INTERNAL AND EXTERNAL AWARENESS

Wesley Sauret

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Philosophy in the College of Arts and Sciences.

Chapel Hill
2014

Approved by:

William Lycan

David Rosenthal

Ned Block

Ram Neta

Matt Kotzen

ABSTRACT

WESLEY SAURET: Internal and External Awareness.
(Under the direction of William Lycan)

In our day-to-day lives we are aware of many different things, from the taste of the coffee we just drank to our own hopes and dreams. The central aim of my project is to understand such awareness. I begin by considering awareness of external items, and argue that representation of an external object is necessary, but not sufficient, for awareness of that object. It is commonly supposed that in order to get a sufficient condition for awareness, the representation of the object must be cognitively accessed by the subject. I give two arguments against this view. The 'content mismatch' argument and the 'capacity mismatch' argument. I argue that cognitive accessibility can allow for both a content and capacity match with external awareness, and develop a plausible neural mechanism underlying cognitive accessibility.

I then extend this theory to account for awareness of mental states. I argue for a nonrepresentational theory where the difference between internal and external awareness depends on what is attended. Finally, I develop a theory of visual attention. I argue that attention is a tool as it is used by subjects to do things. In particular, I argue that attention functions similarly to a highlighter in that they both have the function of making the selected target stand out from the surrounding items. I then ground this theory of attention in neural mechanisms distributed throughout the brain.

ACKNOWLEDGMENTS

I am indebted to many people over the course of this project. First and foremost, I would like to thank William Lycan and Ram Neta for all the years they spent mentoring me. Next, I am very grateful to Ned Block and David Rosenthal for their feedback on my work. I would also like to thank Nate Sharadin for many discussions over the years and for his help in formatting this document. Finally, I would like to thank Crystal Waters for her support through thick and thin.

TABLE OF CONTENTS

1 TRANSITIVE CONSCIOUSNESS

1.1 Introduction

My goal in this introductory chapter is to begin to clear the ground for the arguments of the subsequent chapters by setting out what I take to be the different varieties of consciousness. At a first approximation, we can get a grip on the phenomenon of consciousness by considering cases where a subject is *aware* of an object. For example, suppose a subject sees an eggplant in clear view and thereby becomes aware of the eggplant. In virtue of being aware of the eggplant, the subject is *conscious* of it. Denying this link between consciousness and awareness creates a kind of conceptual tension. It seems incoherent to insist that a subject is aware of the eggplant, while completely unconscious of it (or vice versa). Because of this tight connection between 'consciousness' and 'awareness', I will be treating them as synonyms (see also: Dretske 1993; Rosenthal 1997). In subsequent chapters I will formulate my claims almost entirely in terms of 'awareness' in order to avoid the various hangups that philosophers of mind have about the proper use of the term 'consciousness'. But I intend all of the claims made about the nature of awareness to equally apply to the nature of consciousness.

Both 'consciousness' and 'awareness' can be used transitively and intransitively. The transitive use asserts an awareness relation between a subject of awareness and an object of awareness, e.g., "*Sally* is conscious of *the eggplant*".[1] Intransitive uses can be applied

[1]Despite the important role played by the notion of a 'subject of awareness', it will remain relatively undeveloped. I take it that the minimum requirements for being a subject of awareness are relatively substantive. I do not think that a subject of awareness necessarily needs to be a person, but they do need to be a creature with roughly comparable capacities. It is necessary - but probably not sufficient - for a subject of awareness to have the ability to both perceive the environment and engage in mental operations on the basis of those perceptions.

to a subject or a mental state within them, e.g., "Sally is conscious" or "Sally's belief that the eggplant is purple is conscious". Rosenthal (2005) calls applications to subjects "creature consciousness" and applications to states within them "state consciousness". But this distinction only applies to intransitive forms of consciousness. Transitive consciousness cannot be predicated of a state within a subject without absurdity; e.g., it makes no sense to say that "Sally's visual representation is conscious of the eggplant". Only subjects *as a whole* can be transitively conscious of things. Of course, whenever a subject is conscious of an object they are conscious of that object *in virtue* of being in a relevant mental state. It is just that we ought not consider such states as instances of "transitive consciousness" but rather as the necessary grounds for the subject's being transitively conscious. In the Chapters 3 & 4 we will conduct an in-depth examination of the features that a mental state must have in order to ground transitive consciousness. For now the key point is that these states are not instances of transitive consciousness in and of themselves.

Given the transitive and intransitive uses of 'consciousness', it would be helpful if we could determine the relation between them. Are they fully independent notions, or is one of them to be explained in terms of the other? Since our answer to this question may differ depending on whether we are talking about state consciousness or creature consciousness, we will consider each in turn. For state consciousness, the dominant position in the literature is that intransitive state consciousness is to be explained in terms of transitive (creature) consciousness. Since transitive consciousness is always consciousness of something, what must one be conscious of in order for a state to be intransitively conscious? There are at least two different views. Many have argued that intransitive consciousness applies to a state iff the subject is conscious of the state in question (Armstrong 1980; Lycan 1996; Rosenthal 2005). On this view, an intransitively conscious state is a state that the subject is conscious of. Others have argued that a state is intransitively conscious iff the subject is conscious of what that state is about (Dretske 1993). On this view, a state is intransitively conscious not when a subject is conscious of that state, but rather when a subject is

conscious of what that state is about *in virtue of being in that state*. Although these views nominate substantially different objects of awareness, I think that each captures part of our ordinary usage of the intransitive form of 'consciousness'. However, all that matters for present purposes is that both views agree that intransitive state consciousness is best understood in terms of transitive consciousness. I take it that something along these lines is correct and that intransitive state consciousness is to be grounded in a form of transitive consciousness. Can we extend this thesis to cover cases of intransitive creature consciousness as well?

This is not as easy as one might have hoped because intransitive creature consciousness is used in two distinct ways. The most common way it is used is to mark an occurrent property of a subject, e.g., Sally is currently conscious, but was not when she was under anesthesia. But it is also used to mark a general capacity of a *type* of creature, e.g., humans are conscious but amoeba are not. If we want to ground intransitive creature consciousness in the same way that we grounded intransitive state consciousness, then we will need slightly different specifications for each usage. If we are working with the occurrent usage, then the view is that a subject is intransitively conscious iff they are transitively conscious of at least one thing. If we are working with the capacity usage, then the view is that a subject is intransitively conscious iff they are capable of being transitively conscious of things at least some of the time. I think that in either case, there is something intuitive about the claim that intransitive creature consciousness depends on transitive consciousness.

Part of this is a result of the fact that transitive consciousness is clearly *sufficient* for intransitive creature consciousness. Suppose that Sally has to have some dental work done and as a result is being administered anaesthesia. How do we determine when she has been given enough of the drug to render her unconscious? Typically we will keep administering the drug until she fails to pass a series of tests. Interestingly all of the tests used in such situations involve presenting stimuli to her and determining whether she reacts. That is, in order to succeed on any of the tests she must be transitively conscious of the stimuli that

was presented. Given the adequacy of such tests of determining whether or not a subject is intransitively conscious, it seems clear that, at the very least, being transitively conscious of at least one thing is sufficient for being (occurrently) intransitively conscious.

However, the issue we are really concerned with here is whether or not transitive consciousness *necessary* for intransitive creature consciousness. What considerations can be given in favor of this claim? To begin with, it should be noted that the intransitive use of 'consciousness' is relatively new: its use can be traced back only about two hundred years (Humphrey 1992). While clearly not a decisive consideration, I do think that it is suggestive. But, in order to give a more decisive argument for (or against) the necessity claim, we would need a method for detecting the presence of intransitive creature conscious *independently* of transitive consciousness. Unfortunately there is no theory neutral way of doing this. There are currently two kinds of methods for detecting the presence of intransitive creature consciousness. First, we can test for intransitive consciousness by testing for the presence of transitive consciousness. Second, we can assume that some background theory is true and ask whether that theory suggests intransitive consciousness would be present in the circumstances in question. Since neither method is a theory neutral method for detecting intransitive consciousness independently of transitive consciousness, arguing for (or against) the necessity claim will be difficult.

Of course, that has not stopped people from trying to come up with counterexamples to it. Fred Dretske (1993) suggests that cases of hallucination may be counterexamples to the claim that the occurrent intransitive consciousness can only occur when a subject is conscious of at least one thing. We are asked to suppose that a subject has a visual hallucination as of a purple eggplant on the cutting board where there is, in fact, no such eggplant. Then it is suggested that in this kind of case the subject is occurrently intransitively conscious. And I agree, it does seem like the subject is intransitively conscious in such cases. But one might worry that in such cases the subject is not conscious of anything at all.[2] After

[2]In order to make this objection even stronger, I will be supposing that the subject of the hallucination

all, hallucinations just are the kinds of cases where a subject *seems* to experience an object, but where *no physical object is present*. Thus, "to suppose that hallucination (involving intransitive consciousness) is a consciousness of something would (or so it is feared) commit one to objectionable mental particulars - the sense data that one hallucinates" (1993, p.269). The concern here is an old one. If we insist that the subject is conscious of something in cases of hallucination and we agree that there is no *physical* object present, then we might worry that we are thereby forced to accept that the object is *non-physical*. But I am confident that we can give a satisfactory account of the object of hallucination, without committing ourselves to "objectionable mental particulars".

I will briefly sketch one such view. Perceptual states are representational states.[3] But, a necessary condition for representation is the possibility of misrepresentation. That is, it must be possible for any representation to be about an object that does not exist.[4] Given this, it seems that the right thing to say is that the difference between a veridical perception of an eggplant and an indistinguishable hallucination of one is that in the former case the object of perception exists, while in the latter case it does not. In both cases the subject has an occurrent visual representation that represents an object to be purple, oval, shiny, etc. and the subject is aware of the eggplant in virtue of these represented properties. The two states only differ insofar as the eggplant exists in one case, but not the other. This difference only affects the *success* of the state, not whether it is transitive.

Another purported counterexample comes from the practice of meditation. Some have suggested that there are certain kinds of meditative states where subjects are intransitively conscious, but they are not transitively conscious of anything at all. They are in a state of

---

is not having any other experiences whatsoever. Perhaps all of their other senses have been disabled. Such restrictions are to ensure that the hallucinatory visual experience is the only state available to ground their being intransitively conscious.

[3]This is one of the fundamental assumptions of my entire project. I say a bit more about it in Chapter 3.

[4]Given this fundamental fact about representation, we can begin to see the absurdity of the argument above. If it were correct, every time a genuine representation misrepresented something, there would be a non-physical object that the representation was actually representing even when that representation was not even a *mental* representation.

so-called "pure consciousness". My understanding is that most meditative states are best understood as states where a subject is instructed to inhibit all cognitive activity and to make one's awareness of 'pure sensation' as transient as possible. One is to 'acknowledge' the object of awareness and move on (Sekida 1975). Such cases pose no problem for my view. Although the goal of meditation is to make one's awareness of any particular object as transient as possible, I am skeptical that there exists a kind of meditative state that is so pure as to make one's sensory awareness of objects so transient that there is no remaining awareness of any sensation whatsoever.[5]

I will take it that my responses are sufficient to undermine these purported counterexamples. Given the aforementioned methodological difficulties, I propose that we adopt the most *parsimonious* view. Since we have already adopted the view that transitive consciousness is necessary for intransitive state consciousness, the most parsimonious view would be one that says that transitive consciousness is necessary for all types of intransitive consciousness. This suggests, then, that transitive consciousness is the fundamental form of consciousness. If this is right, then we need to take a closer look at the nature of transitive consciousness. In the remainder of this chapter I will aim to set out the four main varieties of transitive consciousness. These four varieties of consciousness arise from two distinctions that cut across each other. The first distinction marks off the two basic modes of awareness - propositional awareness and nonpropositional awareness - and the second distinction marks off the two different kinds of objects that one can be aware of - internal mental states and external physical objects.

1.2 Varieties of Transitive Consciousness

We will begin with the distinction between propositional and nonpropositional awareness. This distinction maps straightforwardly onto Dretske's (1993) distinction between

[5]Of course, I do grant that there are meditative states that involve no awareness of facts or any cognitive activity with respect the the objects one is momentarily aware of. Thus, some meditative states can be "pure" in an important sense of the term.

consciousness of facts and consciousness of things. Facts are the truth-makers of propositions; they are what true propositions are about. Since facts must be represented propositionally, consciousness of a fact must be grounded in a propositionally structured representation. It is only in virtue of being in a suitable propositional state that a subject is conscious of a fact. According to most contemporary theories of concepts, concepts are the constituents of propositionally structured representations (i.e. thoughts). As a result, propositionally structured representations must have conceptual content.[6] As I understand propositional awareness it just is consciousness of facts so described. We generally describe propositional awareness using the phrase "aware that p", where 'p' is a proposition, e.g., "the eggplant is purple". According to Dretske, 'things' are worldly particulars, like objects and events. Since worldly particulars can be represented in nonpropositional (e.g. iconic) formats, we can be conscious of things in the absence of propositionally structured representations of those things.[7] All that is required for consciousness of a thing is a suitable nonpropositional state about it. As I understand nonpropositional awareness it just is consciousness of things, so described. We generally describe nonpropositional awareness using the phrase "aware of n", where 'n' is a noun phrase, e.g., "the eggplant".

Since one can be conscious of a thing without a propositionally structured representation, consciousness of things is a nonconceptual mode of awareness. This means that a creature can be conscious of a thing without deploying any concepts (or, perhaps, without even having concepts in the first place).[8] The possibility of nonpropositional awareness in the absence of concept possession suggests that nonpropositional awareness is the *phylogenetically basic* form of awareness. Propositional awareness of facts, on the other hand,

[6]For example, the content of the belief that 'the eggplant is purple' consists of (at least) two concepts, e.g., 'eggplant' and 'purple', and these concepts are related by the relation of predication; 'purple' is being predicated of the thing picked out by 'eggplant'.

[7]For more discussion of this point, see: Burge, 2010, pp.537-544.

[8]This is plausible because consciousness of things typically requires that you consciously *perceive* that thing and perceiving something is commonly thought to not require the deployment of concepts. I am not identifying consciousness of things with conscious perception in order to leave open the possibility that a subject can be aware of a thing in virtue of nonconceptual representational formats that are non-perceptual.

requires a certain level of cognitive sophistication. In order to be aware of a fact you must have concepts. If I am aware that the eggplant is purple, then I must have the concept 'purple' and the concept 'eggplant' and apply them to my nonpropositional awareness of the eggplant. If I didn't understand what purple was, then I could not be aware *that* the eggplant is purple even if I could be aware *of* the purple eggplant.

In the process of setting up his own theory of consciousness, Prinz (2012) criticizes the distinction Dretske draws between fact-consciousness and thing-consciousness. Prinz acknowledges that there is an important difference between, e.g., being aware of an armadillo and being aware that it is an armadillo. However, Prinz does not think that we have any reason to suppose that this difference is grounded in a difference in the *variety* of consciousness as opposed to a mere difference in the possible contents of consciousness. It might be that there is only one fundamental variety of consciousness with many different kinds of contents that one can be conscious of.

I agree with the *metaphysical* point being made here. We have no reason to believe that fact-consciousness and thing-consciousness involve substantively different *awareness* relations between the subject and the object of awareness. It might turn out that it is the same relation in each case, just grounded in different kinds of representations with different contents. But, because of the important differences in the nature of these two kinds of representations, I find it helpful to *conceptually* separate awareness relations grounded in propositional representations (i.e. fact-consciousness) from awareness relations grounded in nonpropositional representations (i.e. thing-consciousness). What are these differences?

First, nonpropositional representations typically represent their objects in a map-like way where there is a correspondence between the geometrical or topographic structure of the representation and the structure of the represented object. Propositional representations represent their objects via a correspondence between the syntactical arrangement of the constituent concepts and the syntax of the represented proposition. These two ways of representing are different enough to be construed as being different representational formats

(Cummins 2010). Second, the objects of the two kinds of representations are importantly different. Nonpropositional representations represent particular things, while propositional representations represent propositions. Given these differences, I think we are justified in conceptually separating propositional awareness from nonpropositional awareness.

We will now turn to the distinction between internal and external awareness. This distinction roughly corresponds to the intuitive difference between introspective awareness and perceptual awareness. That is, we can divide the things of which one can be aware into two basic categories: things inside one's mind and things outside it. These objects are importantly different: the ones inside our mind are mental objects, while the ones outside our mind are (typically) physical objects. Given the important differences in the nature of these two types of objects, I think we should institute a conceptual division between cases where a subject is aware of a mental state and cases where a subject is aware of an external object. I will call cases where one is conscious of an external object 'external awareness' and cases where one is conscious of a mental state 'internal awareness'. This distinction exists solely to mark the difference in the ontological category of the object of awareness in each case.

Prinz (2012) criticizes a similar distinction developed by David Armstrong, saying that "the fact that we can be conscious of inner and outer states does not entail that there are two species of consciousness" (p.6). Again, I agree with the metaphysical point that the awareness relations involved are unlikely to be substantially different in each case. However, conceptually separating these two kinds of awareness enables us to leave open the possibility that there are important differences in the necessary conditions for awareness of each type of object. I also think it helps clear up some confusion present in the literature on consciousness. For instance, the higher-order theories of consciousness developed by Armstrong (1980), Lycan (1996), and Rosenthal (2005) are paradigmatic theories of what I call 'internal awareness'. Their goal is to explain the cases where a subject is aware

of an internal *mental* object, which they call a 'conscious state'. On the other hand, so-called 'first-order theories' like those developed by Dretske (1995), Tye (2000), Block (2007), and Prinz (2012) are predominantly theories of external awareness. Their goal is to explain cases where a subject is aware of an external *physical* object. To the extent that they have different explananda, they are not true competitors. While some of the authors listed have gone on to extend their theories to account for the other variety of awareness, we should evaluate such extensions on their own terms and not miss the fact that their original explananda are different. I take it, then, that it will be helpful to keep the two cases separate.

1.3 Conclusion

In conclusion, I think that transitive consciousness is the fundamental variety of consciousness. There are four distinct varieties of transitive consciousness generated by two conceptual distinctions. The first distinguishes transitive consciousness of external objects from transitive consciousness of internal objects. The second distinguishes propositional awareness from nonpropositional awareness. This gives us an exhaustive characterization of the varieties of transitive consciousness: propositional awareness of an external object, nonpropositional awareness of an external object, propositional awareness of an internal mental state, and nonpropositional awareness of an internal mental state.

2 PHENOMENAL CONSCIOUSNESS & ACCESS CONSCIOUSNESS

2.1 Introduction

Now that we have distinguished the four varieties of transitive consciousness, my goal in this chapter is to determine how these varieties of consciousness are related to Ned Block's (1995) distinction between phenomenal consciousness and access consciousness. According to Block, phenomenal consciousness (or P-consciousness) is experience, while access consciousness (or A-consciousness) is a functional role. When you are phenomenally conscious of a content there is something it is like for you to be aware of that content. When you are access conscious of a content it is in a position to be noticed, recognized, reasoned about, and reported by the subject. I will argue that Block's distinction between phenomenal consciousness and access consciousness is best understood as collapsing into the distinction between nonpropositional awareness and propositional awareness, respectively.

2.2 Access Consciousness

Over the years Block has developed two different notions of access consciousness. Early on he defined access consciousness in terms of content being "poised": A representation is access conscious if its content is "poised for free use in reasoning or in rational control of action" (p.238). For Block, a state is poised if it is "ready and waiting". For example, to be poised for attack is to be on the *verge* of attacking (Fn 7, p.245). Thus, in order to be poised for use a state must be *immediately* usable by the mechanisms of reasoning and rational control of action without any further processing. The other restriction is that the representation must be rationally guiding ones behavior. The main argument for restricting

the kind of control involved in access consciousness to instances of rational guidance is that it allows us to rule out certain problematic cases. For example, in cases of blindsight information about the stimulus in the blind field can influence the guesses of the subject. But this kind of influence is not of the right sort to be counted as conscious. It is generally thought that one of the features missing in cases of blindsight is that the information about the stimulus is not available for *rational guidance* of action and speech.

I think that Block is right that the distinctive necessary condition for access consciousness is being immediately usable by cognitive faculties like reasoning and planning. However, several commentators on Block (1995) worried that access consciousness, so described, is not *really* a form of consciousness per se. For example, Natsoulas (1995) worries that "what Block calls 'access-consciousness' amounts to no more than a readiness - a certain representation's being 'poised' - for use in information processing, that is, a readiness that we have no reason to describe as a case of consciousness" (p.264). I take it that a desideratum on a theory of access *consciousness* is the availability of a clear explanation for why the kind of 'readiness' described should count a form of consciousness. How, then, should we understand access consciousness?

Since content must be 'immediately usable' by one's cognitive faculties in order to be access conscious, we can start by noticing that this means that content that is access conscious must be encoded in the working memory. Why is that? Well, working memory functions as the 'hub' of the cognitive system. It is only after being encoded in working memory that propositional states become usable by one's cognitive faculties.[1] Since encoding a representation's content in working memory is a necessary condition for cognitive access to that content, it seems likely that the kind of access involved in access consciousness is 'cognitive access'. It will be instructive, then, to investigate what is involved in having

[1]I think that there is a strong analogy between the way that working memory functions and the way that network hubs function. For example, the hubs in a computer network have the function of efficiently routing information from the source computer to its destination, and the hub cities in the air travel network in the United States function to efficiently route planes to their destinations. Similarly, working memory functions to centralize and then distribute propositional representations to the cognitive faculties.

cognitive access to a content. We will start by considering the following question: does the kind of cognitive access involved in access consciousness require that all the content successfully accessed is conceptual?

As above, I take it that the content of a representation is conceptual if it is represented in a propositionally structured representational format and that the content of a representation is nonconceptual if it is not represented in such a format. Given this, I will argue that *all full-fledged instances of cognitive access require the state to have conceptual content*. I think that much of the plausibility of this claim is comes from the fact that a mental state can only be used as a premise in reasoning if its content is conceptually structured. We would need a radically different kind of model for reasoning and inference if we wanted to allow the use of nonconceptual content to be used as a premise in reasoning (for a review of contemporary models of reasoning, see: Evans 2008). This claim is also derives some of its plausibility from the fact that a mental state cannot be used for rational control of action and speech unless it has conceptual content. Of course, I agree with those who think that nonconceptual content plays an important role in guiding our actions in ordinary cases. I just think that this way of guiding action is not properly understood as an instance of "rational control".[2] Given this, I will be assuming that *concepts are the currency in which cognitive transactions take place*, where 'cognitive transactions' are the processes via which we acquire, store, and use propositional states like belief and knowledge.

Recall that the kind of cognitive access involved in access consciousness makes a state poised for immediate use by the cognitive faculties, such as reasoning and planning. But content is only immediately (or "freely") usable if it can be used without any intervening steps. Nonconceptual content, ex hypothesi, cannot be used by cognitive faculties unless

[2]For example, Aglioti et al. (1995) found that although subjects consciously experience the central circles in the Titchener circle illusion as having different sizes, when they are asked to grasp those circles, their grip sizes were calibrated to the actual size of the circle - not its illusory size. This indicates that the conscious representation of the circle in the ventral stream was not the only representation playing an action guiding role. There is also a (presumably nonconceptual) representation in the dorsal stream that is unaffected by the illusion, which controls grasp size. I think that these kinds of influences on action, while important, are not properly understood as exercises of *rational* control.

it has first gone through the step of conceptualization. Since immediate usability by the cognitive faculties is a necessary consequence of cognitive access to a content, the only kind of content that can count as being successfully accessed is conceptualized content. Block (1995) actually acknowledges a view like this, saying that "some may say that only fully conceptualized content can play a role in reasoning, be reportable, and rationally control action. If so, then nonconceptualized content is not A-conscious" (Fn 9, p.245).

Since the process of conceptualization plays a crucial role in this argument, I should say a bit more about it. Conceptualization is essentially a process via which an object that was represented in a nonpropositional format comes to be represented a propositional format. Such a process is successful just in case the content of the propositional representation "matches" the content of the original nonpropositional representation in some way. There are a couple different ways to understand this process. It might be that one and the same state undergoes some kind of transformation that renders its content propositionally structured. This seems to be the way that conceptualization is conceived in much of the literature. However, I do not think that it is possible for such a transformation to occur. Nonpropositional representations typically represent in a map-like way via geometrical or topographic isomorphism, while propositional representations represent via the syntax of its constituent concepts. Because these representational formats have the content that they have in virtue of different relations to that content, it seems likely that they have very different identity conditions for the persistence of a representation over time. As a result, a 'transformation' from one format to another would be better understood as the destruction of the original map-like representation and the subsequent production of a propositional representation with matching content. Given this, I think that the most plausible way of understanding conceptualization is for it to involve the production of a *numerically distinct* state with propositional content, which usually exists along side the original nonpropositional representation. Since conceptualization is necessary for cognitive access, having

cognitive access to a content requires two things. First, a propositional state must be produced. Second, this new state must be encoded in working memory and made available to one's cognitive faculties. When both of these conditions obtain, one has cognitive access to, and becomes access conscious of, that content.

Unfortunately, things are complicated by the fact that Block (2002) revises the definition of access consciousness. In this revised version of the original paper he argues that "a representation is A-conscious if it is *broadcast* for free use in reasoning and for direct "rational" control of action" (p.208, emphasis added). The reason for this change is that being poised is a dispositional notion, which means that the original definition had the downside of making access consciousness dispositional as well. But since access consciousness is an occurrent phenomenon, it needs to be defined in terms of an occurrent notion like broadcasting. I have no problem with this proposed change as such and agree that it will be useful to ground the kind of poise necessary for access consciousness in a kind of broadcasting.

More recently, however, Block has said that after switching to a global broadcasting model of access consciousness, "the issue of whether the broadcast contents were conceptual or nonconceptual no longer looms large. I don't see why broadcast contents cannot be a mix" (Block 2007, p.539). This seems to conflict with my line of argument above, insofar as it could be taken to imply that nonconceptual content can be broadcast in a way the results in its being cognitively accessed. In order to charitably interpret these comments, I think we need to distinguish two importantly different kinds of broadcasting: broadcasting *to* working memory from other regions, such as the perceptual centers, and broadcasting *from* working memory to the cognitive faculties. The first step of broadcasting - from the perceptual centers to working memory - allows broadcast contents to be nonconceptual. Many of the areas that broadcast to working memory have states that represent their content in the map-like format typical of nonpropositional representations with nonconceptual content. As a result, many broadcast states will have nonconceptual content. This step of broadcasting has the function of making representations available for encoding in working

memory. But, its making them available to working memory does not thereby make them available for use by the cognitive faculties. That requires a second step of broadcasting from working memory to the cognitive faculties. This step of broadcasting that has the function of making representations available for immediate use by the cognitive faculties. Since it has this function, given the arguments above, this variety of broadcasting can only occur if the representation has conceptual content. Notice that in order for a state to be broadcast to one's cognitive faculties, it must first be encoded in working memory, the hub of the cognitive system. This is why broadcasting a state to working memory is insufficient for cognitive access; such broadcasting does not guarantee encoding, it only makes it possible.

Now that we have distinguished those two instances of broadcasting, let us return to Block's suggestion. I think that Block is right to want to think of the kind of poise involved in cognitive access as an instance of broadcasting, so long as we understand that kind of broadcasting as broadcasting to the cognitive faculties. However, nonconceptual content cannot be broadcast in this way, so if we are to maintain the intuition that nonconceptual content can be broadcast, we must understand this as broadcasting to working memory from other parts of the mind. We will return to this second type of broadcasting in the Chapter 4. For now, I am only arguing that any model that allows for the broadcasting of nonconceptual content is most charitably interpreted as a model of cognitive accessibility not cognitive access. Since our project here is to develop a theory of access consciousness, the kind of broadcasting implicated must be necessarily conceptual. That is, it must be broadcasting from working memory to the cognitive faculties. With that settled, let us return to the challenge voiced by Block's critics. How is access consciousness, so understood, a form of consciousness?

I think that the best answer we can give to this question is that access consciousness is grounded in propositional awareness of facts. They are both varieties of consciousness and they both require their content to be conceptual and are, as a result, both grounded in

propositional representations. These similarities might be taken as evidence that all and only cases of access consciousness involve propositional awareness. But I want to resist identifying them. It would be more accurate to say that all instances of access consciousness involve propositional awareness, but not all instances of propositional awareness are going to be cases of access consciousness. My reason for doing this is to allow that one might become aware of a fact without having cognitively accessed the state that grounds that awareness. Nonetheless, understanding access consciousness as consciousness of facts gives us a powerful framework for answering skeptics, like Natsoulas, who worry that access consciousness is not a form of consciousness at all. Since it is obvious that consciousness of facts is a form of consciousness, if access consciousness always involves consciousness of facts, then it is clear that access consciousness is a legitimate variety of consciousness.

In summary, I think that access consciousness is best understood as a special variety of fact-awareness. Access consciousness is the kind of fact-awareness that one has when the representation of that fact that has been encoded in working memory and broadcast to the cognitive faculties. Or in other words, access consciousness is fact-awareness paired with cognitive access to the state that represents that fact. Although I suspect that Block may worry that the view developed here does not really capture what he meant when discussing "access consciousness", we can understand my argument in this section as an argument that the view defended here gives us additional resources for understanding access consciousness, which in turn allows us to vindicate Block's original insight that "access consciousness is a distinctive form of consciousness".

2.3 Phenomenal Consciousness

Phenomenal consciousness is the more intriguing half of the phenomenal/access distinction. Block (1995) defines phenomenal consciousness as 'experience'. Phenomenally

conscious properties are experiential properties and "the totality of the experiential properties of a state are 'what it is like' to have it" (p.230). A necessary condition on a subject's being phenomenally conscious, then, is that they host experiential properties. Experiential properties, as I understand them, are the 'qualitative' properties that 'inhere' in an experience.[3] Block also notes that phenomenal consciousness involves a special kind of presentation to the self: "P-conscious states often seem to have a 'me-ishness' about them; the phenomenal content often represents the state as a state of me" (p.235). Block continues by noting that the 'me-ishness' of an experience is the same even when the states have different experiential properties, e.g., phenomenal consciousness of red and phenomenal consciousness of green have different phenomenal feels, but they are both equally 'self-orientated'. Although Block (1995) does not go so far as to list self-presentation as a necessary requirement for phenomenal consciousness, in this section I will argue that it is.

Flanagan (1997) seems to think that the kind of self-presentation involved in phenomenal consciousness should be understood in terms of access. He argues that "phenomenal consciousness always involves access to whatever we are phenomenally aware of" (p.370). More recently, Levine (2007) has argued that the concept of consciousness itself is constitutively related to the notion of *subjective* access. The kind of experience that constitutes phenomenal consciousness consists of more than the mere instantiation (or 'hosting') of phenomenal properties, it also involves some kind of access to those properties (p.536). I think that Flanagan and Levine are on the right track in insisting that the kind of self-presentation involved in phenomenal consciousness must be a kind of subjective access to experiential properties. Furthermore, I would argue that if there is 'something it is like' to undergo a state, then that state must *feel* a certain way. But surely a feeling of this sort

[3]I am using the phrase 'inhere' as an ontologically neutral term to accommodate both those who think that these properties are properties of the experience itself and those who are convinced by transparency arguments and think that these properties are ultimately properties of external objects. I think both views can be accurately described as views where the properties 'inhere' in an experience.

would only be a form of consciousness if it were a *felt* feeling.[4] This is what grounds the 'me-ishness' of the experience: the experience is for me because *I* feel it. An unfelt feeling would not be a form of consciousness. We should conclude, then, that subjective access to experiential properties is necessary for phenomenal consciousness.

I think that the best way to understand the kind of subjective access necessary for phenomenal consciousness is in terms of being *aware* of experiential properties. Understanding this condition in terms of awareness gives us a nice way of cashing out the intuitive requirement that feeling must be felt in order to be conscious and it makes good on my commitment to 'consciousness' and 'awareness' being synonyms. In more recent work, Block seems to agree:

> Phenomenal consciousness requires Awareness. (This is awareness in a special sense, so in this section I am capitalizing the term.) Sometimes people say Awareness is a matter of having a state whose content is in some sense 'presented' to the self or having a state that is 'for me' or that comes with a sense of ownership or that has 'me-ishness' (as I have called it; Block 1995a). (Block, 2007, p.484)

This passage leads me to believe that I am not neologizing the term by insisting that awareness of experiential properties is necessary for phenomenal consciousness. It is important to see, however, that these two components of phenomenal consciousness - the awareness relation and the experiential properties - are *conceptually independent* parts of an experience. One is a relation and the other is a relatum. Because of this conceptual independence, they can each be studied on their own terms. This shows up in the literature as well. Much of the discussion about phenomenal consciousness, including much of Block's early work, has focused on the nature of experiential properties. Less discussion has been generated over the exact nature of the awareness relation involved in phenomenal consciousness, although that seems to be changing. I would like to pursue this second topic. Exactly what

[4] I am indebted to Dennett (2012) for this way of putting the point. Dennett, of course, thinks that the relevant kind of access is a form of *cognitive* access, which I strongly disagree with. But setting that point of disagreement aside, we both agree that, in general, there must be some kind of access to the object, otherwise it couldn't possibly count as a variety of consciousness.

kind of awareness relation is essential to phenomenal consciousness being the kind of consciousness that it is?

2.3.1 Objects of Awareness

In order to determine the nature of the awareness relation, we must first determine whether the subject is aware of an internal object or an external one when phenomenally conscious. Given how Block defines phenomenal consciousness, it should be clear, at least, that what one is directly aware of is a set of *experiential properties*. But we need to determine whether those experiential properties are properly attributed to an internal object or an external one. To do that we need to examine the experiential properties themselves. There seem to be two importantly different kinds of experiential properties:

1. Experiential properties that are attributed to another object

2. Experiential properties that are not so attributed

Attributed experiential properties are properties that we typically experience as being had by some object other than the mental state in which they inhere. Most perceptible features are attributed experiential properties, e.g., colors, shapes, textures, temperature, smoothness, pitch, timbre, volume, etc. Any remaining experiential properties are experienced as unattributed, e.g., the feeling of depression, drunkenness, some aspects of pain, etc. It seems reasonable to suppose that experiential properties are attributed to another object only if the experience has representational content.[5] I also think it is likely that *all* states have *some* unattributed experiential properties. Even in cases where nearly all the experiential properties seem to be attributed to other objects (e.g. as in a visual experience) we can still become aware some unattributed properties (e.g. the distinctive "visualness" of the

[5] All I need for my discussion here is the relatively weak claim that there are correlations between represented properties (i.e. content properties) and attributed experiential properties. Although at the end of the day I do endorse a form of representationalism, my goal is to frame this discussion as neutrally as possible so that the argument does not depend on accepting this controversial view about attributed experiential properties.

experience). Thus, I think that experiences with representational content have *both* kinds of experiential properties, while putatively nonrepresentational experiences (e.g. moods and emotions) will only have unattributed experiential properties.

When a subject is aware of unattributed experiential properties they are aware of the mental state in which they inhere. That is, every case where a subject is aware of unattributed properties is one where the object of awareness is an internal mental state. Thus, in such cases the awareness relation is properly understood as a variety of *internal awareness*. Since few authors have emphasized the distinction between attributed and unattributed experiential properties, it is commonly supposed that *all* cases where a subject is aware of experiential properties are cases of internal awareness.

I disagree. I want to argue that it is possible for the awareness relation present in some instances of phenomenal consciousness to take the form of external awareness. In particular it seems plausible that the object of awareness is the *external object* when one is aware of the experiential properties attributed to that object. There arguments in favor of this view. To begin with, in order to adequately account for the possibility of perceptual error, theories of perception have to allow that the properties we are aware of in perceptual experience are the properties that our perceptual states *represent the object to have*. That is, perceptual awareness of an external object takes place via awareness of the experiential properties attributed to the object. But very few contemporary authors think that this fact implies that subjects are *actually* aware of their internal mental states in perceptual experience. Instead, it is widely accepted that one can be aware of an external object via awareness of attributed experiential properties.

The second argument proceeds by appealing to the claim that perceptual experience is 'transparent' to external objects. If you find the intuition that perceptual experience is transparent to the objects it is about compelling, then you should think that the object of awareness in such cases is an *external* object rather than an internal one. All we need for this claim is for it to be the case that one is aware of an external object when one is aware

of attributed experiential properties. We need not take on any of the other commitments traditionally associated with stronger versions of the transparency thesis.[6]

Since we have good reason to suppose that awareness of attributed experiential properties involves awareness of an external object, we should reject any interpretation of phenomenal consciousness that implies that subjects can only be internally aware of their experiential properties. While many cases of phenomenal consciousness *do* involve internal awareness of unattributed experiential properties, other cases merely involve external awareness of attributed experiential properties. Thus, when a subject is phenomenally conscious of an eggplant they are aware of the eggplant in virtue of being aware of the visual experiential properties (e.g. purpleness, ovalness, shininess, etc.) that are attributed to the represented eggplant.

2.3.2 Nonpropositional Awareness

I want close this chapter by arguing that the kind of awareness relation involved in phenomenal consciousness - be it internal awareness or external awareness - takes the form of *nonpropositional* awareness. For ease of exposition, I will focus on cases where a subject is phenomenally conscious of an external object. The first consideration in favor of this view comes from noticing that Block thinks that phenomenal consciousness and access consciousness are fundamentally different varieties of consciousness. Since I have argued that access consciousness involves propositional awareness of the accessed content, if we are to retain the intuition that access consciousness and phenomenal consciousness are fundamentally different, phenomenal consciousness cannot involve propositional awareness.

[6]For example, the "moderate" version of the transparency thesis agrees with this minimal claim but goes on to place a restriction on our epistemic access to attributed experiential properties: the *only* way of being aware of attributed experiential properties is as attributed to an external object; one cannot become internally aware of such properties as had by a mental state. The strongest version of the transparency thesis agrees with both of the claims made by the moderate version but goes on to further restrict our epistemic access to experiential properties: the only kinds of experiential properties that one can become aware of are attributed properties; one cannot become aware of unattributed experiential properties (Harman 1990; Tye 2000). I am not interested in pursuing the further restrictions placed on our epistemic access to experiential properties by the moderate and strong versions of transparency.

That means that is must be, by definition, a *non*-propositional form of awareness.

We can reach the same conclusion in another way. Phenomenal consciousness is, according to Block, awareness of experiential properties. Experiential properties are properties that inhere in one's mental states. That is, they are properties of mental objects, which may be attributed to some other object. There are two basic kinds of representational formats: propositional formats, which represent propositions, and nonpropositional formats, which represent worldly particulars and their properties. Since experiential properties are the properties of objects, they are best construed as particular things as opposed to propositions. As a result, awareness of experiential properties is best understood as being grounded in a representation that has a nonpropositional format. Since nonpropositional awareness is based in a nonpropositional format, we have a match in representational format of the two. They both represent particular things without at the same time representing any facts about those things. We should conclude, then, that the kind of awareness involved in phenomenal consciousness is nonpropositional awareness. At first it seems that Block might agree with this claim. For example, Block (1995) says:

> Consider a perceptual state of seeing a square. This state has a P-conscious content that represents something, a square, and thus it is a state of P-consciousness *of* the square. It is a state of P-consciousness of the square even if it doesn't represent the square as a square, as would be the case if the perceptual state is a state of an animal that doesn't have the concept of a square (p.232).

In the case at hand the object of awareness is a square and presumably the square is an external object. In accordance with my suggestion above, Block allows that phenomenal consciousness of a square does not necessarily imply awareness *that* the square is a square. This means that a creature without the concept of a square could still become phenomenally conscious of one. In a footnote in the same paper Block adds that:

> A perceptual experience can represent space as being filled in certain ways without representing the object perceived as falling under any concept. Thus, the experiences of a creature that does not possess the concept of a donut could represent space as being filled in a donutlike way (Fn 4, p.245).

These two passages suggest that, at least when writing his 1995 paper, Block was comfortable with the idea that one can be phenomenally conscious of an object in virtue of being nonpropositionally aware of the experiential properties attributed to that object. Such awareness does not require the possession or deployment of the concepts needed in specifying the content one is conscious of. This leads me to believe that identifying nonpropositional awareness with the kind of awareness necessary for phenomenal consciousness is not terribly revisionary. However, in a response to Michael Tye's commentary on the 1995 paper, Block asks:

> How do we know if P-consciousness is preconceptual? I used the phrase "representational" to describe P-content instead of "intentional" to allow for that possibility, but I have seen no convincing argument to the effect that P-content is preconceptual. Furthermore, there is reason to doubt the preconceptual claim. The specialized modules appear to have lots of conceptualized information. For example, there appears to be information in the face module about people's *occupations* (see Sergent & Poncet 1990; Young 1994a; 1994b). (p.278)

I think that we can respond to Block's worries here. First, following Burge (2010), I want to distinguish between the kinds of categorization and grouping that are properly thought of as the product of conceptual activity and the kinds of categorization and grouping that occur in the construction of a nonconceptual iconic representation. Acts of categorization and grouping are properly thought of as involving the deployment of concepts when those acts are suitably general. That is, concepts must satisfy something like the "generality constraint" (Evans 1982). In which category do the purported counterexamples cited by Block fall: conceptual categorization or nonconceptual categorization?

After looking at the cited papers, I did not see much evidence for that claim that the information about occupation is in the face module per se. The papers cited are about the degree to which some kinds of covert recognition of faces are preserved in subjects with prosopagnosia. The only discussion of occupations in Sergent and Poncet (1990) comes from cases where the subject was unable to identify any faces until she was told that they all

had the same occupation, which then allowed her to correctly guess the shared occupation and identify several of the faces. Such a finding does not require us to suppose that there is occupation information in the face area. It simply indicates that there are some preserved associations that can be reasoned out by the subject after being cued in the right way.

That said, there *are* some interesting findings in these papers. For example, they found that prosopagnosics will learn faces faster when paired with the correct name as opposed to the incorrect one, and that their eye movements are faster to previously familiar faces than unfamiliar ones. But there is no reason to suppose that these kinds of preserved associations are conceptual in nature. The subjects cannot freely deploy this information in order to report the name of the person or that they are familiar and as a result these kinds of associations do not satisfy the generality constraint. As a result, cases of covert recognition must merely involve nonconceptual categorization. This suggests, then, that there is no barrier to accepting Tye's claim that phenomenal consciousness is 'preconceptual' as he puts it.

In the revised version of the paper, Block (2002) raises a different worry about identifying the kind of awareness involved in phenomenal consciousness with nonconceptual content:

> If P-content is non-conceptual, it may be said that P-contents are not the right sort of thing to play a role in inference and guiding action. However, even with non-humans, pain plays a rational role in guiding action. Different actions are appropriate responses to pains in different locations. Since the contents of pain do in fact play a rational role, either their contents are conceptualized enough, or else non-conceptual or not very conceptual content can play a rational role. (p.210)

I think that Block's argument here faces a dilemma: either the action counts as an instance of rational guidance or it does not. If it does not count as a case of rational guidance, then it should be understood as a case of happening to act in accordance with a rational rule without actually being *guided* by the rule. Although we might be inclined to describe this behavior as being "rational", strictly speaking it does not count as an instance of rational

guidance. Since being guided by a rational rule is precisely what is required for the rational control of action, we can see that this interpretation of the case would suggest that the pain is actually playing a *non-rational* role in guiding action. I suspect that this is what is happening in most non-human animals. Their behavior is being guided by the pain that they experience, but they are not following a rational rule. They are merely acting in ways that we judge to be rational, perhaps because those patterns of behavior are conducive to survival. Therefore, on this horn of the dilemma there is no reason to suspect that phenomenal consciousness of a pain must involve conceptual content.

If it does count as a case of rational guidance, then it is a case where the pain has been cognitively accessed by the subject. But, as I have argued above, cognitive access to a state requires that the state's content be conceptualized. When a state is conceptualized a numerically distinct propositional state is formed. This propositional state can then rationally guide action and speech. But notice that this step of conceptualization is something that can happen *downstream* from experience. The fact that it must occur prior to the instance of rational guidance imposes no specific requirements on the content of the experience as such. Therefore, we have no reason to suppose the phenomenally conscious pains must involve conceptual content.

Since I take it that the responses I have given to Block's worries are successful, I think that we should conclude that the kind of awareness involved in phenomenal consciousness should be understood as nonpropositional awareness. This has many benefits. It gives a clean divide between phenomenal consciousness and access consciousness. Access consciousness has conceptual content and takes the form of awareness of facts. Phenomenal consciousness has nonconceptual content is a noncognitive form of awareness with nonconceptual content. I think that we can usefully understand the variety of noncognitive awareness present in cases phenomenal consciousness in terms of nonpropositional awareness of things.

2.4 Conclusion

If my arguments here are successful, then we now have a way of distinguishing the fundamental forms of transitive consciousness. Transitive consciousness can be of two different kind of objects: internal objects and external objects. With respect to each kind of object, there are two modes of awareness: the nonpropositional mode and the propositional mode. In the nonpropositional mode of awareness one is directly aware of the object itself. This is the variety of awareness that phenomenal consciousness, properly construed, requires. If one cognitively accesses the state responsible for their nonpropositional awareness of that object, then they can become access conscious of facts about that object. With these distinctions made, I will now develop my views on the nature of both external awareness and internal awareness.

3 NECESSITY OF REPRESENTATION

3.1 Representation

In this chapter I will evaluate proposals for the necessary conditions for external awareness. In particular, this paper is focused on understanding our capacity for non-propositional awareness of external objects, which provides the necessary backdrop for awareness of facts about those objects. In order to help us uncover the first necessary condition for nonpropositional awareness, we will consider the case of 'occluded Olivia':

> Suppose that Olivia is sitting in a room with good lighting. She is looking in the direction of a picture of an eggplant on the wall, but an experimenter has placed a thick sheet of paper in front of it, blocking it from her view.

Although this may seem like a rather mundane case, but I think it can help us find the first necessary condition for external awareness because we typically think that Olivia will *not* become aware of the occluded picture.[1] If this is right, then understanding how occlusion works can tell us which necessary condition it prevents from obtaining. But, if we are to use cases of occlusion in this way we need to be sure that Olivia is actually unaware of the picture. How can we do this? This is actually a surprisingly difficult question to answer decisively. I will argue here that the best method we have for determining whether or not someone is aware of an object is to solicit a 'subjective' report in concert with certain 'objective' behavioral tests.

A subjective report is simply a report given by the subject that describes what they experienced in a particular trial. If the subject does not report experiencing a particular object, then this suggests that they were not aware of that object. Subjective tests are

[1] I am assuming here that Olivia is not in sensory contact with the picture via any other sensory systems. She is too far away to touch it, cannot hear it, taste it, or smell it.

useful because if a subject does report that they experienced an object, then we can be confident that they did, in fact, experience it. However, since several different cognitive and noncognitive systems are involved in producing a subjective report, a failure to report experiencing an object does not *necessarily* imply that they were not aware of the object in question. It is possible that the subject was actually aware of the object, but that they were unable to report it because some other necessary condition for producing a report failed to be met. For example, it might be that they were aware of the object but they just didn't notice it, so they never reported it. Thus, a failure to report the presence of an object does not guarantee that the subject was unaware of it.

Objective tests aim to measure awareness by using a quantifiable behavioral measure. Although we are interested in objective *tests of awareness*, it is worth noting that most behavioral measures are best suited for detecting the mere presence of information about an object in the mind as opposed to full-blown awareness of the object. Nonetheless, even tests for detecting the presence of information can still be useful. If the subject performs at chance on such a test, then we can be confident that they did not experience the object. This is because performing at chance on these tests means there was not even enough information about the object in the mind for it to be detected, which suggests that they could not have been aware of the object. However, because of how conservative most of these tests are, performing *above* chance on such a task does not guarantee that the subject was aware of the object. Their improved performance may have been caused by some unconscious information about the object. This has lead many researchers to prefer subjective reports over objective tests despite the drawbacks of subjective reports mentioned above (Merikle et al. 2001).

While I agree with Merikle et al. that subjective reports should be one of our preferred tests of awareness, I do think that there is a particular objective test that is somewhat better than the rest at directly measuring awareness: the forced-choice test. In this kind of test the subject is forced to guess whether or not the target object was present in each trial. If they

perform at chance, then we have reason to believe that they were not aware of the stimulus. If they perform above chance, then we have reason to believe that they had some kind of experience of the object. This task is useful because succeeding at it does not require as much information about the object as is needed for issuing a subjective report. I suspect that all the subject only needs to feel like they might have seen something on the trials where the object is present in order to pass the forced choice test.

Now that we know how to test for awareness of an object, how will Olivia perform on these tests? It should be clear that she would report having not experienced the picture of the eggplant and that she would be at chance at guessing whether or not the picture was present behind the sheet of paper on each trial. In cases like this, where both our preferred tests give a negative result, we can safely conclude that the subject was not aware of the object. Since Olivia is unaware of the picture of the eggplant, we can now ask *how* occlusion blocks awareness.

I will argue that occlusion blocks awareness because the sensory systems have the job of representing the environment and occlusion functions to prevent the formation of a representation of the occluded object. I will support each claim in turn. Although there are those who are skeptical of claims that our sensory systems represent the world (e.g., Travis 2004; Brewer 2006; Fish 2009), the scientific consensus seems to be that our sensory systems *do* have the function of representing the world. For example, explanation of behavior at the psychological level frequently proceeds by citing facts about how the subject represents the world (Fodor 1975, 2008). We would have to drastically change our current model of psychological explanation if we decided that the sensory systems do not represent the world. Additionally, at the level of neural processing, the part of the cortex devoted to processing visual information appears to be specialized for extracting information about visual features from the light that falls on the retina (Solomon and Lennie 2007). This kind of feature extraction is best understood as the construction of a representation of those visual features. Finally, Susanna Siegel (2010) and Susanna Schellenberg (2011)

have each given convincing philosophical defenses of the importance of sensory representation. Instead of rehearsing their arguments here I will simply refer you to their excellent papers.

If the sensory systems have the job of representing objects in our environment, why should we think that occlusion prevents the formation of a representation? To answer this we need to consider how representations are constructed in the first place. In each of our sensory systems, the features that we represent the world as having are extracted from the information falling on our sensory transducers. That is the *only* way that we can construct appropriate sensory representations of the world. Now, if we were to block the flow of information from the world to a subset of our sensory transducers (e.g. by closing our eyes), then we will prevent the sensory system connected to those transducers from constructing representations of that portion of the world. In the case at hand, a sheet of paper was used to block the light that is reflecting off of a picture of an eggplant. This effectively cuts off the flow of information from the picture to the sensory transducers on Olivia's retina. As a result, Olivia's visual system receives no information about the picture of the eggplant. In the absence of such information, no visual representation can be formed. Therefore, occlusion prevents a necessary condition for visual representation from obtaining: occlusion blocks the flow of information from the world to the transducers of our sensory systems.

Finally, I think that this fact about occlusion explains why Olivia does not become aware of the occluded object. The argument for this is basically an inference to the only explanation. We know that occlusion blocks the flow of information and as a result prevents the formation of a representation. However, occlusion does not appear to do *anything else*. Thus, if occlusion successfully inhibits awareness of an object and the only direct effect that it has is to prevent the formation of a sensory representation, then we should conclude that awareness is impaired *because of* the absence of that representation.

I have just argued that representations are necessary for awareness on the basis of Olivia's case. But this conclusion also follows from more general considerations about the design of the mind. I think that it is plausible that mental processes (e.g. awareness, attention, memory, reasoning, etc.) are only able to directly interact with other mental items. Since they are unable to directly manipulate external items, such items need a 'mental proxy' to stand-in for them in the mind. Having this proxy then allows mental processes to 'interact' with those objects. So, in order for a subject to become aware of an external object, attend to it, remember it, reason about it, etc. that object must be represented in the mind. This, then, is why we need mental representations. If we are to survive in our environment, our minds must have a way of interacting with non-mental objects. The solution to this general limitation on the mind's functioning is to mentally represent those objects. Connecting this back to the case of Olivia, since occlusion prevents the formation of a mental representation of the picture and awareness is a mental process like any other, she is unable to become aware of the picture.

3.2 Representation is not Sufficient

I have argued that having a representation of a non-mental item is necessary for awareness of that item. Since my project is to examine the necessary conditions for external awareness, we need to ask if this is the *only* necessary condition? In other words, is having a representation of an item *sufficient* for awareness of that item? I will argue that representation is not sufficient for awareness and conclude that there must be at least one more necessary condition for external awareness. There are three arguments for the insufficiency of representation that we will consider here. The first will rely on conceptual analysis of our concept of 'transitive consciousness', the second will argue that there is a capacity mismatch between representation and external awareness, and the third will take the form of a putative counterexample.

3.2.1 Conceptual Argument

Let us begin with the conceptual argument. External awareness is a form of transitive consciousness. That is, all cases of external awareness are cases of consciousness *of* an external object. The interesting thing about transitive consciousness is that it can only be predicated of a subject of a whole. It cannot be predicated of a state within a subject without absurdity; e.g., it makes no sense to say that "Olivia's belief is conscious of the eggplant". Of course, whenever a subject is conscious of an object they are conscious of that object *in virtue* of being in a relevant mental state. Such states are the necessary grounds for the subject's being transitively conscious. The question we want to answer here is: what do these mental states have to be like in order to fulfill the role of grounding transitive consciousness?

We already know one feature the state must have in order to fulfill this role: it must be a representation of the object of awareness. However, since transitive consciousness is a property that can only apply to a *subject as a whole*, the mental state's being a representation does not seem sufficient for it to fulfill its role. It cannot be just any representation sitting there 'spinning frictionlessly in a void'. It needs to be hooked up with the subject of awareness in the right way. Without such a connection to the subject, there would be no reason to think that it can furnish them with consciousness of its content. Whatever this relation turns out to be in the end, the fact that the concept of transitive consciousness requires such a relation implies that mere representation is insufficient for awareness.

Another way to get to the same conclusion is to consider the platitude that when you are aware of an object it registers as 'something to you'. But in order for a representation's content to be something to the subject, presumably it needs to be suitably related to the subject of awareness. As Cohen and Dennett (2011) argue, if a subject's color representations were wholly dissociated from them - simply sitting there in a petri dish with no outgoing connections - then that subject could not possibly be aware of any colors at all. That is, in such a case the colors of one's environment would not be something to them. Both of

these arguments push for the same conclusion. Simply having a representation of an object is insufficient for awareness of that representations content. The representation also needs to, at the very least, be related to the subject of awareness in some suitable fashion. We will consider what this relation amounts to below, but for now all that matters is that these arguments suggest that there must be at least one more necessary condition for external awareness.

3.2.2 Capacity Mismatch Argument

We can also motivate this conclusion by comparing the capacity of sensory representation with the capacity for awareness of external objects. While this is difficult to do with any degree of precision, even rather rough estimates will make it clear that there is a large mismatch in capacity. Let us start with the capacity of sensory representation. From what we know about the functioning of the sensory transducers all the way through the processing hierarchy, nearly everything that hits a transducer gets represented to some degree. In vision this means that all of the objects who's reflected light falls on the retina get represented, for audition all the sound waves that hit the hair cells on the chochlea are represented, and so on. This means that nearly all of the nearby objects end up being represented in one sense modality or another. I want to suggest that this means that the capacity of sensory representation should be measured not in tens but in hundreds.

While the capacity of sensory representation is impressively high, the same cannot be said for the capacity of external awareness. While it is difficult to find a consensus on the capacity of external awareness, Block (2007) estimates its capacity as being somewhere between 8-32 objects at a time (p.489). Others think that the capacity is even lower, either around four items (Dehaene and Changeux 2011) or possibly as low as a single coherent item (Baars 2001). What for our purposes here is that sensory representation has a capacity measure that is at least an order of magnitude larger than the capacity of external awareness. This suggests that we need to place at least one more condition on sensory representation

in order to appropriately restrict the number of representations that are candidates for furnishing the subject with awareness of their content.

3.2.3 Counterexample: Masked Mary

The final argument against the sufficiency of representation for external awareness takes the form of a putative counterexample. I think that there are demonstrable cases where we have good reason to believe that a subject has a representation of an object, but also have good reason to believe that they are not aware of that object. To introduce this counterexample, let us consider the case of 'masked Mary'. Suppose that Mary is sitting in a room with good lighting. She is looking at the location where a picture of an eggplant will be briefly presented. However, an experimenter also briefly presents a scrambled image before and after the eggplant. This technique is called visual masking. In visual masking tasks the object that is being masked is referred to as the 'target stimulus', while the scrambled images (or patterns) that occur immediately before and after it are called the 'forward mask' and 'backward mask', respectively. Although visual masking effects can be found so long as the target stimulus is presented for less than 50 ms, optimal masking conditions are typically closer to 20 ms (Kouider and Dehaene 2007). This briefly presented stimulus is sandwiched between a forward mask and a backward mask. When the masks are presented immediately before and after the target stimulus, they seem to interfere with our ability to become aware of the stimulus.

Since the aim of this task is to prevent Mary from becoming aware of the picture of the eggplant, we had better make sure that she is *actually* unaware of the eggplant. How can we do this? As I said above when discussing the case of occluded Olivia, I think that a combination of the verbal reports given by subjects in these tasks along with the use of certain forced-choice tests should give us a reasonable method for determining whether a subject is aware of the target stimulus. So, what responses do subjects actually give on these tests?

In a review of the literature on masking, Kouider and Dehaene (2007) say that in properly performed masking tasks, subjects will typically report that they did *not* experience the stimulus. But subjective reports are not entirely reliable because it is always possible that the subject failed to report the stimulus because some other necessary condition for report was not met. Is there anything about the setup of the task that suggests that a failure of awareness explains the absence of a report? Yes there is. What makes cases like this special is that the subjects in them are trying to succeed at them. As a result, they are devoting their cognitive resources towards performing the task: they are attending to the anticipated location of the stimulus and are focused on trying to recognize it (or at least recognize that it was present). Nonetheless, despite all this effort, they still report that they did not experience the stimulus. If they are trying to cognitively access information about the stimulus and they fail, presumably this failure is caused by the failure of some necessary condition for cognitive access to obtain. Since cognitive access to information about an object requires that one first be nonpropositionally aware of that object and the other necessary conditions for cognitive access appear to be met in these cases (e.g. they are focally attending to the stimulus), I think it is reasonable to suppose that it is the absence of awareness that is preventing the subject from cognitively accessing and reporting the presence of the stimulus.[2]

In addition to soliciting verbal reports from the subjects, some studies have given their subjects a forced-choice test of awareness. In this test, subjects are presented with a series of trials that randomly alternate between trials where both the target stimulus and the masks are present and trials where only the masks are present. If the subjects are aware of the

[2]This means that the masking cases are importantly different from the distraction cases that are popular in the literature on consciousness. Armstrong (1980) presents the case of a distracted driver and Block (1995) presents the case of an air conditioner operating in the background. In these kinds of cases the subjects seem to be nonpropositionally aware of the object (e.g. the road or the noise made by the air conditioner). So one necessary condition for cognitive access is met. However, they are unable to report anything about the object because they are *distracted*. Presumably distraction occurs when a subject's attention is focused on something else. The distraction cases are paradigmatic instances of false negatives because the negative report is caused by a failure to attend to the object in question, not because they failed to become aware of it.

target stimulus, they will perform above chance on this task, while if they are not, they will perform at chance. In studies performed by Jiang et al. (2006) and Kouider et al. (2006), it was found that the subjects performed at chance, even though they performed above chance on *other* kinds of objective tests. This is why it is important to separate the objective tests that merely test for the presence of information from those that may be testing for awareness.[3] Since both the subjective test and the force-choice test agree that subjects are not aware of the target stimulus in masking tasks, I think that we can safely conclude that masking techniques can prevent subjects from experiencing the masked stimulus. Thus, we should conclude that Mary was unaware of the picture of the eggplant in the case at hand.

Even though Mary is unaware of the eggplant, we do know that she *has* enough information about the eggplant in her mind for it to have detectable influences on her behavior in subsequent tasks. These kinds of unconscious influences are called 'priming effects'. Appealing to the literature on priming can be tricky since the phenomenon has been controversial since its inception. However, the controversy has only been over *which features* of the stimulus can influence the subject's performance in the absence of awareness. For example, early studies on priming appeared to find strong semantic effects. These are cases where the semantic content of the primed word influences your behavior on a subsequent task. So being primed with the masked word CAR might lead you to complete the word stem TR___ as TRUCK more often than when you are primed with a word that is not semantically associated with trucks. Unfortunately, since masking methods have become better at ensuring that subjects are completely unaware of the masked stimuli, these results have been difficult to replicate. Consequently, it is still controversial whether there are cases of semantic priming in the complete absence of awareness.

[3]It is worth noting that the Jiang et al. study used a different masking technique. In their study they used interocular suppression - also called continuous flash suppression - to render a stimulus presented to one eye invisible. Although the masking method differed, they found the same thing. Subjects failed at the forced choice test but performed above chance on other objective tests.

There are some other features, however, that are uncontroversial. There are well established, generally accepted morphological and orthographical priming effects (Kouider and Dehaene 2007). For example, being primed with "THINK" will bias the subject in favor of competing the word stem TH___ as "THOUGHT" because of the morphological similarity between the two and will also bias the subject in favor of completing the word stem as "THIN" because of the orthographical similarity between the two. Since nearly everyone in the field agrees that representations of masked stimuli can make it far enough up the processing hierarchy to have *some* kind of behavioral impact, we have what we need for our purposes here. All of these effects require a coherent representation of the masked stimulus. You cannot have orthographical, morphological, or even semantic effects in the absence of a clear coherent representation of the letters in the masked word. The priming data gives us strong evidence that there is a specific representation of the letters in the stimulus even when the stimulus has been masked. If this is right, then it seems that having a representation of an object is insufficient for nonpropositional awareness of that object.

3.3 Conclusion

I have argued for two claims in this chapter. First that representation is necessary for external awareness. Second, that such representations are insufficient for external awareness. The insufficiency of representation for awareness implies that there must be another necessary condition for awareness. However, we cannot determine what the additional condition is simply based on the arguments given here. All we know right now is that visual masking inhibits this condition from obtaining and that it should involve the representation's being suitably related to the subject. In the next chapter we will consider what proposals there are for this condition.

4 COGNITIVE ACCESS & COGNITIVE ACCESSIBILITY

4.1 Introduction

In this chapter I will evaluate proposals for the second necessary condition for external awareness. One popular proposal is based on the idea that consciousness is fundamentally tied to global information sharing. For example, Stanislas Dehaene (2014) argues that when one is aware of a certain piece of information that the information has entered into an area that makes it available to the rest of the brain. In particular, information becomes conscious when it is broadcast to all our *high-level cognitive faculties*. He argues that, in short, we possess a "mental router" and this router underlies external awareness. I find this background idea very appealing, but it is just that - a background idea. In order to develop this idea into a full-fledged theory of consciousness, a significant amount of detail must be added.

4.2 Cognitive Access

One way of developing a view based on this basic insight is to say that, in order to gain access to the 'mental router' underlying consciousness, the relevant information must be *cognitively accessed* by the subject. Dehaene (2014) defends a version of this view, arguing that:

> Whatever we become conscious of, we can hold it in our mind long after the corresponding stimulation has disappeared from the outside world. That's because our brain has brought it into the workspace, which maintains it independently of the time and place at which we first perceived it. As a result, we may use it in whatever way we please. In particular, we can dispatch it to our language processors and name it; this is why the capacity to report is a key feature of a conscious state. (p.303)

The key idea here is that there is a storage area connected to (or preceding) the 'mental router', which can encode and maintain representations for long periods of time. The best proposal for this kind of storage space is working memory, since encoding in working memory is necessary for the processes describes (e.g. issuing a report). Since Dehaene thinks that information must be stored in working memory prior to being broadcast to our cognitive faculties, we can see that he thinks that cognitive access is necessary for external awareness. I will call all theories that nominate cognitive access as the distinctive secondary condition for external awareness 'cognitive theories of awareness' since they think that the representation of a stimulus must have gained a certain amount of 'cognitive traction' with the subject in order for the subject to become nonpropositionally aware of the represented object.

Many who endorse a cognitive theory of consciousness were inspired by the framework developed by Bernard Baars. Baars (1988) argued that the mechanism that determines which objects a subject becomes aware of involves competitive interactions amongst the specialized modules that represent those objects. The winner of the competition gains access to a 'global workspace': a non-modular system with a limited capacity that allows specialized modules to exchange information. Once a representation is broadcast within the global workspace, the subject becomes conscious of the object represented by that state. Baars argued that the global workspace is grounded in the activity of *subcortical* neurons in the brainstem and the thalamus. In earlier work, Dehaene and colleagues objected to this way of grounding the global workspace, arguing that a system with a larger informational capacity is needed in order to share the kind of highly detailed information subjects are aware of (Dehaene 2009). They suggested that such a system should be grounded in the activity of *cortical* neurons that project between brain areas, which they called the global neuronal workspace (GNW).

Those who endorse a version of the global workspace theory frequently say that they are proposing it as a theory of perceptual awareness of external objects. For example, Dehaene et al. (2006) say that the question they want to answer is: "How do we consciously perceive a visual stimulus?" and, in a more recent review, Dehaene and Changeux (2011) say that they are concerned with transitive consciousness, e.g., "I was not conscious of the red light" (p.200). At the same time, they seem to operationalize external awareness in terms of *reportability*. For example, Dehaene et al. say that consciousness is "associated with a distinct internal space, buffered from fast fluctuations in sensory inputs, where information can be shared across a broad variety of processes including evaluation, verbal report, planning and long-term memory" (2006, p.205). The view seems to be that the kind of cognitive access necessary for reportability is the second necessary condition for awareness of external objects.

The bold nature of this claim has led some (e.g., Block 2002, 2007) to interpret the GNW theory as a theory of propositional awareness rather than nonpropositional awareness. Although I suspect that, in the end, this may be the most charitable way of understanding their claims, the proponents themselves are rather adamant that they aim to explain what is going on in cases where a subject is aware *of* a red stimulus, not ones where a subject is aware *that* the stimulus is red. It seems to me that the most straightforward interpretation of the theory in light of these passages is that the GNW *is* a theory of nonpropositional awareness of external objects. They just happen to think that reportability and cognitive access are necessary conditions for this kind of awareness.

Although the GNW theory is one of the more popular cognitive theories of consciousness, it is not the only one. Some philosophers, like Daniel Dennett, accept views very similar to the GNW but prefer to understand the phenomenon in more metaphorical terms, e.g., as 'fame in the brain' or 'cerebral celebrity'. Those who favor the higher-order thought theory of consciousness frequently talk as if cognitive access is necessary for external awareness. For example, Lau and Rosenthal (2011) and Brown (2011) both argue that cognitive

access is necessary for external awareness and seem to think that this requirement is based on the necessity of a higher-order thought for external awareness.[1] Even though I will largely be focusing on the global workspace theory for the remainder of this section, the arguments developed here are intended to apply to all cognitive theories.

4.3 Content Mismatch

Although I think there is some promise to the background idea that consciousness is tied to the "brain-wide information sharing", I do not think that cognitive access to information is necessary for its being shared in this way. More broadly, I think that any view that argues that cognitive access is necessary for nonpropositional external awareness is misguided. There are two key arguments that we will consider against these kinds of cognitive theories. The first argument cites a mismatch between the type of content involved in cognitive access and the type involved in external awareness. The second argument cites a mismatch between the capacity of cognitive access and the capacity of external awareness.

Let us begin with the argument for a mismatch in content. The content of a mental state is how it represents the environment as being. There are two broad categories of mental content: conceptual content and nonconceptual content. Following Tye (2000), I will define nonconceptual mental content as content that can be attributed to a subject without the subject's exercising (or even possessing) the concepts used in specifying the content (p.62). More specifically, I will be distinguishing conceptual content from nonconceptual content in terms of the *representational format* of the content (Cummins 2010). The format of a representation is determined by the structure of the representation, which in turn determines the way that the representation represents its object. Conceptual contents are

[1]This interpretation comes from their rejection of Block's overflow argument. They would have no reason to reject this interpretation if they did not think that cognitive access is necessary for external awareness. However, it is not certain that this is the view they actually endorse since Rosenthal (2005) has also argued for an interpretation of the masking experiments where subjects *are* externally aware of masked stimuli. But if you hold a view that wildly liberal, then you should definitely think that external awareness overflows cognitive access. I see no way of making these two interpretations consistent.

represented in propositional formats, where the syntactical arrangement of the constituent concepts determine the content of the representation. Nonconceptual contents likely have a variety of different representational formats. However, the one we will be focusing on here is the map-like representational format typical of perceptual representations. Map-like representational formats are not propositionally structured. Instead, the content of these representations is determined by the geometrical or topographical structure of the representation. Such representations are said to be 'structurally isomorphic' to the represented object (for further discussion see: Burge 2010, pp.537-544). By distinguishing conceptual from nonconceptual contents in this way we have a concrete proposal for how each kind of content works.

I have argued elsewhere that cognitively accessed contents are necessarily conceptual (Chapter 1). In brief, a cognitively accessed content is, by definition, a content that is 'immediately usable in reasoning or the rational control of action'. However, reason cannot operate on content represented in nonconceptual formats. Thus, only already conceptualized contents are immediately usable by the relevant processes. Therefore, cognitively accessed contents are necessarily conceptual contents.

The contents involved in external awareness, however, are not necessarily conceptual. That is, I think it is *possible* for external awareness to be grounded in person-level nonpropositional representations whose content represents the object of awareness in a map-like way. It need not be the case that external awareness always involves nonconceptual content, or even that it usually involves such contents in order for this argument to succeed. All I need is for it to be possible. Now, many different arguments have been given for the claim that person-level perceptual contents can be based in nonpropositional representational formats (Dretske 1995; Bermúdez 1995; Peacocke 1998; Tye 2000; Bermúdez 2007), but I will not rehearse them here. Elsewhere I have argued that there is a strong divide between awareness of things and awareness of facts and that these different kinds of awareness are grounded in different kinds of representational states: perceptual states

and doxastic states, respectively (Chapter 1). I will take it, then, that there can be instances of external awareness grounded in nonconceptual representations of things. If this is right, then we have reason to believe that there is a content mismatch between cognitive access and external awareness.

4.4 Capacity Mismatch

The other argument against cognitive theories of awareness is that the capacity of cognitive access is much smaller than the capacity of external awareness. In order to establish a mismatch in capacity we need somewhat precise estimates of the capacity of cognitive access and external awareness. Luckily there has been a significant amount of research relevant to the capacity of cognitive access, which will allow us to get a more precise estimate of the capacity of cognitive access. The notions of cognitive access and working memory are at times used interchangeably in the literature, but I do not think this is right. There is more involved in cognitive access than working memory encoding (and maintenance). As a result, I think that we can best understand the relation between these two faculties by seeing working memory encoding as necessary (but not sufficient) for cognitive access. As a result, it will nonetheless impose a bottleneck on the number of items that can be simultaneously cognitively accessed. This means that if we want to determine the capacity of cognitive access we should proceed by looking at the research on the capacity of working memory. The vast majority of the research on the capacity of working memory has been carried out using either visual or auditory tasks. Interestingly the same limit turns up on both kinds of tasks. It is generally agreed that the capacity of both visual working memory (Sperling 1960; Sligte et al. 2008; Brady et al. 2011) and auditory working memory (Gilchrist et al. 2008) is around four items or fewer. The big controversy is currently over whether or not the capacity limits found in these studies are just two instances of hitting the same domain-general capacity limit, or whether there are several domain-specific capacity limits, which would allow one to encode and maintain up to four items per sensory system.

The evidence is mixed. The early work done using the dual task paradigm suggested that there was little to no interference when performing verbal and visuospatial tasks simultaneously. This led many to conclude that each sensory system has its own domain-specific bit of working memory with its own capacity limit (Baddeley 1986). Unfortunately this early work hasn't held up as well as it might have because the early implementations of these dual tasks didn't properly separate iconic memory from working memory. In order to dissociate the high capacity sensory-specific memory from working memory, masks must be used in order to reset the sensory memory. However, the masks must be delayed long enough for the first round of encoding to occur. The purpose of the mask is simply to prevent the subject from re-accessing sensory memory after the original encoding finishes. Cowan and colleagues have performed several experiments using this methodology, including experiments using both visual and auditory stimuli in the same task. The results suggest that the four item limit is, in fact, a domain-general limit on working memory encoding and maintenance (Cowan 2001; Saults and Cowan 2007; Cowan 2010).

While we can get a somewhat precise estimation of the capacity of cognitive access, it is more difficult to get a reasonable estimate of the capacity of external awareness. For example, the capacity of awareness for simple *visual* objects has been estimated to be in the range of 8-32 items (Block 2007). Unsurprisingly, this estimation has been challenged by the cognitive theorists who typically argue that the capacity of external awareness is four items or fewer (Baars 2001; Dehaene et al. 2006; Dehaene and Changeux 2011). Nonetheless, if the capacity of working memory turns out to be a domain-general limit of four or fewer items, then there is a powerful argument available for establishing a capacity mismatch between working memory and external awareness. This is because although there is also significant controversy over the capacity of external awareness, there are certain circumstances where it would be exceedingly revisionary to deny that one is aware of more than four items at a time. In particular, if we *sum* the number of items that a subject is aware of simultaneously across all of their different senses, then it seems clear that the capacity

of external awareness must be larger than four. It would only take awareness of *a single item* in each of the five senses to exceed the capacity of cognitive access.

I find the intuition that I can be aware of at least one item per sense incredibly compelling. However, until someone designs a task able to test the capacity of awareness across every sense, we have no resources for supporting this intuition other than our own introspective judgments. In my own case, for example, I can confidently judge that I am currently aware of: the feelings in my fingers as I type, the sound of music playing, the way my computer screen looks, the taste of a mint, and the smell of coffee. Of course since our sense of taste and smell play less central roles in our daily lives than the other senses, we are aware of items via those senses less frequently. Nonetheless, we can get the same result by supposing that we are aware of two items each in vision, audition, and touch. Note that I am not saying that I can produce this judgment without cognitively accessing each item, but rather that it seems to me that I am directly aware of all of them in a single introspective act where I broadly distributing my attention inwardly. I hope that you will share this judgment after doing the same.

Although I think this is the most powerful form of the capacity mismatch argument, we can also formulate versions of the argument limited to a single sensory system. Doing so will allow us to sidestep the debate over whether or not the capacity limits on working memory are domain-specific or domain-general, and will (hopefully) allow us to avoid the downsides of appealing to introspection to establish the capacity of external awareness. Much research has been done on the capacity of visual working memory and the capacity of visual awareness, so that is the sensory system that we will focus on. For many years, Ned Block (1995, 2007, 2008, 2011, 2014) has forcefully argued that the capacity of visual awareness exceeds the capacity of visual working memory. We will consider the evidence he cites for this claim, most of which was gathered using the 'partial report' paradigm popularized by George Sperling (1960).

4.4.1 Partial Report Paradigm

In an attempt to determine whether or not subjects can see more than they can report, Sperling used a task with three stages. First, a fixation point was presented to the subject. Second, a 3x4 character array was briefly flashed on the screen for around 50 ms. Third, after the array was gone, the subject had to report the characters. Initially, Sperling had simply prompted the subjects to report as many characters as they could from the array (the full report condition). He found that subjects were typically able to accurately report only about four of the characters in the array (4.3 characters on average). It was clear from these results that there was a bottleneck *somewhere*, which ultimately limited the number of objects that subjects could report. But where is the bottleneck located? Is the bottleneck is located in the information about the stimulus that is accessible by the subject or is it located in the subject's ability to report the characters?

In order to decided between these two interpretations, Sperling changed the way that the subjects were instructed to report the characters. Instead of reporting all of the characters that they could, subjects were presented with an auditory tone that cued which row they should report from. There were three possible tones - a high tone, a middle tone, or a low tone - each of which corresponded to one of the three rows. For example, when a subject heard the high tone that was their cue to report as many of the characters as they could from the top row. The use of an auditory cue to solicit a report was a major improvement over past implementations of the partial report paradigm, as it allowed Sperling to precisely control the time at which the cue was presented, which, in turn, allowed him to quantify the amount of information about the stimulus available over time. Using this task, Sperling found that, after being trained up on the task, the subjects in the partial report condition could accurately report most of the characters in a given row (3.1 characters out of 4 on average), no matter which row was cued. Since subjects were able to report roughly 76% of the cued characters in the partial report condition, compared to only 36% in the full report condition, Sperling concluded the bottleneck must be located in the subject's ability

to report the characters in the array, suggesting that subjects have *more* accessible visual information than they can report, even in brief presentations.

This limited interpretation is generally accepted, even by the cognitive theorists. The results of partial report tasks indicate that there are two kinds of memory. The first is a transient, high capacity form of memory (i.e. 'iconic' memory) that contains all of the information about the stimulus that is accessible by the subject. The second is a more durable, low capacity form of memory (i.e. working memory) that contains the information the subject will report. Where opinions sharply divide is on the status of the information in the high capacity 'iconic' memory. Is this information conscious or unconscious? That is, does this information take the form of a visual experience as of the stimulus or is it merely unconscious visual information?

Sperling concluded that the information takes the form of a visual experience of the stimulus, saying that this kind of "short-term information storage has been tentatively identified with the persistence of sensation that generally follows any brief, intense stimulation. In this case, the persistence is that of a rapidly fading, visual image of the stimulus" (p.26). Evidence in support of this hypothesis comes from two sources. First, Sperling found that when the subjects were presented with a bright white post-exposure field, instead of the usual black field, that the accuracy of both partial reports and full reports went down significantly. This kind of interference strongly suggests that the accessible information "depends on a persisting visual image of the stimulus" (p.27). Second, subjects in Sperling's partial report task "report that the stimulus field appears to be still readable at the time a tone is heard" (p.20). This is true in all partial report tasks. Subjects generally report experiencing all the characters in the stimulus while it is present and for a short time after it has left the screen. This fits squarely with my own experience in partial report tasks and I predict you will also agree, if try the task for yourself.

If this is right, then subjects in partial report tasks experience nearly all of the characters they are presented with and this experience persists in iconic memory allowing them

to give a partial report in response to a cue that arrives after the array has left the screen. The bottleneck only comes into play when the subjects have to encode and maintain information about the characters in working memory so that they can issue a report. Thus, the results of the partial report paradigm are often taken to support a mismatch between the capacity of visual awareness and the capacity of visual working memory. This interpretation of the partial report paradigm has been widely accepted (Dretske 1981, 2006; Block 1995, 2007, 2011; Lamme 2003, 2006; Tye 2006, 2009; Prinz 2012). But, since there is significant variation in the views of those who accept this interpretation, I should be careful to note that I am merely claiming here that the subjects are nonpropositionally aware of the specific shapes of nearly all the characters in the array.[2]

It is important to see how limited this claim is. I am not claiming that the subjects are propositionally aware of the identities of all the characters or of any other facts about them. All I am claiming is that they are nonpropositionally aware of the specific shapes of the characters. If subjects are, in fact, nonpropositionally aware of the specific shapes of 9-12 of the characters in the array but they can only cognitively access and report about 4 of them, then there is a capacity mismatch between these two faculties. The 5-8 remaining characters are instances of nonpropositional awareness in the absence of cognitive access. Since similar conclusions have been reached using partial report tasks with different stimuli and different experimental setups (e.g., Landman et al. 2003; Sligte et al. 2008, 2009), I think that we can conclude that cognitive access is *not* necessary for external awareness.

4.4.2 Objections

As expected, those who favor cognitive theories of consciousness object to interpreting the partial report data as supporting a capacity mismatch (Kouider et al. 2010; Cohen and Dennett 2011; Phillips 2011; Brown 2011; Lau and Rosenthal 2011). Since the data

[2]In some cases the presentation of the array is so brief (or it has been otherwise interfered with) that it seems unlikely that there is a high detail representation of the characters anywhere in the visual system. In these cases, subjects are unlikely to be nonpropositionally aware of the specific shapes of all of the characters.

indisputably show that the partial report condition is superior to the full report condition, the cognitive theorists still need an account of this superiority. The only way to do this is by allowing that there is a highly detailed representation of the stimulus *somewhere* in the subjects' visual systems. Because they agree that there is such a representation, the only move available to them for denying the purported mismatch in capacity is to deny that the highly detailed information in iconic memory takes the form of a visual experience of the stimulus. That is, they argue that the information in iconic memory is unconscious. But this move faces an obvious difficulty: the subjects unanimously report that they were aware of the entire array, even after the array leaves the screen.

In order to overcome this difficulty, the cognitive theorists give the following response. First, they agree that the subjects are correct when they report that they were aware of some features of the stimulus besides the reported characters. But, they argue that the subjects' visual experience of the stimulus before the cue is actually grounded in a separate representation *with significantly less detail* than the one present in iconic memory. As a result, the subjects' experience of the stimulus actually has a much lower level of detail than the subjects claim it does. They argue that the reason that the subjects nonetheless report having a high detail experience of the stimulus is that they *overestimate* the amount of detail that they were aware of because of an illusion caused by the absence of attention. This illusion has been referred to as the 'inattentional inflation of subjective experience'. The exact account of the illusion - and the low detail representations that ground it - differs depending on which version of the cognitive theory one accepts.

The exact account of the illusion differs depending on which kind of cognitive theory one accepts. Specifically the higher-order theorists argue that the low detail representation is a generic representation of the unattended characters (Cohen and Dennett 2011; Lau and Rosenthal 2011; Brown 2011), while the global workspace theorists argue that it is a representation of character fragments (Kouider et al. 2010). On both views, subjects are actually aware of these low-fi representations of the characters in the array before the

cue. After the cue, they attend to the cued characters and become aware of the specific representations of them, allowing them to report those characters. They also report having an experience of the specific details of the entire array because of an illusion caused by not attending to them. In effect, the subjects mistake the generic/fragment contents that they were aware of for specific ones.

On the fragment proposal put forward by Kouider et al., in the absence of attention the character fragments are left unbound and as a result the subjects can, at best, be only aware of the character fragments. They offer some evidence for the claim that subjects are only aware of character fragments in unattended locations, which comes from a previous study performed by the same group. Kouider and Dehaene (2007) used a partial report task where the subjects were occasionally asked if any of the characters in the uncued rows were unusual. They found that subjects typically failed to report noticing instances where a character was rotated or replaced with an atypical symbol (e.g. a happy face), suggesting that they were merely aware of fragments of the character.

4.4.3 Reply

I have some doubts about these alternate explanations of the data. First, I want to address the evidence marshaled by Kouider et al. in favor of the fragment illusion. The task used by Kouider and Dehaene (2007) is similar in the relevant respects to a 'change blindness' task. But the phenomenon of change blindness does not force us to accept to any conclusions about what subjects are aware (or unaware) *of*. It is perfectly compatible with these findings that subjects were aware of the specific details of the thing that changes (or is unusual). These findings merely demonstrate that subjects are frequently unaware *that* that thing changed (or was unusual). But a failure to be aware of a particular fact does not imply anything about the perceptual experience of the thing itself. So this evidence does not favor of the fragment illusion any more than it favors the mismatch argument.

Second, with respect to the generic illusion argued for by the higher-order though

theorists, it is not clear to me how one could nonconceptually represent a generic alphanumeric character. But this would need to happen in order for their response to work unless they think that the subjects' awareness of the characters is *always* conceptual. But it is implausible that all awareness of external things is necessarily conceptual. The fragment illusion argued for by the GNW theorists fairs better here, as it is easy to see how one would nonconceptually represent character fragments.

But my main concern with the explanations given by the cognitive theorists is that positing generic (or fragment) representations as an explanation of the subjects' purportedly errant judgments will result in an *overload of working memory*. Why? Well, according to the cognitive theorist's own theory, these low detail representations must be encoded and maintained in visual working memory in order for the subject to be aware of them. But remember that, in the partial report condition, subjects can report 3.1 of the cued characters on average. This means that roughly 3 out of 4 slots are filled in visual working memory. This means that there is not much room for encoding the generic (or fragment) representations. Of course, things may not be quite as bad as they seem since the fidelity of the items to be stored in working memory *does* make a difference to how much can ultimately be stored (Brady et al. 2011). So perhaps it is possible that they are so low detail that they can fit in alongside all the encoded items that are reported. In the best case scenario it will be a rather tight fit, with room for little else.

What about in the uncued, free-report condition? In that condition the subjects were able to report 4.3 characters on average, which means that they had encoded and maintained just about the maximum number of items that visual working memory can hold. How are the generic/fragment representations going to be encoded and maintained in this case? There is no room for them, no matter how small they are. If there are no generic/fragment representations encoded in visual working memory, then the subject will not be aware of generic/fragment contents. If they are not aware of these contents, then there is no basis for the 'inattentional inflation of detail' to occur. The subjects are simply wrong when they say

that they were aware of the uncued characters in any way. The cognitive theorists have two options for dealing with the capacity limits faced in the free report condition: (i) they have to say that the subjects were aware of nothing other than the specific reported characters, or (ii) they will have to argue that the capacity of visual working memory is larger than four items. I do not think that either option is very plausible.

Nevertheless, for the sake of argument, let's suppose that the capacity of visual working memory turns out to be five items instead of four. In this case, if the capacity limit also turns out to be a *domain-specific* limit, then I think that the alternate response developed by the cognitive theorists may end up being empirically adequate. That is, if it turns out that there are separate working memory stores for visual items, auditory items, olfactory items, gustatory items, and somatosensory items, and each of those stores can hold five items, then I have no strong objections to the ability of the generic/fragment illusion to explain the results of the partial report paradigm. If, on the other hand, we expand the capacity of working memory to five items but the capacity limits on working memory are *domain-general*, then I think that the cognitive theory is still sunk. The capacity of working memory would be overloaded if the subjects in this task were aware of a *single* additional item. But what about the auditory cue? Where is that being stored? It must have been encoded and maintained in order for the subjects to successfully encode the letters in the correct row. And what of their bodily sensations? It seems exceedingly unlikely that they were completely unaware of their body during the task. After all, the original partial report task required the subjects to *write down* the cued characters. It would be incredibly revisionary to deny that the subjects were aware of the feelings in their hand as they wrote down their response. There are simply too many items that the subjects were plausibly aware of for working memory to encode and maintain them all if the capacity limits are domain-general.

Again, for the sake of argument, let us suppose that it turns out that the capacity limits on working memory are domain-specific. In this case, the cognitive theorists would be able

to explain how all these other items were encoded in different working memory stores, allowing the subject to be aware of them. This would mean that we have two empirically adequate explanations of the partial report paradigm. How do we decided between them? Choice between empirically adequate alternatives is difficult. It will often come down to the theoretical virtues of the respective theories. In the case at hand the relevant features are the comparative simplicity of the theories and how revisionary each theory is.

I think a view that implies a capacity mismatch does better on both counts. Such a view will only posits a single representation (in iconic memory) of the unreported characters and say that this representation is the one that grounds the subject's awareness of those letters. Cognitive theories, on the other hand, must posit at least two representations to explain the same data: an unconscious representation (in iconic memory) that explains the superiority of the partial report condition and the various generic/fragment representations (in working memory) that ground the subject's low fidelity awareness of the unreported characters. The mismatch theory is also less revisionary than the cognitive theory because it is able to take the subjects' reports at face value and avoids implying that the subjects are undergoing illusions caused by inattention. Given both of these virtues, I think that we should conclude that a theory that implies a mismatch in capacity between cognitive access and external awareness is preferable to one that does not. But even getting to this point required several charitable assumptions and I think it is very unlikely that all of them will hold up under scrutiny. So I think that we have ample reason to conclude that there is a mismatch in capacity between these two faculties.

4.4.4 Conclusion

The success of the content mismatch and capacity mismatch arguments suggests that we ought to abandon the proposal that cognitive access is the second necessary condition for external awareness. If we must abandon this proposal, what other options do we have that might better account for the various features of external awareness? In order to help us

cut the list of options down to size, I will use the arguments I have presented to determine what kinds of features that an acceptable proposal must have. First, I gave a conceptual argument against the sufficiency of representation, which concluded that, in order for a representation to furnish the subject with awareness of its content, it must bear *some suitable relation* to the subject of awareness. This is the first requirement for an acceptable proposal - it must require that the representation bear a relation to the subject of awareness. Second, I gave two different capacity mismatch arguments. The first argued that the capacity of sensory representation is much higher than the capacity of external awareness and the second argued that the capacity of cognitive access is smaller than the capacity of external awareness. This gives us the second requirement for an acceptable proposal - we must have no reason to suspect that there is a mismatch in capacity between external awareness and the proposed condition. Finally, I presented a content mismatch argument against the necessity of cognitive access for external awareness. I argued that external awareness can be grounded in nonconceptual content, while cognitive access cannot. This gives us the third requirement for an acceptable proposal - the proposed condition must allow for nonconceptual content to ground awareness. What options do we have that can meet these three requirements?

4.5 Cognitive Accessibility

Let us start by revisiting an idea suggested by Stanislas Dehaene. In presenting his theory of consciousness, Dehaene argues that consciousness is fundamentally tied to *information sharing*. I noted that there were several different ways to develop this idea and that Dehaene developed it in a way that required information to be cognitively accessed prior to being broadcast in the way that grounds external awareness. But, as we have seen, this kind of proposal faces insurmountable difficulties. Nonetheless, I think that Dehaene was on the right track in emphasizing the importance of this kind of information sharing. Perhaps there is another way of developing a theory based on this basic insight that can

avoid the objections leveled against cognitive theories of consciousness?

The most straightforward way to do this would be to accept the basic outline of his view. However, we could replace the competition for encoding in working memory with an additional broadcasting mechanism. This mechanism broadcasts representations from various input modules (e.g. the perceptual centers) to working memory. By accepting the basic outline, we agree that information broadcast from working memory to the cognitive faculties is conscious. By adding a second broadcasting mechanism that connects the input modules to working memory we add another locus of consciousness. There are now two mechanisms that broadcast conscious content: the one that broadcasts *to* working memory and the one that broadcast *from* working memory. The broadcasting mechanisms have different functions corresponding to their different targets. Broadcasting to working memory has the function of making representations accessible to working memory, while broadcasting from working memory has the function of making representations available for use by our cognitive faculties. In other words, broadcasting to working memory is necessary for cognitive accessibility, while broadcasting from working memory is necessary for cognitive access.

By saying that broadcasting to working memory is necessary for external awareness we are, in effect, requiring that a representation be cognitively accessible in order for it to furnish the subject with awareness of its content. That is not to say that content stops being conscious if it is actually accessed, just that the minimal condition for awareness is accessibility. This gives us a new proposal for the second necessary condition for external awareness: a representation must be broadcast in a way that makes it accessible to working memory in order for it to furnish the subject with awareness of its content. I will call this proposal the "accessibility view". I think it has a lot going for it. I will argue that it has the resources to explain the two key examples in this paper and is able to meet the three desiderata described above. Let us start by considering how it can explain the case of masked Mary and the results from partial report paradigm.

In the case of masked Mary, the accessibility view is able to agree that the data indicate that visual masking does not prevent the information about the stimulus from being fully processed, which means that subjects like Mary have a representation of masked stimulus. According the accessibility view, visual masking prevents the representation of the stimulus from being broadcast to working memory, rendering it cognitively inaccessible. Because it is not cognitively accessible, it is not in a position to furnish Mary with awareness of the stimulus. Thus, this proposal is able to vindicate our intuition that Mary is unaware of the masked stimulus.

In the partial report condition of the Sperling task, the accessibility view has a leg up on the competition since it is generally agreed that the information in the high capacity 'iconic' system is *accessible* to the subject. This is because the subjects in the task are able to report almost all of the characters in any of the cued rows. If some of the characters were inaccessible, then they could not be cognitively accessed after their row was cued, which would in turn prevent the subject from being able to report the presence of that character. But, according to the accessibility view, if the information in the high capacity system is cognitively accessible, then it furnishes the subject with awareness of its content. Happily this corresponds with the subjective reports of the participants: they unanimously report being aware of the entire stimulus - both during the exposure and for a short time after it leaves the screen.

The accessibility view is also better positioned to meet the three requirements described above. Recall that I argued that any adequate theory of external awareness must require that the representation of the object of awareness bear *some* kind of relation to the subject of awareness, and that there must be a match in both the capacity and content of external awareness and proposed condition. By proposing cognitive accessibility as a necessary condition for external awareness, we meet the first requirement. A representation is cognitively accessible when it bears the accessibility relation to the subject. The hope is that this accessibility relation is strong enough to ground the subject's being transitively

conscious of the representation's content. What about the second and third requirements?

I have already argued that external awareness can be (and often is) grounded in representations with nonconceptual content. Given this, in order to establish a content match between the two, it must be the case that nonconceptual content can become cognitively accessible as well. Since I have already argued that cognitively accessed contents must be conceptual, we should expect cognitively accessible contents to be *conceptualizable*. That is, the only requirement placed on accessible contents is that they be able to be conceptualized when accessed. Since nonconceptual content can be conceptualized, we should expect that nonconceptual content can be accessible to working memory. This is one of the advantages of adding a broadcasting step prior to working memory: it allows unencoded, unconceptualized contents to be broadcast and thereby furnish the subject with awareness of that content. We have good reason to believe, then, that there is a match in the contents one can be externally aware of and those that can become cognitively accessible.

Arguing for a match in the capacities of external awareness and cognitive accessibility is a much more difficult task. Part of the reason for this is that it is difficult to precisely calculate the capacity of external awareness. It has been suggested to be around 8-32 *visual* items at once (Block 2007), which I agree is probably in the right ballpark. I suspect that the total number of items a subject can be simultaneously aware of is much higher as we are the kinds of creatures that have several different sensory systems and all of these systems are engaged in the project of representing the properties of external items. Even though we may be able to get a very rough estimate of the number of items that a subject can be externally aware of, it is nearly impossible to get a theory neutral estimate of the number of mental states that are cognitively accessible at any given time. The problem is that we currently have no methods for directly measuring whether a mental state is cognitively accessible. Our only option is to look at what our best theories of cognitive accessibility say about the matter. But doing this will beg the question one way or the other, depending on the view selected. Because it is so difficult to build a case for a positive match in the

capacities of cognitive accessibility and external awareness, we might be better off simply considering whether or not there are any good reasons to expect the presence of a *mismatch* in capacity.

I do not think there are any reasons for concern here, provided that we have an adequate definition of cognitive accessibility. If the definition proves to be inadequate, then it will face obvious counterexamples. For example, Dehaene and Naccache (2001) define accessible representations as representations that *can* be amplified by attention. But, Block (2007) rightly argues that cognitive accessibility, so defined, has a much higher capacity than external awareness, resulting in cases of where a representation can be attended but still not result in awareness (p.492). An example of such a case can be found in Jiang et al. (2006). They found that by using continuous flash suppression they could suppress awareness of a nude image. They also found that the subject's attention was either attracted to or repelled from the location of the nude, depending on the gender of the nude and the gender and sexual orientation of the subject. For instance, in the case where the image was a nude female and the subject was a heterosexual male, the invisible nude image attracted their attention to the location of the nude. This indicates that there can be states (e.g. the representation of the nude) that are amplified by attention, but still do not result in awareness. As a result, I agree that when cognitive accessibility is defined in the way that Dehaene and Naccache (2001) define it, the capacity of accessibility ends up being much larger than the capacity of external awareness. This simply tells us that we need a better way of defining cognitive accessibility.

Unsurprisingly, my proposal is that we should define cognitive accessibility in terms of content that has been broadcast to working memory. Unfortunately this does not help us much because it is not clear what the *psychological marker* of broadcasting to working memory would be (without consulting theory under consideration). Thus, it is hard to directly detect instances where a representation has been broadcast, but that is precisely what

would be required to finding evidence of broadcasting without awareness. Given this difficulty, instead of aimlessly looking for counterexamples at the psychological level, I think we will have more success by descending to the neural level. But, in order for this move to be productive, the information carried by coalitions of neurons must necessarily correlate with the content of the mental states realized by those coalitions at the psychological level. As a result, facts about what is happening at the neural level can only help us in this task if the mental representations that we care about are grounded in the activity of the coalitions of neurons that carry the relevant information.[3] If this assumption is correct, then we should be able to make some progress in this task by considering the necessary conditions for the broadcasting of the information carried by a neural coalition.

4.5.1 Neural Mechanism of Accessibility

One proposal is that the strength of the neural coalition underlying a mental representation plays a decisive role in whether or not that representation ends up furnishing the subject with awareness of its content (Farah 1994). In cases where the strength of the neural coalition is sufficiently high, the information the coalition carries about the stimulus is broadcast to working memory. Once this occurs, the mental representation of the stimulus becomes cognitively accessible and the subject becomes aware of the represented stimulus. In order for a proposal like this to work, three conditions must obtain. First, there must be a rather large number of different neural coalitions present in one's sensory systems at any given time. This is plausible based on what we know about sensory processing. Second, there must be variation in the strengths of the different coalitions. This is also plausible given that the environment contains a wide variety of stimuli of different intensities. Third,

[3] In making this claim I do not intend to endorse the identity theory *or* functionalism. Both theories agree that there are realizers and functional descriptions of them. They differ in which one they identify the phenomenon with. Functionalism identified the phenomenon with the functional description, while the identity theory identifies it with the realizer. I think that you can agree with everything I say here, regardless of your preferences, since you should be able to supplement the view developed here with your preferred identity claim.

there must be some kind of 'test' that is applied to these coalitions, which results in the broadcasting of the ones that are 'sufficiently strong'. There are two different kinds of test one might use here. First, you might use a threshold that tests the *absolute strength* of the coalitions relative to a baseline. If they are stronger than the baseline, then they are broadcast to working memory. Second, you might use a competition that tests the *relative strength* of the coalitions. If a coalition is the strongest of the current competitors, then it wins the competition and is broadcast to working memory. Which kind of test fits best with what we know about broadcasting and external awareness?

On the first option, all that matters is getting above some absolute strength threshold. As a result, there are theoretically no restrictions here on how many states (or how few) can be broadcast at any given time. On the second option, even if all the states are super strong, only the strongest will be broadcast. Similarly, if they are all incredibly weak, some state will still be broadcast become accessible even if in ordinary circumstances none of them would win. Given these implications, I think it is clear that we should go with the approach that uses a threshold that tests the absolute strength of the coalition. After all, one of our goals here is to establish a match in the capacity of cognitive accessibility and external awareness. Using a competitive test would undermine that match, while a threshold test would allow for a match, even though it would not necessarily guarantee it.

If the strength of a neural coalition is going to determine whether or not a representation is broadcast, then we need to know what features of a coalition determine its strength. There are two components typically cited as influencing the signal strength of a neural coalition: the average firing rate of the neurons in the coalition and the degree of synchrony amongst those neurons. In general, the higher the firing rate and the more synchronous their firing is, the stronger the neural representation of the feature. Although having a high firing rate (30-80 Hz) is a necessary condition for being broadcast, it is not sufficient. We know this because there is ample evidence from research on visual masking (Kouider and Dehaene 2007), visuospatial neglect (Rees et al. 2000; Vuilleumier et al. 2001), attentional

blink (Luck et al. 1996), and binocular rivalry (Brown and Norcia 1997), which shows that there can be strong signals in the sensory cortices that do not result in awareness.

What is interesting is that further research in many of these areas has indicated that even though the neural signals are similar between cases where a subject is aware of a stimulus and cases where a subject was not, there were significant differences in the degree of synchrony across these conditions. In particular, it has been found that invisible masked stimuli have lower synchrony in the gamma band than visible stimuli (Summerfield et al. 2002). It has also been found that stimuli that go unnoticed in attentional blink tasks have lower early synchrony in the gamma range (Fell et al. 2002). Finally, it has been found that the image that becomes suppressed in cases of binocular rivalry have lower synchrony in the gamma band right before the switch (Doesburg et al. 2005). Curiously, I was unable to find any studies examining the role of gamma synchrony in visuospatial neglect, but given the wealth of evidence from the other areas of research, I think that we are safe in predicting that there would be a difference in gamma synchronization in visuospatial neglect as well.

The general trend that we can extract from these studies is that in cases where a subject is aware of a stimulus we find increased synchrony at gamma frequencies in the coalition carrying information about that stimulus, while in cases where the subject is not aware of the stimulus we find decreased synchrony those same frequencies. I think that we should conclude that in order for information about a stimulus to be broadcast to working memory (and in turn furnish the subject with awareness of it), the coalition of neurons carrying information about that stimulus must have a sufficiently high average firing rate (e.g. in the gamma band, between 30 and 80 spikes per second), and they must have a *high degree synchrony*. This requirement makes sense from a design perspective. There has been some theoretical work done that suggests that high levels of synchrony in the gamma band promotes the transfer of information to other areas (Salinas and Sejnowski 2001). This gives us another reason to expect that gamma synchrony would assist in broadcasting of

information long distances to the regions underlying working memory.[4] It also favors a threshold view over a competitive view. Since one of the main motivations for nominating synchronous firing as a necessary condition for accessibility is that high levels of this kind of firing promotes communication between far flung regions of the brain and there will be a minimum level of synchrony that must be attained before there is any improvement in the transfer of information, we should expect that there is a baseline degree of synchrony required before it impacts the communication of information.

We can think of this mechanism as working similarly to two people trying to communicate with each other over a great distance. Just as when you gradually turn up the volume on your megaphone the message eventually gets loud enough for the other person to understand, so too increasing the synchrony of a neural coalition in the gamma band eventually results in the communication of information to distant brain regions. Because of this similarity, we should think of theories that use strength-based thresholds for determining which information is broadcast as "volume control" theories. The name originally comes from work by Daniel Dennett (1978) and was taken up by Christopher Hill (1988). I think it is particularly apt for this kind of theory.

But where does the "control" part come into play? How is the 'volume' of a neural coalition turned up or turned down? There are several different processes that have been implicated in increasing synchrony and firing rates. One process is the kind of recurrent feedback that Lamme (2003) argues is necessary for external awareness. I think that Lamme is right that recurrent feedback plays an essential role in helping many coalitions reaching the threshold for broadcast, but it is not clear to me that it is necessary in every case

[4]Given this model of broadcasting to working memory, we should expect that when gamma synchrony is sufficiently high that the neural coalition is in a position to cause changes in the regions responsible for working memory. Such changes would be evidence that the stimulus is beginning to be encoded in working memory. Interestingly it has been found that the neural coalitions encoding a stimulus can become *phase-locked* with neural coalitions in other parts of the brain, particularly in frontal cortex. This phase-locked synchronization across distant portions of the brain has been suggested to promote the flow of information between them (Varela et al. 2001). So perhaps phase-locking at gamma frequencies between the working memory regions and the neural coalitions broadcasting their content is the necessary condition for working memory encoding to take place.

because it is not the only way for a coalition with a high strength to come into existence. Another process is the modulation caused by attention. Attention has been demonstrated to increase both the firing rates of neural coalitions and their synchrony in the gamma band (Chapter 6). This means that attention can also play a role in promoting the broadcasting and encoding of the information carried by neural coalitions. I think it is quite likely that some stimuli are so intense that the encoding of their features produces a coalition with sufficient strength to be immediately broadcast. No help needed. Some evidence for this claim comes from the fact that subjects can give responses about the gist of a briefly presented (around 20 ms) scene as fast as 260 ms after its presentation (Rousselet et al. 2005). This is reaction time is faster than top-down attention could be deployed to the whole scene and faster than the time it would take for a non-negligible amount of recurrent processing to take place, which suggests that some stimuli are sufficiently strong to produce neural coalitions that broadcast their content without assistance.

If any neural coalition with sufficiently high synchrony in the gamma band can broadcast its content, can there be a match in capacity between broadcasting to working memory and external awareness? It should be clear that it is not too restrictive. It should definitely be possible for 30-40 clusters of neurons to hit that level of synchrony simultaneously. The only worry is whether or not such a condition is sufficient for awareness. That cannot be determined at present, but if it does prove to be sufficient I think that will mean that we need to start the search for a third necessary condition for external awareness. It will not undermine the proposal that cognitive accessibility is necessary for external awareness.

4.6 Conclusion

In conclusion, I have argued for two claims. First, I have argued that cognitively accessing a representation is not necessary for external awareness, on the basis of the content and capacity mismatch arguments. Second, I have argued that cognitive accessibility is the

other condition necessary for external awareness. I then discussed the mechanism underlying cognitive accessibility, arguing that it only occurs once a representation's content has been broadcast to working memory. Since the best way to understand this kind of broadcasting is at the neural level, I agued that the strength of the neural coalition grounding the representation determines whether or not the information is broadcast. When a neural coalition reaches the baseline strength required to pass the threshold for broadcasting, it is sent to working memory.

5 INTERNAL AWARENESS

5.1 Introduction

In this chapter I will evaluate proposals for the conditions necessary for internal awareness of one's own mental states. Unfortunately, much of the literature relevant to this topic has focused more on the process of introspection and how it produces justified introspective judgements than on internal awareness per se. It is quite common to find projects that are only interested in understanding introspection for the sake of epistemology, particularly evident in the literature on self-knowledge. My project here is not an epistemological one. It is a metaphysical project aimed at understanding the conditions necessary for introspective awareness.

Elsewhere I have argued that there are two varieties of awareness: nonpropositional awareness of objects and propositional awareness of facts (Chapter 1). Introspection typically involves *both* varieties of awareness. In introspecting a state one first becomes nonpropositionally aware of the state's experiential properties. That is, one has an 'introspective experience' of that state. One then applies introspective concepts to the experience in order to form judgments about it, resulting in awareness of the facts so judged. I think, then, that all acts of introspection involve internal awareness. But I do not necessarily think that the converse is true. It seems likely that there are instances of internal awareness that are not properly understood as being introspective.[1] Since nonpropositional awareness of a state's experiential properties is the foundation of introspection, that will be the variety of

[1]This is particularly evident if you think that introspection is an intentional act undertaken by a subject. Since one can become internally aware of a mental state involuntarily (e.g. an intense pain that you would prefer to ignore), to the extent that introspection is a voluntarily initiated act of awareness, cases of involuntary internal awareness would not be properly understood as being introspective.

awareness I focus on here.

Elsewhere I have argued that there are two conditions necessary for nonpropositional awareness of an *external* object. First, the subject must have a nonpropositional representation of the relevant object (Chapter 3). Second, that representation must be cognitively accessible (Chapter 4). When both conditions are met, the subject becomes nonpropositionally aware of the represented object. More formally:

> For all external objects O, a subject is aware of O only if they have a cognitively accessible representation of O.[2]

Since I think it is preferable to adopt of set of views that hang together nicely, my strategy for developing a theory of internal awareness will be to consider the different options available for *extending* this theory of external awareness to cover the case of internal awareness. I will consider two ways of doing this. The first theory will ground the difference between internal awareness and external awareness in a representational difference. In particular, it will extend the theory by removing the scope restriction when quantifying over the possible objects of awareness. That is, it will say that in order to become aware of any object one must have an accessible representation of it. The second theory will ground this difference in awareness in a nonrepresentational difference. In particular, it will extend the theory by arguing that a difference in the way that one allocates one's attention explains the difference between internal and external awareness. Ultimately, I will argue that the nonrepresentational extension should be preferred.

---

[2]I am using a conditional instead of a biconditional in this formulation in order to hedge my bets against possible counterexamples. I am convinced that both of these conditions are necessary for awareness. I am not quite as convinced that there are no other necessary conditions for awareness. Perhaps there is a third condition that will be revealed by a clever counterexample. That said, even though I am only committed to the necessity of these two conditions, I will typically talk as though the two conditions are jointly sufficient as well.

5.2 Representational Extension

The most straightforward way to extend the accessibility theory of external awareness to cover cases of internal awareness would be to simply remove the scope restriction when quantifying over the objects one is aware of. Instead of only quantifying over *external* objects, we can quantify over *all* objects. By removing the scope restriction, the view says that in order to be aware of an object - regardless of the ontological status of that object - one must have a representation of it and that representation must be broadcast to working memory in a way that makes it cognitively accessible. In the cases where the object one is aware of is a mental state with representational content, the accessible representation that furnishes awareness of it is a *higher-order* representation. Since awareness of things (as opposed to facts) is grounded in nonpropositional representations of those things, internal awareness of a mental state must itself be grounded in some kind of nonpropositional representation. Although perceptual representations are the clearest examples we have of nonpropositional representations, I do not think that the nonpropositional representation deployed in this theory takes the form of a perceptual representation per se.

Since this theory says that an accessible nonpropositional higher-order representation is necessary for internal awareness, it will be helpful to compare it to the two standard versions of the higher-order representation theory of internal awareness. In particular, the theory under consideration here is quite similar to the higher-order perception theory defended by Armstrong (1980) and Lycan (1996). However, calling it the higher-order 'perception' theory is a bit of a misnomer since neither author argued that you literally perceive your mental states. Instead the discussion typically proceeded in terms of 'internal monitoring' and attending. I think the most charitable interpretation of what they say about internal awareness is that they think that there must be a higher-order nonpropositional representation in the internal monitor in order for the subject become aware of the represented mental state. As a result, we both agree that one could be aware of a mental state even if they did not possess the concepts required to specify the correctness conditions for the higher-order

representation of that mental state. If this is right, then the theory developed here is very similar to the views developed by Armstrong and Lycan.

The other variety of the higher-order representation theory of internal awareness is known as the higher-order thought theory. This view has been ably defended by Rosenthal (2005), who argues that a subject must have a propositional representation of a mental state in order to become aware of that mental state. Let us start with the positives. I like this view as a theory of the conditions necessary for awareness of *facts* about one's mental states. In fact, it is the exact view that I adopt. However, the view is typically proposed as a theory of the conditions necessary for awareness of mental state *themselves*. The problem is that propositional representations represent *propositions*, not particular things. Of course, one might be able to become aware of a particular thing *indirectly* via their awareness of a proposition about that thing. For example, I might become indirectly aware of Paris, by having the thought that "Paris is the capital of France". However, such cases are not direct enough to be considered cases where a subject is aware of the object that the proposition is about. Only nonpropositional representations have the function of directly representing particular things. Since mental states are particular things, direct awareness of them must be grounded in a nonpropositional representation.

But, even if we suppose that propositional representations could directly represent particular things, I still think that there are significant differences between the view developed here and the higher-order thought theory. The higher-order thought theory implies that all cases of internal awareness are cases where one is aware of facts about the thing one is aware of. But that implication makes it difficult to understand cases of introspective confusion. In such cases one is aware of the experiential properties of the mental state, but has no idea what to make of them. Such cases, taken to the extreme, are cases of nonpropositional awareness of a particular thing (and its properties) in the absence of awareness of any facts about it at all. Such occurrences are possible on the view I develop, but not on the higher-order thought view. This suggests that the higher-order extension of the accessibility theory

is much closer to the higher-order 'perception' theory than the higher-order thought theory. That said, there is a feature of the view developed here that is unlike either of the traditional higher-order theories.

The truly distinctive feature of this version of a higher-order theory is that it restricts the number of higher-order representations that are in a position to furnish the subject with awareness of a mental state. The usual way of developing a higher-order theory of internal awareness is to nominate a particular kind of representation as the variety of representation essential for internal awareness and stop there. But for the same reasons that simply having a representation of an object is insufficient for external awareness, simply having a representation of a mental state will be insufficient for internal awareness. Having a representation of the right type that is just floating around in one's mind does not make one aware of the represented mental state. Instead, that representation must be related to the subject in the right way in order to confer awareness of its content. In particular, the representation be cognitively accessible. It must have been broadcast to working memory and ready to be encoded. By placing this further requirement on the higher-order representation, I think that this theory is better positioned to accommodate the conceptual argument from Chapter 3. In order for a subject to be transitively conscious of an object, the representation of that object must be suitably related to the subject of awareness. By requiring cognitive accessibility for consciousness of a mental state, we have a chance of satisfying that requirement. As a result, I think that this version of the higher-order representation theory is the most plausible version of the theory available.

5.3 Nonrepresentational Extension

Instead of grounding the difference between internal and external awareness in terms of a difference in representational content, we could ground it in some nonrepresentational difference. What nonrepresentational faculty might be able to play such a role in generating internal awareness? I think that attention is a good candidate for the job. If someone were

to asks you to focus on some portion of your phenomenological field, how would you do this? If you are like me, you would do this by directing your attention inwards. As a result of cases like this, I want to suggest that attention plays a role in the initiation of internal awareness. In particular, I think that attention plays a key role in the kind of internal awareness one has when introspecting. It is no coincidence, then, that attention, like introspection, can be voluntarily controlled. Finally, as argued in Sauret and Lycan (2014), attention is not a representational faculty. Attention makes changes to pre-existing representations and such changes are frequently made on the basis of other representations, but attention does not in and of itself represent what is attended to. I will have more to say about this in the next chapter. Given all of this, how would a view like this work?

We should start by considering what kind of difference in attention is going to ground the difference between internal awareness and external awareness. My suggestion is that the relevant difference in attention would be in the way one *allocates attention*. In ordinary cases of external awareness, attention is allocated by default to the experiential properties attributed to the external object. For example, when a subject is aware of a purple eggplant, attention is allocated to the attributed experiential properties - e.g., purpleness, ovalness, and shininess - and the subject thereby becomes aware of the eggplant.[3] However, we can *reallocate* attention to also include the unattributed experiential properties of the mental state. Attending to the unattributed experiential properties results in their becoming cognitively accessible, putting them in a position to be conceptualized and for judgments to be formed about them. Most importantly, by becoming cognitively accessible, the subject becomes aware of the accessible experiential properties. Since unattributed properties are not attributed to anything else, one necessarily experiences them as being had by the mental state in which they inhere. Since a mental state is the object of awareness in this case, awareness of unattributed experiential properties is properly classified as a case of internal

[3]I do not mean to imply here that I think that attention is necessary for external awareness. I do not. On the view being presented here, attention would only be necessary for internal awareness.

awareness.

It should now be clear how a difference in the allocation of attention can make a difference in the type of awareness one has. Attending experiential properties results in their being broadcast to working memory, making them cognitively accessible. This, in turn makes the subject aware of those properties. When those experiential properties are attributed, it is a case of external awareness, when those properties are unattributed, it is a case of internal awareness. In cases like perceptual experience, attention is by default allocated to attributed experiential properties, but can be turned to the unattributed ones. Since the present view accounts for the difference between internal awareness and external awareness in terms of differences in the allocation of attention it is a version of the 'attention view' (AV) defended in Sauret and Lycan (2014). This view, in short, says that when attention is allocated to unattributed experiential properties, the subject becomes directly aware of them; when attention is not so allocated, the subject is unaware of them.

The higher-order representation theories of internal awareness provide a nice contrast on this point. On those views, subjects can only become *directly* aware of attributed experiential properties. Let me explain. If you want to become aware of an external object, you must first represent that object and attribute certain experiential properties to it. It is in virtue of being aware of those properties that one is aware of that object. But if one wants to become aware of the unattributed properties of that state one cannot just attend to those properties and become directly aware of them. No, you must instead represent the state all over again and attribute experiential properties to it. It is only by being aware of those higher-order attributed properties that one might become aware of the original unattributed properties of the mental state. The only reason one would need a higher-order representation is if one were unable to become directly aware of unattributed experiential properties. Because of this contrast, one way of deciding between these two theories will be based on whether or not you think that we can become directly aware of unattributed experiential properties. That said, how does this proposal work for the attributed experiential

properties?

5.3.1 The Transparency of Experience

I have already argued that when attention is turned towards attributed experiential properties one attends to - and becomes aware of - the *external objects* that those properties are attributed to. But, that means that attending to these properties results in a case of external awareness, which creates a tension in the view. In order to resolve this tension we need to decide between two options: either there must be some *other way* of attending to attributed experiential properties that allows the subject to become internally aware of them, or one simply cannot to become internally aware of them. Unfortunately, neither option is particularly compelling. The first option seems to be a nonstarter. Given what we know about the way attention works from our best scientific theories, we have no reason to believe that there are multiple ways of allocating attention to the same property. You either attend to it or you do not. At best you might be able to get a difference in the *degree* of attentional influence, but there is no reason to suppose that a difference in the degree of influence will translate into a difference in the kind of object one is aware of. The second option is less problematic, but still somewhat worrisome as it is commonly supposed that one can be internally aware of *all* the experiential properties of one's mental states.

Even though the second option clashes with this intuition, I think that we have to go with it. Adopting this view means that when attention is allocated to the attributed experiential properties of a mental state, one is aware of the objects that those properties are attributed to. There is no way to become internally aware of them as had by the mental states in which they inhere. Although this view is revisionary, it is not as crazy as it might sound. After all, it is common in the literature on perceptual phenomenology to claim that perceptual experience is *transparent* to the objects the experience is about. The intuition that drives the transparency thesis is that it *seems to us* as if we are directly aware of external objects and their properties, rather than being aware of the properties of our experiences.

As Harman (1990) puts it:

> When Eloise sees a tree before her, the colors she experiences are all experienced as features of the tree and its surroundings. None of them are experienced as intrinsic features of her experience. Nor does she experience any features of anything as intrinsic features of her experiences. And that is true of you too ... Look at a tree and try to turn your attention to intrinsic features of your visual experience. I predict that you will find that the only features there to turn your attention to will be features of the tree. (p.667)

I think that Harman is right that it we are unable to attend to the attributed experiential properties *as* properties of one's experience. The *only* way to attend to (and become aware of) attributed experiential properties is as attributed to an external object. Some transparency theorists also go on to claim that the *only* kinds of experiential properties that one can become aware of are attributed properties (Harman 1990; Tye 1995). But we definitely do not want to accept this further claim. Doing so would, in effect, eliminate the possibility of internal awareness entirely. All awareness would be external awareness, a view that is almost certainly wrong. Even ignoring that implication, this thesis cannot be right. Briefly, even in the best case for the view - visual experience of an object - one can become aware of unattributed experiential properties. For example, in having a visual experience one is aware of the visual properties attributed to the object, but one can also become aware of the 'visualness' of that experience. But 'visualness' is not a property that is attributed to any object. As a result, the view cannot be right.

Even though the extreme transparency thesis is wrong, I still think that the moderate thesis should be accepted. It fits nicely with my own experience to say that one can only become aware of attributed experiential properties as had by the objects to which they are attributed. Given this, how does the nonrepresentational extension work? Well, one is typically just externally aware of the attributed experiential properties of a state, however, when one turns one's attention to the unattributed experiential properties one becomes internally aware of the mental state that has those properties. But this raises a question. Since most

states have both kinds of properties, what happens when one attends to both kinds of experiential properties at once? The answer is that the dichotomy between internal awareness and external awareness should not be taken to imply that an experience must be wholly composed of one type of awareness. Instead, it is typically the case the we are aware in a *mixed* state of awareness, where one is both internally aware of unattributed experiential properties and externally aware of attributed properties.

I think that this fits nicely with the way that our states actually appear to us. Pains are a particularly useful example because being aware of a pain essentially involves being in a mixed state of awareness. Some of the experiential properties of a pain are experienced as attributed to a certain body part. I do not simply have a pain, I have a pain *in my leg*. Other experiential properties of a pain are not experienced as attributed. For example, the affective aspects of the pain, which motivate you to favor that leg, are not experienced as attributed to the leg itself. Instead, they simply make you resistant to using it. Despite these different components present in cases of pain, when one is aware of a pain, one is aware of all these different experiential properties simultaneously. It takes careful introspective work to divide out the different components of the phenomenology of pain.

5.3.2 Empirical Commitment

One concern that I have about this way of extending the accessibility theory is that, given the theory of attention I defend elsewhere (Chapter 6), it ends up making a relatively substantive assumption about the neural implementation of mental states. This is because what attention does at the neural level when it selectively modulates some property is it directly modulates the firing pattern of the subset of neurons that carry information about that property. This is, for example, how it works with perceptual states. Perceptual states have a map-like representational structure where each subset of the map is responsible for representing the experiential properties attributed to that portion of space. This makes it easy to selectively attend to any of the properties that an object is represented to have. You

simply modulate the neurons responsible for encoding that property.

Although I have been largely focused on how we become internally aware of perceptual states, I do think that one can become internally aware of occurrent propositional attitudes like beliefs and desires. Because of these commitments, if we are going to make the attention view work as a theory of internal awareness in general, each type of mental state must be instantiated in a way that allows attention to selectively modulate the unattributed experiential properties of that state. In order to allow for this, all mental states must be implemented in a way that structurally decomposes into basic units. This does not necessarily mean that it must be map-like, but it does mean each subset of neurons must be responsible for encoding a particular experiential property, whether that property is attributed or unattributed. If all mental states are instantiated in the brain in this way, then it should be possible for a subject to selectively attend to any subset of their experiential properties at will. And, as a result, it should be possible for any of the experiential properties that a mental has to be broadcast to working memory and become cognitively accessible.

5.3.3 The Acquaintance Theory

The nonpropositional theory developed here bears certain similarities to the acquaintance theory of internal awareness. Since Russell (1912) is the source of the acquaintance theory, we will begin by looking at his characterization of it. Russell says that "we shall say that we have *acquaintance* with anything of which we are directly aware, without the intermediary of any process of inference or any knowledge of truths" (p.46). The metaphysical relation of acquaintance, then, is one that can obtain only when there is no mediation between the subject and the object. Russell specifically singles out inference and knowledge of truths as objectionable forms of mediation, but I suspect that he also would have objected to representation as well. Mediation by nonconceptual representations would prevent acquaintance just as easily as mediation by inference. So, the definitive claim of the

acquaintance theory is that in introspection we are *directly* aware of the experiential properties of our states. When we reflect on one of our current sensations, an itch perhaps, we are directly acquainted with the itch itself.

The attention-based theory developed here agrees that awareness of experiential properties is similarly direct. Awareness of the experiential properties of one's mental states is not mediated by representations, conceptual activity, or steps of inference. Because they are similarly direct theories of awareness, I want to suggest that we should categorize the attention view as a version of the acquaintance theory of internal awareness. In doing so I do not mean to imply that it adopts the baggage customarily associated with acquaintance views. I only mean that we can help illuminate the features of the attention view by thinking in terms of being acquainted with the experiential properties of one's mental states.

5.4 Engineering Constraints

The higher-order extension seems to be the default view. It simply involves iterating the view already argued for elsewhere. If the view is correct for the external case, then we should assume that, ceteris paribus, it will work for cases of internal awareness as well. In order to determine whether the ceteris paribus clause holds here, we need to ask why a representation is necessary in cases of external awareness. The answer seems to be that our mental faculties cannot reach out into the world and interact with physical objects. In order to overcome this limitation, we must have a sensory representation of those objects that stands in for them as a proxy of sorts. Anything that the mind needs to interact with has to be represented by a mental state. We can think of this as an engineering constraint on how the mind functions: mental processes can only interact with mental items. This engineering constraint explains why, in cases where a subject is aware of a physical object outside their mind, a representation of that object is necessary for awareness of it. Awareness is itself a mental relation between the subject of awareness and the properties of a mental state.

Facts and propositions are non-mental items, regardless of whether the fact is about a

physical thing or a mental thing. A fact about a mental state might be called a meta-mental fact, but it is not itself ontologically mental. As a result, all facts must be propositionally represented in order for them to have an impact on the subject. Similarly, external objects in one's environment are almost by definition non-mental. As a result, in order to become aware of an external object, we must mentally represent that object and then bear an awareness relation the represented properties. The only exception to this rule occurs when the item is already mental. Mental states and their properties are, by definition, already present in the mind. As a result, there is no need to re-present them. The subject can already directly interact with them without any stand-ins or proxies.

This asymmetry in the nature of the objects of awareness produces an asymmetry in the requirements for awareness of those objects. In order to be aware of a fact, you must have a suitable propositional representation of that fact. Similarly, in order to be aware of an external object, you must have a suitable nonpropositional representation of that object. However, one does not need a nonpropositional representation of a mental state in order to become aware of that mental state. A mental state is, by definition, already in the mind. They are already in a position to be presented to the subject; they do not need to be *re*presented. Positing a nonpropositional representation of a mental state would unnecessarily complicate the picture.

Given this important difference between the requirements for awareness of mental items as opposed to non-mental items, I think that we should conclude that the ceteris paribus clause does not hold here. All other things are *not* equal between the two cases. Non-mental items must be mentally represented in order to get a foothold in the mind, while mental states are already present within the mind. Therefore, we need not assume that the higher-order representation view is the default view. In fact, it is likely the reverse. Considerations of simplicity suggest that we should expect representation to be used only when needed. As a result, we should expect the nonrepresentational view to be the default view.

5.5 Appearance vs. Reality

The argument in the last section was not a decisive argument against the higher-order representation theory of internal awareness. It simply concluded that the burden of proof should not be placed on the nonrepresentational view to justify the absence of a representation. Instead the burden should be placed on the higher-order theory to justify the presence of a seemingly unnecessary representation. Even though I think that this argument is successful in shifting the burden of proof, I will nonetheless take it upon myself to give an argument for why we ought not include a representation in our theory of internal awareness.

Any theory that uses a representation of an object to mediate the subject's awareness of that object will necessarily have to distinguish between the appearance of an object and the reality of that object. The appearance of an object depends on the properties it is represented to have, while the reality depends on the properties it actually has. Where ever there is a distinction between appearance and reality, there is room for error. As a result, higher-order representation theories of internal awareness imply the possibility of error in the act of awareness itself. That is, according to the higher-order representation theory, the properties you aware aware of your mental state as having (the appearance) may not actually be the properties that your mental state has (the reality).

Of course, a theory's imply a distinction between appearance and reality is a virtue when it comes to explaining awareness of external objects. After all, these kinds of errors are relatively common in perceptual awareness. As a result, any suitable theory of external awareness needs to be able to explain how and why such errors occur. This explains, in part, why representational theories of external awareness are so plausible - they have a ready answer for what happens when appearance and reality diverge. Our purpose here, however, is to evaluate theories of internal awareness. Is there also an appearance and reality distinction when it comes awareness of one's own mental states, as implied by the higher-order theories?

I do not think there is. It is commonly thought that one cannot be wrong about their mental states. I suspect that the reason that the infallibilist intuition is so widespread is that most people implicitly believe that there is no difference between appearance and reality with respect to the experiential properties of one's mental states. The way that the experiential properties seem to you when you are nonpropositionally aware of them is the way that they *are*. There is no room for error. You can neither hallucinate having a mental state that you do not have nor be the victim of an introspective illusion with respect to the properties of a mental state that you do have. The appearance is the reality. Interestingly, some contemporary philosophers agree with this claim. For example, Christopher Hill (1991) says that "there is no appearance/reality gap in the case of sensations" (p.127). While this claim is restricted to the special cases of sensations, I see no barrier to extending the thesis to cover all cases of internal awareness. When one is aware of the experiential properties of their states, one is aware of the reality of those experiential properties. I find this intuition very compelling.

Some extend this thesis about nonpropositional awareness to cover propositional awareness as well. That is, they think that the propositional states like belief, judgment and knowledge cannot be wrong with respect to mental states. A frequent response to the infallibilist intuition with respect to *self-knowledge* is to point out all of the cases where our introspective judgments are wrong. Schwitzgebel (2008) presents several cases where it seems that our judgments about our mental states go wrong. For example, suppose that your spouse mentions to you that you seem to be angry about being stuck doing the dishes again. You reflect on your current phenomenology and come to the conclusion that you are not angry about having to wash the dishes. However, suppose that it turns out that you are wrong about this. You might even come to agree that you were angry after the fact. There are a litany of cases like these where our introspective judgments end up erring, which suggest that such an extension of the basic view would be misguided. However, none of these cases pose any problem for the core view that one cannot be wrong about the

experiential properties that one is aware of nonpropositionally. Facts about the reliability of nonpropositional awareness do not in and of themselves imply anything about the reliability of propositional awareness of mental states, and vice versa. I think that propositional states are necessarily error prone because they are composed of concepts and an application of concepts implies the possibility of error.

In conclusion, I find the intuition that there is no distinction between appearance and reality for nonpropositional awareness of our mental states very compelling. If it turns out to be correct, then we ought to resist generalizing my theory of external awareness by simply removing the scope restriction. This move creates a higher-order representation theory of internal awareness, which places an objectionable form of mediation in nonpropositional awareness of mental states. Instead, we should generalize the theory by adopting the attention view. Internal awareness is produced when unattributed experiential properties become cognitively accessible as are result of attention being directed towards them.

5.6 Awareness is Acquaintance

I noted earlier that the attention view of internal awareness was similar enough to the acquaintance theory to be considered a version of the theory. This was because both the attention view and the acquaintance theory think that awareness of experiential properties is unmediated by inference, representation, or conceptual activity. As Russell puts it, "we shall say that we have *acquaintance* with anything of which we are directly aware, without the intermediary of any process of inference or any knowledge of truths" (1912, p.46). But even though the metaphysical relation of acquaintance plays an essential role in Russell's theory, we do not get much of a positive proposal for what it amounts to. For the most part, Russell sticks to characterizing it as a direct relation free of any objectionable form of mediation. The closest we get is the suggestion that there is a tight connection between acquaintance and awareness.

This suggestion, however, is a good one. I think that the metaphysical relation of acquaintance is a plausible proposal for grounding the awareness relation present between subject an object in cases of internal awareness. When one is aware of an experiential property, one is acquainted with that property. Elsewhere I have suggested that in order to defeat the typical arguments for representational theories of awareness (e.g., Lycan 2001), we need to have a proposal for what awareness amounts to if it doesn't amount to representation. The answer proposed here is that nonpropositional awareness amounts to acquaintance with experiential properties. I think this is true for both internal awareness and external awareness. In cases of external awareness, one is directly acquainted with the attributed experiential properties and is thereby aware of the object to which those properties are predicated. It is the very nature of attributed experiential properties that acquaintance with them passes through them and to the object they are attributed to. But, recall my warning at the start of this chapter. This is a metaphysical project. I am not introducing the metaphysical relation of acquaintance in the service of some epistemic upshot. In fact, I doubt that there are any particular epistemic upshots from being acquainted with something. I am proposing that we understand awareness as acquaintance because of how it illuminates the functioning of the mind, not because of how it might secure a special kind of self-knowledge.

6 VISUAL ATTENTION

6.1 Introduction

In this paper I will present a theory of attention. In particular, since most of the neuroscientific research on attention has focused on the effects of attention on visual processing, the theory proposed will, strictly speaking, only be proposed as a theory of visual attention. Nonetheless, despite the lack of empirical data on the the areas relevant to attention in the other senses, I fully expect that the framework proposed here can be extended into a general theory of attention. Since a comprehensive examination of the literature on just *visual* attention would require its own book-length discussion, this chapter will only be a brief overview of my views - a sort of prcis for how that book-length discussion might go. Given these aims, I will approach the topic by first arguing for a view about how attention works at the psychological level and then investigating the options available for grounding this conception of attention in a neural mechanism. To get us started, we will consider the metaphysical question: what is attention?

6.2 Attention is a Tool

I think that it is productive to think of the faculty of attention as a 'cognitive tool'. In what ways is attention like a tool? Clearly not in every respect as most tools are artifacts but attention is clearly not an artifact. Nonetheless, there are some important similarities that are worth considering. For example, I think that attention is similar to other tools in that it is the kind of thing that it is in virtue of its function. And, just as the use of tools typically involves the modification of the target that the tool was used on, I think that attention also involves a modification of the representation of the target attended to. Finally, attention,

like other tools, can be *used* by a subject to complete a wide variety of tasks where the performance of that function might be helpful.

What kind of tasks does attention play a role in? Although this is far from comprehensive, people have argued that attention plays a key role in: remembering things, tracking objects, becoming conscious of things, and even securing demonstrative reference to things. Given how diverse that list is, one might become concerned that a single faculty could not be used to perform all these different jobs. This might suggest, in turn, that perhaps attention is not a uniquely identifiable mental process. Christopher Mole (2011) has formalized this concern into an argument against the possibility of grounding attention in a uniquely identifiable mental process. He calls it the 'explanatory over-burdening' argument. If this argument is right, then my project is in trouble. Luckily, I think I can show that the argument is misguided.

The key to understanding how we can locate a unique process underlying all these different uses of attention is to see that attention is not just any kind of tool. It is a *general purpose* tool. General purpose tools are used by subjects to assist in the performance of a variety of tasks, but do not have any of those tasks as a part of their essential functions. For example, an ordinary general purpose knife is a tool that has a very limited function. It has the function of cutting things. However, this function can be usefully deployed to assist the subject in completing a variety of tasks, from opening boxes to eating dinner. The important thing is that even though the knife plays a key role in performing these different tasks, it does not have the performing of *any* of them as a part of its function. Similarly, I think that attention is a general purpose tool that performs a limited function. It just turns out that this limited function is incredibly useful and can be used to assist in the execution of a variety of tasks, including those listed above. What, then, is the function of attention?

I think that attention functions similarly to another tool we are all familiar with: a highlighter. Highlighters have the limited function of emphasizing selected text. There are two key parts of this. First, some text must be selected. Second, the selected text is

emphasized. Technically speaking, the highlighter is only performing its function in the second step. The first step is best understood as a necessary prerequisite for the highlighter's performing its function. It is a necessary step because a highlighter emphasizes text by increasing its prominence *relative* to the other bits of text nearby. Given this, a highlighter can only successfully perform this function if a *subset* of the text is selected. As a result, a selection process must occur prior to its successfully performing its function. Once the selection step has been completed, the highlighter can then perform its function of emphasizing the selected text relative to the neighboring text.

Similarly, I think that attention has the psychological function of emphasizing selected representations.[1] Just as with the highlighter, the performance of this function can be broken down into a selection step and an emphasis step. An object must first be selected as the object to be attended and then attention can emphasize the representation of that object. Here too, the successful completion of the selection step is a necessary prerequisite for attention to perform its actual function of emphasizing a selected representation.

6.2.1 Selection

Even though the selection step is merely a prerequisite for the use of these tools, I think that an examination of how this process works in each case will help illuminate certain experiences we all have with the deployment of attention. How, then, does the selection step work when using a highlighter? Ordinarily this step is guided by a voluntary decision made by the subject using it. In particular, it is guided by a decision regarding which text is *relevant* to the subject's current goals. The text deemed relevant is selected for highlighting. Similarly, we can voluntarily select the target of attention. Subjects are able to decide which features are relevant to their current task and then deploy attention to

[1] I should note that a representation is only needed when the item being attended to is an *external* item. Internal mental items do not need to be represented in order to be attended. They are already in the mind. See Chapter 5 for more on this. Since the vast majority of the literature on attention is on the deployment of attention to external items, that is the case I will be focusing on here.

the items with those features. But because there are so many different tasks one might want to perform, nearly any feature could end up being relevant to performing the task at hand. As a result, subjects must be able to flexibly update the list of which features are deemed relevant.

The most straightforward way of implementing a capacity like this would be to store an 'attentional template' in a temporary storage system (Duncan and Humphreys 1989). Such a template would contain all the information about which features are currently relevant. Following the analogy with the highlighter, the attentional template would be voluntarily set by the subject prior to the deployment attention and would not be considered part of the faculty of attention per se. The template would simply have a role in biasing the processing unit that controls the deployment of attention (which *is* a part of attention per se) in favor of task-relevant items.

Although the selection step is typically under voluntary control, the selection of a target for attention can also be guided involuntarily. This is one of the places where the analogy breaks down. The selection process governing the usage of a highlighter cannot, in ordinary circumstances, proceed involuntarily. Why, then, should we suppose that attention can be guided in this way? Let us consider the following case. Suppose you set yourself the task of reading a difficult paper in a cafe. Given that task, the noises around you are typically irrelevant to your performance of the task and so would not be part of the currently active attentional template. Nonetheless, I expect that everyone has experienced cases where there was, e.g., a sudden loud crash of a mug breaking on the ground that results in your attention being pulled away from the paper. In cases like this, the sudden onset of a loud noise could be said to 'capture' your attention. Capture occurs when a particularly 'salient' item is selected for attention even though it does not fit with the current attentional template.

In order for information about which items are salient to influence the selection process, we must suppose that there is a processing unit somewhere that is continuously tracking the relative salience of all of the objects represented by the subject. This salience processing unit would have the role of biasing the deployment processing unit in favor of salient items. Given a mechanism like this, we can account for cases of capture by supposing that when the salience of an object is sufficiently high, it can swamp the bias from the attentional template and result in the deployment of attention towards the salient item. This gives us a nice outline for how attention is controlled. There are two sources of input to the deployment processing unit: a voluntarily chosen attentional template in a temporary storage system, and a processing unit that automatically calculates the relative salience of items in the environment. The deployment processing unit then weighs these biasing signals and allocates attention on that basis. Like the case of the highlighter, once an object has been selected, attention can finally perform its function of emphasizing the selected item.

6.2.2 Emphasis

When using a highlighter, the emphasis step occurs in virtue of the highlighter's ability to favorably adjust the context of the selected text, so as to make the selected text 'stand out' from the rest. As I suggested above, since the emphasis bestowed is comparative, a highlighter can only successfully perform its function if only a portion of the possible targets are selected. I think that attention functions analogously to a highlighter: it emphasizes representations by making the selected representation 'stand out' from the other representations in the mind. Of course, saying that the representation 'stands out' from its neighbors is merely a metaphor. I think we should ground this metaphor in terms of the relative 'strengths' of the representations involved. A representation 'stands out' from the crowd when it is stronger than all the other representations. We could say, then, that attention has the function of increasing the 'strength' of the selected representation relative to

the rest. How, then, do we determine the 'strength' of a representation at the psychological level?

My proposal is that we can ground the psychological strength of a representation in its ability to connect with other faculties in the mind. Relatively isolated representations are 'weak', while those that are integrated into multiple regions are 'strong'. On this view, since attention has the function of increasing the relative strength of a representation, it could perform this function by increasing the representation's connectivity relative to the neighboring representations. Since we typically attend to objects in order to think about them and to use them to guide intentional actions, increasing the connectivity of a state would typically involve increasing its connectivity with the system that governs thought and intentional action, i.e., with the cognitive system. This suggests, then, that attention has the function of increasing the strength of the selected representation, by way of increasing its connectivity with other faculties in the mind, particularly the cognitive system.

The proposal detailed here for how an item is selected and subsequently emphasized gives us a useful framework for understanding attention. In the remainder of this chapter I will develop a proposal for the neural mechanism that underlies attention, so understood. The hope is that by understanding the neural mechanisms underlying the mental faculty of attention, we will be able to more firmly ground the 'highlighting' function of attention discussed here.

6.3 Neural Correlates of Attentional Emphasis

I argued above that attention is a cognitive tool and that the use of any tool proceeds in two steps. First, the user must select an object on which to use the tool. Second, the tool is used and it performs its function on the selected object. We will begin by considering the second step. What is the neural signature of attention performing its function? That is, what kinds of changes in neural processing are found when attention is deployed and why should we expect that those changes will ultimately result in an increase in the connectivity

of the selected representation?

Before we can answer these questions, we will need to make some controversial assumptions. First, we will need to assume that the mental representations that are selected by attention at the psychological level are grounded in 'coalitions of neurons' in the relevant brain regions. Second, we will need to assume that there is a necessary correspondence between the information carried at the neural level and the mental content of the representation at the psychological level. Since we are focused on neural signature of visual attention, it must be the cast that the contents carried by visual representations are grounded in the information carried by coalitions of the neurons in visual cortex. If such a correspondence obtains, then consideration of the effects of attention at the neural level can help inform our theory of attention at the psychological level.

6.3.1 Effects on a Single Neuron

The first major finding is that attention can increase the *firing rate* of neurons that respond to the attended stimulus. What does this mean? Well we should probably cover a little background in order to make it clear. Neurons in visual cortex tend to selectively respond to particular stimulus property, e.g., color, orientation, direction of motion, etc. The response of an individual neuron to variations in their preferred property can be plotted as a 'tuning curve' for that individual neuron. Each neuron only responds to a small portion of the visual field called their 'receptive field'. The main output from a neuron is the action potentials that it fires (also called 'spikes'), which typically occur several times a second. The preferences of the neuron are determined by which property results in the highest firing rate.

Although all of the results agree that attention can increase firing rates, the exact pattern of facilitation differs depending on the stimulus used. Typically, attention is studied by measuring the firing rate of an individual neuron in response to a *single stimulus* inside that neuron's receptive field. In particular, they are interested in the change in firing rate when

attention is present versus when it is absent. However, since there is significant variability in the way that individual neurons respond to the same stimuli, single-neuron studies typically average the response of a neuron across many trials in order to bring out the variations specifically caused by differences in the allocation of attention.

One of the main findings is that attention proportionally increases the responses to all stimuli, resulting in an overall 'multiplicative' scaling of the neuron's tuning curve. This kind of scaling is typically small but significant, in the range of a 5-20% increase in the rate of firing. For example, in visual area 4 (V4) responses to all orientations increase by approximately the same proportion when attention is directed toward the stimulus (McAdams and Maunsell 1999) and in the middle temporal area (MT) attention to a stimulus proportionally increases the neural response to all directions of motion (Treue and Trujillo 1999). Early models that aimed to account for these effects were called 'gain modulation' models of attention because they suggested that attention had the function of simply increasing the sensitivity of the neuron across the board.[2]

Although the gain modulation model of the effects of attention is appealingly simple, there are some findings that are difficult to reconcile with that model. For example, attention has also been associated with scaling of *contrast* response functions for neurons in V4 (Reynolds et al. 2000; Williford and Maunsell 2006). More importantly, it has long been known that shifting attention between *two stimuli* inside the receptive field of a single neuron can result in much more significant shifts in firing rate (Moran and Desimone 1985). In these studies the recorded neuron has two stimuli placed in its receptive field: one that is preferred (i.e. it strongly drives the neuron when presented alone) and one that is not preferred (i.e. it produces little to no response when presented alone). When both appear in the receptive field at the same time and attention is not present, the neuron's response

[2] 'Gain' is a term from electrical engineering, which describes the ability of a circuit to amplify the power or amplitude of a signal. 'Gain modulation', then, describes the ability of attention to amplify the firing rate of a neuron (or population of neurons).

is the *weighted average* of its response to each stimulus individually. However, when attention is directed towards the preferred stimulus the neural response is greatly enhanced and when attention was directed towards the non-preferred stimulus the neuron's response is significantly suppressed. This strong modulation of firing rate found from shifting attention between preferred and non-preferred stimuli has been replicated in subsequent studies (Reynolds et al. 1999; Ghose and Maunsell 2008).

It seems that attention to a single stimulus is generally associated with an increase in the neuron's response, but when two stimuli are used attention is found to either increase or decrease the neuron's response, depending on which stimulus is attended. Recently, it has been suggested that the attention-related changes in firing rate seen in both conditions can be explained by a 'response normalization' mechanism (Reynolds and Heeger 2009; Lee and Maunsell 2009). Normalization was first used to account for the way that neurons respond when presented with multiple stimuli in their receptive fields. Normalization is able to explain why, for example, the firing rate of a neuron in these cases is the weighted average of its response to each of the stimuli individually. The normalization model of attention suggests that what attention does is strengthen the excitatory drive or the suppressive drive fed into the normalization mechanism. This will ultimately result in either an increase or decrease in firing rate depending on which drive has been strengthened more. Happily, such a mechanism is able to account for all the results discussed so far. This suggests, then, that the mechanism via which attention modulates the firing rate of individual neurons is by biasing the functioning of normalization mechanisms already active in the processing of visual information.

I suggested earlier that attention has the function of increasing the connectivity of the representation of the attended stimulus. How do these findings relate to that proposed function? We have discovered that the deployment of attention has a dual effect on the activity of individual neurons. First, it increases the firing rates of the neurons that prefer the features of the attended stimulus. Second, it suppresses the firing rates of the neurons who

do not prefer the features of that stimulus. I have suggested elsewhere that we can determine the 'strength' of a neural coalition by looking at the average firing rate of the neurons in that coalition. If attention increases the firing rates of neurons carrying information about the attended stimulus relative to those that do not, then it thereby increases the strength of that neural coalition. Since neurons with higher firing rates are better able to make an impact on downstream areas, it would make sense, then, that strong neural coalitions are in a better position to connect with distant regions in the brain.

6.3.2 Effects on a Population of Neurons

Although the effects of attention have been primarily studied at the level of individual neurons, several studies have begun looking at the effects of attention on the activity of whole populations of neurons. The most important finding for our purposes here is that attention has been found to increase the *synchrony* of populations of neurons responding to the attended stimulus (Fries et al. 2001). Attention seems to cause the neurons responding to the attended stimulus to fire in unison. Interestingly, although this feature is readily apparent at the population level, it was obscured by the methods used to study the effects of attention on individual neurons. By averaging across trials in order to average out variability in neural response, they also averaged out information regarding the precise timing of each spike, which is necessary for noticing synchronization.

An interesting feature of these results is that attention does not result in the synchronization of a population of neurons at just any temporal interval. Instead, they preferentially synchronize their firing with gamma oscillations in the local field potential (LFP). What are these gamma oscillations? Oscillations at different frequencies are based in extracellular voltage fluctuations, which arise from summed electrical activity in a population of neurons. These oscillations can be measured on the scalp by EEG and in the brain by an electrode measuring the LFP. Both the EEG signal and the LFP can be decomposed into different frequency components, allowing a precise consideration of the role of oscillations

in each frequency band.[3] When neural networks are activated, the power in higher frequencies increases, particularly in the gamma range (30-80 Hz). This has led some to argue that a prominent gamma rhythm is the signature of an engaged network. What this means is that the production of gamma oscillations in LFP are not necessarily intrinsically computationally relevant. They may merely be a byproduct of the activity in the network as a whole (see: Jia and Kohn 2011, for an overview of gamma oscillations). But, even though the gamma oscillations themselves may not have a particular functional role, they seem to have been co-opted for coordinating the firing of populations of neurons. However, it is important to note that the synchronization of a population of neurons with gamma oscillations occurs in several different parts of the brain, completely independently of attention. Attention does not uniquely induce gamma synchrony. There tends to be a degree of temporal coherence across the neural population even in the absence of attention. What attention does do is *increase* the synchronization with gamma oscillations of the neurons encoding the attended stimulus and *decrease* (but not eliminate) the synchronization of the neurons encoding the unattended stimuli (Fries et al. 2001; Bichot et al. 2005).

Interestingly, Womelsdorf et al. (2006) found that we can even predict the speed of change detection on the basis of gamma synchronization. This demonstrates that the presence of one of the typical behavioral measures of attention (improved reaction time) can be predicted on the basis of the presence of one of the primary neural signatures of attention (gamma synchronization). This suggests, in turn, that gamma synchronization is facilitating neural communication in a way that ultimately results in improved reaction times. Given these findings, perhaps increased gamma synchrony can help ground attention's function of increasing the connectivity of attended representations. There is evidence that is can increase connectivity as it has been demonstrated that spikes arriving simultaneously have a greater impact there than unsynchronized spikes (Usrey et al. 1998; Salinas

[3]These components are delta ¡ 4 Hz, theta 4-8 Hz, alpha 8-12 Hz, beta 12-30 Hz, gamma 30-80 Hz, and high-gamma ¿ 80 Hz.

and Sejnowski 2001). Therefore, if sensory neurons carrying information about a stimulus synchronize their firing, the downstream areas are more likely to have a strong representation of that stimulus. Thus, increased synchrony seems to be an ideal mechanism for enhancing the connectivity of a selected representation at the expense of others.

More recently it has been argued that the improved connectivity caused by gamma synchronization can be further enhanced by the presence of gamma synchrony at both the source *and* the target location. That is, instead of just having gamma synchronization on the input side, if you have gamma synchrony on the output side as well, particularly when they are in phase with each other, communication is even more effective. This hypothesis, called the 'communication through coherence' (CTC) hypothesis, has been defended recently by Pascal Fries (2005, 2009). As Bosman et al. (2012) put it, "rhythmic activity in a target group entails corresponding fluctuations in postsynaptic membrane potentials and postsynaptic shunting, which render input most effective if it is consistently timed to the peaks of depolarization, i.e., if it is synchronized with the target" (p.875). The idea behind the CTC is that when a population of neurons manage to selectively synchronize their firing with their downstream target, they 'block' the other competing representations from controlling the firing of that target population of neurons. If this proposal is correct, then it is specifically the synchronization of firing rates *across areas* that results in increased connectivity. I think that CTC is an interesting proposal with some strong, but not conclusive, evidence behind it. For now I will concentrate primarily on the role played by gamma synchronization on the input side, but it is important to keep in mind that this may only improve connectivity if it results in gamma synchronization on the output side.

6.3.3 Conclusion

There are two key features of the neural signature of attentional emphasis. First, there is an increase in firing rates for neurons encoding the attended stimulus and a decrease in firing rates for the other neurons. Second, there is increased synchronization amongst the

populations of neurons encoding the attended stimulus, particularly in the gamma band, and decreased synchronization amongst populations of neurons encoding the unattended stimuli. Both of these effects seem to have the function of increasing the strength of the neural representation of the attended stimulus and the promotion of its connectivity with distant regions.

6.4 Neural Correlates of Attentional Selection

Now that we have considered the neural correlates of the emphasis step, we will turn to considering the correlates of the mechanisms involved in the selection step. I suggested above that there are three processing units that play an essential role in the selection of a target for attention. First, we need to find the area in charge of the actual deployment of attention. This area will contain the processing unit that weighs the different bias signals and determines where attention is ultimately allocated. Second, we need to find the area that grounds the salience processing unit, which has the function of biasing the deployment unit in favor of salient items. Third, we need to locate the storage area that carries the attentional template, which has the function of biasing the deployment unit in favor of task-relevant items. I will present evidence regarding the grounding of each unit in turn.

6.4.1 Deployment Processing Unit

The job of the deployment processing unit is to determine where attention should be deployed based on the incoming information about object salience and task-relevance. Since the deployment unit determines where attention is deployed, it would be the causal source of the modifications that a representation undergoes when being emphasized. What evidence is there concerning the causal sources of attentional modulation in visual cortex? There are three classes of evidence concerning the causal origins of attentional modulation in visual cortex: correlational, anatomical, and causal. We will examine each of these in turn.

To begin with, we have correlational studies, which can be subdivided into two groups. First, there are the brain imaging studies that have found activation in certain regions in humans when performing attentional tasks. Second, we have neural recording studies in monkeys. The imaging studies have found several areas are correlated with attentional effort outside of the visual cortex, but most of the areas correlated with attention are not consistently found across studies. One region in particular stands out as being consistently activated during attentionally demanding tasks is the frontal eye field (FEF), which is located in Brodmann area 8 near the prefrontal cortex (Ungerleider and Kastner 2000). Gregoriou et al. (2009) performed a direct recording study where they simultaneously recorded the activity of neurons in FEF and in V4. They found that attention to a stimulus inside of the receptive fields of neurons in both areas lead to enhanced oscillatory coupling between the FEF and V4, particularly at gamma frequencies. Interestingly, this is the kind of synchronization across brain regions that the CTC hypothesis argued was important for the transmission of information. This coupling appears to be initiated by the FEF, which changes oscillatory frequencies 8-13ms before V4 does, across a range of oscillatory frequencies. Considering the known conduction velocities and synaptic delays between these two areas, the time-shifted coupling is of the right delay to be able to optimize the postsynaptic impact of spikes from the FEF on V4. This study demonstrates that there is a correlation between gamma synchrony in FEF and V4. It also suggests that FEF may be the cause of those changes in V4.

Next we have the anatomical evidence. Anatomical studies looking that the interconnectivity of various brain regions have determined that the FEF is both directly connected to the majority of visual cortex and also indirectly connected to it via its direct connections to the lateral intraparietal area (LIP) and the superior colliculus (SC) (Ungerleider and Kastner 2000). These connections mean that it is anatomically possible for the FEF to be the source of attentional modulation.

Finally, we have direct causal evidence that induced changes in FEF can cause changes

in V4 that are remarkably similar to those occurring with the deployment of attention. Moore and Armstrong (2003) performed an experiment in which a microstimulation electrode was placed near FEF neurons that had overlapping receptive fields with neurons in V4 that they were recording from. They found that microstimulation of the FEF neurons enhanced the gain of the stimulus-evoked responses from the V4 neurons. Crucially, this enhanced the gain of the V4 neurons in the same manner as the presentation of an attentional cue to attend to the stimulus in their receptive field. The same kinds of changes have been found in V4 by administering a D1 receptor antagonist to FEF (Noudoost and Moore 2011). This indicates that the results obtained by Moore and Armstrong (2003) were not an artifact of the use of a microstimulation electrode. It also demonstrated that dopamine plays a crucial role in the deployment of attention. Together these three sets of evidence strongly suggest that the FEF is the causal source of attentional modulation, suggesting, in turn, that FEF is a good proposal for the 'deployment unit' required by our psychological model of attention.

6.4.2 Salience Processing Unit

The job of the salience processing unit is to calculate the relative saliency of the objects represented by the subject. Since I have argued that the salience unit plays a crucial role in the capturing of attention, the area that calculates saliency should also be active prior to the deployment unit so that can be the causal source of the information that results in the redeployment of attention to the salient object. What area seems like it can play the role needed by the theory?

The lateral intraparietal area (LIP) is the most likely candidate for having the function of tracking the salience of represented objects. Part of the reason for this is that neurons in LIP do not have a fixed preferences. That is, unlike the neurons in visual cortex, they do not have a preferred stimulus that remains the same over time. Instead these neuron appear to respond more strongly when the stimulus at that *location* is more *salient*, regardless of

the particular features that make it salient (Bisley and Goldberg 2003).[4] There are two parts of this claim. First, it seems that LIP represents items by their location, which means that we can think of the salience signal as a 'map' of locations of varying salience. Itti and Koch (2001) describe it as "a scalar, two-dimensional map whose activity topographically represents visual saliency, irrespective of the feature dimension that makes the location salient" (p.198). Second, the map in LIP represents the general salience of an object, not any particular feature. This helps explain why, for example, Bisley and Goldberg (2006) found that neurons in LIP respond strongly to briefly flashed stimuli regardless of their other properties.

If there are salience maps in LIP, in order for them to play the role of the salience processing unit posited by my theory, these maps must be able to influence the deployment of attention. Happily there is some causal evidence that induced changes in the map in LIP can cause changes the bias signal sent to the deployment unit. Mirpour et al. (2010) found that stimulating LIP biases visual search toward the corresponding part of the visual field, presumably because such stimulation directly increases the represented salience at that location.

The salience maps in LIP are able to quickly (80 ms) represent the location of a 'pop-out' stimulus (Ipata et al. 2006). Interestingly, it seems that such signals can be suppressed after learning that those stimuli are not task relevant. That is, 'top-down' signals regarding the relevancy of the pop-out item seem to be able to reduce the activity of the LIP neurons representing the salience of that item. This means that the salience map is not determined solely by the visual input itself. In subsequent research multiple 'top-down' influences have been found to affect the salience map in LIP. For example, it can be influenced learned rules about the task-relevance of the object at that location (Ipata et al. 2009) and by the reward value of the object at that location (Anderson et al. 2011). Interestingly, the salience of a

[4]For example a neuron in LIP would respond equally strongly to the sudden onset of a red letter amongst blue letters as a blue letter amongst red letters. The features do not matter in and of themselves, only the relative salience.

stimulus can be suppressed after being closely examined and determined to be irrelevant at the moment (Mirpour et al. 2009). These results indicate that the salience maps in LIP are formed on the basis of both the actual salience of the object at that location and information from the attentional template about what is relevant. This means that subjects have a degree of control over the ability of salient stimuli to capture attention.

It seems, then, that LIP contains maps that represent the salience of objects in each part of the visual field. However, in order for LIP to play the role proposed for the salience unit, it must occupy the right place in the 'flow chart of the mind'. In particular, it must be active *prior* to the deployment module (FEF) in cases of attentional capture and in cases of the voluntary selection of an item for attention (if LIP is active at all) it must be active *after* the deployment module has initiated attentional modulation. Buschman and Miller (2007) have done work to address exactly this issue. They found that when a pop-out stimulus is the target in a task, the signal representing the location of the target appeared first in LIP and then in FEF (and dlPFC). This suggests that in cases of capture the signals flow from visual cortex to LIP, and then on to the deployment unit in FEF. In the case where there was no pop-out target and the subject simply had to search for the target on the basis of the task instructions, they found that the signal representing the location of the target turned up first in FEF (and dlPFC) and then in LIP. This suggests that the two areas are connected in the proper fashion to work as required by the model.[5]

6.4.3 Attentional Template

Earlier I noted that there are three processing units that need to be grounded in regions of the brain: the deployment unit, the salience unit, and the storage area that holds the

[5]I should note that what is frequently called 'bottom-up attention' in the literature is actually two importantly different phenomena: the automatic facilitation of salient stimuli and the capture of top-down attention. I have proposed that the salience maps in LIP explain the phenomenon of attentional capture. I *do not* think that this process is also involved in the automatic facilitation of stimuli often described under the heading of 'bottom-up' attention. In fact, I do not think that this kind of automatic facilitation really deserves to be called 'attention'. I consider this kind of automatic facilitation to play a role similar to other kinds of recurrent feedback emphasized by Lamme (2006).

attentional template. We have seen how the first two can be grounded, but how can we ground the final unit that has the function of biasing the deployment unit in favor of task-relevant items?

The evidence indicates that dorsolateral prefrontal cortex (dlPFC) is associated with the control of attention (Knudsen 2007). For example, Lebedev et al. (2004) used a task that required attending to one location and remembering another and found that, although dlPFC neurons signaled both locations, the majority of them signaled the attended location. Neurons in dlPFC are a good candidate for the site of the temporary storage system that holds the attentional template because, like the neurons in LIP, they are not tuned to fixed stimulus features. Instead, dlPFC neurons adapt their responses to represent currently relevant information, such as information about categories (Freedman et al. 2001). We also have knockout evidence for the dependence of task-relevant information on dlPFC. When dlPFC is inactivated with muscimol there is specific disruption of the subject's ability to successfully complete visual search tasks, but not simple detection tasks that can proceed via capture (Iba and Sawaguchi 2003).

Besides the direct anatomical evidence that dlPFC is a plausible site for the storage area that holds the attentional template, we should also think that this area plays this particular role because dlPFC has also been implicated as the site of another important mental faculty: working memory. Working memory is the canonical temporary storage system in the mind. It has a limited capacity and it makes all of the encoded information available to various cognitive faculties. Since working memory is a temporary storage system and it is grounded in dlPFC, we have independent reason to suppose that dlPFC stores the currently active attentional template.

Working Memory and Attention

There has also been a significant amount of research on the relation between attention and working memory at the psychological level, some of which is worth discussing given

our goals here. Soto and colleagues have conductive a series of interesting experiments suggesting that *any* information present in working memory can bias the selection of an attentional target. That is, information that is stored in working memory but is not part of the 'attentional template' created for the execution of the task at hand can still bias the selection of a target (Soto et al. 2008). This is an unexpected finding. Nonetheless, it makes sense if we suppose that the connection between working memory and the deployment unit is *hardwired*. That is, the simplest way for an attentional template in working memory to influence the deployment of attention would be if whatever information is being maintained in working memory automatically biased the selection process in the deployment unit.

There is also an important relation between *working memory load* and the ability of items in working memory to bias the selection of the attentional target. Lavie (2005) has shown that a subject's ability to filter out irrelevant stimuli during selection depends on the processing load in working memory. As the working memory load increases, the influence of the attentional template deceases. Interestingly, Soto and Humphreys (2008) found that the bias from *irrelevant* information maintained in working memory is also reduced as working memory load increases. This implies that working memory load has a uniform effect on reducing the influence of each individual item present in working memory. It does not matter whether the item stored in working memory is the 'attentional template' or an irrelevant distractor, the more items there are in working memory the less influence they each have on the selection of a target. It is as if they mutually inhibit each other.

The psychological data also indicates that the selection of the target of *visual* attention need not be based in an attentional template that is in a visual format. For example, Potter (1975) demonstrated that the recognition of a visual stimulus in a rapid visual stream can be facilitated just as much by advance verbal cueing of its meaning as by visual cueing of its appearance. The effect of a working memory distractor on search is effective even when the working memory stimulus is encoded verbally (Soto and Humphreys 2007). This effect can also be moderated by the *semantic relationship* between an item held in working memory

and a distractor in a search task (Moores et al. 2003). Together these studies suggest that information held in working memory can bias the selection of a target either directly on the basis of its content, or indirectly on the basis of conceptual associations between that information and the target.

In conclusion, we have good reason to suppose that working memory (and the area that contains working memory, dlPFC) is crucially involved in the selection of a target for attentional deployment. We found that we can ground the attentional template in a voluntarily encoded representation of any format and that this template can bias the deployment module in favor of task-relevant items. What was interesting was that this kind of biasing is not limited to the attentional template. Any information present in working memory biases the selection process, even when that information favors an task-irrelevant object.

6.4.4 Conclusion

I have argued that the three processing units crucially involved in the selection of the target of attention can each be grounded in the activity of a relevant region of the brain. The deployment unit is based in FEF, the salience unit is based in LIP, and the attentional template is located in working memory which is itself located in dlPFC. There is sufficient empirical evidence both of their individual involvement but also that they are wired together in the ways necessary to ground the theory of attention proposed here.

6.5 Implications

I have argued in this chapter that attention, strictly speaking, is a process that starts with the deployment unit grounded in FEF and terminates in the modulation of the activity of the neurons in visual cortex, particularly V4 and MT. The effect that attentional modulation has on neural activity is to increasing the firing rates of the neurons encoding the attended features and to increase their synchrony with gamma oscillations. Elsewhere in I

have suggested that attention plays a number of roles in both internal and external awareness. I will briefly list the main claims here and then explain how the theory of attention developed in this chapter can fulfill all those roles.

I made five claims regarding the role played by attention. The first one is that attention is one of the mechanisms of adjusting the 'volume' of a state (Chapter 4). That is, I suggested that attention can increase or decrease the strength of a neural coalition. The second one appears in the same chapter, when I agree with Block that attention is neither necessary nor sufficient for external awareness. The third commitment is that attention *is* necessary for cognitive access to a piece of information (Chapter 3). Finally, I made a fourth and a fifth commitment in my chapter on internal awareness (Chapter 5). First, I claimed that attention is able to make previously inaccessible unattributed experiential properties cognitively accessible. Second, I agreed with the intuition behind the transparency thesis, claiming that we cannot attend to attributed experiential properties qua property had by the mental state in which it inheres. How can the faculty of attention detailed here consistently play all of these different roles?

First, my discussion on how attention emphasizes representations made it clear that attention does have the function of strengthening neural coalitions encoding the attended feature. This will enable it to play a role in controlling the 'volume' of a state. Second, I have argued elsewhere that cognitive accessibility occurs once a coalition passes a strength-based threshold and that passing this threshold makes the subject aware of the accessible contents. Given this, in order for attention to be neither necessary nor sufficient for external awareness, attention must be able to strengthen a neural coalition encoding some feature, but still fail to make it strong enough to pass this threshold. The empirical evidence suggests this can happen (e.g., Jiang et al. 2006) and there is no barrier to its happening given the mechanism I have identified as underlying attention.[6] Next, as I have argued elsewhere, neural coalitions can be strengthened by other processes (e.g. bottom-up facilitation and

[6] Although I do acknowledge that cases like this may be somewhat uncommon.

recurrent feedback) enough to pass the threshold for accessibility without needing to also be strengthened by attention. As a result, the theory developed here can fit the requirements of being neither necessary nor sufficient for cognitive accessibility and external awareness.

Third, this commitment might seem to create a tension with my commitment that attention *is* necessary (but not sufficient) for cognitive access. But I do not think it does. The reason for this is that in order for information to become cognitively accessed it must pass a different, much higher, threshold for coalition strength. Attention can fail to strengthen a coalition enough to pass this threshold, making it insufficient for cognitive access. But, it turns out that there is an interesting feature about the design of the mind. Even though in principle there is nothing stopping a state from passing this threshold in the absence of attention, in practice it is not possible. In other words, the threshold for cognitive access seems to have been set so high that even when all of the other ways of strengthening a neural coalition are active the coalition is still not strong enough to pass the threshold. It must be helped by attention. Because of this, attention is necessary for cognitive access.

Fourth, in my chapter on internal awareness I argue that when unattributed experiential properties become cognitively accessible the subject becomes aware of those properties. I then suggest that attention is the faculty that makes these previously inaccessible experiential properties strong enough to become cognitively accessible. What is interesting about this claim is it implies that when a representation becomes cognitively accessible in cases of external awareness, really all that is accessible are the content properties. The unattributed properties remain inaccessible. In order to make those properties accessible, the neurons encoding those properties must be strengthened by attention.[7] As long as these experiential properties are grounded in structurally decomposable states of the brain, then on the current theory of attention there is no barrier to directing attentional modulation to them and strengthening them enough to become cognitively accessible.

[7]That is not to say that they must be voluntarily attended, we are all familiar with cases where an intense mental state, like a sharp pain, capture our attention even though we would prefer to ignore that state.

Fifth, since attention, on my view, only does one thing - it strengthens neural coalitions - it is not possible to attentionally modulate the same coalition in two different ways. It can be done to different degrees, but not in completely different ways. This means that my theory of attention leaves no room for claims that one can attend to attributed experiential properties in more than one way. You cannot. Happily this corresponds quite nicely with the transparency thesis that I also argue for in the same chapter. One cannot attend to attributed experiential properties except as had by the objects to which they are attributed.

It seems, then, that my theory of attention can fulfill all the roles I have attributed to it. This is a good thing. My view is internally consistent. But this fact can also be used as an argument in favor of this theory of attention. Since no other theory of attention can accommodate all of these roles for attention, if my arguments elsewhere are persuasive enough to lead us to conclude that attention does play each of these roles, then we should think that my theory of attention is preferable to the other theories for that reason. In other words, the other uses to which I have put attention in the other parts of the dissertation jointly argue for this theory of attention over its competitors because it is the only one that can accommodate them all.

REFERENCES

Aglioti, S., DeSouza, J. F., and Goodale, M. A. (1995). Size-contrast illusions deceive the eye but not the hand. *Current Biology*, 5(6):679–685.

Anderson, B. A., Laurent, P. A., and Yantis, S. (2011). Value-driven attentional capture. *Proceedings of the National Academy of Sciences*, 108(25):10367–10371.

Armstrong, D. M. (1980). What is consciousness? In *The Nature of Mind and Other Essays*. University of Queensland Press.

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

Baars, B. J. (2001). A biocognitive approach to the conscious core of immediate memory. *Behavioral and Brain Sciences*, 24:115–116.

Baddeley, A. (1986). *Working Memory*. Clarendon Press.

Bermúdez, J. L. (1995). Nonconceptual content: From perceptual experience to subpersonal computational states. *Mind and Language*, 10(4):333–369.

Bermúdez, J. L. (2007). What is at stake in the debate on nonconceptual content? *Philosophical Perspectives*, 21(1):55–72.

Bichot, N. P., Rossi, A. F., and Desimone, R. (2005). Parallel and serial neural mechanisms for visual search in macaque area v4. *Science*, 308(5721):529–534.

Bisley, J. W. and Goldberg, M. E. (2003). Neuronal activity in the lateral intraparietal area and spatial attention. *Science*, 299(5603):81–86.

Bisley, J. W. and Goldberg, M. E. (2006). Neural correlates of attention and distractibility in the lateral intraparietal area. *Journal of Neurophysiology*, 95(3):1696–1717.

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

Block, N. (2002). Concepts of consciousness. In Chalmers, D. J., editor, *Philosophy of Mind: Classical and Contemporary Readings*. Oxford University Press.

Block, N. (2007). Consciousness, accessibility, and the mesh between psychology and neuroscience. *Behavioral and Brain Sciences*, 30(5-6):481–548.

Block, N. (2008). *Proceedings of the Aristotelian Society*, 108(1pt3):289–317.

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

Block, N. (2014). Rich conscious perception outside focal attention. *Trends in Cognitive Sciences*, 18.

Bosman, C. A., Schoffelen, J.-M., Brunet, N., Oostenveld, R., Bastos, A. M., Womelsdorf, T., Rubehn, B., Stieglitz, T., De Weerd, P., and Fries, P. (2012). Attentional stimulus selection through selective synchronization between monkey visual areas. *Neuron*, 75(5):875–888.

Brady, T. F., Konkle, T., and Alvarez, G. A. (2011). A review of visual memory capacity: Beyond individual items and toward structured representations. *Journal of Vision*, 11(5).

Brewer, B. (2006). Perception and content. *European Journal of Philosophy*, 14(2):165–181.

Brown, R. (2011). The myth of phenomenological overflow. *Consciousness and Cognition*, 21(2):599–604.

Brown, R. J. and Norcia, A. M. (1997). A method for investigating binocular rivalry in real-time with the steady-state vep. *Vision Research*, 37(17):2401–2408.

Burge, T. (2010). *Origins of Objectivity*. Oxford University Press.

Buschman, T. J. and Miller, E. K. (2007). Top-down versus bottom-up control of attention in the prefrontal and posterior parietal cortices. *Science*, 315(5820):1860–1862.

Cohen, M. A. and Dennett, D. C. (2011). Consciousness cannot be separated from function. *Trends in Cognitive Sciences*, 15(8):358–364.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, 24(1):87–114.

Cowan, N. (2010). The magical mystery four: How is working memory capacity limited, and why? *Current Directions in Psychological Science*, 19(1):51–57.

Cummins, R. (2010). *The World in the Head*. Oxford University Press.

Dehaene, S. (2009). Neuronal global workspace. In Bayne, T., Cleeremans, A., and Wilken, P., editors, *The Oxford Companion to Consciousness*. Oxford University Press.

Dehaene, S. (2014). *Consciousness and the Brain: Deciphering How the Brain Codes Our Thoughts*. Viking.

Dehaene, S. and Changeux, J.-P. (2011). Experimental and theoretical approaches to conscious processing. *Neuron*, 70(2):200–227.

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

ences, 10(5):204–211.

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

Dennett, D. (2012). A phenomenal confusion about access and consciousness. Presented at the Turing Consciousness conference.

Dennett, D. C. (1978). Why you can't make a computer that feels pain. *Synthese*, 38(3):415–456.

Doesburg, S., Kitajo, K., and Ward, L. (2005). Increased gamma-band synchrony precedes switching of conscious perceptual objects in binocular rivalry. *Neuroreport*, 16(11):1139–1142.

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

Dretske, F. (1993). Conscious experience. *Mind*, 102(406):263–283.

Dretske, F. (1995). *Naturalizing the Mind*. MIT Press.

Dretske, F. (2006). Perception without awareness. In Gendler, T. S. and Hawthorne, J., editors, *Perceptual Experience*. Oxford University Press.

Duncan, J. and Humphreys, G. (1989). Visual search and stimulus similarity. *Psychological Review*, 96(3):433–458.

Evans, G. (1982). *The Varieties of Reference*. Oxford University Press.

Evans, J. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology*, 59(1):255–278.

Farah, M. J. (1994). Visual perception and visual awareness after brain damage: A tutorial overview. In Umilta, C. and Moscovitch, M., editors, *Consciousness and Unconscious Information Processing: Attention and Performance 15*, pages 203–236. MIT Press.

Fell, J., Klaver, P., Elger, C., and Fernandez, G. (2002). Suppression of eeg gamma activity may cause the attentional blink. *Consciousness and Cognition*, 11(1):114–122.

Fish, W. (2009). *Perception, Hallucination, and Illusion*. Oxford University Press.

Flanagan, O. (1997). Conscious inessentialism and the epiphenomenalist suspicion. In Block, N., Flanagan, O., and Güzeldere, G., editors, *The Nature of Consciousness: Philosophical Debates*. MIT Press.

Fodor, J. A. (1975). *The Language of Thought*. Harvard University Press.

Fodor, J. A. (2008). *Lot 2: The Language of Thought Revisited*. Oxford University Press.

Freedman, D. J., Riesenhuber, M., Poggio, T., and Miller, E. K. (2001). Categorical representation of visual stimuli in the primate prefrontal cortex. *Science*, 291(5502):312–316.

Fries, P. (2005). A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. *Trends in Cognitive Sciences*, 9(10):474–480.

Fries, P. (2009). Neuronal gamma-band synchronization as a fundamental process in cortical computation. *Annual Review of Neuroscience*, 32(1):209–224.

Fries, P., Reynolds, J. H., Rorie, A. E., and Desimone, R. (2001). Modulation of oscillatory neuronal synchronization by selective visual attention. *Science*, 291(5508):1560–1563.

Ghose, G. M. and Maunsell, J. H. R. (2008). Spatial summation can explain the attentional modulation of neuronal responses to multiple stimuli in area v4. *Journal of Neuroscience*, 28(19):5115–5126.

Gilchrist, A. L., Cowan, N., and Naveh-Benjamin, M. (2008). Working memory capacity for spoken sentences decreases with adult ageing: Recall of fewer but not smaller chunks in older adults. *Memory*, 16(7):773–787.

Gregoriou, G. G., Gotts, S. J., Zhou, H., and Desimone, R. (2009). High-Frequency, Long-Range Coupling Between Prefrontal and Visual Cortex During Attention. *Science*, 324(5931):1207–1210.

Harman, G. (1990). The intrinsic quality of experience. *Philosophical Perspectives*, 4:31–52.

Hill, C. S. (1988). Introspective awareness of sensations. *Topoi*, 7(1):11–24.

Hill, C. S. (1991). *Sensations: A Defense of Type Materialism*. Cambridge University Press.

Humphrey, N. (1992). *A History of the Mind: Evolution and the Birth of Consciousness*. Simon and Schuster.

Iba, M. and Sawaguchi, T. (2003). Involvement of the dorsolateral prefrontal cortex of monkeys in visuospatial target selection. *Journal of Neurophysiology*, 89(1):587–599.

Ipata, A. E., Gee, A. L., Bisley, J. W., and Goldberg, M. E. (2009). Neurons in the lateral intraparietal area create a priority map by the combination of disparate signals. *Experimental Brain Research*, 192(3):479–488.

Ipata, A. E., Gee, A. L., Gottlieb, J., Bisley, J. W., and Goldberg, M. E. (2006). Lip responses to a popout stimulus are reduced if it is overtly ignored. *Nature Neuroscience*, 9:1071–1076.

Itti, L. and Koch, C. (2001). Computational modeling of visual attention. *Nature Reviews Neuroscience*, 2:194–203.

Jia, X. and Kohn, A. (2011). Gamma rhythms in the brain. *PLoS Biology*, 9(4):e1001045.

Jiang, Y., Costello, P., Fang, F., Huang, M., and He, S. (2006). A gender-and sexual orientation-dependent spatial attentional effect of invisible images. *Proceedings of the National Academy of Sciences*, 103(45):17048–17052.

Knudsen, E. I. (2007). Fundamental components of attention. *Annual Review of Neuroscience*, 30(1):57–78.

Kouider, S., de Gardelle, V., Sackur, J., and Dupoux, E. (2010). How rich is consciousness? the partial awareness hypothesis. *Trends in Cognitive Sciences*, 14(7):301–307.

Kouider, S. and Dehaene, S. (2007). Levels of processing during non-conscious perception: A critical review of visual masking. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 362(1481):857–875.

Kouider, S., Dehaene, S., Jobert, A., and Le Bihan, D. (2006). Cerebral bases of subliminal and supraliminal priming during reading. *Cerebral Cortex*, 17(9):2019–2029.

Lamme, V. A. (2003). Why visual attention and awareness are different. *Trends in Cognitive Sciences*, 7(1):12–18.

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

Landman, R., Spekreijse, H., and Lamme, V. A. (2003). Large capacity storage of integrated objects before change blindness. *Vision Research*, 43(2):149–164.

Lau, H. and Rosenthal, D. (2011). Empirical support for higher-order theories of conscious awareness. *Trends in Cognitive Sciences*, 15(8):365–373.

Lavie, N. (2005). Distracted and confused?: Selective attention under load. *Trends in Cognitive Sciences*, 9(2):75–82.

Lebedev, M. A., Messinger, A., Kralik, J. D., and Wise, S. P. (2004). Representation of attended versus remembered locations in prefrontal cortex. *PLoS Biology*, 2(11):e365.

Lee, J. and Maunsell, J. H. R. (2009). A normalization model of attentional modulation of single unit responses. *PLoS One*, 4(2):e4651.

Levine, J. (2007). Two kinds of access. *Behavioral and Brain Sciences*, 30(5-6):514–515.

Luck, S. J., Vogel, E. K., and Shapiro, K. L. (1996). Word meanings can be accessed but not reported during the attentional blink. *Nature*, 383:616–618.

Lycan, W. G. (1996). *Consciousness and Experience*. MIT Press.

Lycan, W. G. (2001). A simple argument for a higher-order representation theory of consciousness. *Analysis*, 61(269):3–4.

McAdams, C. J. and Maunsell, J. H. R. (1999). Effects of attention on orientation-tuning functions of single neurons in macaque cortical area v4. *The Journal of Neuroscience*, 19(1):431–441.

Merikle, P. M., Smilek, D., and Eastwood, J. D. (2001). Perception without awareness: Perspectives from cognitive psychology. *Cognition*, 79(1):115–134.

Mirpour, K., Arcizet, F., Ong, W. S., and Bisley, J. W. (2009). Been there, seen that: A neural mechanism for performing efficient visual search. *Journal of Neurophysiology*, 102(6):3481–3491.

Mirpour, K., Ong, W. S., and Bisley, J. W. (2010). Microstimulation of posterior parietal cortex biases the selection of eye movement goals during search. *Journal of Neurophysiology*, 104(6):3021–3028.

Mole, C. (2011). The metaphysics of attention. In Mole, C., Smithies, D., and Wu, W., editors, *Attention: Philosophical and Psychological Essays*, pages 60–77. Oxford University Press.

Moore, T. and Armstrong, K. M. (2003). Selective gating of visual signals by microstimulation of frontal cortex. *Nature*, 421:370–373.

Moores, E., Laiti, L., and Chelazzi, L. (2003). Associative knowledge controls deployment of visual selective attention. *Nature Neuroscience*, 6:182–189.

Moran, J. and Desimone, R. (1985). Selective attention gates visual processing in the extrastriate cortex. *Science*, 229(4715):782–784.

Natsoulas, T. (1995). How access-consciousness might be a kind of consiousness. *Behavioral and Brain Sciences*, 18:264–265.

Noudoost, B. and Moore, T. (2011). Control of visual cortical signals by prefrontal dopamine. *Nature*, 474:372–375.

Peacocke, C. (1998). Nonconceptual content defended. *Philosophy and Phenomenological Research*, 63(2):381–388.

Phillips, I. B. (2011). Perception and iconic memory: What sperling doesn't show. *Mind and Language*, 26(4):381–411.

Potter, M. C. (1975). Meaning in visual search. *Science*, 187(4180):965–966.

Prinz, J. (2012). *The Conscious Brain*. Oxford University Press.

Rees, G., Wojciulik, E., Clarke, K., Husain, M., Frith, C., and Driver, J. (2000). Unconscious activation of visual cortex in the damaged right hemisphere of a parietal patient with extinction. *Brain*, 123(8):1624–1633.

Reynolds, J. H., Chelazzi, L., and Desimone, R. (1999). Competitive mechanisms subserve attention in macaque areas v2 and v4. *The Journal of Neuroscience*, 19(5):1736–1753.

Reynolds, J. H. and Heeger, D. J. (2009). The normalization model of attention. *Neuron*, 61(2):168–185.

Reynolds, J. H., Pasternak, T., and Desimone, R. (2000). Attention increases sensitivity of v4 neurons. *Neuron*, 26(3):703–714.

Rosenthal, D. M. (1997). A theory of consciousness. In Block, N., Flanagan, O. J., and Guzeldere, G., editors, *The Nature of Consciousness*. MIT Press.

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

Rousselet, G., Joubert, O., and Fabre-Thorpe, M. (2005). How long to get to the gist of real-world natural scenes? *Visual Cognition*, 12(6):852–877.

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

Salinas, E. and Sejnowski, T. (2001). Correlated neuronal activity and the flow of neural information. *Nature Reviews Neuroscience*, 2:539–550.

Saults, J. S. and Cowan, N. (2007). A central capacity limit to the simultaneous storage of visual and auditory arrays in working memory. *Journal of Experimental Psychology General*, 136(4):663–684.

Sauret, W. and Lycan, W. G. (2014). Attention and internal monitoring: A farewell to hop. *Analysis*, 74(3):363–370.

Schellenberg, S. (2011). Perceptual content defended. *Noûs*, 45(4):714–750.

Schwitzgebel, E. (2008). The unreliability of naive introspection. *Philosophical Review*, 117(2):245–273.

Sekida, K. (1975). *Zen Training: Methods and Philosophy*. Weatherhill.

Sergent, J. and Poncet, M. (1990). From covert to overt recognition of faces in prosopagnosic patient. *Brain*, 113(4):989–1004.

Siegel, S. (2010). Do visual experiences have contents? In Nanay, B., editor, *Perceiving the World*. Oxford.

Sligte, I. G., Scholte, H. S., and Lamme, V. A. F. (2008). Are there multiple visual short-term memory stores? *PLoS ONE*, 3(2):e1699.

Sligte, I. G., Scholte, H. S., and Lamme, V. A. F. (2009). V4 activity predicts the strength of visual short-term memory representations. *Journal of Neuroscience*, 29(23):7432–7438.

Solomon, S. G. and Lennie, P. (2007). The machinery of colour vision. *Nature Reviews Neuroscience*, 8(4):276–286.

Soto, D., Hodsoll, J., Rotshtein, P., and Humphreys, G. W. (2008). Automatic guidance of attention from working memory. *Trends in Cognitive Sciences*, 12(9):342–348.

Soto, D. and Humphreys, G. (2008). Stressing the mind: The effect of cognitive load and articulatory suppression on attentional guidance from working memory. *Perception and Psychophysics*, 70(5):924–934.

Soto, D. and Humphreys, G. W. (2007). Automatic guidance of visual attention from verbal working memory. *Journal of Experimental Psychology: Human Perception and Performance*, 33(3):730–737.

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

Summerfield, C., Jack, A. I., and Burgess, A. P. (2002). Induced gamma activity is associated with conscious awareness of pattern masked nouns. *International Journal of Psychophysiology*, 44(2):93–100.

Travis, C. S. (2004). The silence of the senses. *Mind*, 113(449):57–94.

Treue, S. and Trujillo, J. C. M. (1999). Feature-based attention influences motion processing gain in macaque visual cortex. *Nature*, 399:575–579.

Tye, M. (1995). *Ten Problems of Consciousness*. MIT Press.

Tye, M. (2000). *Consciousness, Color, and Content*. Bradford Books.

Tye, M. (2006). Content, richness, and fineness of grain. In Gendler, T. S. and Hawthorne, J., editors, *Perceptual Experience*. Oxford University Press.

Tye, M. (2009). *Consciousness Revisited: Materialism Without Phenomenal Concepts*. MIT Press.

Ungerleider, L. and Kastner, S. (2000). Mechanisms of visual attention in the human cortex. *Annual Review of Neuroscience*, 23(1):315–341.

Usrey, W. M., Reppas, J. B., and Reid, R. C. (1998). Paired-spike interactions and synaptic efficacy of retinal inputs to the thalamus. *Nature*, 395:384–387.

Varela, F., Lachaux, J.-P., Rodriguez, E., and Martinerie, J. (2001). The brainweb: Phase synchronization and large-scale integration. *Nature Reviews Neuroscience*, 2:229–239.

Vuilleumier, P., Sagiv, N., Hazeltine, E., Poldrack, R. A., Swick, D., Rafal, R. D., and Gabrieli, J. D. E. (2001). Neural fate of seen and unseen faces in visuospatial neglect: A combined event-related functional mri and event-related potential study. *Proceedings of the National Academy of Sciences*, 98(6):3495–3500.

Williford, T. and Maunsell, J. H. R. (2006). Effects of spatial attention on contrast response functions in macaque area v4. *Journal of Neurophysiology*, 96(1):40–54.

Womelsdorf, T., Fries, P., Mitra, P. P., and Desimone, R. (2006). Gamma-band synchronization in visual cortex predicts speed of change detection. *Nature*, 439:733–736.