succeed. So I conclude that current empirical evidence provides strong support for the claim that attention is necessary and sufficient for making perceptual representations conscious, and no convincing evidence to reject that claim.

3.3 Alternative Interpretations of the Evidence

I might rest my case here, but there is one more challenge that deserves attention. This one was put to me forcefully in a commentary that David Chalmers delivered in response to a related paper of mine (Chalmers, 2008). I will not do justice to all the moves that Chalmers makes, but I will address the main thrust of his critique.

Chalmers attacks the evidence offered in support of the view that attention is necessary for consciousness.

I offered two main sources of evidence for the conclusion that consciousness requires attention. Healthy subjects fail to report stimuli when attention is divided (inattentional blindness) and damage to attention centers leads to neglect. Chalmers argues that both sources of evidence are open to an alternative interpretation. Subjects may have conscious experiences that they simply fail to report. In neglect, subjects may be incapable of reporting stimuli in their blind fields precisely because they cannot attend. Perhaps attention is required for reporting, but not for experience itself. If so, neglect provides no evidence for the conclusion that consciousness requires attention. It discussing inattentional blindness, Chalmers notes that there is an alternative interpretation of these studies: subject may suffer from inattentional amnesia (Wolfe, 1999). They may consciously experience the surprise stimuli and instantly forget them. We know that visual memory isn't very good, so the subjects' retrospective reports are not necessarily reliable.

Let me reply to these two alternative interpretations in turn. First consider neglect. The evidence that people with neglect do not experience things on the left is not restricted to self-report. They fail to consciously see objects presented on the left or, in some cases, the left half of objects presented centrally. Neglect can be demonstrated in many ways. When asked to copy line drawings, features on the left are ignored; when asked to form mental images of familiar places, landmarks on the left are forgotten; when asked to bisect lines, the midpoint is placed too far to the right; when asked to find target letters in a group of distractors, targets on the left are missed. Thus, they fail to demonstrate consciousness on both implicit and explicit measures. In the absence of positive reason to think these individuals are conscious, I think this widespread constellation of behaviors should be taken as evidence that they are not.

What about inattentional blindness? Chalmers says this might better be regarded as inattentional amnesia, but I see little reason to explain the results that way. Mack and Rock (1998) consider this possibility and devise a clever experiment to rule it out. If unattended stimuli were consciously seen, then they should able to integrate with other conscious stimuli presented consecutively and proximately to produce illusory motion. But they exert no such influence. Chalmers replies that motion perception involves binding, and binding is associated with attention. Thus, the absence of illusory motion is explained by the absence of attention, not the absence of consciousness. But this reply only serves to support the AIR theory. Conscious experiences are characteristically bound, and if attention is required for binding, that is another powerful reason for thinking that attention coincides with consciousness.

Moreover, there are at least three more reasons to reject the inattentional amnesia story. First, the main argument for taking the story seriously is that visual memory doesn't last very long, and Mack and Rock's stimuli are quickly presented. But some inattentional blindness studies using longer durations. The most dramatic case in in a study by Most et al. (2005). They have subjects count how many times animated black or white letters bounce against the side of a computer screen. While

this is going on a red cross travels across the center of screen, taking a full five seconds to reach the other side. The cross differs from the other letters in color, luminance, shape, and trajectory, but it is still missed by many subjects. Why would subjects *forget* a stimulus that is so different from everything else and present for such a long time. The inattentional amnesia account simply makes no sense here.

Second, inattentional amnesia fails to explain the phenomenal difference between Sperling cases, on the one hand, and Mack and Rock's studies. In the Sperling studies, subjects are sure they have seen a letter array, even if they can't make out the letters. In the Mack and Rock studies, subjects have no idea there has been a surprise stimulus. They would be at chance if they had to guess whether something was presented. The Sperling seem like a situation where something is seen and forgotten from the subjects' point of view. The Mack and Rock cases do not seem this way to the subjects. That difference cries out for an explanation. If both cases were just forms of amnesia, they should seem similar to subjects, but they do not. The best explanation, I submit, is that the inattentional case results in blindness, where as the Sperling case, where attention is available, does not.

Third, inattentional amnesia is difficult to square with the other studies that I mentioned at the outset. In the attentional blind paradigm and load-induced blindness, subjects are told to search for a target. They are strongly motivated to recall when the target has been seen. Despite this, targets go undetected. Again, we can ask, why would subjects who are looking for a target forget that they saw it?

What I am offering here is an argument to the best explanation. It is always possible that people with neglect or subjects in psychological studies are having conscious experiences that they cannot demonstrate using subjective or objective measures. I think this possibility is a bit like skeptical hypotheses. We cannot refute it, but we should not take it very seriously. We should not postulate conscious experience in cases where we have no strong evidence that there is experience. The standard interpretation of neglect and inattentional blindness studies are preferable to the alternatives that Chalmers would have us consider.

4. Conclusion

It has been known for a long time that perception can occur without consciousness. In this chapter, I've tried to identify that factor that makes a difference. Philosophical attempts to distinguish conscious and unconscious perception have often hinged on armchair arguments that do not withstand empirical scrutiny. I tried to illustrate that with my discussion of the HOR theory. In its place, I suggested that consciousness arises when and only when intermediate-level representations are modulated by attention—what I call the AIR theory. Attention is the difference maker. This conclusion is based on empirical research, but has also come under recent empirical attack. I deflected what I take to be the most powerful objections. Perhaps some further finding will force a revision of the view, but, for now, it looks like we have found the key to consciousness. Attend and you will see.

References

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

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

Chalmers, D. (2008). Is there Consciousness Outside Attention?: Comments on Jesse Prinz. Address given at the annual SPAWN conference, University of Syracuse, August 2008.

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

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

De Gelder, B., Morris, J. S., and Dolan, R. J. (2005). Unconscious fear influences emotional awareness of faces and voices. *Proceedings of the National Academy of Sciences*, 102, 18682-18687.

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

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

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

Hamm, A. O., Weike, A. I., Schupp, H. T., Treig, T., Dressel, A., and Kessler, C. (2003). Affective blindsight: intact fear conditioning to a visual cue in a cortically blind patient. *Brain*, 126, 267-275.

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.

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

Koch, C., and Braun, J. (1996). Towards a neuronal correlate of visual awareness. *Current Opinion in Neurobiology*, 6, 158–164.

Koch, C., and Tsuchiya, N. (2007). Attention and consciousness: two different processes. *Trends Cog. Sci.* 11, 16-22.

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

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

Li, F. F., VanRullen, R., Koch, C., and Perona, P. (2002). Rapid natural scene categorization in the near absence of attention. *Proceedings of the National Academy of Sciences*, 99, 8378 - 8383, 2002

Lycan, W. G. (2001). A Simple Argument for a Higher-Order Representation Theory of Consciousness. *Analysis*, 61, 3-4.

M. Bar (2000). Conscious and non-conscious processing of visual object identity. In Y. Rosetti and A. Revonsuo (Eds.) *Dissociations: Interaction between dissociable conscious and nonconscious processing* (pp. 153-174). Amsterdam: John Benjamins.

Macdonald J. and Lavie N. (2008). Load induced blindness. *Journal of Experimental Psychology: Human Perception and performance.* 34, 1078-1091.

Mack, A. and Rock, I. (1998). *Innatentional Blindness*. MIT Press.

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

Marr, D. (1982). *Vision*. San Francisco, CA: Freeman.

Marshall, J. C., and Halligan, P. W. (1988). Blindsight and insight in visio-spatial neglect. *Nature,* 336, 766 – 767.

Mitchell, J. F., Stoner, G. R., and Reynolds, J. H. (2004). Object-based attention determines dominance in binocular rivalry. *Nature*, 429, 410–413.

Most, S. B., Scholl, B. J., Clifford, E., and Simons, D. J. (2005). What you see is what you set: Sustained inattentional blindness and the capture of awareness. *Psychological Review, 112*, 217-242.

Prinz, J. J. (2000). A neorofunctional theory of visual consciousness. *Consciousness and Cognition*, 9, 243-59.

Prinz, J. J. (2005). A neurofunctional theory of consciousness. In A. Brook and K. Akins (Eds.), *Cognition and the brain: Philosophy and neuroscience movement* (pp. 381-396). Cambridge: Cambridge University Press.

Prinz, J. J. (forthcoming). *The conscious brain*. New York, NY: Oxford University Press.

Raymond, J. E, Shapiro, K. L., Arnell, K. M. (1992). Temporary suppression of visual processing in an RSVP task: an attentional blink? *Journal of experimental psychology. Human perception and performance*, 18, 849–60.

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

Rock, I., and 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 N. Block, O. Flanagan, and G Güzeldere, (Eds.), *The Nature of Consciousness: Philosophical Debates* (pp. 729-753). Cambridge, MA: MIT Press.

Salinas, E., and Sejnowski, T. J. (2001). Correlated Neuronal Activity and the Flow of Neural Information. *Nature Reviews Neuroscience*, 2, 539-550.

Simons, D. J., and Levin, D. T. (1997). Change blindness. *Trends in Cognitive Science*, 1, 261-267.

Sohal, V.S and Huguenard, J.R. (2005) Inhibitory coupling specifically generates emergent gamma oscillations in diverse cell types. *Proceedings of The National Academy of Science*, 102, 18638-18643.

Sperling, G. (1960). The Information Available in Brief Visual Presentations. *Psychological Monographs,* 74.

Weiskrantz, L. (1986). *Blindsight: A case study and implications*. Oxford: Oxford University Press.

Winkielman, P., Berridge, K. C., and 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.

Zeki, S. (1993). *A vision of the brain*. Oxford: Blackwell.