THE FUNCTIONAL CONTRIBUTIONS OF CONSCIOUSNESS

DYLAN LUDWIG

A DISSERTATION SUBMITTED TO THE FACULTY OF GRADUATE STUDIES IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

GRADUATE PROGRAM IN PHILOSOPHY
YORK UNIVERSITY
TORONTO, ONTARIO

July 2022

Dissertation Abstract

Most existing research programs are occupied with the difficult question of what consciousness *is*, overlooking what the more interesting and fruitful research question: what does consciousness *do*? My dissertation develops a philosophical method for identifying the functional capacities that conscious experience contributes to information processing systems.

My strategy involves systematically consolidating and interpreting a range of psychological and neuroscientific research in order to compare conscious and unconscious processing in different psychological domains, namely, vision, emotion, and social cognition. I also defend the principle of *functional pluralism*: given that conscious experiences presumably form a relatively diverse class in the natural world, we should expect them to facilitate a diverse range of functions in different psychological domains. My pluralist account implies that we will be able to amass a collection of functional markers that can guide future ascriptions of experience to all sorts of natural and artificial systems. Understanding consciousness' functional profile should also ultimately help us answer the general but elusive question of what consciousness is as a feature of psychological systems.

After laying out the general framework and critically evaluating prominent theories of consciousness in the first chapter, I begin the process of identifying FCCs in particular psychological domains. In my second chapter, I identify some candidate functional markers of consciousness in the functionally-complex domain of visual perception, including the processing of semantic information inherent in more informationally-complex visual stimuli, increased spatiotemporal precision, and representational integration over larger spatiotemporal intervals. My third chapter discusses the domain of emotional processing, where I argue that experience facilitates the inhibition of, the conceptualization of, and flexible response to emotionally valenced representational content. In my fourth chapter, I review a range of bias-intervention strategies that explicitly draw on the functional resources of conscious experience. In my final chapter, I draw some conclusions about the nature of consciousness based on my functional analysis. I introduce what I call a Local Workspace Theory, arguing that consciousness is at least in part characterized by a high degree of representational complexity afforded by the structural mechanisms that realize it and reflected in the psychological functions that it facilitates.

To Lauren and Elliot; always on my mind.

Acknowledgements

Special thanks to Muhammad Ali Khalidi, Jacob Beck, Kristin Andrews, Alice MacLachlan, Regina Rini, Lorraine Code, Lauren Edwards, Sam Clarke, Ned Block, Henry Jackman, Jesse Prinz, and Dale Stevens

Table of Contents

List of Figures

“Oh, unconscious cerebration! You will have to give the wall to your conscious brother”. -Bram Stoker (Dracula, 1897)

Chapter 1: Introduction to the Functional Contributions of Consciousness

Abstract

The most widely endorsed philosophical and scientific theories of consciousness assume that it contributes a single functional capacity to an organism's information processing toolkit. However, conscious processes are a heterogeneous class of psychological phenomena supported by a variety of neurobiological mechanisms. This suggests a plurality of functional contributions of consciousness (FCCs), in the sense that conscious experience facilitates different functional capacities in different psychological domains. In this chapter, I first develop a general methodological framework for isolating the psychological functions that are characteristic of conscious experience. I then show that the leading theories—Global Workspace Theory, Information Integration Theory, Attended Intermediate Representation Theory, and First- and Higher-order Representationalist Theories— all fail to acknowledge what I call *functional pluralism*, and are theoretically restricted as a result.

1. What Consciousness Does

When philosophers and scientists study consciousness, they tend to ask questions about *what it is*. After centuries of philosophical thought aimed at discovering its fundamental nature, only a loose cluster of descriptive definitions has emerged: consciousness is "experience" or "awareness", it seems to exist uniquely as a "subjective" or "first-person" phenomenon, and it is characterized roughly by "what-it's-like" to be the subject of different kinds of experiences (e.g. Block 2010, Siewert 2013, Schwitzgebel 2016). These descriptive definitions are necessarily but unfortunately vague, and much subsequent philosophical work has been devoted to trying to render them more precise. Likewise, the science of consciousness is often focussed on providing descriptive accounts, the goal being to explain what conscious experience *is* as a feature of neurobiological and/or psychological systems (e.g. Lamme 2006, Dehaene 2014). These paths of inquiry have failed to achieve widespread consensus, however, and a viable overarching theory of consciousness has yet to be established. Descriptive definitions generally seem to have limited potential for explaining consciousness' place in the natural world.

Another important kind of explanation has to do with identifying *functions*. In cognitive science, the notion of "function" has a long and contentious history. Despite this, there are some shared ideas about what functions are and why picking them out is helpful in explaining the workings of psychological systems. Functions are generally thought of as individual contributions to the "goals" of a system; distinct causal components of the processes internal to the system (e.g. Maley & Piccinini 2017). This does not necessarily need to be understood mechanistically, in the strict and narrow sense of being cashed out in the terms of "hard sciences" like neurobiology, chemistry or physics. Functions can be pitched purely at a psychological or computational level of description, and can include anything from parsing sentences in linguistic comprehension, to valencing representations in emotional evaluation, to computing binocular disparity in visual perception. If particular psychological functions can in fact be shown to correspond with particular neurobiological mechanisms, robust functional explanations of psychological phenomena can be established.

One frontline of philosophical debate has been whether or not the functions of information processing systems need to be explained and individuated etiologically; that is, in terms of either evolutionary or developmental history (for classic etiological accounts, see, e.g. Millikan 1984, Dretske 1988, Neander 1991). Some philosophers assume instead that an "ahistorical" account of function is possible (e.g. Cummins 1975, 1983). Without wading too deeply into these waters, I think these can ultimately be thought of as compatible projects. Etiological accounts are more explanatorily demanding, in that they try to paint a bigger picture by explaining functions in relation to certain selection pressures or learning objectives. However, this bigger picture doesn't seem to be necessary in every case in which we want to describe a function in some complex system (Maley & Piccinini 2017). It might be that "proper functions", which are the result of temporally extended adaptations on etiological accounts, are just a subclass of the kinds of functions that characterize the causal workings of a system, the total set of which are the targets of ahistorical functional analyses. Griffiths (1993) explicitly argues for this kind of compatibility between the different approaches to understanding function, where each has a different explanatory aim but both are concerned with elucidating the same kinds of causal processes. Alternatively, it might turn out that these approaches simply have different explanatory targets altogether and can (or should) be pursued independently. In what follows, I will employ the less restrictive ahistorical notion of function, and as such will typically avoid making specific claims about evolutionary history. This is primarily because I take it to be extremely difficult to isolate viable causal explanations over such vast expanses of time. Maley and Piccinini (2017) similarly argue that "the causal histories that ground functions on [etiological] accounts are often unknown (and in many cases, unknowable), making function attribution difficult or even impossible" (p. 238). However, the mere fact that organisms have evolved some psychological capacity—like conscious experience—ought to inspire questions about its functional role(s).

Existing theories and experimental paradigms don't always approach consciousness in terms of identifying its function or functions; that is, in terms of *what it does* for information processing systems. In fact, there is still much skepticism about the possibility of identifying the functional role(s) that conscious experience plays in psychological processes. In its most extreme form, the epiphenomenalist still denies that consciousness is causally efficacious at all. But even more mainstream approaches assume that consciousness itself isn't appropriately explained in functional, causal terms; only the psychological processes that happen to be conscious can be explained functionally. Some philosophers, for example, think that it is precisely because consciousness cannot be explained functionally that the so-called "hard problem" arises (Chalmers 1995, Lycan 2019). At the other extreme, some philosophers have argued that consciousness is inextricable from its functional role (e.g. Cohen & Dennett 2011).

Typically, though, theories of the nature of consciousness at least implicitly suggest that it plays a particular role, or confers some kind of processing advantage to the system. That is, functional claims are built into the different kinds of accounts already on offer, whether or not they are made explicit (see below for a survey of some of the leading theories). One need not adopt a full-blown functionalist metaphysics, according to which functional role exhausts what it means for a process to be conscious, in order to appreciate that consciousness has a functional role. But it seems that progress in the philosophy and science of consciousness requires that we take the project of giving an account of its psychological function(s) seriously. At this point, the burden of proof must be shifted to those who continue to deny that consciousness plays any functional role in information processing systems. This denial has specific metaphysical implications, and the epiphenomenalist still owes us a plausible explanation of how this particular psychological property alone could be a feature of the ongoing internal dynamics of psychological systems and yet play no functional role in psychological processing. Moreover, if epiphenomenalism were right, then all past, present, and future attempts to bring empirical science to bear on the problems of consciousness would be hopeless. In this case, we would be forced to collectively abandon the project of developing a scientifically informed theory of consciousness, and its broad and varied naturalistic research program that has already made progress in a variety of ways, both theoretical and practical. In contrast, an increasing number of prominent researchers endorse and have fruitfully pursued the general idea that conscious experiences contribute functionally to an organism's information processing toolkit (e.g. Dehaene et al. 2006, Seth 2009, Cohen & Dennett 2011, Frith & Metzinger 2016, Lamme 2020, Boyle 2020, Birch 2020), although substantial disagreement remains about what these functional capacities are.

Trying to isolate what it is that consciousness does, or what it contributes functionally to a system, makes for a more tractable project than other methodological approaches, primarily because consciousness is best understood as a feature of dynamic psychological *processes*[1]. Functional accounts have a methodological advantage over attempts to explain "creature" or "state" consciousness, for example, because the former reflect the active, living, causal nature of conscious psychological phenomena. A wide range of disparate psychological processes have been given illuminating functional explanations: attention, belief, emotion, perception, etc. Attention is assumed to play particular functional roles within perceptual processing, for example, including orienting the perceiver in visual space in order to optimize subsequent perceptual and motor interactions with a visual scene (Zhou et al. 2016). It is uncommon to

[1] In what follows, I'll speak of conscious "processes" as opposed to "states", because I think this better captures the target of inquiry in the cognitive sciences. If consciousness is indeed best understood as part of dynamic psychological processes, all the more prima facie reason to think that we can appropriately talk about its function(s).

simply pursue a descriptive project for these psychological constructs; indeed, descriptive definitions usually come only after an analysis of function, if at all. Centering the discussion around attentional "states" is at best theoretically uninteresting, and at worst metaphysically confused. The constructs that emerge from functional analyses pick out different distinguishable aspects of complex psychological processes. A single psychological process can involve, say, emotionally-driven perception or attended doxastic contents, each aspect playing a distinct role in a complex causal story. The same is plausibly true for subjective experience: it ought to be construed as simply another theoretically and empirically distinguishable aspect of certain psychological processes that provides its own functional contributions to the workings of the system.

Determining consciousness' functions therefore involves treating awareness itself as a causal variable, the manipulation of which would mean a different causal outcome (Woodward 2017). I think that intuitive examples of the advantages of conscious over unconscious processing abound: my awareness of the sound prompted me to go investigate its source; my experience of a delicious meal renders it vivid in memory; my awareness of my recurring anxiety made me seek professional help. If the subjects of these stories had not had particular conscious experiences, things would have turned out differently. However, an adequate theory of the function(s) of consciousness must be precise about what it is that conscious experience is doing in each particular psychological process. Theoretical frameworks must be clear about what it means for a particular function to be associated with conscious experience, and empirical paradigms need to carefully tease out the causal story by way of precise experimental manipulations.

1.1 Degrees of Dissociation

There is growing recognition of the need for a systematic account of what I call the functional contributions of consciousness or "FCCs" (e.g. Frith & Metzinger 2016, Lamme 2020). This is methodologically related to the search for Neural Correlates of Consciousness (NCCs), in that both research programs hope to establish *markers* that can be used as operationalizable proxies for experience. The most common method for isolating the characteristic markers of conscious processes on both paradigms is to compare them with unconscious processes (Baars 1988); either to isolate differences in their neural implementation (NCCs) or to isolate the different representational and behavioural capacities they facilitate (FCCs). In neurotypical subjects, these comparisons are often made possible with masking or suppression techniques that can render perceptual (typically visual) stimuli unconscious on controlled trials (Yang et al. 2014). The comparative paradigm that characterizes the majority of consciousness research has also been largely informed by neuropsychological findings, as pathological or atypical conscious processing can provide a stark illustration of the functional capacities and limitations of unconscious processing (Koch 2004). Impairments of visual function, such as agnosia or visual neglect, provide unique opportunities for assessing the functionality of the visual system in the absence of typical conscious experiences (see Lahav 1993, Ogden 2005, Verdon et al. 2010). The general methodology of the comparative paradigm is to hold all else equal (e.g. on a perceptual recognition task) except for whether or not experience is present (e.g. because of masking or as a result of pathology), the functional consequences of which provide strong reasons for thinking that there are psychological functions that organisms can only carry out if they are conscious.

A crucial methodological point that is often overlooked, however, is that viable philosophical and scientific research on consciousness depends on the identification of markers that do not *dissociate* from conscious experience. One conceptual tool that philosophers have employed for millennia is the notion of necessary and sufficient conditions, the logic of which can be leveraged here. A structure or function is *necessary* for consciousness if it is present in every instance of conscious experience. A structure or function is *sufficient* for consciousness if unconscious processes are never associated with this feature; in other words, the presence of the feature is enough to assume the presence of consciousness. These conceptual points ultimately suggest different degrees of potential dissociation between a candidate structural or functional marker and conscious experience.

If a psychological function or neurobiological mechanism is both necessary and sufficient in this way, it can be considered a marker of consciousness in the strongest sense, and ultimately provides a robust operationalizable proxy for experience that can be employed in subsequent research. Such an intimately associated marker would always and only be present when a system is conscious, and so could ground extremely confident ascriptions of conscious experience across tasks, individuals and species. However, it seems unlikely that there will be many markers that fall into this class (if any), given that conscious experiences form a relatively disparate, heterogeneous category in human beings, let alone across species. In other words, we should expect proposed markers to fail the test of necessity. More generally, philosophers have argued that combined necessity and sufficiency might just be too strict—or too essentialist—as a definitional criterion, and as such will consistently fail to establish viable constructs in the sciences (e.g. Boyd 1999, Khalidi 2013).

The assumption of *functional pluralism*—the hypothesis that consciousness contributes a variety of different functions to information processing systems—implies that FCCs are most likely to fall into the next class. If a candidate marker is sufficient for consciousness (no unconscious processes have this feature) but not necessary (it is not present in every instance of conscious experience), then there is still a relatively robust sense in which it is a genuine marker that can be used in subsequent research. Another way of saying this is that the purported marker is still *unique* to consciousness, because the marker is associated with consciousness to the extent that it is never a feature of unconscious processing. This is true even if the marker fails to generalize. Pluralistic accounts emphasize the crucial point that there are a variety of different kinds of conscious experience, both within individual humans (e.g. visual experiences versus emotional experiences) and across species (e.g. visual experiences versus sonar experiences). This intuition seems to be reflected in Nagel's (1974) influential recognition that a bat's experiences are likely fundamentally different from our own, presumably at least in part because they play different information processing roles based on the specific needs of the organism. Functional pluralism is a general theoretical commitment to the idea that consciousness is likely required for carrying out a range of different functions in different biological systems, implemented by diverse neurobiological mechanisms[2] (setting aside the possibility of conscious experience in artificial systems). In other words, there is no

[2] Similarly, Koch et al. (2016) conclude their seminal review by stating that genuine NCCs also fall into this class: a variety of neural structures and their processing dynamics are sufficient, but none of them alone are necessary, for experience (i.e. structural pluralism). See also Dale et al. 2012.

reason to assume that the same markers will be present in every instance of conscious experience in the natural world. If this is right, it gives us prima facie reason to be skeptical of any monolithic theory that assumes a certain structure and/or function to be broadly generalizable as a marker of consciousness.

Crucially though, the most widely endorsed theories proceed as though we should expect there to be only a single FCC; in other words, that consciousness contributes only one unique functional advantage to information processing systems. Systems that fail to exhibit these singular markers are consequently denied ascriptions of conscious experience on theoretical grounds. Little attention has been paid to the hypothesis that consciousness might facilitate a variety of functions in different information processing domains. I think it is time to update this central working assumption, and take seriously the idea that consciousness is a multifunctional psychological attribute**.** In this vein, it has been argued independently (e.g. Godfrey-Smith 2017) that it at least makes evolutionary sense for consciousness to evolve in response to a range of different selection pressures. Different natural systems plausibly face different adaptive challenges and learning objectives, suggesting that conscious psychological processing would be valuable for different reasons in different systems, based on their relevant processing demands and component functions. Theories that adopt functional pluralism as a working hypothesis, while necessarily being more complex, are thus ultimately poised to explain a wider range of target phenomena. In turn, theories that fail to adopt functional pluralism are limited in their explanatory scope.

When we look at the details, conscious visual perception is functionally very different from a conscious emotional episode or being conscious of one's social biases, which suggests that important philosophical distinctions can be made among kinds of conscious processes and their component functions. There is evidence, for example, that consciousness facilitates such diverse functions as the processing of motion information over larger spatiotemporal intervals in visual perception (Faivre & Koch 2014), the inhibition of response in emotional processing (Ginot 2015), and a handful of bias intervention strategies in the domain of social cognition (Lai et al. 2014). These findings already exhibit the fruitfulness of functional pluralism as a working hypothesis. Even though there is a tendency to reify consciousness into a single phenomenon, some philosophers have attempted to make certain distinctions among kinds of conscious experiences in order to capture some of these differences; for example, sensory versus cognitive phenomenology (Strawson 2011), or phenomenal versus access consciousness (Block 1995). However, we are far from consensus with regards to the kinds of distinctions that are the most appropriate. But one ought to be sceptical that these existing distinctions are fine-grained enough to be truly informative about how consciousness functions in specific psychological domains, as they risk illicitly lumping together importantly different phenomena; for example, lumping conscious emotional episodes and conscious vision under some broad category. Paying more attention to the domains that emerge in the neural and psychological sciences is likely to be a better guide to the proper distinctions to be made in terms of the functions of different kinds of conscious processes. And although it might seem as though there is the possibility of identifying an indefinitely large list of different conscious processes with a similarly large number of functions (e.g. one for each color experience), it is not unreasonable to expect that natural clusters of FCCs will emerge in each processing domain (e.g. within vision or color processing more broadly).

It is certainly possible that a marker is necessary for consciousness (it is present in every instance of experience) but not sufficient (unconscious processes also have the feature in question). This state of affairs constitutes a much weaker sense in which the candidate feature is a marker of consciousness; rather, such purported markers dissociate from consciousness in a way that is problematic for ongoing research. If a purported marker is shown to be necessary but not sufficient, further conceptual and empirical work is needed in order to show its particular relationship to conscious experience. Take recurrent neural processing, or feedback among neural populations, for example. There seems to be emerging consensus that this property of neural interaction is a necessary structural feature for generating consciousness (setting aside the functions it might be associated with): every case of consciousness is correlated with feedback activity between neural populations (e.g. Koch 2019). However, there are also reasons to believe that recurrent processing of this sort is a ubiquitous structural feature of complex information processing systems, and hence that some unconscious processes might also be underpinned by recurrent neural activity. This ultimately means that more work is needed in order to identify the conditions under which recurrent processing generates conscious experience. Various candidates have been proposed in order to answer this challenge to recurrent processing accounts; for instance, that consciousness requires feedback activity at certain activation thresholds (Fisch et al. 2009) or between specific neural populations in sensory cortices (Lamme & Roelfsema 2000). Such specifying conditions might be attainable, and might yield a marker of consciousness that is both necessary and sufficient, albeit in a qualified sense. However, in some instances it might also be theoretically or practically challenging, if not impossible, to identify these qualifying conditions. If this is the case, the presence of the necessary feature alone could not be used to distinguish conscious and unconscious processes, and would therefore be a marker of consciousness only in a weak, “promissory” sense. The candidate marker dissociates from experience in an important way, and therefore could not be used as a reliable proxy for consciousness in subsequent research.

Finally, many purported markers will turn out to be neither necessary (not present in every instance) nor sufficient (unconscious processes also have the feature in question) for conscious experience. In this case, the purported marker is not a marker at all, as its presence or absence can never be used to distinguish between conscious and unconscious processing. This represents the highest degree of dissociation possible, and research programs that champion markers that fail tests of both necessity and sufficiency perhaps ought to be significantly reformulated. These different degrees of possible dissociation between a candidate function and conscious experience are summarized in Figure 1.

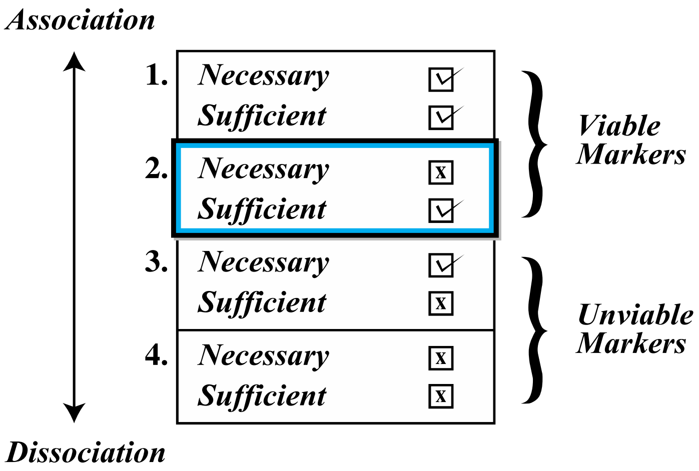

Figure 1. Structures and/or functions that are both necessary and sufficient for experience (1) would be highly associated, and thus extremely reliable markers. However, functional pluralism predicts that this is an unlikely state of affairs given that conscious processes form a diverse class in the natural world. Structures and/or functions must be sufficient for (i.e. unique to) experience in order to allow us to distinguish conscious from unconscious processing, but there is no reason to expect them to be necessary (2); hence, we should expect viable markers to fall into this category. Structures and/or functions that are necessary but not sufficient for experience (3) dissociate from experience in the sense that they cannot reliably be used to distinguish conscious and unconscious processing; they fail as viable markers. Finally, structures and/or functions that are neither necessary nor sufficient for experience (4) represent the highest degree of dissociation, and are therefore the least viable markers.

Frameworks that integrate the search for FCCs with the search for NCCs are especially promising for uncovering markers that can guide future consciousness research. Functional theories benefit from incorporating analyses of the underlying neural mechanisms, and neurobiological models become more theoretically valuable when they take psychological function into account. This obviously makes matters more complicated though, as it requires us to navigate the complex relationships between psychological functions and neurobiological structures. A detailed taxonomy of the different functions that are sufficient for or unique to different kinds of conscious processes, focussed on particular features of their corresponding neural implementation, has not been satisfactorily achieved. I think this is exactly the direction in which consciousness research should go, as such a taxonomy would ultimately provide a diverse arsenal of markers that could be used to ground ascriptions of consciousness in a wide range of systems. It should be noted that this ought to be a fruitful endeavor even if one remains committed to an epiphenomenalist or dualist metaphysical framework: reliable correlations between structural and functional elements of the system and conscious experience are informative regardless of the ontological relationship between them.

I endorse the motivating assumption that a general theory of the nature of consciousness—still the holy grail of consciousness research—should only come after, and should be abstracted from, a detailed account of its functional contributions and their structural counterparts. That is, having a clearer picture of the range of functional contributions that consciousness makes to a system's processing capacities offers crucial clues as to what it is about the nature of consciousness such that it occupies those functional roles and not others. This means that any attempt at theoretical unification under abstract principles must be sensitive to the range of different functions that we uncover in our domain-specific analyses. This research program is still in its infancy, however, and much important work needs to be done in collectively determining and taxonomizing consciousness' functions before we start the process of abstracting general principles. As we systematically explore the different functional contributions of consciousness, we might find either a) important convergences of function that can indeed ground more abstract and general claims about consciousness, or b) important divergences of function that flesh out the pluralistic framework. A successful functional model might even prompt significant revisions to our understanding of consciousness itself as a unified kind, resulting in more scientifically appropriate, function-relevant constructs that reflect principled distinctions among different sorts of conscious processes. But in order to make progress here, we'll first need to look closely at individual psychological tasks within different processing domains, and compare the functional capacities of conscious and unconscious processing. We can try to supplement these functional comparisons by isolating mechanisms in the brain that plausibly constitute the physical substrates of different conscious experiences, using facts about structure to guide and constrain claims about function when possible. After all, approaches that can integrate diverse methods of explanation generally tend to produce the most compelling theories.

2. Consciousness and the Brain

It is fairly common practice to look to the brain[3] in order to supplement our understanding of a particular psychological function. Using facts about structure as a guide to function has been fruitful in many diverse instances across the sciences. However, the relationship between structure and function is complex, perhaps especially in the cognitive sciences, and it has inspired much philosophical attention as a result. Historically, the multiple realizability of psychological functions posed a major challenge for any attempt to reduce them to neurobiological mechanisms (Fodor 1974, Kim 1992). More recently, the notion of "neural reuse" (Anderson 2014, Barack 2016)—which is essentially the flipside of multiple realizability—has been invoked to emphasize the fact that the same neural structure can support a range of psychological functions. This further complicates attempts to map psychological functions onto structures in the brain. These discussions have inspired a growing acknowledgment of intricately "crosscutting" taxonomies employed in the cognitive sciences (Khalidi 2017), involving complex ontological relationships between phenomena described at different levels of organization in the system (Wimsatt 1994).

Despite this complexity, it doesn't follow that mapping structure-function relationships in order to reinforce psychological explanation is hopeless. Many psychological constructs have been illuminated by facts about their neural implementation. We can, for example, learn a lot about the relationship between attention and perception by looking at the neural architecture that facilitates the top-down modulation of gain in sensory neurons (Wu 2017). Similarly, we have achieved a better understanding of emotional processing by looking at the amygdala's patterns of reactivity and connectivity (Barrett & Wager 2006). We can also map psychological

[3] I focus on human consciousness and human brains for several reasons, including the questionable ethics surrounding the study of non-human animals in the cognitive sciences, the ability to leverage report in human subjects as one experimental foothold into conscious experiences, and the sheer complexity of the human nervous system, which increases the likelihood of there being diverse kinds of conscious experience to investigate. However, this project does have important applications to the study of non-human animal minds, and I will draw on existing research using non-human animals when it is relevant to understanding consciousness in humans.

functions to neural structures even if the former are multiply realizable; albeit not in a way that warrants strict reduction. Aizawa (2017) points out, for instance, that vision scientists reliably provide mechanistic accounts of normal color vision in humans, despite significant diversity in the details of each individual's visual system, including differences in the photopigments in the cones of the retina as well as in the structure of the pre-retinal crystalline lens.

The idea of neural reuse can also be leveraged in order to navigate the complex relationships between structure and function, again, without arriving at strictly reductive accounts of psychological processes. Anderson (2014) argues that, so long as we are sensitive to a) the fact that a structure's underlying causal properties (perhaps best construed in terms of its "neuroscientifically relevant psychological factors", or NRPs) enables it to support a number of different functions, and b) the fact that functions emerge from the concerted activity of different coalitions that are dynamically established between neural structures, structure-function mappings are justified and indeed illuminating. The discovery of Broca's area, for instance, was pivotal in understanding the details of various linguistic functions, and yet the same neural region has been shown to be involved in a wide range of cognitive functions, including action recognition and the production of mental imagery (Anderson 2014, p. 4), depending largely on its dynamic network connectivity. One moral to be drawn from all this is that structure-function mappings in cognitive science require more specificity than has sometimes been assumed, and will likely yield only highly local and particular mappings of specific processes. Moreover, these mappings will need to consider the complex and dynamic connectivity between neural structures that is necessary for implementing particular psychological functions. In other words, sweeping generalization, wholesale reduction, and strict localization will fail to capture the intricacies of how neural mechanisms support various psychological phenomena.

Given the value in seeking explanations that integrate different levels of analysis, most existing theories of consciousness look to the brain for clues about its nature and function. In fact, the leading theories—Global Workspace theory, Information Integration theory, Attended Intermediate Representation theory, and First and Higher Order Representationalism—typically appeal in some way to neurobiological facts in order to defend their models. But once again, we need to be cautious about the inferences we draw about the functions of conscious experience from facts about neurobiological structure.

The multiple realizability of psychological phenomena needs to be taken seriously in the study of consciousness. On one hand, it might lend support to the idea that different kinds of nervous systems, and perhaps even artificially designed systems, could potentially support conscious experience. There might be nothing especially significant about the human brain—except for the sheer magnitude of neural connectivity—in terms of its ability to realize consciousness, so long as other systems have similar causal features. This is an outstanding theoretical and empirical issue, however, and it is far from achieving consensus. On the other hand, and more relevant to the current project, multiple realizability suggests that different structural networks in the same brain might realize different instances of conscious psychological processing. This suggestion, in turn, does seem to be borne out by the empirical data (Koch et al. 2016). It is an open issue whether or not facts about NCCs will help to determine whether the proper taxonomical strategy is to lump or split these different conscious processes, but I think this will have to wait until a more detailed account of their functional similarities and differences emerge.

The lessons of the neural reuse framework must also be acknowledged in the study of consciousness. Neural reuse suggests that the same brain region can support conscious or unconscious processes, depending both on the region's underlying functional properties (or NRPs) and its dynamic connectivity with other neural structures. This suggestion has also been borne out by the empirical data: the leading scientific theories of consciousness acknowledge that their proposed NCCs include neural structures that support a range of different psychological functions, including unconscious ones (e.g. Lau & Passingham 2007, Boly et al. 2017). In other words, there is growing recognition of structural and functional overlap between conscious and unconscious processes. The amygdala has been found, for example, to perform the same representational valencing in both conscious and unconscious emotional processing networks (Diano et al. 2017). This common finding will be particularly helpful in developing an effective comparative approach to isolating FCCs.

Once again, the moral is that consciousness research needs to focus on particular neurobiological networks underlying particular conscious processes, because the diverse structures that support different conscious experiences likely do so by dynamically interacting in different neural coalitions. Looking closely at the particular structural and functional overlap of neural coalitions that support particular unconscious and conscious processes allows us to sharpen the standard comparative methodology into a more precise process- or task-specific analysis. Researchers typically compare conscious and unconscious performance on some set of tasks, mapping neural activation and assessing psychological performance along the way, ultimately so that they can draw general conclusions about conscious experience. We can update this comparative method by subtracting more "locally" shared structural and functional components from the additional structural and functional components that are only present in particular instances of conscious processing. This should allow us to tease out a variety of functional contributions that are unique to consciousness in different task-domains, and ground them in the functional properties of the relevant neural structures. I hypothesize that, given the diversity of conscious processes and their neural correlates, we should expect to uncover a variety of psychological functions that are only present when a particular process is conscious.

Besides the need to navigate complex structure-function relationships, another reason for caution must be made explicit here. Namely, there are problems surrounding a kind of functional subtraction that is rooted in the "assumption of pure insertion"—or the assumption that each interacting component of a network in the brain will have a cleanly distinguishable function, like individual functional building blocks that can be reassembled in various ways to produce mere conjunctions of function (Poldrack & Yarkoni 2016). By analogy, it would be a mistake to think that we could subtract sodium from an NaCl molecule in order to cleanly carve up its individual functional properties. The different components of molecular compounds combine in ways that are nonlinear, in the sense that they are more than the mere sum of their parts. Given that psychological functions also necessarily emerge in the interactions between neural structures, neuroscience must also explicitly work to identify significant network properties; that is, the type of interactivity between component structures is crucial for understanding the functions that they facilitate. In terms of mapping functions to these different structural partnerships then, it will likely turn out that progress will require the identification not only of NRPs for each component region, but also for each significant kind of network property. This sort of suggestion has been made in other areas of cognitive science. Mahon (2015, p. 80, 102), for example, has argued that understanding the neural underpinnings of conceptual representation will require treating the connectivity between different modal systems as its own "unit of analysis".

As previously mentioned, one network property that is consistently implicated in producing conscious experience is recurrent processing, or feedback activation between certain neural regions. Neuroimaging has consistently revealed the significance of recurrent processing circuits in facilitating conscious experience (e.g. Lamme & Roelfsema 2000, Dehaene et al. 2006, Koch et al. 2016, Auksztulewicz et al. 2012). Lamme's (2006) work on recurrent processing in the visual system even inspired a formal theory of consciousness built around the significance of feedback activity. Indeed, the proposed NCCs in most of the leading theories of consciousness implicate recurrent processing in some way. It is important to take seriously the idea that there is something crucial about recurrent processing circuits for the generation of conscious experience, despite the need to distinguish the properties of recurrent processing circuits that underpin conscious information processing from any that underpin unconscious information processing. Recurrent processing circuits have some common features, which might be plausibly construed as the underlying causal or functional properties (i.e. NRPs) of this particular kind of network configuration (Douglas & Martin 2007). These include gain modulation, extraction, selection, and amplification of feedforward neural signals. There seems to be something significant about maintaining information in recurrent circuits for further signal processing, likely between very particular cortical regions, for the generation of conscious experience. This fact should ultimately guide and constrain the ongoing functional analysis of consciousness.

In line with the pluralist picture being developed here, we ought to take seriously the hypothesis that the NRPs of recurrent processing circuits facilitate different functions for different conscious processes in different information processing domains. For instance, it is an intuitively plausible hypothesis that gain modulation will be useful in different ways for different processing tasks; facilitating different processing capabilities in visual perception, emotional processing, and social cognition, for example. In contrast, existing accounts typically latch on to one isolated neural circuit involving recurrent activity as necessary and sufficient for conscious experience, in order to supplement their assumption that consciousness plays a singular functional role in the psychological system.

In the remainder of this chapter, I survey the leading accounts of consciousness, and will argue that their proposed functional markers are not generalizable across different conscious experiences and their distinct neural correlates (i.e. they are not necessary), which is to be expected given the assumption of functional pluralism. This ought to clear the deck for future work aimed at isolating genuine functional markers that are in fact unique to (i.e. sufficient for) consciousness, despite their lack of generalizability. Only after this sort of analysis should we begin to draw general conclusions about the nature of consciousness from the specific kinds of functional capacities it facilitates.

3. Existing theories

3.1. Global Workspace Theory

Global Workspace Theories (GWTs) are widely assumed by both philosophers and scientists to provide a plausible general account of consciousness. At the theoretical core of GWTs, conscious experience is argued to be inextricably linked with the cognitive notion of "access" (Dehaene 2014). Information enters into conscious awareness only when it is actually[4] accessed in a way that makes it available for use by a range of different processing subsystems. Consciousness is the "global broadcasting" of information throughout the psychological system, amplified and maintained in a limited-capacity, but highly connected, centralized "workspace" (Baars 1988). On this model, consciousness plays a "supervisory" role, in the sense that the global workspace allows flexible control over the information processing in its jurisdiction. This global informational access is argued to be definitive of all conscious psychological processing.

[4] And not merely potentially. See Stoljar (2019) for a critique of the idea that conscious access should be defined dispositionally.

The notion of access requires careful unpacking. At one point in its development, the global workspace model seemed to equate access with attention (e.g. Dehaene et al. 2006, Kouider & Dehaene 2007). These earlier iterations proposed that conscious experience requires that two conditions be met: a) an incoming signal must be strong enough that it can potentially be registered by the appropriate access mechanisms—ruling out, say, stimuli that are presented too quickly and too closely together as in masking paradigms (see Enns & Di Lollo 2000)—and b) *attention* must be directed towards the signal in order for the wider system to actually access it. However, in later iterations (e.g. Dehaene & Changeux 2011, Dehaene 2014), access is explicitly distinguished from attention. On one hand, attention is argued to be insufficient for access, as attention often operates on unconsciously processed information that remains inaccessible (Koch and Tsuchiya 2007). On the other hand, attention might not even be necessary for access, because on some interpretations of certain experimental results (e.g. Wyart & Tallon-Baudry 2008), access occurs in the absence of attention[5]. On most accounts, however, attention and access remain closely related, and it is typically assumed that attention is either a “prerequisite” (Dehaene & Changeux 2011) or else indeed a necessary condition (Stoljar 2019) for access. The crucial point though is that access is intended to be more psychologically complex than attention according to GWTs, involving both the relevant uptake mechanisms (which either typically or necessarily involve attention) *and* downstream broadcasting to different subsystems for use in specific tasks. Rooted in this account of access, GWTs propose a fairly straightforward descriptive account: “consciousness is brain-wide information sharing” (Dehaene 2014).

GWTs rely heavily on facts about neural architecture in theorizing about the nature of consciousness. Global informational access is thought to be realized by very particular neurobiological networks (Dehaene & Changeux 2011). The central workspace is implemented by long-range excitatory neurons that form recurrent corticocortical connections between prefrontal, parietal and cingulate regions. In addition to this, access to and dissemination of information is made possible by recurrent thalamocortical loops that link the central workspace

[5] Although it may be that access occurs in the absence of *overt* attention, but not *covert* attention. Without alternative suggestions for specific access mechanisms that do not involve attention at all, this interpretation of the data seems plausible.

to various specialized processing networks. These proposed NCCs were identified using the same common contrastive methodology that builds on the pioneering work of Baars (1988), and has become fairly standard in the field. A variety of stimuli are presented to subjects, some of which reach conscious awareness and some of which don't. For example, it is common practice to briefly present words and images to subjects that are either readable/visible or unreadable/invisible depending on whether or not they are masked by other stimuli (Dehaene & Naccache 2001). Subjective reports of the awareness of the stimulus (often supplemented with confidence measures) are typically relied upon to determine the threshold of conscious awareness, and establish minimal contrasts between conscious and unconscious processing (Dehaene & Changeux 2011). Various imaging techniques, including fMRI, EEG and MEG, are then used to monitor brain activity, ultimately in order to distinguish the neural correlates of conscious versus unconscious processing. Although many aspects of this experimental paradigm remain controversial, GWT theorists have taken the correspondence between activity in anterior cortical regions and subjective reports of awareness as evidence for their claim that global informational access is the essence of conscious processing.

The GWT framework is uniquely explicit about the function that it assumes consciousness contributes to information processing systems; that is, global informational access is often explicitly formulated in functional terms. In his book, Dehaene is clear about the fact that the GWT model construes consciousness as fulfilling "a particular operational role" (Dehaene 2014, p. 103). He writes,

> "Consciousness has a precise role to play in the computational economy of the brain—it selects, amplifies, and propagates relevant thoughts…Although unconscious processing can be deep, conscious access adds an additional layer of functionality. The broadcasting function of consciousness allows us to perform uniquely powerful operations." (Dehaene 2014, p. 103)

Note that GWTs posit a single functional contribution that consciousness makes to an organism's psychological toolkit: *the* broadcasting function of consciousness. Even though it enables a variety of "uniquely powerful operations" downstream, consciousness itself facilitates one singular functional capability—namely, accessing information and keeping it "active" in the flexible, brain-wide broadcasting network—that grounds all of these secondary capabilities. Dehaene proposes, for instance, that resolving ambiguities in visual processing, keeping information in working memory, and applying flexible strategies in problem solving, are computational by-products that are made possible by the function of global access (Dehaene 2014). And even though Baars' (1988) initial formulation suggested a "great number of useful roles played by consciousness" (p. 347), he too refers to a "most fundamental function" (p. 348) and assumes that any other proposed functions all "really belong to the entire GW system" (p. 347). In other words, there is no pluralism of function being assumed here, in the sense that consciousness enables a variety of distinct FCCs in different processing domains. Rather, consciousness is construed as a system-wide phenomenon that confers unique but centralized information processing capacities as a result of its fundamental causal role in global informational access. Considered as candidate markers, any downstream functional by-products require some theoretical work to link them back to the primary marker of global broadcasting. Global broadcasting therefore is still taken to be the sole direct marker of consciousness, which we might find evidence for in some downstream operation that the global workspace enables. Thus, the flexible executive function enabled by global informational access alone is assumed to be both necessary and sufficient for, and hence a strong functional marker of, conscious experience.

The core assumption that consciousness facilitates flexible executive processing has deep historical roots (e.g. Umilta 1988, Jack and Shallice 2001). Interestingly, the flexible processing that global informational access enables has been argued by some to be characterized, at least in part, by cognitive mechanisms that allow the "chaining" together of simpler processes into more complex wholes. Sackur and Dehaene (2009) employed a simple multi-step arithmetic operation in order to provide empirical evidence that while individual elementary operations can proceed without awareness, conscious experience is required to "chain" more than one operation together. It is thought that this is at least one way in which consciousness enables a kind of processing flexibility that is otherwise absent.

The evaluation of GWT's singular functional claim involves determining whether or not the model offers a generalizable FCC; that is, whether every conscious process is best characterized as implementing the global broadcasting function; whether this function is a necessary feature of every conscious experience. There are reasons to think that GWT fails to offer an FCC that can be generalized to capture all instances of conscious processing because some conscious processes do not seem to involve global informational access. On most theories of self-deception, for instance, information that individuals are consciously aware of can sometimes fail to be accessed by the wider system in some way, such that it fails to influence behaviour, emotional assessment, and the rational formation of belief (Mele 2001). For example, a mother might fail to acknowledge that her son is engaged in criminal behaviour despite bailing him out of jail for robbery and theft on numerous occasions. Also, although GWT assumes a modular architecture for the specialized subsystems that feed into the global workspace (Baars 1988), perceptual illusions could in principle be construed as cases in which information that we are conscious of (e.g. that the two horizontal lines in the Muller-Lyer illusion are equal in length) fails to be truly globally accessed, as it is not freely used by the systems that produce our visual experience of the illusion. Admittedly, these potential counterexamples will strike many as controversial, and therefore require much more theoretical elaboration if they are to count against GWT's necessity claim.

More convincingly perhaps, there are reasons to think that conscious perceptual processes are not always accessed by the wider system because high capacity perceptual (e.g. visual) experiences facilitated by sensory systems are plausibly more informationally abundant than what can be processed by the low-capacity global access system (Block 1995, 2011, Tononi et al. 2016). Support for this kind of "overflow" comes in part from theoretical considerations about neural and psychological architecture, as well as intuitive interpretations of introspective data. But some empirical considerations seem to provide fairly concrete evidence of conscious experiences that overflow what subjects access. On one historically influential paradigm, Sperling (1960) constructed a simple but powerful experimental design, in which only a subset of stimuli that enter visual consciousness (and are perhaps stored in "iconic" memory) seem to be accessed cognitively (e.g. by working memory). Twelve items (e.g. letters) in a grid are shown to subjects, and typically about 3-4 of them can be freely recalled. What is significant, however, is that subjects can typically recall any row of items that is cued by the experimenters after the initial presentation, suggesting that the entire grid enters visual consciousness, but only a subset are subsequently accessed for report depending on the nature of the cue. Block (2011, 2014) and others compellingly argue that the best interpretation of this data is that much of the information that is processed consciously by the visual system remains unaccessed, despite being accessible. Tononi et al. (2016) make a similar point, arguing that "the information that specifies an experience is much larger than the purported limited capacity of consciousness" (p. 457).

There are still certainly opponents of this sort of perceptual richness in the GWT camp and beyond, and this issue remains hotly contested. Gross & Flombaum (2017), for example, argue that post-stimulus cueing actually prompts the subject to sample from probabilistic representations that are initially constructed as inferences about the noisy signal they are presented with, and that there is no need to invoke conscious experience at this earlier stage of visual processing. Philips (2011), in contrast, interprets the results of Sperling-style experiments as evidence of postdictive perceptual modulation, according to which the auditory stimulus presented immediately after the visual array actually changes how the array is visually perceived (i.e. what information is privileged). Many opponents also appeal to phenomena like inattentional and change blindness—namely, the failure to recognize significant modifications to a visual scene—as evidence that consciousness is limited to what is accessed by the global broadcasting system (Cohen & Dennett 2011, Dehaene 2014). The apparent richness of visual perception is even argued by some to be illusory, brought about by a subject's inflation and overestimation of the available visual information outside of conscious access (e.g. Odegaard et al. 2018). This is assumed to be further supported by a poverty of input mechanisms in the retina at the periphery of vision, which is presumably where we might be confident in locating unaccessed visual information (Lau & Brown 2019).

However, Sperling's general experimental framework has been modified and updated in various ways, and continues to yield similar results. Bronfman et al. (2014), for example, cued subjects to attend to a particular location on the grid in a letter identification task, and then subsequently asked them to report different features of the other letters in the grid—namely, whether their color distribution had high or low diversity—that were outside of focal attention.

Subjects were able to distinguish between high and low color diversity despite having their attentional resources focussed elsewhere. The most plausible interpretation of this updated experimental design is that the visual system extracts summary statistics from the experienced elements of a visual scene independently of cognitive access (Usher et al. 2018). Opponents of overflow have to posit that the conscious judgment of color diversity is based entirely on unconscious representations of the elements of the scene; in other words, that we experience the relation (i.e. color diversity) without experiencing the relata (i.e. individual color elements). To test this, Usher et al. (2018) ran a Sperling-like experiment, which revealed that subjects were as good or better at discriminating which letters were presented outside of focal attention (i.e. individual elements) as they were at indicating whether they were the same or different (i.e. summary statistics). They take this as compelling evidence that experiencing and discriminating a second-order statistical summary outside focal attention requires differentiated conscious representations of the individual elements (Usher et al. 2018, p. 6). This ultimately suggests that consciously processing the “gist” of a visual scene (this term is employed by Koch 2019, and others) does not seem to require cognitive access.

Usher et al. (2018) go on to argue that conscious visual perception likely overflows reportability because the overflowing stimuli are typically task-irrelevant and so, despite being represented, differentiated and mined for summary statistics, the individual elements remain uncategorized, unconceptualized, and hence not made available for report. To illustrate this point, imagine your experience of walking through a well-stocked grocery aisle, focussed on finding the pickles. If asked to report whether you passed by salad dressings, you are likely able to, as this simply requires that you extracted a summary or gist of the visual scene outside of focal attention. If asked whether you saw a particular bottle of salad dressing though, you will likely be unable to, because you were not looking for salad dressing, and so your visual system did not waste resources on individuating them in detail and categorizing them as you passed by. This task-irrelevance provides an elegant explanation of why subjects often fail to report elements of the visual scene outside of focal attention in inattentional and change “blindness” studies.

It is clear that more sophisticated and compelling philosophical interpretations of these data continue to be offered. In a recent paper, Jake Quilty-Dunn (2020) argues that these Sperling-like experiments can be explained further by taking the different representational formats of iconic memory versus visual working memory into account. He argues that the data suggest that iconic memory—a perceptual phenomenon—encodes rich, feature based, holistic representations that draw on statistical properties of stimuli, while visual working memory—a cognitive phenomenon—encodes object-based discursive representations, due to the different representational formats employed by each system. Further, he argues that if the only representations that enter consciousness are those discursive representations that are accessed by working memory, then we get the odd conclusion that visual representations that employ iconic formats are never conscious. This is not only intuitively absurd, but it directly contradicts generally acknowledged principles of visual perception, like the topographical isomorphism of visual representations. It seems much more plausible, contrary to GWTs, that iconic visual representations are conscious and informationally abundant, while subsets of this information can be, but need not be, taken up by access mechanisms for further downstream (i.e. cognitive) processing.

There are several other lines of reply to overflow sceptics. First, it seems that the richness of retinal input in the periphery has in general not been fully appreciated (Block 2019a); in fact, Tyler (2015) argues that this assumption of inadequate color sensitivity in the periphery, for example, is a “widespread misconception even among vision scientists”, as mechanisms are in place that can support sophisticated color processing outside of the fovea. Second, it seems odd to assume that we are systematically deceived about the richness of our own experiences, and Block (2019b) argues that cases of purported subjective inflation in the lab are plausibly best understood as tapping into perceptual judgments (e.g. confidence) rather than perceptual experiences themselves. Third, the claim that all conscious experiences require cognitive access (or at least partial access-see e.g. Kouider et al. 2010) simply seems to invoke an implausible cognitive and neural architecture. GWTs imply that every element of a particular instance of experience—which can be staggeringly complex (e.g. feeling elated at the sights and sounds of a vast expanse of natural beauty)—requires a corresponding access mechanism, which places architectural demands on the access system beyond what is psychologically and neurally plausible. Finally, there is the sheer introspective implausibility of the claim that we are only conscious of what we access, which implies a fairly limited scope of experience at any given moment, given the low-capacity of access mechanisms. Koch (2019, p. 38) similarly notes: "when I attend to a particular location or object, intently scrutinizing it, the rest of the world is not reduced to a tunnel, with everything outside the focus of attention fading away". The hypothesis that conscious visual experience overflows cognitive access clearly has broad explanatory power, and is generally coherent with a range of philosophical and scientific considerations.

One final line of criticism that is often made about the generalizability of GWTs concerns the theory's proposed NCCs (e.g. Koch et al. 2016, Boly et al. 2017, Storm et al. 2017). The idea is that if prefrontal cortex and related structures are necessary for realizing the global workspace, then we should in turn be able to appeal to PFC activity to determine whether the global workspace is a necessary feature of consciousness more generally. Because the evidence seems to be pretty definitive that conscious experiences occur in the absence of PFC involvement, however, GWTs do not seem to represent a generalizable theory of consciousness. Patients with bilateral frontal lobectomy or severe prefrontal trauma, for example, report preserved conscious experience, and their behavioural interactions with the world appear normal even to trained neurologists (Boly et al. 2017). In addition, electrical stimulation of posterior (i.e. sensory) cortical regions can elicit specific experiences, which is not the case for more anterior regions of the cortex (Koch et al. 2016). It is also intuitively plausible that non-human animals that lack the kind of PFC found in human brains nevertheless have conscious experiences (Tononi 2012a). Finally, there is evidence that the PFC and related regions can carry out functions associated with the workspace—including cognitive control, response inhibition, task switching, conflict monitoring and error detection—entirely unconsciously (van Gaal & Lamme 2012a, Kouider & Faivre 2017). These neurobiological considerations reinforce the criticism of the GWT model, if global informational access is indeed realized in large part by the PFC.

Empirical and theoretical accounts, therefore, seem to be converging on the idea that it is necessary to posit a range of psychological processes that are conscious but are not being cognitively accessed and globally broadcast throughout the system, and rich (e.g. visual) perceptual experiences are a paradigm example. This suggests that at least some conscious processes seem to have psychological functions that are distinct from global informational access, which in turn opens up the conceptual space for ascriptions of consciousness to systems that lack these functional and structural features. The fact that these cases are incompatible with GWTs reflects their failure to endorse functional pluralism. Block's (1995) distinction between access and phenomenal consciousness was intended as a way of capturing a distinction between the experiential aspects of consciousness and a particular functional role that characterizes *some* conscious processes. Whether or not this will turn out to be the most theoretically and empirically useful delineation of "kinds" of consciousness, the notion of conscious access helps us pick out the relationship between some conscious experiences and certain cognitive capabilities. Instances of conscious experience that are not accessed remain unexplained on the GWT framework, however. Again, a detailed, process specific analysis is required to tease out exactly what some of the FCCs of these conscious processes are.

We could also evaluate GWTs in terms of whether or not they offer a candidate function that is unique to experience; that is, whether the proposed function is sufficient for consciousness. In other words, if something like global informational access can be carried out unconsciously, this is a clue that GWTs are on the wrong track altogether. Frith and Metzinger (2016) argue that there is no obvious reason why global informational access should be associated with subjective experience, given that this kind of access is important for a wide range of psychological processes, including those that occur unconsciously. Research on unconscious perception (e.g. Block & Philips 2016), subliminal priming (e.g. Hassin et al. 2007, Kiesel et al. 2007) and the phenomenon of blindsight (e.g. Alexander & Cowey 2010), for example, all assume the availability of unconsciously processed information for a range of downstream processes (e.g. discrimination tasks, detailed motor responses, etc.). Research on implicit social bias (e.g. Oswald et al. 2013) similarly assumes that unconsciously constructed and maintained representations seep deeply into the system and influence a wide range of psychological processes—everything from perception to explicit reasoning to behaviour—in a way that is consistent with some form of wide-ranging access.

These sorts of research paradigms suggest that the very same subsystems that are implicated in the global workspace model—namely, evaluative systems, long-term memory systems, attentional systems, perceptual systems and motor systems (Dehaene & Changeux 2011)—dynamically and flexibly share information that is processed unconsciously. As a result, there is at least a sense in which access is not unique to, or is not sufficient for, consciousness, and hence is a poor candidate for characterizing what it is that consciousness contributes to information processing systems. It will likely turn out that a certain *kind* or *degree* of access and flexible processing is only possible with conscious experience. However, these considerations complicate GWT's attempt to identify consciousness with accessibility for downstream processing.

Despite its falling short as a generalizable theory of consciousness and its psychological functions, there are still insights to be drawn from the GWT framework. Again, it will plausibly turn out that access for global broadcasting—of a certain kind or degree perhaps—is a unique functional contribution of *some* conscious processes, regardless of whether or not GWTs can be generalized into a unified account. That is, it is possible that some forms of global access can only be performed consciously, rendering it one viable FCC among many. Additionally, GWTs also emphasize the idea that recurrent activity, where information is "kept active" by reciprocally and dynamically communicating neural structures, is crucial for consciousness, even if they fail to recognize that this likely applies outside of a centralized workspace and wide-ranging broadcast network. For the present purposes though, it is evident that GWTs are not compatible with the idea that conscious processing facilitates utterly different FCCs in different psychological domains—especially FCCs that do not involve global informational access—which limits the applicability and plausibility of the overarching theoretical framework.

3.2 Higher Order Thought Theory

There is another widely endorsed theoretical framework that is conceptually similar to GWT. Higher order thought (HOT) theories also assume that what renders representational content conscious, say an instance of visual perception, is the additional processing of that content by some additional cognitive mechanism (Rosenthal 2005, Genaro 2018, Lau & Brown 2019). HOT theory takes the target of explanation here to be mental "states" understood as discrete representations. The theory endorses a sort of neo-Freudianism, in that it assumes that the same state with the same content can pass freely between consciousness and unconsciousness. Conscious experience is brought about when the subject of a first order mental state "becomes *aware of* being in that state"; a claim that has been formalized in the Transitivity Principle (Rosenthal 2005). There are various proposals for how best to characterize the relevant higher order mechanisms; for example, one recent account suggests that consciousness arises because the brain is continuously and unconsciously learning to representationally redescribe its own activity to itself (SOMA theory, see Cleeremans et al. 2020). Regardless of how this is fleshed out, the key difference with GWT is that the cognitive mechanism that HOT theorists think is both necessary and sufficient for consciousness is construed as a second order (i.e. meta-representational) "inner" awareness—sometimes described as self-consciousness (e.g. Siewert 2013)—as opposed to signal boosting first order representational content for global broadcasting (Brown et al. 2019).

Another difference between HOT and GWT approaches has to do with their proposed NCCs. While most GWT theorists assume that recurrent activity involving the PFC constitutes the neurobiological underpinnings of the cognitive mechanisms that consciousness facilitates, HOT theorists have not seemed to reach consensus about this. Genaro (2018), for instance, denies that the PFC is necessarily a neural correlate of the higher order mechanisms that bring about conscious experience, while seminal papers (e.g. Lau & Rosenthal 2011, Brown et al. 2019) assume that the PFC is a likely candidate for underwriting the relevant higher order representations. However, the motivation for HOT theories is not to reduce or even correlate conscious experiences to neurophysiological processes, but rather to reduce them to particular psychological and representational processes. As such, HOT is a philosophical framework that is not necessarily coupled with claims about how consciousness is realized in the brain, and so the appeal to structure as a guide to claims about function can be bracketed here.

Contrary to GWT, many versions of HOT are also explicit in claiming that consciousness is not appropriately described in functional or causal terms. In his book outlining the HOT framework, Rosenthal (2005) writes,

> “There is a temptation to think that consciousness has a function, so that the consciousness of our beliefs and desires makes a difference to our reasoning and planning. But it is likely that what makes such a difference is only the content of these states, and not their being conscious.” (Rosenthal 2005, p. 17)

In more recent work as well (e.g. Lau & Rosenthal 2011), it is argued that “awareness might not have a special utility,” but that ultimately HOT theory is “not committed to the prediction that awareness has no function whatsoever, it is neutral with respect to this issue” (p. 368). HOT’s scepticism, or at least agnosticism, about the functional role of conscious experience seems to be rooted in its motivating assumption that conscious and unconscious representations are functionally alike, which would certainly imply that consciousness adds nothing to a mental state’s functional profile.

However, other HOT theorists are sympathetic to the functional project. In a recent discussion of threat response that assumes the HOT framework, LeDoux and Daw (2018), for instance, argue that “human conscious experiences have a decisive role in behaviour that is conceptually, psychologically and neurally distinct from the processes that control reflexes, fixed reactions, habits, contingency-dependent learned behaviours and even behaviours on the basis of implicit forecasting” (p. 278). Once again, explicit formulations of the functional implications of HOT reflect the familiar assumption that consciousness’s primary functional role is ultimately to facilitate flexible executive processing. This is made explicit on the SOMA version of HOT theory (Cleeremans et al. 2020), according to which the function of consciousness is to redescribe the brain’s activity to itself (via meta-representation), which enables flexible adaptive control over the organism’s behaviour. The idea that consciousness facilitates “deliberative explicit decision-making and behavioural control” (p. 279) and is highly correlated with, or perhaps even inextricable from, flexible executive control functions and the more complex cognitive and behavioural capacities they facilitate clearly has a deep history in consciousness research (e.g. Umilta 1988, Norman and Shallice 1986, Hommel 2007, van Gaal et al. 2012, Earl 2014). This functional claim aligns with GWT's assumption that when information is consciously accessed and globally broadcast, the wider system can exert deliberative control over a range of cognitive and behavioural processes.

This means that HOT either dodges the issue of function altogether, and is therefore ill-equipped to isolate viable FCCs, or else is subject to some of the same criticisms as GWT, as the same sort of functional consequence is proposed for both access and meta-representational mechanisms. As we have already seen, it is unlikely that all conscious processes involve the kind of higher order representation thought to bring this flexible processing about; for example, high capacity perceptual systems plausibly overflow limited capacity higher order systems. Because of this, HOT also fails to generalize as a theory of the functions of consciousness. HOT theorists might appeal to the methodological difficulty in empirically studying purported phenomenally conscious processes that one is not having a higher order thought about, but this doesn't necessarily show that such processes don't exist, but rather reveals deep methodological issues for all consciousness research that relies on report paradigms (see Usher et al. 2018 and Block 2019b for novel solutions to this problem). The overflow claim, however, is rooted in facts about the processing capacities of certain psychological subsystems and their neural substrates, as well as plausible interpretations of repeatedly employed empirical paradigms. Further, HOT is also subject to the architectural and introspective implausibility arguments, as well as the argument from plausibly conscious organisms that lack a functioning PFC (if HOT's are indeed supported by these structures). As it stands, HOT also fails to provide an account of FCCs that does justice to the plurality of conscious processing.

Despite its failure to generalize, meta-representation is a notion that is likely defined so as to be sufficient for consciousness; that is, organisms engaged in the kind of meta-representation described by HOT, and the kind of flexible control over information processing that it facilitates, probably have corresponding conscious experiences. Thus, both GWT and HOT models ultimately propose a kind of flexible executive processing as the sole FCC, albeit underpinned by different specific cognitive mechanisms. Once again though, HOT's sufficiency claims require careful consideration of an adequately large number of specific instances of meta-representation-enabled flexible control coupled with evidence of experience, both in human beings and across species, in order to secure them as viable markers of consciousness. The hope is that employing the kind of nuanced, process specific approach suggested here will help to flesh out which particular conscious experiences among many involve the functional advantages that cognitive processes like meta-representation confer.

3.3 First Order Representationalism

For many philosophers, First Order Representationalist (FOR) accounts of consciousness are preferred, because they are thought to avoid the problems that higher order theories face (e.g. Tye 1995, Lycan 2001, 2019). On these views, conscious experiences are determined by representational content. That is, there is nothing more that needs to be explained about consciousness beyond giving an account of the representational contents of a particular conscious experience. Given the widely accepted idea that there are unconscious representations as well (e.g. Augusto 2013, Mattiassi et al. 2014, Poldrack 2020), some FOR theorists, like Dretske (1995) and Tye (1995), argue that the difference between conscious and unconscious representations has to do merely with the manner in which a representation is used by the system (e.g. they are "poised" for use by central cognition, see Bourget & Medelovici 2013). Others claim instead that consciousness can be explained by appeal to a distinctive sort of representation, say the representation of something like "sensory qualities," understood as properties of objects and events that are eligible for being present in experience (e.g. Lycan 2019). On Lycan's (1996) account, for instance, conscious processes are essentially representational processes that track those properties of external objects that are possible objects of experience. Like HOT, FOR theories are not committed to claims about NCCs (except for Recurrent Processing Theory discussed below), so once again these issues can be bracketed.

As theoretically elegant as reducing consciousness to representation might seem, long-standing debates about just how to properly understand the nature of conscious and unconscious representational content have muddied these waters significantly. The major theoretical rift lies between externalist and internalist accounts of content. Much ink has been spilled debating these positions, and a thorough summary is beyond the scope of this project.

But briefly, internalist representationalism assumes that intrinsic facts about representing systems are sufficient for explaining and individuating representational contents (e.g. Shoemaker 1994, Kriegel 2002). Because internalists take representational content to be fixed by facts that are intrinsic to representing systems (e.g. content supervenes on intrinsic properties of the brain), internalist representationalism implies that consciousness can be accounted for entirely by understanding how representations are constructed by the system for use in the specific processing tasks that it carries out. Several arguments and thought experiments—most famously, Putnam’s Twin Earth scenario (1975)—challenged internalism on the grounds that duplicate systems might plausibly represent different contents based on facts external to them, and as a result, many representationalists have adopted an externalist framework. Externalist representationalism assumes instead that content is determined by facts about how representing systems are causally and informationally hooked up to the world. Theorists have appealed, for example, to the “transparency” of representational content upon introspection to further argue that content is determined by facts about the objects and events that systems come to stand in causal and informational relations with (Harman 1990, Tye 1995). The transparency argument claims that upon introspection of our perceptual experiences, we “see right through” to the properties of the external objects themselves, suggesting that there is nothing more to perception than what is caused by these external properties (Harman 1990). Externalist representationalism, as a theory of consciousness, therefore implies that conscious experiences depend upon facts about the external world that the system happens to represent.

It should be stated that FOR theories are typically not explicit about consciousness’ functional role, presumably because they were proposed as alternatives to accounts that appeal to functional properties in their explanations of the nature of phenomenal experience (i.e. Functionalism). However, representationalism is at least conceptually compatible with the idea that conscious representations occupy specific functional roles in psychological systems, as this recognition need not lead to a wholesale reduction of experience to functional role (Lycan 1996). Thus, we can conjecture that if conscious experiences are reducible to representational contents, then the function of consciousness might be derivable from the function of representational contents. While there is one sense in which representations play a variety of functional, causal or computational roles—indeed on some views, perhaps representations are necessary for the functioning of all psychological processes—there is another sense in which representationalist theories imply a unitary functional contribution that the ability to represent makes to information processing organisms. On some externalist representationalist accounts, for instance, the function of representing is roughly to establish a particular tracking relation (e.g. Dretske 1995, Tye 1995) whereby information bearing processes in a system track objects and events in the external world. Thus, the function of consciousness might also be to facilitate a particular kind of tracking relation, perhaps specifically tracking the unique sensory qualities of objects and events that are "eligible" for being experienced (Lycan 1996). In contrast, internalist representationalism assumes that rather than necessarily enabling an informative relation to the external world, representations are in the business of facilitating certain information processing capacities in complex systems. This would suggest that the function of conscious experience is to facilitate a subset of representational processes, perhaps distinguished by their unique computational role and/or biological substrates. In the absence of consensus on the issue of just what sensory qualities or computational roles are included in the subset that distinguishes conscious from unconscious processing, the precise function of consciousness remains indeterminate.

Thus, the major challenge for this kind of functional interpretation of representationalism, besides overcoming the internalist/externalist impasse, is saying what it is that distinguishes conscious from unconscious representation. This will ultimately help us determine what it is that consciousness adds to a subject's information processing capacities. At first glance, the proposals already on offer are not as illuminating as one would hope. Saying that conscious representation is in the business of representing the kinds of properties of objects that are capable of being conscious, for example, is largely uninformative. The same is true of attempts to distinguish unconscious from conscious representation by appealing to functional or computational role; that is, by claiming that conscious representations are the ones that play a role in conscious processes. As they stand, these are circular explanations, and as such, they do not satisfactorily specify what it is that consciousness uniquely contributes to the functional capacities of information processing systems. The crucial point here though is that although these theories gesture towards monolithic accounts of consciousness' function based on monolithic accounts of the nature and function of representation, they are perfectly compatible with the pluralistic picture. The failure to recognize functional pluralism represents a significant limitation of these theoretical frameworks. There is clear explanatory power in adopting the idea that conscious representations facilitate a variety of functions in information processing systems, and no prima facie reason to rule it out on the representationalist model. The representations that function in visual processes are much different than those than function in emotional processes, and such stark differences in functional make-up strongly suggest a plurality of FCCs, rather than a single one.

On another interpretation of the representationalist research program, however, we need not assume that the functions of consciousness map cleanly onto the functions of certain kinds of representations. It might be that representations have a certain general function (e.g. tracking, supporting computations), as well as a function specific to conscious representations (e.g. tracking sensory qualities, supporting conscious computations), while consciousness itself has a different function altogether. In other words, given the conceptual compatibility of the representationalist picture with a range of different possible functional theories of consciousness, representationalism might be best understood as neutral with regards to FCCs. This further supports the claim that FOR is perfectly compatible with a pluralist picture, to the extent that conscious representations plausibly play a range of different roles in a variety of different of different information processing domains.

The result of this analysis suggests that representationalism must either a) infer the functional contribution(s) of consciousness from the functional role of certain kinds of representations, which typically results in a unified claim but is in fact perfectly compatible with functional pluralism, or b) remain neutral on the function of consciousness, again leaving room for a pluralist interpretation. In light of this, I submit that a fruitful avenue for representationalist theories of consciousness is to explore this potentially valuable theoretical alliance. Exploring the different representational capacities that conscious experience facilitates in different domains helps to make determinate exactly how the functions of certain kinds of representational phenomena can inform our understanding of the functions of certain kinds of conscious processes. One might even find a bridge between externalist and internalist representationalism in the functional pluralist model, as it might turn out that some conscious representations function importantly to link organisms informationally with their environments (e.g. tracking moving objects in visual perception) while others function to support certain crucial computations (e.g. modulating conceptual content in order to inhibit an intense emotional episode). While this is obviously speculative for now, it should serve to illustrate the importance of explicitly acknowledging functional pluralism, and how the failure to do so leaves the representationalist research program limited in its explanatory power.

Another sort of FOR account that needs to be explicitly mentioned here is Recurrent Processing Theory (RPT). RPT is typically construed as a general FOR theory that seeks to explain how representations become conscious by appealing directly to facts about the underlying neural mechanisms. The core idea of RPT is that sensory information becomes conscious when it is processed in recurrent feedback loops between primary and secondary sensory areas of the cortex (Lamme 2006), and perhaps above a certain threshold of activation (Fisch et al. 2009). Indeed, the theory was first developed to explain conscious perceptual states, but has also been extended to explain other conscious processes. van Gaal and Lamme (2012), for example, argue that recurrent processing is also necessary to establish consciousness outside of perceptual systems, including cognitive processes like decision making and planning that implicate more anterior neural regions.

RPT is a theory that is mostly concerned with isolating NCCs. It is not always explicit about the function(s) that it takes conscious experience to facilitate; namely what function recurrent processing is contributing to processes that would otherwise be unconscious. However, the overall framework is similarly poised to accommodate a pluralistic account of the functional contributions of consciousness, so long as it is explicit about the fact that different recurrent processing networks plausibly facilitate different conscious processes with different functional contributions. Given the importance of recurrent processing to most theories of consciousness, especially those that look to the brain for insight into its functions, a pluralistic account of FCCs will likely be developed within a broad RPT framework. Although, as mentioned previously, if unconscious processes also sometimes correspond with recurrent neural activity, then more work is needed to specify the conditions under which it realizes conscious processing, which is likely no small feat.

3.4 Attended Intermediate-Level Representation Theory

Jesse Prinz's (2012) Attended Intermediate Representation (AIR) Theory also proposes that conscious experiences only arise when information is accessed by higher level cognitive mechanisms. What is distinctive about AIR is that a) the representations that can enter consciousness are restricted to contents that are "intermediate" in the perceptual processing hierarchy, and b) attention—defined as the capacity to make information available for encoding in working memory—is the specific mechanism that is proposed to be necessary and sufficient for bringing those representations into conscious experience (Prinz 2012). Intermediate level perceptual representations alone (e.g. of objects) can be conscious when they are attended to, but not lower level (e.g. orientation and luminance values) or higher level representations (e.g. concepts).

This is also a theory that proposes very particular NCCs in support of its conceptual framework. On this model, the neural signature of all conscious processing is the synchronization of gamma band activity across posterior and anterior regions of the brain (Prinz 2012). The idea that gamma band synchrony is unique to conscious psychological processes has been largely discredited by the discovery that unconscious processes, like early stages of visual processing and multisensory integration (e.g. Misslehorn et al. 2019), exhibit the same kind of synchronized gamma activity. Furthermore, gamma synchrony fails to reliably correlate with subjective reports of conscious experience (see Aru et al. 2012). Koch et al. (2016) conclude that gamma synchrony is neither necessary nor sufficient for consciousness. For these reasons, AIR's claims about the structural correlates of consciousness will not be able to further support its claims about the function of consciousness.

Like representationalism though, because AIR reduces consciousness to a kind of representation, the function(s) of consciousness can in some sense be derived from the function(s) of these particular kinds of representations, namely attended sensory representations of mid-level properties. Prinz explicitly endorses the idea that consciousness is in the business of making certain perceptual information—specifically, information about "whole objects with rich surface details, located in depth, and presented from a particular point of view" (2012, p. 79)—available for use by working memory, which subsequently enables the system to select from a range of inputs that motor systems use for guiding interaction with the world (p. 169). This account applies to other sensory modalities besides vision as well, as there is evidence that all sensory modalities are similarly hierarchically structured, and Prinz argues that all conscious perceptual information is similarly "perspectival" in nature. AIR therefore also assumes a single unique functional contribution of consciousness, derived from a general account of its nature: "consciousness makes information available for decisions about what to do, and it exists for that purpose" (p. 203). This account represents a principled modification of familiar "cognitive" theories of consciousness like GWTs and HOT.

As a result, there are familiar reasons why AIR fails to generalize as a theory of the functional contribution of consciousness. On one hand, like any theory that identifies consciousness with higher level mechanisms such as attention, claims about function will not generalize to cases of phenomenally conscious states that are nevertheless unattended (e.g. rich perceptual experiences)]. On the other hand, the notion of cognitive phenomenology—the seemingly obvious idea that we are conscious, not just of sensory processes, but of cognitive processes as well (e.g. Strawson 2011)—is explicitly denied by this theory, and so no attempt has been made to explain what consciousness might add to tasks at "higher levels" than intermediate level perceptual processing. It is also not clear how to make sense of other purported forms of conscious experience on this view, including interoceptive experiences of hunger and emotional experiences of anger. The failure to once again establish a plausible candidate for the single FCC suggests that a more fine-grained account that takes seriously the possibility of functional pluralism is required.

There are also once again reasons to believe that the function being ascribed here is not unique to, or sufficient for, conscious experience. Several lines of research suggest that the sort of perceptual information at the heart of AIR can be used to guide action even when that information is not consciously experienced. Research on the phenomena of blindsight, for example, supports the idea that unconsciously processed visual information can guide interaction with objects. Ajina and Bridge (2017) review evidence that suggests that higher level visual information about shape and form can be processed and used to guide action in patients with significant damage to V1, and that this capacity can increase with training. Similarly, much work on visuo-motor transformation (e.g. Crawford et al. 2011) has exposed the complex ways that perceptual information about objects is used to guide action wholly unconsciously. Taken together, the implication is that AIR's candidate function of "attention-for working-memory use" is neither necessary nor sufficient for, and so dissociates from, conscious experience.

3.5 Information Integration Theory

Finally, Information Integration Theory (IIT), developed and endorsed by neuroscientists like Giulio Tononi and Christof Koch, focusses on a different aspect of informational content in order to explain the nature of conscious experience. In essence, IIT claims that "consciousness corresponds to the capacity of a system to integrate information" (Tononi 2004), which is thought to be "a fundamental property of any mechanism that has cause-effect power on itself" (Koch 2019, p. 79). Key to this proposal is a quantitative measure of the degree of both differentiation (understood as the number of possible informational states in the system) and integration (understood in terms of functional and anatomical interactivity) of information in complex systems. This value is represented symbolically as the Greek "Phi", which is thought to provide a reliable measure of the "amount" of consciousness present in the system. Conscious experience is thus defined by IIT as "a maximally integrated conceptual structure" (Tononi 2012b, p. 3016), where concepts are understood technically as "cause-effect repertoires" realized by neural populations that are distributed in nested hierarchies throughout the thalamocortical system. Crucially, this measure of integration also reflects the irreducibility of information to its component parts. More precisely, Phi-values capture the irreducibility of the cause-effect repertoires of the neural "complexes" that realize an experience's particular qualitative features to the cause-effect repertoires of its subsystems (Tononi et al. 2016). For example, an experience of a blue book is irreducible to, and has different causal properties than, the experiential components "blue" and "book-shaped". Consciousness, therefore, is ultimately realized by information integrating mechanisms in the brain that can effectively match the extrinsic causal structure of the world with the intrinsic causal structure of the system by combining distinct informational elements into maximally unified experiential wholes (Tononi et al. 2016).

This theoretical framework is explicitly intended to illuminate neuroscientific research on consciousness. By working backwards from a set of proposed “axioms of experience” (Tononi & Koch 2015, Tononi et al. 2016), IIT infers the requisite characteristics of the neural correlates of consciousness. The neural complexes that realize consciousness must have the right causal properties in order to generate and sustain the intrinsic, structured, specific, and unitary nature of experience. On this view, no single set of interacting neural regions will alone be necessary for conscious experience. Rather, information is integrated across the thalamocortical system in a highly dynamic manner, where each unique experience corresponds with activity in a highly specified set of interconnected (largely posterior) neural regions and their unique cause-effect repertoires. It is important to note that IIT is not committed to the idea that the PFC and the higher order mechanisms it is thought to underwrite are necessary for conscious experience[6]. This distinguishes it from many of the other leading theories, both because of its rejection of cognitive requirements for consciousness—which in turn implies that consciousness is a ubiquitous phenomenon in the natural world as opposed to being restricted to organisms that can engage in the appropriate higher order cognitive functions—and its acknowledgement that conscious processes form a heterogeneous class of psychological phenomena supported by a variety of neurobiological mechanisms.

IIT is also sensitive to the importance of recurrent processing for conscious experience. On this account, systems that process information in an entirely feed-forward manner do not have the right causal dynamics for generating consciousness. That is, feed-forward networks cannot be part of the kinds of complexes that enable the system to exert causal power on itself, because input and output layers only causally interact with mechanisms external to the system itself (Tononi & Koch 2015, Koch 2019). This is true even of feed-forward systems that perfectly

[6] Tononi et al. (2016) explicitly endorse a version of Block’s overflow hypothesis.

match the functional properties of systems with feedback or recurrent activity (i.e. "unfolded" networks, see Doerig et al. 2019): the latter alone can be said to be maximally integrated (i.e. have a significant Phi-value) based on the causal dynamics internal to the system. On this view, then, recurrent neural activity is structural property that is at least necessary for conscious experience, and a variety of different recurrent processing networks realize a variety of different kinds of conscious experience.

The IIT framework appeals primarily to the causal profiles of the neurobiological mechanisms that integrate information into unified conscious experiences, which seems to imply that consciousness itself is an important feature of the causal dynamics of the system. With regards to the specific function of consciousness, Tononi et al. (2016) argue that, evolutionarily speaking:

> Organisms with brains of high [Phi] should have an adaptive advantage over less integrated competitors because they can fit more concepts (that is, functions) within a given number of neurons and connections. (2016, 458)

Koch (2019) also emphasizes the fact that integrated information ultimately makes individuals "more adept at exploiting regularities in a rich environment" (p. 124), suggesting that it is in an adaptive sense that consciousness plays a functional role in the lives of organisms. In this way, the intrinsic cause-effect power that characterizes integrated information helps to "shape function in a broader sense" (p. 122), allowing organisms to more effectively and efficiently interact with complex environments.

Despite this claim about the adaptive function of consciousness, it has been argued that IIT is best construed as an epiphenomenalist account. Indeed, Koch explicitly states that "in [a] strict sense, experience has no function" (2019, p. 121) seemingly on the shaky grounds that one can have an experience without any ongoing information processing, or without it "doing anything". And IIT's response to unfolding arguments—namely, that you can have two systems with equal functional capacity but only one is conscious—seems to suggest that consciousness adds no functionality to information processing systems. Drawing out this interpretation, Liu (2020) argues that IIT can only point to "natural associations" between Phi-values and patterns of behaviour over evolutionary time scales, and in turn fails to explain the role that integrated information has in causing specific behaviours in individual organisms (Liu, 2020, p. 460). On this interpretation, conscious experience (understood as integrated information) is merely a by-product of the ongoing activity of complex organisms that are engaged in adaptive behaviours, but itself cannot be said to play a causal, functional role in the ongoing information processing carried out by individual organisms.

Instead of attempting to resolve the conflict between epiphenomenalist and functional interpretations of IIT, we can consider the implications of adopting either of them. On one hand, it seems like IIT theorists have opened the door for the epiphenomenalist interpretation by denying a relationship between consciousness' adaptive value over evolutionary time and its functional efficacy throughout the life of an organism. The idea on an epiphenomenalist interpretation again is that experience itself does nothing because it is merely a by-product of the system's ongoing generation of adaptive behaviours, even if those same behaviours confer evolutionary advantages to the survival of organisms—this is why organisms with higher Phi-values are better adapted to their environments. But if this interpretation is right, then all the worse for IIT's ability to contribute to the general project of providing a naturalistic account of consciousness, and to the specific project of identifying functional markers that can act as operationalizable proxies for experience in future consciousness research. That is, on an epiphenomenalist interpretation, IIT fails entirely as a naturalistic theory of consciousness.

On the other hand, IIT's proponents seem to have explicitly rejected an epiphenomenalist interpretation of their account, and they appear very much concerned with the project of formulating a naturalistic account of consciousness. The claim that consciousness confers adaptive advantages by allowing organisms to better exploit regularities in rich environments is precisely an attempt to incorporate experience into the broader causal, functional history of the organism. Conscious systems therefore have a processing advantage over unconscious (e.g. unfolded) systems with equal functionality, because the former are more efficient and effective at performing those functions given how they integrate information. It seems then that the only way to make sense of the claim that information integration confers this sort of evolutionary advantage to organisms is to assume that it does so by virtue of the fact that it plays a particular functional role in the ongoing information processing internal to the system. It is true that Phi-values merely reflect the internal causal dynamics of the system, but these internal dynamics are inextricable from the representational and behavioural capacities they facilitate. Koch (2019) assumes as much, noting that "any one conscious experience contains a compact summary of what is most important to the situation at hand…this precis *enables* the mind to call up relevant memories, consider multiple scenarios, and ultimately execute one of them" (p. 124, emphasis added). In feature integration, for example, it is the *experience* of objects as having bound features that is most important, and which facilitates more complex, adaptive interactions with them. In this way, the information integration that comprises a particular conscious experience (e.g. of a blue book) is the very same the integrated information that is brought to bear functionally on the task at hand (e.g. finding the blue book).

On this interpretation, it seems that IIT offers a fairly clear account of consciousness' functional contribution to information processing systems. The proposal is that consciousness maximizes or optimizes informational content in a system because it integrates it into a unified whole that is more than the sum of its parts. And so even though IIT is in many ways unique compared to the other theories surveyed here, it too seems to suggest that consciousness facilitates a single functional contribution in information processing systems. It is important to note that on this interpretation, the proposed function of consciousness is, perhaps unsurprisingly, similar to the functional contribution proposed by GWT and HOT: experience facilitates a sort of flexibility in processing that allows organisms to better exploit regularities in a rich environment. All other functions made possible by this are again construed as mere functional by-products of information integration. The more integration, and therefore consciousness, in a system, the richer the informational content that can be brought to bear on a variety of psychological tasks. This means that consciousness functions to increase (i.e. maximize) a system's intrinsic informational and causal capacity, the functional consequences of which can ultimately be thought of as a marker of experience. Crucially, this FCC is construed as being both necessary and sufficient for ascribing conscious experience to a system.

There are two potential avenues for responding to the functional interpretation of IIT, and the claim that the processing advantages made possible by information integration are a necessary functional marker of every conscious experience. The first is to marshal evidence that integration is not in fact a necessary feature of every conscious experience, in line with functional pluralism. In defense of the claim that information integration is not necessary for consciousness, Brogaard et al. (2021) survey empirical research that seems to provide examples of failures to integrate stimulus features in conscious visual experience. Treisman et al. (2006), for instance, found that subjects were better at judging the proportions of separate features (i.e. colors and letters) than they were at judging the proportion of conjunctions (i.e. colored letters) in a briefly presented visual scene. In the same vein, Neri and Levi's (2006) experimental work suggests that subjects are better at detecting either orientation or color independently in peripheral vision than they are at detecting combinations of orientation and color (which fits nicely with Usher et al.'s 2018 experimental results discussed earlier). Other studies have provided further evidence of these sorts of "unbound experiences" in more ecologically valid settings, such as failures to bind colors to sets of objects that are moving too quickly across the visual field (e.g. Arnold 2005, Holcombe 2009). These studies imply that some conscious experiences are not maximally integrated, and therefore do not involve the functional (i.e. informational and causal) advantages that such integration is purported to facilitate. This sort of work therefore provides compelling counterexamples to IIT's claim that all experiences are marked by this functional feature.

However, even if it can be shown that there is still the relevant maximal integration in these apparent cases of unbound experience, there is a second line of response to IIT's singular functional claim. This involves instead recognizing that the functional contribution of consciousness is described far too abstractly on this model, and as such is not adequately specified in such a way that it picks out the functional contributions that consciousness makes on particular information processing tasks. The idea here is that even if we grant that information integration of some kind or degree is a necessary feature of every conscious experience, the particularities of how integration provides a functional advantage still need to be spelled out for different psychological task-domains. The informational and causal advantages of integration are plausibly much different in the domain of visual processing than they are in the domain of emotional processing, for example; integration might facilitate a particular kind of representational capacity for vision (e.g. increased ability to track motion), while facilitating a particular kind of regulatory function for emotion (e.g. inhibition, conceptualization). On this reading, IIT can be thought of as compatible with a functionally pluralistic model of consciousness because the functional contributions afforded by integrated information will be distinct across different tasks, although much more work would be needed to tie specific forms of integration to different functional capacities in different information processing domains. However, IIT fails to make functional pluralism explicit, while simultaneously acknowledging a kind of structural pluralism about conscious experience, which again limits the applicability of the theoretical framework.

The claim that consciousness is in the business of information integration becomes even less plausible if enough evidence of unconscious information integration can be marshalled. Before briefly reviewing evidence for this, it is worthwhile to note again that some researchers are careful to differentiate distinct kinds or domains of information integration, including, for instance, spatiotemporal, multisensory, and semantic integration (Mudrik et al. 2014). This is significant, again because there seems to be no a priori reason to assume that integration will facilitate the same causal or informational capacities in every psychological domain. This point opens up the possibility that certain forms of integration might be performed by unconscious processes but not others, or that certain domains of integration are better handled by conscious processes. However, if any unconscious processes can be plausibly described as having integrated informational content (i.e. a positive Phi value), then IIT would fall short of explaining what it is that is unique about consciousness, and hence what function it contributes to information processing systems.

The unconscious binding of sensory information, either within a single modality or across multiple modalities, seems to provide preliminary evidence of unconscious integration. There is evidence, for example, that subjects unconsciously integrate multisensory information in ways that affect bodily self-consciousness (Salomon et al. 2017). Other research suggests that subjects unconsciously integrate audiovisual stimuli on congruency priming tasks, but only after subjects learned the relation between stimuli consciously (Faivre et al. 2014). Supporters of IIT, however, might respond that these informational states are too low level to count as genuine conceptual integration as defined by their framework. Evidence of higher level unconscious information integration will therefore strengthen the case against IIT.

Mudrik et al. (2014) provide a thorough review of a range of studies that seem to show genuine unconscious information integration. One of the more striking pieces of evidence comes from the domain of "high level" semantic integration (Mudrik & Koch 2013). Within the continuous flash suppression (CFS) masking paradigm, for instance, a target visual image presented to the non-dominant eye can be suppressed from conscious awareness when a high contrast image is presented repeatedly to the dominant eye. Subjects report no awareness of the target image during CFS. However, there is evidence of fairly sophisticated unconscious semantic integration of the invisible image, based on what are known as congruency effects. Essentially, incongruent images (e.g. a basketball scene where the ball has been replaced by a watermelon) reliably break through the CFS-induced "cloak of invisibility" faster than congruent images when the contrast between the images is slowly reversed (Block & Philips 2016). This seems to provide evidence of conceptual and contextual integration of an object and its background in the absence of conscious awareness.

The explosion of literature on implicit bias over the last few decades also provides robust evidence of unconscious information integration. Many forms of implicit bias are understood as the unconscious binding of a particular object of thought or perception with emotional valence (e.g. Healy et al. 2015). Again, within the stimulus masking paradigm, the activation of the amygdala in response to subliminally presented images of Black faces correlates reliably with behavioural indicators of implicit prejudice (Amodio 2014). This research is generally accepted to provide evidence that complex thoughts and behaviours are influenced by the unconscious emotional valencing of percepts and concepts, which seem to require the integration of distinct informational contents into a unified, irreducible whole.

There are a variety of experimental paradigms that suggest other forms of unconscious information integration in different psychological domains. In the domain of spatial integration, for example, there is a wealth of evidence suggesting that multiple features (e.g. components of facial identity or natural scenes) are combined unconsciously, because “the response induced by a combination of features is different from the summed responses evoked by each of the features separately” (Mudrik et al. 2014, p. 491). The emerging picture is that unconscious processing can at least perform some genuine forms of information integration thought to be under the jurisdiction of conscious processing alone. Assuming that we can assign a Phi value to capture information integration in unconscious processes, IIT fails as a theory of consciousness, and therefore of its unique functional contributions.

The existence of unconscious information integration is a major obstacle to the claim that conscious experience is to be defined as the quantity and quality of information integration in a system. But the theory is also insensitive to the fact that, even if there are kinds or degrees of integration that are unique to consciousness, it is plausible, given the variety of processes that occur consciously, that this either does not capture the full functional profile of consciousness or already entails a kind of functional pluralism. Again, a more nuanced survey of the existing empirical work is required to establish a precise account of the true FCCs. But there should be sufficient reason to doubt that these kinds of monolithic theories, as they are presently formulated, can capture the diversity of conscious psychological processes and the functional capacities that they contribute to psychological systems.

4. Conclusions

The leading philosophical and scientific theories are insensitive to the possibility that consciousness contributes a wide range of functions to information processing systems. But there is good reason to be sceptical of monolithic theories and wholesale claims in this regard, as these accounts reliably fail to generalize as explanations of consciousness and its functions. In general, there is evidence that the functional contributions of consciousness are too functionally disparate to be subsumed under a single theoretical construct. Some theories also fail to appreciate the functional capacities of unconscious processes, and therefore propose candidate markers that strongly dissociate from conscious experience. There is still much to be gleaned from these prominent theories going forward, however. Systematic analyses of the kinds of access, meta-cognition, and information integration that are truly unique to consciousness, for example, would be important contributions to its emerging functional profile. But I suggest that the new frontier of consciousness research ought to involve the identification and categorization of a wide variety of FCCs, as predicted by functional pluralism. This will require us to methodologically shift focus away from overarching theoretical frameworks to the systematic analysis of particular tasks and particular processing capacities within distinct psychological domains. It is a rich hypothesis that conscious experience contributes different functional capacities in different processing domains (e.g. visual perception versus emotional processing versus social cognition), and it has the potential to prompt significant revisions to existing taxonomies of consciousness.

Chapter 2: Functions of Consciousness in Visual Processing

Abstract

It is a plausible philosophical hypothesis that conscious experience contributes different functional capacities in different psychological domains (i.e. functional pluralism). This chapter begins to tease out some of the functional contributions that consciousness makes to visual processing in human beings. Drawing on a range of psychological and neurobiological research, I discuss both semantic and spatiotemporal processing as specific points of comparison between the functional capabilities of the visual system in the presence and absence of conscious awareness. I argue that consciousness contributes a cluster of particular functional capacities to visual processing, enabling, for example, the processing of semantic information inherent in more informationally-complex visual stimuli, increased spatiotemporal precision, and representational integration over certain spatiotemporal intervals. Such an analysis ultimately yields a plurality of "local" functional markers that can be used to guide subsequent work in the philosophy and science of consciousness. These domain-specific functional contributions, however, are not captured by leading monolithic accounts of consciousness like Information Integration and Global Workspace Theories. The different FCCs identified here ultimately offer clues towards general models of consciousness, suggesting that visual experience is characterized by *informationally-rich* representations made possible by recurrent neural activity in the visual processing stream.

1. Building Visual Awareness:
The Basic Structure and Function of Visual Perception

Visual perception has long been the object of intensive analysis in Western philosophy (e.g. Locke 1689, Russell 1912, Burge 2010), and the primate visual system is perhaps the most thoroughly studied sub-system in the psychological and neural sciences (e.g. Hubel and Wiesel 1968, Van Essen et al. 1992, Hilgetag et al. 2016). The integration of research across these disparate intellectual disciplines has resulted in an impressively detailed functional account of visual processing in human beings, reinforced by a remarkably thorough understanding of its underlying neurobiological mechanisms. David Marr (1982) took the visual system to be a paradigm target of explanation for cognitive science, and we have made astounding progress towards unravelling its mysteries since the publication of his influential work on the topic. In this section, I will briefly summarize some of the core principles that have emerged from this research, which form the conceptual backdrop for thinking about visual perception and the functional roles of conscious experience.

One core theme that has emerged from the relevant empirical research is the idea that visual processing is fundamentally hierarchical in nature, which is reflected in both functional and anatomical organization (e.g. Movshon & Simoncelli 2014, Wilson & Wilkinson 2015, Siegle et al. 2021). What begins with the transduction of light into patterns of electrochemical activity by photoreceptor cells reaches its apex in sophisticated capacities for recognition, categorization and simulation supported by densely interconnected cortical networks. Throughout the visual processing hierarchy, neural populations relay information downstream about a range of different visual properties that they selectively respond to. These properties include location, orientation, motion direction, color, and spatial frequency (Bednar & Wilson 2016). Input at each successive stage is pooled, allowing statistical information to be extracted (or perhaps abstracted) in the construction of representations of objects and events. When this information reaches its initial destination in the cerebral cortex at V1 (also known as the primary visual or striate cortex)**,** elegantly structured multidimensional topological "maps" of patterned neural activation facilitate preliminary processing and integration, before signalling further downstream to neural populations that respond to increasingly complex features of visual stimuli (Movshon & Simoncelli 2014). The hierarchical nature of visual perception means that more informationally-demanding functions are carried out at successive levels in the visual system.

Another way of understanding the visual hierarchy is in terms of the kinds of representations that are constructed and employed at different levels of processing. It should be noted that some philosophers deny on conceptual grounds that earlier stages of visual processing are representational in nature (e.g. Orlandi 2014, Bourget 2010). This is because many philosophical definitions insist that representations a) have conditions under which they can be said to be inaccurate or to misrepresent their causes, b) are sufficiently “decoupled” from proximal input and therefore cannot be understood as simple causal covariation, and c) guide the behaviour of the system (Orlandi 2014). It is argued that none of these conditions are met by processes carried out early in the visual system, such as the initial encoding of orientation information from patterns of light captured by photoreceptors in the retina. In fact, some accounts go so far as to claim that the contents of genuine representations are “grounded” in the properties of phenomenal consciousness, and so the notion of unconscious, low level representation in the visual system is conceptually confused (see e.g. Kriegel 2013).

Despite these philosophical reservations, it has been a fruitful practice in the sciences to appeal to the concept of representation in order to describe activity widely throughout the hierarchy (e.g. Poldrack 2020). The idea that processes early in visual system traffic in representations is supported, I think most plausibly, by the constructive nature of visual processing. The early visual system must actively construct usable information out of the input signals that are transduced in the retina by drawing on and integrating previously encoded regularities. Retinal input alone is not enough to determine the object or event that was its distal cause, and so processes immediately downstream begin constructing the most plausible interpretation of what that cause might be. As Eric Kandel puts it, “if the brain relied solely on the information it received from the eyes, vision would be impossible” (Kandel 2012, p. 203). The constructive nature of the visual system seems to secure the relevant accuracy conditions (i.e. these constructions can be more or less accurate) as well as the appropriate decoupling from proximal input (i.e. constructions go beyond what is provided by proximal input). Finally, the phenomena of blindsight and related neuropsychological findings offer plausible evidence that, despite damage to the primary visual cortex, the processing of low level stimulus features at earlier stages in the hierarchy can guide categorization and discrimination behaviours at the system level (Brogaard 2011).

It therefore seems that even on highly restrictive definitions, early visual processes appear to satisfy the conditions for being genuinely representational, and there is no reason to assume that representations are necessarily conscious. Again, the emerging consensus in cognitive science is that representations are a crucial explanatory posit for making sense of the activity of the visual processing hierarchy: orientation maps are ultimately integrated into representations of shape; processes tracking the dynamics of motion feed into representations of the animacy of objects; and monocular reconstructions of the visual field are compared in order to calculate representations of depth, for example. The important takeaway is the idea that each successive processing stage in the hierarchy employs more informationally-rich representations, in the sense that they track higher quantities (e.g. more detailed shape information) and more diverse qualities (e.g. simultaneous shape and motion) of the properties of stimuli, ultimately in the service of increasingly complex visual functions.

A second core theme that has emerged in the scientific study of visual perception is the idea that different subsystems, presumably supported by extrastriatal areas of the cortex or processing regions downstream from V1—are responsible for carrying out different functions. Goodale and Milner (1992), for example, hypothesized that V1 neurons project to two functionally distinct visual "streams", the ventral and dorsal, thought to be responsible for conscious object recognition and unconscious visuo-motor interaction, respectively. Although this strict segregation of the two streams has been called into question (e.g. Wu 2014, Rosetti et al. 2017), several other downstream cortical regions are assumed to support distinct aspects of visual processing. Areas in the Inferior Temporal Cortex, for example, have been shown to support semantic and cross-modal association, which are thought to be crucial for abstract conceptual thought (Buckner & Krienen 2013, Bonner & Price 2013, Binder 2016). Similarly, the middle temporal visual region (MT) has historically been thought to be crucial for motion processing, although more recent work suggests rather that the area's primary functions include specific forms of integration and segmentation of inputs from V1 (Born & Bradley 2005). As usual, there are reasons to be skeptical of strict selectivity and localization, in the sense that implies a clean one-to-one mapping from psychological function to neurobiological structure. A cortical region like the FFA, for example, which was once thought to selectively respond to faces, has now been shown to play a role in a variety of psychological functions, responding also to cars, birds and other types of visual stimuli, likely as a result of the different "neural coalitions" it establishes with other structures (Anderson 2014). Still, there seems to be sufficient evidence that the visual processing labour carried out by regions downstream from V1 is divided into a variety of relatively function-specific neurobiological networks. This ultimately suggests that there are a number of distinct functional components of visual perception that are carried out by a variety of independent but obviously interacting subsystems.

Finally, our intensive study of the visual system has suggested that feedback connections are a central architectural feature of the visual processing hierarchy: higher levels establish complex functional and anatomical inputs to lower levels (Liang et al. 2017, Nurminen et al. 2018). These recurrent processing circuits are found throughout the visual processing hierarchy, and are thought to facilitate a range of processing capacities, including the modulation, amplification, prediction and contextualization of bottom-up signals in the visual system (Federer et al. 2020). This "tuning" of signals at lower levels of the processing hierarchy seems in fact to be necessary for the increased capacities for information extraction that are involved in more sophisticated perceptual functions like object recognition and categorization. This is because it ultimately allows task relevant information to be selected for further processing at the expense of task irrelevant information (Wu 2017).

According to one hypothesis with significant empirical and theoretical support (e.g. Lamme 2006, Sikkens et al. 2019), feedback connections in the visual cortex are necessary for the generation of conscious experience. One major theoretical and empirical challenge that arises here is the fact that recurrent processing might turn out to be a relatively ubiquitous feature of complex information processing systems, and some unconscious visual processing might also involve neural populations with feedback connectivity (see Koivisto et al. 2010). In other words, recurrent processing might not be sufficient for conscious experience, even if it is necessary. This means that carefully assessing the properties of particular networks that facilitate particular neurological and psychological functions is crucial. The challenge is to home in on which specific recurrent processing networks are in fact sufficient for conscious visual experience; in other words, we need to isolate the qualifying conditions under which recurrent processing is associated with consciousness. Looking closely at these underlying structural properties should help us supplement our analysis of the FCCs in this domain.

This detailed structural and functional mapping of the human visual system has subsequently influenced a range of philosophical debates surrounding the nature of vision, including the extent to which cognition influences perception (e.g. Vetter & Newen 2014), the informational richness of perceptual content (e.g. Ludwig 2020), and the role of attention in visual processing (e.g. Wu 2017). Given that it has also historically been treated as paradigmatic of phenomenal content in philosophical discussions (e.g. McDowell 1994, Tye 2000, Prinz 2011), it should be no surprise that visual processing has also become the primary foothold into philosophical and scientific issues surrounding consciousness. Most of the leading experimental paradigms use visual stimuli to study particular conscious experiences (e.g. Dehaene & Changeux 2011, Block 2019, Lamme 2020), and several specific hypotheses about consciousness can be tested with carefully designed visual processing tasks. This is due in part to the fact that robust relationships between specific visual stimuli and certain patterns of neural activation are relatively well established, and so researchers can fix crucial variables and perform fairly precise experimental manipulations in order to isolate the operant neural mechanisms (for a thorough review, see Koch et al. 2016). Moreover, psychophysical tools used to probe the functional-psychological components of visual processing (e.g. priming, adaptation paradigms) have been carefully developed over many decades of research, and are consequently well established in the field (e.g. Kominsky & Scholl 2020). Finally, minimal thresholds for conscious experience can be identified using subjective report, no-report, and perhaps even no-cognition paradigms (see Brascamp 2015, Block 2019), allowing for precise structural and functional comparisons between conscious and unconscious visual processing (discussed in more detail below).

It should be emphasized here that there has been much debate about the relationship between attention and conscious experience, especially within the domain of vision, which muddies these waters a bit (e.g. Prinz 2012, Tsuchiya & Koch 2014). The case could certainly be made that all leading theories of consciousness assume that attentional capacities stand in complex relations to conscious experience. Much work has also been done to explore how attention modulates visual experience in a variety of ways (e.g. Carrasco 2011). However, consciousness and attention doubly dissociate as psychological phenomena (e.g. Van Boxtel et al. 2010), and I assume that very few theorists would propose a wholesale reduction of one to the other. As such, I maintain that attention is a complicated cognitive capacity (or set of capacities) that deserves is own systematic analysis, as does conscious processing, and that understanding them in in their own right as psychological constructs will ultimately be the best way to understand the relations between them. Therefore, I will intentionally avoid some of these important issues related to attention, although the discussion below will directly bear on the issue of prising attention and other executive functions apart from experience in experimental and clinical settings.

Ultimately, the goal here is to look to structural and functional organization in order to pinpoint exactly *where*, *how*, and most importantly *why* consciousness enters the visual processing hierarchy. Consciousness, as a general target of inquiry, is best understood as a *property* of particular psychological (e.g. visual) *processes*: some processes have the property that they are subjectively experienced, while other processes lack this property. Like other psychological properties, there is good reason to think that consciousness contributes functionally to the overall workings of the psychological system: the presence or absence of consciousness makes a difference, causally speaking. This point is becoming orthodoxy in consciousness research (e.g. Deheane et al. 2006, Cohen and Dennett 2011, Frith and Metzinger 2016, Lamme 2020, Boyle 2019, Birch 2020). But in order to understand precisely what it is that consciousness *does* in information processing systems, we have to first look closely at particular functions employed for specific processing tasks in order to isolate specific that truly require, are sufficient for, or are unique to, consciousness.

A more "localized"—that is, domain- or task-specific—contrastive analysis of unconscious and conscious processing helps us tease out the functional contributions that consciousness makes to particular information processing capacities, adding nuance to our understanding of the functions of, and ultimately the nature of, experience. I argue that there are a cluster of functional contributions that consciousness makes to human vision, and that these are different from the functional capacities it confers in other processing domains. In the rest of this chapter, I will first survey some of the most compelling evidence of the extent and limits of unconscious visual processing, focussed on spatiotemporal and semantic visual processing tasks as specific points of comparison. I will then review and interpret a range of comparative research in the psychological and neural sciences, in order to identify some candidate psychological functions in the domain of visual processing that are likely unique to conscious experience. My analysis suggests that conscious experience facilitates a) the processing of semantic information inherent in more informationally-complex visual stimuli, b) increased spatiotemporal precision, and c) representational integration over certain spatiotemporal intervals. In the final section, I will draw out some preliminary implications that this functional analysis has for theories of the nature of visual consciousness.

2. Appreciating Unconscious Visual Processing

Contrasting conscious and unconscious processing requires that we first get the best possible picture of the functional capacities of the visual system in the absence of awareness[7]. Although it remains controversial in certain philosophical discussions (e.g. see the debate in Block and Philips 2016), many cognitive scientists have endorsed the existence of unconscious visual perception (e.g. Kouider & Dehaene 2007, Tamietto & de Gelder 2010, Quilty-Dunn 2019). The idea is that a remarkable amount of visual processing seems to happen unconsciously, and to such an extent that it seems plausible to ascribe paradigmatically perceptual capacities to human subjects that remain wholly outside of conscious experience. And while it is not entirely necessary for understanding conscious visual perception that there be unconscious processing capacities that meet some preestablished philosophical criteria for being genuinely 'perceptual'—and so I won't labour this specific point in what follows—employing this terminology can help us appreciate what the relevant research has revealed about the functional capabilities of unconscious processes in the domain of vision.

It is important to appreciate that the causal properties of the neurobiological structures that support unconscious visual processing are fairly well understood. The retina, optic tract, lateral geniculate nucleus of the thalamus, and visual cortical regions are all implicated in the unconscious processing of visual information (e.g. Lamme & Roelfsema 2000). Crucially though, neuroimaging research and theoretical interpretation suggests that even though there is

[7] A similar point is often made about the search for neural correlates of consciousness (e.g. see Breitmeyer 2015).

complex feedback connectivity at the lowest levels of the hierarchy (e.g. Drinnenberg et al. 2018), unconscious processing in the visual system is typically realized by rapid, feedforward "sweeps" (FFS) of activation throughout these structural networks, during which information about a stimulus is rapidly communicated strictly from lower to higher levels (i.e. bottom-up) in the hierarchy (Dehaene et. al. 2006, van Gaal & Lamme 2012). This commonly identified neural signature of the visual system during unconscious processing will likely continue to be a major point of comparison with the neurobiological correlates of conscious vision. The current project certainly will not resolve all the issues surrounding the structural correlates of unconscious vision and the role of feedback processing in conscious experience, but it is good practice to use emerging clues about the activity of underlying neurobiological structures to constrain and guide our claims about function whenever possible.

Several empirical paradigms bear on the functional capacities of unconscious visual processing. I will focus on just a couple of clusters of evidence in order to offer a thorough examination of each; one that emerges from laboratory settings and careful experimental manipulation and another that emerges from the study of particular pathological impairments to visual processing.

2.1. Continuous Flash Suppression

There are two dominant experimental paradigms for studying unconscious visual processing: visual masking and interocular suppression. Masking techniques allow a rapidly presented visual stimulus to be kept from conscious awareness due to the presentation of a second stimulus closely before and/or after the target. Subjects are subsequently asked to perform a range of different tasks that indirectly probe the extent to which this visual information has been registered by the visual system despite failing to reach consciousness (Breitmeyer & Ogmen 2006, Kouider & Dehaene 2007). Visual masking studies are also often performed while subjects are undergoing some form of neuroimaging (e.g. fMRI, MEG, EEG, single-cell recordings) aimed at identifying and contrasting the neural correlates of conscious and unconscious visual processing. Indeed, masking techniques played a significant role in the discovery that unconscious processes are typically marked by FFS neural activity (e.g.

Fahrenfort et al. 2007). Although masking has a long history in the vision science community, a new form of interocular suppression is quickly becoming the primary method for comparing the structural and functional elements of conscious and unconscious visual perception.

In general, interocular suppression techniques exploit a unique feature of binocular vision; namely, "the reflexive suppression that occurs when different images are simultaneously presented to the two eyes" (Yang et al. 2014, p. 1). When different stimuli are shown to each eye separately but at the same retinal location, the visual system fails to fuse them into a single percept, due to the natural constraint that dictates that two different objects cannot occupy the same location in space. Instead, the two stimuli "compete" for subjective awareness; that is, only one can be consciously seen at a time. There is typically a relatively spontaneous alternation between which stimulus reaches conscious awareness at a given moment and which remains supressed from awareness (Kovacs et al. 1996). Despite this, the primary advantage of interocular suppression over masking techniques is that stable input to the visual system can remain unconscious for longer periods of time. Modifications of this general experimental set-up have allowed researchers to more precisely identify when a stimulus is in fact consciously experienced and when it is not, which can ultimately provide compelling evidence of both the neurobiological correlates and the functional capabilities of unconscious visual processing.

One form of interocular suppression in particular has emerged as the primary psychophysical tool for comparing conscious and unconscious visual processing: Continuous Flash Suppression (CFS). Mudrik et al. (2011) employ CFS, and describe it as follows:

> In CFS, distinct color images (Mondrians) presented successively at approximately 10 Hz to one eye can reliably suppress the conscious awareness of an image presented to the other eye for a relatively long duration. (Mudrik et al. 2011, p. 765).

Essentially, CFS is a unique modification of the interocular suppression paradigm that involves the controlled suppression of one monocular input from awareness by rapidly and repeatedly presenting a high-contrast color-patterned image, or Mondrian, to the other eye, which dominates visual awareness. Stimuli like Mondrians have traditionally been presented to subjects on computer screens, although some researchers have begun to use CFS to supress real objects in the environment from awareness with the help of augmented reality goggles (Korisky et al. 2019). Masking and suppression techniques like CFS are also typically paired with other tools of experimental psychology like priming or the induction of adaptation aftereffects (Yang et al. 2014). The idea is that evidence of priming or adaptation involving visual stimuli that are suppressed from conscious awareness suggests that those stimuli are being processed unconsciously in ways that are paradigmatically perceptual. These tools are therefore used to probe the extent of unconscious visual processing.

A further modification of CFS, known as Breaking Continuous Flash Suppression (b-CFS), involves slowly reversing the contrast between the two images to determine the precise moment that the target image finally “breaks through” suppression and becomes consciously seen (Stein & Sterzer 2014). Interestingly, reliable patterns emerge here; for example, some classes of stimuli reliably break through suppression faster than other classes of stimuli. The theoretical interpretation here is typically that the initial unconscious processing of the invisible stimulus can boost its input signal, somehow preferentially “empowering” it to rise to the level of conscious experience (Yang et al. 2014, p. 5). The underlying principle is that more “meaningful” stimuli are brought into consciousness more quickly by way of unconscious processing. This accounts for some of the observed variability between subjects in b-CFS experiments, as “meaningfulness”—understood here as the degree of perceptual or cognitive salience—depends on a range of contextual variables like individual learning history and experimental design (see figure 2).

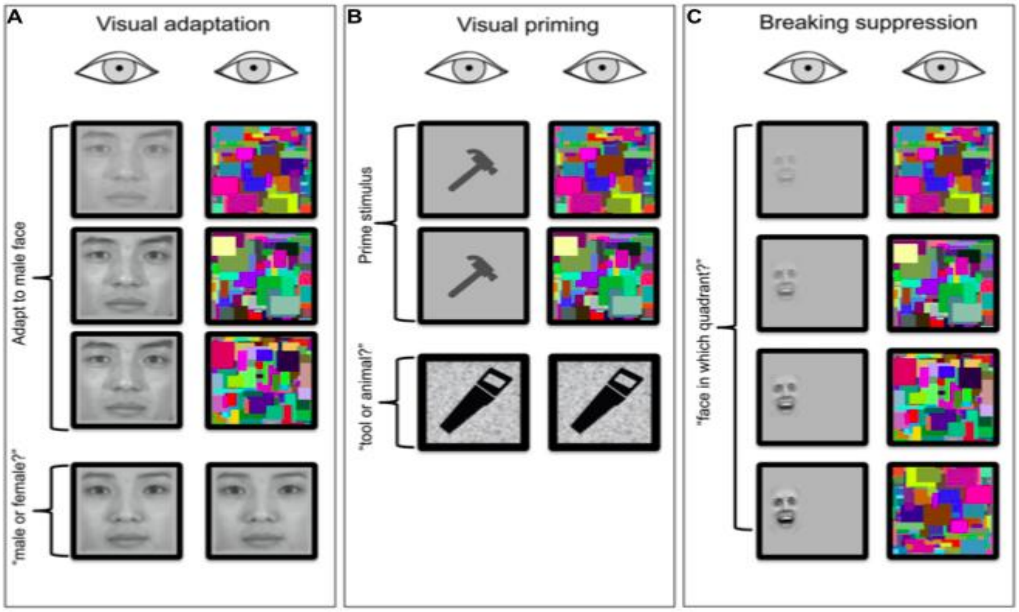

Figure 2. Yang et al. (2014) schematic representation of the three main CFS paradigms. A) CFS paired with the visual adaptation paradigm. B) CFS paired with visual priming paradigm. C) Breaking-CFS paradigm.

Like any empirical paradigm, the results of CFS studies need to be carefully assessed before anything philosophically relevant can be extracted from them. Accordingly, several prominent philosophers and scientists have critically reviewed key aspects of the CFS paradigm (see Yang et. al 2014, Block and Philips 2016). There are long standing methodological concerns surrounding both the subjective and objective measures that are used in the lab to verify whether or not stimuli are consciously perceived, and these certainly also arise in CFS research. Despite acknowledging unconscious visual processing, Ian Philips (2018), for example, has consistently raised the concern that subjective report, one of the primary methods used to determine whether a stimulus is consciously seen or not on CFS trials, faces the "problem of criterion". The central concern here is that it is always possible to interpret subject's reports as reflecting a conservative response bias, which opens up the possibility that subjects are perhaps at least partially aware of stimuli that they report being unaware of on some trials.

There are also challenges in interpreting the behavioural measures that are used to objectively probe awareness, such as when subjects are asked either to detect the presence of a target or to discriminate between a target and a decoy. This is because, a) some argue that chance level performance on these tasks does not necessarily entail that stimuli were unconscious, given that “the absence of evidence is not the evidence of absence” (Altman and Bland 1995), and b) above chance performance doesn’t necessarily entail that stimuli were consciously seen, given that some of these behavioural tasks can be accomplished in the absence of awareness. Furthermore, subjective and objective measures of awareness can dissociate in laboratory settings, pointing to different conclusions about the conscious experiences of a subject (Yang et al. 2014). This all seems to make it very difficult to be certain whether or not awareness is truly absent during specific CFS trials.

Several responses to these methodological concerns have emerged in both theory and practice. In terms of subjective measures of awareness, report paradigms now commonly involve graded perceptual awareness scales in an attempt to nuance subjects’ responses. Subjective reports are also increasingly combined with measures of confidence. Some studies employ “post-decision wagering”, for example, in which a subject’s confidence in their reports is indexed by the amount of money that they were willing to bet on their accuracy (e.g. Persaud et al. 2007). Other paradigms rely on subjects’ metacognitive judgments without introducing the risk/loss aversion that comes with wagering (Maniscalo & Lau 2012**).** These supplemental methods can help researchers control for potential report biases and motivate subjects to respond without conservative response criteria. Although this is speculative, it seems counterintuitive that subjects would develop a conservative response bias on these sorts of tasks in the first place, given the lesser costs associated with false positives (e.g. I thought I saw the tiger) than with false negatives (e.g. I didn’t see the tiger) when it comes to conscious visual perception.

It is important to note some further complications surrounding subjective report in consciousness research. As discussed in the previous chapter, experience is often quantitatively richer than what subjects can report on at a given moment (e.g. Block 2011, 2014). This means that subjects may be conscious of a stimulus that they have not accessed by the mechanisms underlying verbal report. In other words, the ability to make a subjective report may not be a necessary condition for awareness. To be as clear as possible, the claim is typically that perceptual experiences can remain potentially reportable (say with the appropriate cue, as in the pioneering study by Sperling 1960) while not being actually accessed for report by subjects.

Perceptual stimuli that are wholly unavailable for report, i.e. that are *unreportable* because they are presented too rapidly for instance, are not typically assumed to be part of this conscious "overflow" in perception (Block 2011, 2014). On the other hand, this also means that the cognitive processing underlying subjective report is a potential confounding variable for debates about whether or not either access for global broadcasting or some meta-representational processing is necessary for perceptual consciousness above and beyond processing in perceptual systems, especially in the context of the search for NCCs.

As such, many researchers have developed "no report" paradigms, where subjective report is used to eventually calibrate objective markers like eye movement patterns and pupil dilation that can be objectively linked to visual awareness (e.g. Frassle et al. 2014). There are still potential confounds here though, as even without explicit report, merely noting a change in the contents of perceptual awareness may produce cognitive effects that are unrelated to the conscious experience itself, but that still accompany it. Psychologists like Jan Brascamp et al. (2015) and philosophers like Ned Block (2019) have thus advocated for an even more rigorous "no cognition" paradigm. Using binocular rivalry, no cognition trials involve the standard binocular alternation between stimuli, except that subjects cannot detect the change due to the nature of the stimuli used (e.g. randomly and sporadically moving patterns of the same color dots), which eliminates any confounding cognitive processing. It remains to be seen how effectively CFS can accommodate a no cognition paradigm. Regardless, much work has been done to supplement subjective report in consciousness research, which must ultimately be acknowledged as an extremely valuable foothold into the study of conscious experience.

In terms of other standard objective measures of awareness, scientists who employ CFS to study conscious and unconscious visual processing have recognized that different measures, like detection and categorization tasks, tap into different stages of processing. This means that apparent failures to detect the presence of a stimulus might be good evidence that it did not reach consciousness, whereas apparent failures to categorize a stimulus based on some feature might still occur either with minimal conscious awareness of the whole stimulus, or in cases where certain (e.g. low level) stimulus features break through the suppression but not others (e.g. higher level features). One solution, therefore, is to employ several independent and increasingly stringent measures of awareness—both subjective and objective—within a single study, and to avoid generalization across these measures (e.g. Yokoyama et al. 2013, Gelbard-Sagiv et al. 2016). Another way to strengthen objective measures of awareness is to integrate them with the distinct measures used to probe the extent of visual processing (e.g. priming, adaptation, etc.) within a single experiment, so that both task design and levels of attention and motivation remain constant. These distinct tools that are used to assess the extent of visual processing on a given CFS trial, like priming or inducing adaptation effects, are much less controversial in the cognitive sciences, and indeed have become popular methods in a wide variety of experimental paradigms. In summary, while there is always room for continued refinement of the methods employed, it is generally agreed that CFS is a viable and illuminating experimental design for rendering visual stimuli unconscious and assessing the extent of unconscious processing of visual information.

Given this arsenal of experimental tools and the massive body of relevant empirical literature, there is a robust enough research program here to begin to get a sense of the kinds of visual processes that can occur unconsciously. Research using CFS has revealed a range of higher level visual processes that do not require consciousness. Gelbard-Sagiv et al. (2016) provide a recent summary of some of the most compelling evidence (see also references therein):

> Remarkably, several recent studies demonstrated that many high-level processes can take place even when the stimuli are invisible: observers were found to read and process the meaning of words, process semantic incongruencies in written sentences and visual scenes, perform arithmetic operations, categorize faces and other objects, process emotions, and exercise executive functions, in the absence of perceptual awareness.

While the authors suggest that we be cautious about some of these results, as some experimental paradigms might not be sensitive to the possibility that some stimulus features are being suppressed from awareness while others are not, CFS has delivered some fairly robust evidence of the functional capacities of unconscious visual processing.

2.1.1 Unconscious Visual Integration

Interestingly, some labs have used this technique to challenge specific core assumptions of some of the most prominent theories of consciousness. Information Integration Theory (IIT), for example, assumes that conscious experience is required for the integration of disparate information into a unified whole (Tononi et al 2016). In contrast, Liad Mudrik and colleagues have dedicated years of research to uncovering the kinds of integration of visual information that can be accomplished without subjective awareness, often employing CFS techniques (for review, see Mudrik et al. 2014). It has been shown, for example, that unconscious visual processes can facilitate: a) the association of visually presented words, even with fairly large temporal integration windows of up to 78 seconds (Reber & Henke 2012), b) the integration of disparate features of visual stimuli, even with relatively high spatial integration windows (Oriet & Brand 2012)) and c) high level semantic and syntactic integration of visually presented words and numbers (Sklar et al. 2012).

Looking closely at a particular experiment can help illustrate this empirical finding regarding unconscious information integration. Plass et al. (2014) developed a CFS study that tested the extent of unconscious audiovisual integration, relying both on the logic of priming studies and the common finding that visual information about mouth movements can influence auditory processing of words. They found that lip movements that were rendered invisible with CFS still facilitated performance on tasks that required subjects to categorize a spoken word (e.g. as either a tool word or non-tool word), when the target word was the same as that articulated by the suppressed lip movements. The study employed both subjective measures of awareness—namely, reports on whether the face was visible in addition to reports on the location of a circular probe near the mouth—and objective measures of awareness—namely, subjects were asked to indicate the color of a translucent ellipse placed over the mouth region of the suppressed face. This study provides compelling evidence that unconsciously processed visual information was integrated with auditory and linguistic information when performing the word categorization task, as is standard in speech perception (see Venezia et al. 2015).

IIT theorists might deny that this unconscious visual processing counts as genuine information integration as defined by their research program, in the sense that involves combining distinct "conceptual" structures into a single unified representation (Tononi et al. 2016). More specifically, this priming effect might either be considered too low level or informationally simplistic to capture the kinds of cause-effect repertoires underlying information integration theories of consciousness, or it might instead be construed as merely serial processing of distinct representational elements without integration. However, both of these objections are thwarted by the fact that the audiovisual processing in question fits perfectly well with IIT's definition of integration, when construed as a graded notion as it was intended to be. That is, the resulting causal and computational resources are more than the summed total of the resources provided by the component representations (Mudrik et al. 2014), and this is simply not the case with merely serial processing. Moreover, this sort of multi-modal perceptual integration specifically facilitates speech recognition and language comprehension, which are generally assumed to be "higher" information-processing achievements. This is just a sample of the research aimed specifically at uncovering the extent to which unconscious processes can integrate visual information, but it brings us closer to identifying the upper limits of the functional capabilities of unconscious vision more generally. Unconscious visual processing seems to be capable of at least some genuine forms of information integration. Put differently, information integration turns out to be insufficient for conscious experience.

2.1.2 Accessing Unconscious Visual Information

The same kinds of challenges have been made to assumptions held by the Global Workspace Theory (GWT), and specifically its claim that consciousness' functional role is to facilitate wide-ranging access to information (Dehaene 2014). A variety of studies have shown that unconsciously processed visual information is available to the same processing subsystems identified in the Global Workspace model (Dehaene & Changeux 2011); namely, evaluative systems, long-term memory systems, attentional systems, language systems, and motor systems. For instance, much work has been done to understand the role of unconsciously processed visual information in the guidance of action (e.g. Brogaard 2011, Goodale & Milner 2013). The visual system constructs representations in preparation for visually guided action that are available to motor and decision-making systems despite their failing to reach awareness (e.g. Bargh & Morsella 2008). Even the most basic visuomotor tasks might require that visual information processed outside of awareness be freely used by systems that predict, compare and execute intended actions. CFS studies also repeatedly reveal that unconsciously processed words affect semantic processing networks, and can even influence problem solving strategies (e.g. Zabelina et al. 2013). Finally, a variety of research paradigms suggest that unconsciously processed visual information is available to systems responsible for emotional or evaluative assessment (e.g. Fang et al. 2016, Diano et al. 2017); in fact, visual stimuli are the primary experimental tool used to explore unconscious emotional processing (e.g. Morris et al 1998, Williams et al. 2004, Mendez-Bertolo et al. 2016). In general, these experimental paradigms suggest that unconscious visual information is also accessed for use by a relatively wide range of downstream processing subsystems.

One study by Sklar et al. (2012) illustrates this sort of challenge to GWT. In one experimental set-up, relatively complex mathematical problems (e.g. three-digit subtraction equations) were suppressed from awareness using CFS. Both an objective forced-choice measure and a nuanced subjective measure consisting of direct questions about the trials were used to establish subject's lack of awareness of the suppressed math problems. After the suppressed primes were presented, subjects were asked to pronounce out loud a target number that was either the correct or incorrect solution to the suppressed equations. The researchers found a significant priming effect in reaction times to correct responses, suggesting that "the primed equation was mentally accessed (that is, that the equation had been solved)" (Sklar et al. 2012, p. 19616), even though the visual information remained unconscious. Subliminal priming of this sort continues to be developed as a valuable tool for studying a wide variety of unconscious influences on thought and behaviour (see Elgendi et al. 2018).

GWT theorists have ruled out this interpretation of unconscious priming effects in vision, based on their definition of unconscious processing as “a condition of information inaccessibility” (Dehaene et al. 2006, p. 3), according to which unconscious information cannot facilitate task performance in this way (Bussche et al. 2008). But any facilitation in task performance indicates at least some form of access to unconscious visual information by decision- and action-guiding systems, whether or not the access is “global” in the relevant sense. It should also be noted that the ‘global’ access proposed by GWT is not taken to be absolute, as information in the global workspace need not be accessible by every system—GWT is thought to be compatible with modularity and informational encapsulation, for instance (Dehaene et al. 1998). This ultimately suggests again that we adopt a graded notion of access, and there is no obvious reason to deny that this kind of cognitive operation crosses the conscious/unconscious divide. Evidence that unconscious visual information is used by systems that process mathematical equations, like other unconscious priming studies, provides compelling reason to doubt that access is the (sole) functional contribution that conscious experience makes to visual perception. Neither integration nor access are therefore unique to (or sufficient for) conscious experience, and so are not likely candidates for picking out the functional contribution that consciousness makes in the domain of vision.

To summarize, in order to isolate what it is that consciousness contributes functionally to visual perception, it is necessary to get the best picture possible of the upper limits of unconscious visual processing capacities. It is becoming a widespread theoretical assumption that “feature extraction, categorization, some interference and inference occur regardless of whether one is conscious of the visual stimulus or not” (Lamme 2020). The CFS research programs outlined above help to reveal a central point that has been emerging over the last few decades of consciousness research: unconscious visual processing is functionally impressive, but it is also functionally limited. Unconscious vision is itself functionally hierarchical and can take on increasingly demanding information processing tasks (Breitmeyer 2014), and yet it maxes out at a certain level of functional complexity, at which point, the resources of consciousness are presumably recruited. The functions of access and integration exhibit this point in an interesting way: some kinds of access and integration can be carried out unconsciously while others cannot. This means that integration and access must be understood as graded notions that in some ways dissociate from conscious processing, and therefore cannot be common denominators that exhibit the function of conscious experience in visual perception.

It is important to keep in mind the possibility that the specific limits on unconscious visual processing depends on a variety of factors, and so are likely to vary across different individuals, different task demands, and presumably different species. Nevertheless, while unconscious visual processes have an astonishing arsenal of functionality, perhaps even enough to establish genuine perception (e.g. say on the grounds of criteria like perceptual constancy, see Block in Peters et al. 2017), their limits suggest that conscious processes contribute functional resources that are otherwise lacking. This specific point about visual processing is significant both for a) general theories that continue to deny that consciousness adds any functionality to the psychological system (e.g. Hassin's [2013] "Yes It Can" principle which states that unconscious processes have all the functional capability of conscious ones), and b) theories like IIT and GWT that mistakenly inflate the functional contributions of consciousness at the expense of appreciating the functional capacities of unconscious visual processing.

2.2. Visual Neglect

Another major source of evidence for unconscious visual processing comes from studying individuals who have experienced significant damage to the visual system. Neuropsychological analyses of pathological conditions have been a significant source of insight in the history of cognitive science, and once again this is seen paradigmatically in the domain of vision. Disorders like blindsight, visual object agnosia and visual neglect have continued to provide unique access into the workings of conscious and unconscious visual processing. Some cases of visual spatial neglect offer a particularly colorful illustration of the functional capacities and limitations of unconscious visual processing.

Visual hemineglect is a particular neuropsychological disorder that is caused by cellular damage in areas of the posterior parietal cortex and surrounding structures, typically as a consequence of strokes or aggressive tumors (e.g. Ogden 2005). The resulting deficit involves a lack of visual awareness in the region of the visual field that is contralateral to the damaged neural tissue (Berti & Rizzolatti 1992). In many instances, patients even seem to neglect the corresponding space of visual imagery drawn from long-term memory, like one side of a familiar street they are asked to recall (e.g. Bisiach & Luzzatti 1978), suggesting that the lack of awareness cannot be explained away as a mere deficit in attentional mechanisms. Despite this lack of awareness in hemineglect patients, there is evidence that visual information in the damaged visual field is processed unconsciously to some extent. Striking dissociations between subjective measures of awareness and objective measures of performance are well documented in hemineglect patients**.**

The first body of evidence for unconscious visual processing in patients with hemineglect is more anecdotal in nature. Several researchers have captured unusual and theoretically fascinating behaviours that result from these specific deficits in conscious visual perception (see figure 3). In one famous case (Halligan & Marshall 1988), researchers presented two separate line drawings of a house to a hemineglect patient. In one of the images, the left side of the house (appearing in the affected region of the visual field) was on fire. The subject reported that there was no difference between the two images, suggesting that they were not consciously aware of the burning side of the house. Some might argue that the subject was at least partially aware of the neglected stimulus, although the extent of cortical damage caused by aneurysm and resulting haemorrhage, and well as broader patterns of behaviour (e.g. when asked to bisect horizontal lines, the patient was typically over 50% to the right of true centre), provide strong reasons to reject this interpretation (Halligan & Marshall 1998, Ogden 2005). Nevertheless, when subsequently asked which house they preferred to live in, the subject reliably indicated that they preferred the house that was not on fire. This has been taken by many cognitive scientists as compelling evidence of fairly sophisticated processing of the part of the visual image that the subject reported not being consciously aware of. Over the subsequent decades, a variety of these sorts of atypical behaviours and similarly unusual confabulations provide prima facie reason to assume that although not consciously perceived, visual stimuli in neglected areas of visual space in hemineglect patients are still processed unconsciously (e.g. Verdon et al. 2010, Li & Malhorta 2015).

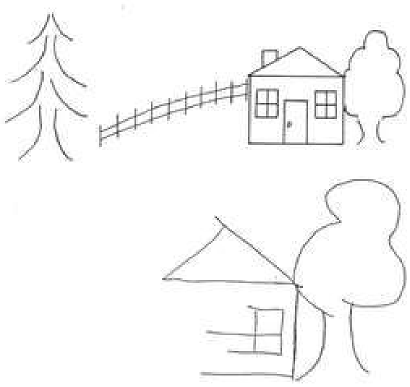

Figure 3. Typical performance for a hemineglect patient (bottom) when asked to copy a model image (top). From Ogden (2005).

The second, and likely more convincing, body of evidence comes from additional experimental tools that are specifically designed to test the extent to which visual information is being processed in hemineglect patients. Like in CFS research, psychophysical tools such as priming can be used to probe unconscious visual processing under pathological conditions. Berti and Rizzolatti (1992), for example, ran a standard priming study on patients with unilateral visual hemineglect. Their experiment showed that primes presented to the neglected portion of visual space facilitated task performance, even on trials where paired stimuli belonged to the same conceptual category despite being physically dissimilar. Once again, the extent of neural and behavioural pathology strongly indicates that visual consciousness was truly disrupted in the subjects. The authors concluded that “patients with neglect are able to process stimuli presented to the neglected field to a categorical level of representation even when they deny the stimulus presence in the affected field” (Berti & Rizzolatti 1992, p. 345). Similarly, Nakamura et al. (2012) used a priming paradigm to assess the extent to which words are processed in hemineglect patients. They also found that primes facilitated task performance even when presented to the neglected part of the visual field.

As a result, Brogaard et al. (2020) have recently argued that research on unconscious processing in hemineglect patients casts further doubt on “integrative” models like IIT and GWT as viable theories of consciousness and its function. They draw on evidence that certain visual illusions, like Kanizsa-style amodal completion illusions that require integration of visual information with “amodal” assumptions about objects in the world, still occur in subjects with hemineglect who deny having an experience of half of the available visual cues (Vuilleumier & Landis 1998). Because this appears to be genuine integration of visual information in the absence of awareness, integration is insufficient for conscious experience. The authors argue that the illusion only occurs when all the visual elements are integrated, and so the only viable interpretation is that despite failing to reach awareness, visual information is integrated in such a way as to play a constitutive role in establishing what becomes phenomenological content (Brogaard et al. 2020).

This sort of research further supports the picture that is emerging in experimental contexts: although functionally impressive, there are specific limits on unconscious visual processing. Sprenger et al. (2002), for instance, found that subjects with hemineglect had specific deficits in colour processing in the neglected parts of their visual field. Furthermore, much theoretical and empirical work has been done, for example, to try to understand the extent to which visual information is processed in patients with similar pathological conditions like “blindsight”, which is caused by damage to primary visual cortices. Alexander & Cowey’s research (2010), for example, suggests that when faces, colors, shapes and patterns are presented to blindsight patients, only “simple” stimulus features like luminance are processed unconsciously, as those features alone appear to be driving performance on perceptual discrimination tasks. Once again, several distinct research paradigms in neuropsychology are converging on the compelling idea that there are functional limitations in unconscious processing that are plausibly related to the increasingly complex information processing demands that certain stimuli make on the visual system. Specific performance failures under pathological conditions constitute further compelling evidence that consciousness is recruited for tasks that are more functionally complex.

3. Comparing Conscious Visual Perception

The same empirical tools that are used to investigate unconscious visual processes have been used to compare them with conscious ones. Once a minimal threshold for awareness is established, by leveraging report and/or appealing to objective indices of conscious experience, precise structural and functional comparisons can be made. The next step then is to look closely at some particular information-processing tasks in order to get the clearest picture possible of the similarities and differences in function between conscious and unconscious visual processing. This is necessary for isolating specific functional capacities that are unique to conscious visual processing. Ideally, we can then use these functional comparisons to isolate any common factors across distinct tasks, and only then begin to abstract a general philosophical account of consciousness, by asking what it is about experience such that it plays the role(s) that it does in visual processing. Building such a pluralistic functional profile will indeed be a conceptually nuanced project. There is wealth of research exploring very specific but at least conceptually unrelated visual functions like motion and color processing, many individual differences in processing capability have been observed between subjects, and studies that use similar experimental paradigms can either produce different interpretations of the same results, or produce different statistical results altogether. All of this will make drawing any substantial generalizations difficult. However, beginning to engage with this complexity by closely examining a few particular processes seems to be the only way forward for the philosophy and science of consciousness.

In the next sections, I'll look at semantic and spatiotemporal processing carried out by the visual system, in order to more precisely compare functional capabilities on these different kinds of tasks both in the presence and absence of conscious experience. Although my approach will be different from traditional contrastive analyses in that its primary target is a functional and not a structural comparison, clues about NCCs and other neurobiological properties will be drawn upon to guide and strengthen claims about function.

3.1 Semantic Processing in Vision

One central set of functions carried out by the human visual system is the processing of semantic information. This occurs, for example, both when we extract meaning from written language and when we recognize conceptual relations between bits of visual imagery. Semantic processing is a particularly good place to start building a functional profile for conscious experience in visual processing because the use of subliminally presented written words and images to prime performance on semantic tasks is extremely common in consciousness research. It has become one of the main battle grounds for debates about the extent and limits of unconscious visual processing (Kouider and Faivre 2017), which makes it particularly suitable for the kinds of contrastive analyses that dominate the field.

A variety of masking and binocular rivalry studies, including experiments using CFS, have probed conscious versus unconscious semantic processing by using written words to prime behavioural responses. Robust unconscious priming effects with written words are generally well established both in psychological (e.g. Jiang et al. 2007, Costello et al. 2009, Reber & Henke 2012, Armstrong & Dienes 2013) and neuroimaging research (e.g. Nakamura et al. 2007, Axelrod et al. 2014). Zabelina et al. (2013) for example, found that subjects performed better on compound association word problems, where three seemingly unrelated words (e.g. pine, crab, sauce) form familiar compounds with a solution word (e.g. apple), when the problem words were presented as subliminal primes during CFS. Crepaldi et al. (2010) similarly found that morphologically similar word pairs (e.g. “fell” and “fall”) showed significant priming effects on a standard masking paradigm only when the words were conceptually related (e.g. so not “bell” and “ball”), suggesting unconscious processing of semantic relationships. And, several studies have shown that more emotionally laden words reliably emerge from suppression faster than emotionally neutral words, which researchers have taken as compelling evidence of high level semantic processing of written language in the absence of visual awareness (e.g. Yang & Yeh 2010, Sklar et al. 2012).

However, there seem to be limits to the extent to which we can unconsciously process the semantic content of written words. One recent neuroimaging study (Nakamura et al. 2018), for example, found unconscious semantic priming effects only if primes and targets were separated by no more than two words in a sequence. Other masked priming studies suggest that the extent of unconscious priming effects with unpracticed word pairs depend significantly on their associative strength and semantic similarity (e.g. Van Den Bussche et al. 2012, Ortells et al. 2013). The point here is that although the visual system can extract meaning from written language in the absence of awareness, these capacities do in fact max out, either, for example, if the stimuli are too informationally complex (e.g. there are too many represented elements to process) or if the conceptual relations among words are not salient enough from previous learning history.

Probing semantic processing in conscious and unconscious vision using images has produced a similar pattern of results, providing even more evidence of the functional resources that consciousness contributes to visual processing. A variety of independent research programs have revealed that processing certain semantic relationships in visual imagery requires the subject to have conscious experiences. By leveraging combinations of suppression techniques like CFS and behavioural measures like semantic priming, evidence is mounting that tasks that require the perceptual discrimination of images based on the basic level category (Rosch 1978) that they belong to (e.g. snake or spider, tool or animal) cannot be accomplished when those images are unconsciously processed by the visual system (Cox et al. 2018, Hesselmann et al. 2016, Stein et al. 2020). Koivisto & Rientamo's (2016) study explicitly probed the kinds of semantic categorization that occur in the presence and absence of consciousness (see figure 4). They found unconscious priming in superordinate categorization tasks only (e.g. animal vs. non-animal), whereas no unconscious priming effect was observed when categorization tasks relied on basic level categories (e.g. horse vs. non-horse). They take this as evidence that unconscious representations in the visual system, which are presumably processed in a rapid feed-forward manner throughout the visual hierarchy, are much coarser than conscious ones, thus limiting the kinds of discriminations they can support. Interestingly, unconscious priming effects with visual imagery are typically still observed in these studies when subsequent discrimination tasks rely on low level information like shape rather than conceptual information, which is consistent with other research showing low level feature-driven facilitation of performance in the absence of awareness (e.g. Koivisto & Grassini 2018).

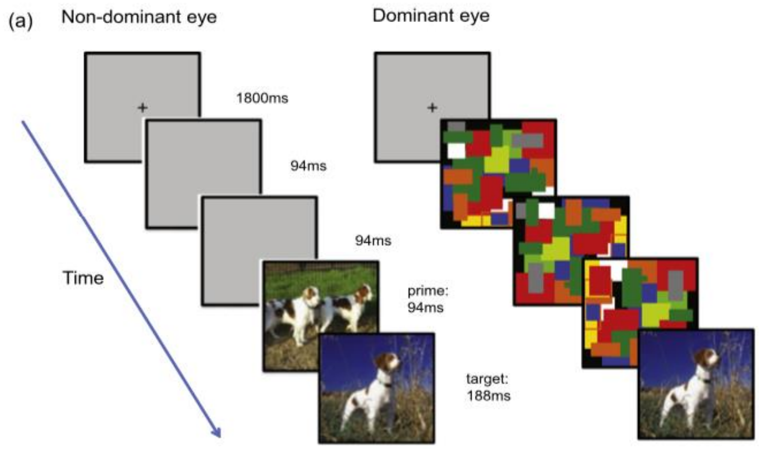

Figure 4. Koivisto & Rientamo's (2016) schematic representation of the CFS/semantic priming paradigm.

In summary, there appear to be robust unconscious semantic priming effects when primes are written words that remain under a certain threshold of representational complexity, but only limited unconscious priming effects when primes are rich visual images and when tasks go beyond mere discrimination by superordinate category or shape. These sorts of empirical results need to be carefully interpreted in order to extract their underlying significance for philosophical theories of consciousness. How do we best describe the functional contributions that consciousness is making to visual processing of semantic information?

It is not a simple task to characterize the relevant functional roles here. Semantic processing is complex, and plausibly involves a range of interacting computational capacities. The visual system's role, however, can be isolated in principle and understood independently. Briefly, semantically-laden visual stimuli activate linguistic processing mechanisms (presumably supported in part by structures in the anterior temporal lobe region of the ventral visual stream) as a result of learned associations between those stimuli and some set of abstracted referential content. In the case of written language, associations between words and their referential content become deeply entrenched throughout development. Crucially, written language relies on extremely simple visual stimuli: a few straight lines appropriately arranged is all that is needed to convey semantically relevant information, once those semantic relationships are learned. If unconscious processes are limited in their representational capacities, that is, they are only capable of trafficking in coarser representations with limited informational detail, then it makes sense that semantic processes can be activated unconsciously by words given the simplicity both of the stimulus itself and the internal processes required to represent it.

This in turn suggests that semantic tasks that rely on more informationally detailed representations require conscious experience. Koivisto and Rientamo (2016) assume as much, arguing that discrimination tasks that require more fine-grained representational capacities require conscious experience. There is a certain level of representational complexity needed to capture the appropriate stimulus features of a visual image in order to carry out certain semantically driven tasks. Here, complexity simply refers to the number of informational elements that comprise a given representation, or the quantity of distinct qualitative dimensions of a stimulus that are captured in a representation. Whereas representations of words track simple features like line orientation, processing visual imagery requires the combination of a variety of potentially task relevant visual features like color, overall shape including depth, and motion (especially if the semantic task requires categorization in terms of animacy). In this sense, the emerging picture is that semantic processing tasks that require representations of stimulus features at a certain level of informational complexity can only be accomplished when subjects are consciously aware of that stimulus. That is, consciousness appears to contribute increased functional resources to the visual processing of semantically relevant stimuli by facilitating more informationally-complex representations that are required for extracting meaning from more informationally-complex written language and visual scenes.

Semantic processing by the visual system is also a good place to start to address the individual differences observed in CFS research and related experimental paradigms. While this remains speculative until future research addresses these questions more directly, it is plausible that individual differences in performance on conscious and unconscious semantic processing tasks in vision are the result of individual differences in the ways that individuals encode meaning. Differences in individual learning histories, for example, will likely affect the stability of certain semantic relationships. In other words, more contact with certain semantic relations in the world (e.g. the meanings of words in a particular semantic domain) might increase the salience of these relations in subsequent episodes of visual processing. Much like how a skilled musician or athlete will eventually relegate much of their perceptual-motor processing to unconscious mechanisms, so too familiarity with certain semantic domains is likely to change the way that information is consolidated into memory systems, and ultimately how it is drawn upon at later times (e.g. Lupyan et al. 2020) At the very least, there is a testable hypothesis in this vicinity that unconscious processing of semantic relationships between elements of a visual image might be more robust as a result of greater familiarity. Future research might also look for patterns of difference between different age groups or different linguistic communities. This would also support the assumption that consciousness is required for processing more novel stimuli, which might ultimately bottom out in representational complexity, if novel stimuli require richer representational resources. Despite making it harder to draw a clean line between conscious and unconscious functions across individual humans, let alone species, these individual differences add support to the pluralist claim that conscious experience occupies different functional roles in different systems based on individual processing demands.

3.2 Spatial and Temporal Processing in Vision

Another central set of functions carried out by the visual system involves processing information about space and time. This is especially relevant to visually-guided action, where detailed spatial and temporal maps are constructed in order to plan, predict and guide even the most elementary bodily movements (e.g. reaching for and grasping objects, see Crawford et al. 2011). Representing spatial and temporal information in vision also sometimes requires integrating spatially or temporally disparate elements in order to carry out a cognitive or behavioural task effectively (e.g. tracking motion). Once again, spatiotemporal visual processing is fairly common in CFS and related research, and thus provides relatively stable grounds for comparison between the functions of conscious and unconscious visual processing.

Even though it is doubtful that the dorsal visual stream operates entirely unconsciously as was once assumed (e.g., see Wu 2020)**,** research into visuomotor transformation does provide compelling evidence that much spatial and temporal information processed in vision remains unconscious, and yet continues to guide behaviour in sophisticated ways. The visual system creates, maintains and updates representations of the positions of our bodies and other objects in "egocentric" space, as well as representations of the absolute size of objects and relations in "allocentric" space, that don't always enter into conscious experience (see Brogaard 2011 for review). This is most obvious in low-complexity cases like reflex, as well as high complexity cases like "flow", where much of the action planning and execution that relies on visual information remains unconscious. These are impressive feats of unconscious visual processing. In order to isolate what it is that consciousness contributes functionally to the visual processing of spatial and temporal properties, then, we need to compare performance on tasks where the key variables can be manipulated along their relevant dimensions.

Several specific studies have, for example, shed light on consciousness' role in enhancing spatial and temporal resolution in vision. Koivisto et al. (2014), employ a go/no go animal/non-animal categorization task paired with EEG recording in order to investigate the categorization capacities of conscious versus unconscious vision. Their results represent compelling evidence that rapid categorization can occur when masking disrupts recurrent processing and stimuli are not consciously perceived. And yet, when stimuli are unmasked and recurrent processing is established and maintained, a) the "clarity" or grain of visual categorization increases, b) categorization for unclear images gets faster and more accurate, and c) a greater electrophysiological difference is observed when categorizing animal versus non-animal stimuli. These functional advantages are explicitly construed by the authors as contributions that conscious awareness—facilitated by recurrent processing coalitions in the visual system—makes to the processing of spatial resolution. Along the same lines, Diano et. al (2017) argue that subcortical, feedforward (and presumably unconscious) processing of visual stimuli results in low frequency representations that trade detail for rapid categorization based on global properties. This contrasts with recurrent processing networks in the cortex that generate high spatial frequency representations in order to extract fine details from a visual scene for finer grained categorical distinctions. There seems to be growing consensus that unconscious processes simply traffic in coarser, and therefore less spatially and temporally detailed representations of visual stimuli.

Another variable that can be manipulated in this regard is the extent to which spatial and temporal integration is needed in order to carry out specific visual functions. Faivre and Koch (2014), for example, used CFS to compare performance on a task that requires the integration of motion-relevant visual information over increasing temporal periods. In one experimental set-up they used “apparent dot motion” stimuli, where multiple dots flashing in succession across a screen are reliably perceived as a single dot in motion. They employ an adaptation paradigm on both conscious and unconscious trials in order to assess and compare the extent of visual processing in each case. In both conscious and unconscious experimental conditions, they found adaptation effects in response to the presentation of the stimuli, suggesting paradigmatic perceptual capacities across the conditions. Crucially though, these adaptation effects only occur on unconscious trials when the “temporal integration windows”—or the temporal “distance” between the individual represented elements that need to be integrated—were sufficiently short (i.e. when the successive dot flashes were 100 ms apart). In contrast, integration over longer temporal windows (i.e. 400ms, 800ms, 1200ms) only occurred when subjects were consciously aware of the stimuli. This suggests that although some spatial and temporal integration of motion information is possible in the absence of awareness, integrating over larger temporal distances seems to depend on conscious experience.

Mudrik et al. (2014) review various such attempts to discover the capacities and limitations of unconscious spatiotemporal integration, which they take to be important for isolating the kinds of spatiotemporal processing tasks that require consciousness. Despite evidence that unconscious visual processing occurs on relatively spatiotemporally complex inputs (e.g. facial identity, natural scenes, etc.), the authors suggest that at a certain threshold, visual elements distributed across space and time likely cannot be processed together by the visual system without conscious awareness. One possible explanation they offer for this functional difference is that this threshold in unconscious processing represents a limit on ensemble encoding, understood as a subject’s ability to extract summary statistics from arrays of simultaneously presented visual stimuli. The assumption is that there is an important relationship between this psychological capacity and conscious experience, and ultimately that conscious processing is indeed required for integration over a certain threshold of complexity.

Finally, some familiar visual illusions depend on the influence of particular contextual elements that are spatially or temporally distributed. This means that another way to manipulate key variables here is to vary the amount, distribution and kind of contextual visual information that is drawn upon by the visual system in processing certain illusory stimuli. To this end, Harris et al. (2011) used CFS to selectively suppress contextual information from awareness while presenting subjects with different visual illusions. They found that simultaneous brightness illusions, in which identical stimuli look differently shaded due to differences in surrounding luminance, persists in the absence of awareness. In contrast, they found that Kanizsa-style contour illusions, in which the visual system represents illusory surface and edge information because of cues provided by shapes in the surrounding context, did not persist when *all* the surrounding cues—and not simply *half* of the cues as in the case of hemineglect patients discussed above—were suppressed from awareness (see figure 5). A plausible interpretation of this observed functional limitation of unconscious visual processing is that simultaneous brightness requires the processing of less spatially distributed information (i.e. merely two points of comparison) than the Kanizsa-style illusions do (i.e. typically three or four spatially distributed shapes that provide contour cues). That the latter illusion persists when only half of the cues are unconsciously processed in hemineglect patients but does not persist when all of the cues fail to reach awareness due to suppression techniques gives us an even more precise idea of the functional contributions of consciousness to these visual processes. More recently, Chen at al. (2018) similarly found that the Ebbinghaus illusion persists even when the surrounding context is selectively suppressed from awareness using CFS, whereas the Ponzo illusion does not. Again, the interpretation given here is that Ebbinghaus illusions are the result of lower level processes like contour interaction that require less spatially distributed information, whereas Ponzo illusions require holistic processing of much more spatially distributed contextual information. This all seems to support the hypothesis that one of consciousness' functions in vision is to facilitate more distributed, and hence in some ways more representationally demanding, spatial and temporal processing.

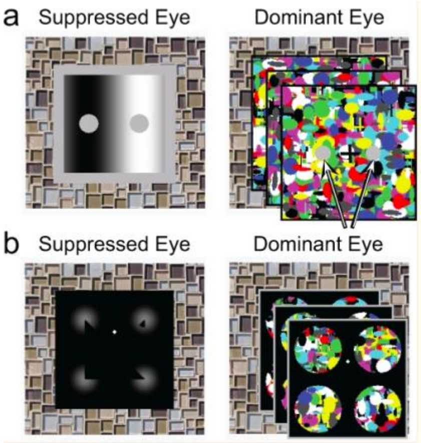

Figure 5. Harris et al. (2011) schematic representation of the CFS/visual illusion paradigm. A) Simultaneous brightness illusion under CFS. B) Kanizsa-style illusion under CFS.

4. Structural Evidence

These comparisons of function can be reinforced by what we know about the neurobiological mechanisms that carry them out. The emerging functional profile of consciousness in vision fits well with the idea that recurrent processing plays a key role in generating visual awareness. Recurrent (or re-entrant, or feedback) processing is a network property of neural populations, "where neurons mutually influence each other through direct, bidirectional interactions" (O'Reilly et al. 2013). It is characterized by stable patterns of connectivity and communication between particular neural regions at different levels of the visual cortical hierarchy, enabling higher level populations to modulate the activity of lower level populations (Tapia & Beck 2014, Sikkens et al. 2019). Some describe the onset of recurrent neural processing in the visual system metaphorically as an "ignition", or a rapid, non-linear increase of firing activity in response to small gradual changes in visual stimulus (e.g. breaking CFS suppression). This produces "intense and long lasting" neural co-activation that seems to be systematically associated with conscious experience (Fisch et al. 2009, Noy et al. 2015).

There seems to be a rare general consensus among scientists and naturalistic philosophers that recurrent processing of some sort is a necessary (although likely not sufficient) structural requisite for conscious experience (e.g. Lamme 2006, Dehaene et al. 2006, Lau and Rosenthal 2011, Faivre & Koch 2014, Sikkens et al. 2019). It should be noted again that it is likely not sufficient for consciousness because recurrent processing is a ubiquitous structural feature of complex systems, and thus some unconscious processes are likely also underpinned by recurrent neural mechanisms. This means that in order for recurrent processing to serve as a reliable marker of consciousness, we need to narrow down the qualifying conditions that link it to consciousness (e.g. the neural populations involved, the kind or level of activation required, the resulting representational structure, etc.). Various such proposals have been put forward, however the issue can obviously not be resolved here. As such, claims about the role of recurrent processing at this point should be charitably interpreted as the modest claim that there is at least one structural property of complex systems that is always required for the generation of conscious experience, even if it is also important for some unconscious processes.

Some of the deeper philosophical worries about how to precisely understand the metaphysical relationship between this important structural property and the psychological processes it enables inevitably arise here. For example, it might be that recurrent processing either a) is merely correlated with, b) is causally relevant to, or c) *just is* (i.e. is the reductive base of) conscious experience**.** I tend to favour a broadly anti-reductionist physicalist metaphysics of mind, according to which consciousness is entirely a product of activity in nervous systems (and perhaps artificial systems with sufficiently similar causal capacities), and yet there are features of consciousness as a psychological property that are not appropriately explained at the level of neural activity, thus restricting a wholesale reduction. These issues lie beyond the scope of the present project. The idea that recurrent processing is a feature of the brain that is importantly linked with consciousness should be compatible with most (at least non-dualistic) metaphysical theories of mind. Still, this structural property provides important insights into the nature of consciousness as a feature of complex systems.

First, note that recurrent processing and its association with conscious experience reflects the lessons of overarching frameworks in the philosophy of neuroscience like “neural reuse” or “neural recycling” that advocate for more nuanced accounts of the complex relationships between neural structure and psychological function (Anderson 2014, Barack 2019). According to these frameworks, the same neural structure can perform a variety of functions depending on the communicative partnerships it establishes with other structures

(Cabeza et al. 2018). This implies that the same neural structures can in principle support both conscious and unconscious processing, depending on the coalitions of recurrent activity that are established. Another way of putting this is to say that a structure (e.g. a region of the Anterior Temporal Lobe) is limited in functional capacity (e.g. the visual discriminations it can support) until it "ignites" recurrent circuit activity with some other structure(s) in the hierarchy.

There is still lively debate about which particular kinds of recurrent processing networks generate conscious visual experience; namely, whether more "localized" recurrent processing within sensory cortices or rather more "global" recurrent processing by way of long-ranging thalamocortical loops is necessary for visual awareness. This debate is directly relevant to claims about the functional contributions of consciousness. Those who remain skeptical of the claim that merely local recurrent processing is required for visual information to enter consciousness are typically committed to more architecturally demanding theoretical frameworks like Global Workspace and Higher Order theories. The real point of contention is whether or not functions supported by anterior structures like the pre-frontal cortex are necessary and/or sufficient as markers of conscious experience. On one view, conscious visual experience only arises once visually represented information is subject to further processing by higher cognitive functions (e.g. attention, meta-representation) carried out by the PFC and surrounding anterior regions (e.g. Dehaene et al. 2006, Lau and Rosenthal 2011, Odegaard et al. 2017, Brown et al. 2019). In contrast, many prominent theorists have begun to endorse the idea that these functions and their supporting structures are neither necessary nor sufficient features of conscious visual perception (van Gaal & Lamme 2012, Block 2014, Koch et al. 2016, Tononi et al. 2016, Boly et al. 2017). "Recurrent Processing Theory" (RPT) in particular has come to denote a related set of theoretical positions that endorse the basic tenet that recurrent processing networks within sensory cortices are the "true" NCCs (Lamme 2010).

There seems to be overwhelming evidence that the recurrent processes responsible for visual awareness take place among regions in the visual cortical hierarchy, and not higher cortical regions like the prefrontal cortex. Of course, some researchers continue to interpret the data as pointing to a key role for such anterior structures in the generation of conscious experience (e.g., see Michel & Morales 2019). However, in their seminal review, Koch et al.

(2016) survey a wealth of NCC research (e.g. neuroimaging contrasts, neuropathological evidence, etc.) that suggests that recurrent activity among a variety of neural structures in the sensory cortex is sufficient for conscious experience, while none of these particular recurrent networks alone seem to be necessary. Some of the most compelling evidence against the view that PFC and other anterior structures, as well as the functions they support, are necessary for consciousness comes from neuropsychological observation of pathological conditions. The fact that individuals with severe damage to these anterior regions (in some cases they are entirely removed, see e.g. Brickner 1952) seem to still maintain sensory consciousness, should be fairly compelling evidence that they are not necessary structural requisites of experience (Boly et al. 2017). In fact, NCC research seems to be pointing to a "Posterior Hot Zone" (Boly et al. 2017), or a diverse range of recurrent networks within the sensory cortices that are in fact sufficient for conscious experience.

This means that while recurrent processing in general might be a necessary structural requisite for consciousness, no *particular* recurrent processing network is necessary; that is, several specific recurrent processing networks are likely sufficient for producing conscious experience, including perhaps some in the PFC.[8] Such a structural pluralism, however, casts doubt on monolithic theories like GWT and HOT that assume that a particular cognitive function carried out by a particular neural structure must always be present when a given process is carried out consciously, and instead lends support to a pluralist account of the functions of consciousness.

Crucially though, we are still in need of a more satisfying answer to the question of exactly *why* we would expect these biological networks to facilitate the psychological functions that they do. In other words, we need to identify exactly which causal properties of recurrent networks are significant for conscious experience. There is a fruitful sort of reflective equilibrium at play here, where functional and structural accounts can be mutually informative in our quest to understand of the nature of conscious experience as a psychological property. One relevant causal property of recurrent processing networks seems to be, at least in the

[8] Whether or not this is a case of multiple realizability depends on how unified a concept of consciousness remains after its functional plurality is properly understood.

visual system, the modulation, and most importantly the selective amplification and stabilization, of visual information. A similar point is made by several researchers: Hupé et al. (1998, 2001) argue that feedback in the visual system "amplifies and focuses activity of neurons in lower visual areas"; Douglas and Martin (2007) contend that recurrent processing facilitates gain modulation, extraction, selection, and amplification of feedforward neural signals; Fairve and Koch (2014) suggest that recurrent processing is likely required for "maintaining a signal" over time; and Tapia and Beck (2014) review a range of other theoretical accounts of how to make sense of the causal dynamics of recurrent processing (e.g. minimizing prediction error, attentional modulation, "frame-and-fill" models, etc.). This reflects the common intuition that consciousness in some way involves keeping information "active" for use. The key point is that if the claim that recurrent processing networks are necessary (and under the appropriate qualifying conditions, sufficient) for consciousness is correct, then experience is at least partially characterized by this kind of enhanced signal processing; that is, we might think consciousness *is* in some important sense a kind of amplification and stabilization of information.

These causal properties of recurrent processing networks in the visual system can supplement our understanding of the functional contributions that consciousness makes to visual processing. The selective amplification and stabilization of information that recurrent processing facilitates is plausibly required for the construction of more informationally-rich or complex representations of visual stimuli that characterize conscious experience**.** Koivisto et al. (2014), again, found that while subjects could categorize objects in natural scenes even when masking disrupted feedback, the resolution and "clarity" of perceptual representation was enhanced when feedback was not disrupted. This is consistent with models of visual processing according to which feedback connections use the lower cortical areas they communicate with as "active-blackboards" (e.g. Bullier 2001). On these models, higher level computations actively and selectively modulate the response patterns of lower level neural populations, typically by amplifying some informational details at the expense of others, which ultimately changes the nature of the input back to higher areas. Indeed, a variety of research programs have probed the functional capacities afforded by recurrent processing networks. There is a lot of work, for example, showing that while functions like texture boundary detection are possible on a purely feedforward computation, other functions like surface and figure-ground segregation require the processing dynamics inherent in feedback connectivity (e.g. Scholte et al. 2008, O'Reilly et al. 2013). This is obviously an important clue for teasing out what it is that consciousness contributes to visual processing, if recurrent processing is indeed importantly associated with consciousness. Ultimately, the causal properties of recurrent networks seem to enable more complex representational resources, which helps to explain why consciousness contributes the functions that it does to visual processing.

Interestingly, this structural evidence might help us explain the significance of the "Breaking Continuous Flash Suppression" paradigm (see, e.g. Mudrik et al. 2011). Several CFS studies rely on the fact that some stimuli "break through" the controlled suppression faster than others, although there is much disagreement about how to interpret these findings. Yang et al. (2014), for example, say that the unconscious processing of certain stimuli "boosts" its input signal, thereby "empowering" it to emerge into consciousness. On the model being sketched here, these claims can be interpreted as pointing to the process by which the feedforward processing of some visual information signifies something "meaningful" to the subject, which draws on the resources of recurrent processing and the functions of consciousness it supports for further processing of that signal. Thus, it is empowered to rise to the level of consciousness in the sense that the output of feedforward processing signals the relevance of the stimuli and the need selectively amplify and stabilize our representations of it for further psychological applications. A lot more can be said here, as these hypotheses deserve further exploration, but the preceding remarks at the very least provide a theoretical framework for making sense of this common empirical finding.

One final point to reiterate here is that, given the possibility that there are recurrent neural networks that do not on their own generate conscious experience, there is more work to be done explaining why conscious experience correlates with recurrent processing between the specific areas of the brain that it does. In other words, there is more to be said about why recurrent processing is necessary and sufficient for consciousness only when it is established between specific cortical regions. Tapia and Beck (2014) propose that the recurrent networks that generate visual experience are typically established between sub-regions of V1 and functionally specialized extrastriatal areas like V5/MT+, and future neuroimaging work should move towards more precision in this regard. Further, Sikkens et al. (2019) point out that "not all feedback (not even within the cortex) may play the same role", supporting the idea that the plurality of specific recurrent processing networks might involve different functional capacities. More precisely, it might turn out that when different extrastriatal areas feedback to lower levels like V1, they support different functional contributions of consciousness, even within the domain of visual processing. A fascinating avenue for future research would be to look more closely at specific neural partnerships in order to make our understanding of the particular functional capacities they facilitate more precise. The future of the science and philosophy of consciousness might therefore involve understanding how particular (e.g. visual) experiences result from the consolidation of activity carried out by relatively functionally independent recurrent processing networks within particular regions of the cortex (and presumably both within and across perceptual modalities).

4.1 Lessons from Artificial Intelligence

One place to look for supplemental evidence of the functional significance of recurrent processing is in computational models of the visual system using artificial neural networks. Artificial visual systems have become impressively sophisticated, and are able to perform an increasingly wide range of visual recognition and categorization tasks. Despite this, there are still significant limitations in artificial neural networks compared to the primate visual system, plausibly in part because the former are traditionally characterized by feedforward architectures, while the latter make use of extensive feedback connectivity. Krieman and Serre (2020) provide a detailed summary of the limitations of feedforward architectures and the advantages of feedback mechanisms in artificial neural networks. Feedforward (or bottom-up) processing is sequential, and so relies on computations like filtering and normalization in order to recognize or classify a visual stimulus. When artificial neural networks introduce recurrent connectivity, however, they are able perform these tasks more accurately, flexibly and efficiently. Recurrent neural networks (RNNs) can, for example, solve harder recognition problems than feedforward systems, such as correctly categorizing objects that are heavily occluded, thanks to the additional computational resources it endows the system with. Once again, the claim here is merely that this is a structural property that is necessary for carrying out the psychological functions that seem to be closely associated with consciousness; that is, there are likely to be other important structural properties of complex systems that only taken together are jointly sufficient for conscious experience. Indeed, identifying these additional properties is an important next step for research into the natural and artificial structural requisites of consciousness.

Further, applying category labels to visual inputs, is "but one of many visual routines needed for scene understanding" (Krieman and Serre 2020, p. 223). Pattern completion and perceptual grouping, for instance, are visual tasks that feedforward architectures struggle to accomplish. The ability to extrapolate or generalize to novel visual stimuli is another crucial task for visual systems, and it seems that feedforward architectures are computationally impoverished in this regard. Finally, while various spatial relations can be easily identified by feedforward networks, other relations (e.g. same-different) seem to test their computational limits. The authors conclude that "feedback mechanisms may be the key component underlying human-level abstract visual reasoning" (Krieman and Serre 2020, p. 231). While more work needs to be done to understand precisely the causal properties of RNNs and their similarities to and differences from biological recurrent networks, the leading interpretation in AI research is that top-down influences allow visual systems to apply linguistic categories, symbolic reasoning, prior "experience", and a sort of "common sense knowledge" to the processing of complex visual stimuli. This research lends further support to the idea that recurrent processing confers specific processing advantages to the human visual system.

One challenge to the significance of recurrent connectivity for visual processing that has been put forward is called "the Unfolding Argument" (Doerig et al. 2019). Appealing to the properties of artificial neural networks, the argument is that "any recurrent network can be unfolded into a feedforward network implementing the same function" (Doerig et al. 2019, p. 52). In other words, it seems like the computational advantages of recurrent processing can be recreated in systems that have no feedback connectivity whatsoever, the implication being that feedback activity is somehow actually unnecessary for carrying out the kinds of functions associated with consciousness. However, rather than being a challenge to recurrent processing theories of the functions of consciousness, the idea of unfolding actually speaks to the need to replicate the computational properties of RPNs. Moreover, Krieman and Serre (2020) argue that:

> Unfolding a highly recurrent network to create a deeper feedforward network makes a commitment to a specific architecture and a given number of computational steps… Recurrent connections offer the flexibility to potentially vary the depth of processing across tasks, without the need to change the architecture for each task. (p. 226).

The point here is that the resulting “unfolded” systems are far too inefficient and inflexible to be biologically plausible, which again, speaks directly to the functional advantages of recurrent processing, rather than against its significance to the dynamics of human visual processing.

5. Functional Contributions of Consciousness: Visual Perception

The goal here was to begin to isolate some of the functions that consciousness contributes to the domain of visual processing, or at least formulate some testable hypothesis in this direction. Again, this will ultimately require identifying *where*, *how* and *why* consciousness enters the visual processing hierarchy, metaphorically speaking. After surveying the relevant empirical work and drawing out the key theoretical implications, we are in a position to offer some preliminary remarks in this direction.

The most difficult aspect of the problem seems to be saying exactly *where* consciousness enters the visual processing hierarchy. Differences in both individual learning history and experimental design make it difficult, if not impossible at this point, to draw a hard line in the visual system with unconscious processing on one side and conscious processing on the other in terms of their functional characteristics. But this result should be neither surprising nor troubling, given both the diversity and flexibility of the structures and functions that characterize human visual processing. In fact, exploring this shifting ontological boundary between unconscious and conscious vision will continue to strengthen our models of the functional contributions of consciousness. At this point, however, I think it is illuminating to acknowledge that the kinds of limited functional capacities that unconscious visual processes carry out are importantly related to the functions carried out by conscious visual processes; namely, there is a sort of spectral continuity here. That is, we typically find a relationship of degree between the functional capacities of unconscious versus conscious visual processing (e.g. spatiotemporal integration and resolution, semantic processing). According to the model on offer here, unconscious visual processing is in the business of constructing coarse representations that are useful within a certain limited range of informational complexity. This in turn implies that consciousness enters the picture or boost processing capacity wherever a certain visual task requires informational richness that cannot be captured by coarse unconscious representations, which is clearly a multifaceted and highly context dependent set of perceptual circumstances.

We are also in a better position to answer *how* it is that consciousness enters the hierarchy, and recurrent processing is at the centre of this story. Feedback mechanisms in the brain seem to allow information to be selectively amplified and stabilized, and the resulting representational and functional capacities are strongly associated with conscious perceptual experience. Again, more work is needed in order to identify why it is that recurrent processing between very particular regions in the sensory cortices gives rise to the particular conscious experiences they do, with the particular functional advantages they confer. Regardless, this general explanation hangs together nicely with the emerging account of *why* it is that consciousness enters the visual processing hierarchy; namely, to facilitate more informationally "rich" representations required for more demanding visual functions. Another way of saying this is that consciousness facilitates the processing of more "complex" stimulus features. Complexity and richness are intended as flexible terms that quantify a variety of "dimensions" of information processing demands placed on the visual system and the representations it employs**.** It might refer, for example, to the level of spatial resolution required to discriminate or categorize a particular visual input, the temporal distance between two related visual stimuli, or the conceptual subtlety of semantic content inherent in a visual scene. A host of research directly supports this general point about the different dimensions that comprise stimulus complexity. Song and Yao (2016), for example, found that the ability to discriminate images that were suppressed by CFS decreased systematically as stimulus complexity increased. Simple oriented gratings were easier to discriminate than faces or houses when processed unconsciously, when other factors like luminance were held constant. These findings provide insights into the function of conscious perception and offer an experimental approach for mapping out the scope and limits of unconscious processing.

By way of further clarification, it might also be said that the richness of conscious visual experience is twofold. On one hand, conscious visual processing appears to be characterized by representations that are more informationally abundant in sheer quantity (e.g. which is reflected in arguments for conscious overflow, such as Block 2014). On the other hand, conscious visual processing appears to be characterized by representations that are more abundant qualitatively, in terms of the diversity of the kinds of visual properties that can be represented simultaneously. This is reflected in arguments that consciousness is required for certain feats of perceptual integration (e.g. Faivre & Koch 2014). These quantitative and qualitative processing advantages enable a level of representational complexity unavailable to unconscious processes.

Given all this, we can now point to some specific candidate FCCs in the domain of vision. In particular, conscious processes appear to facilitate:

1. Increased capacities for semantically interpreting visual stimuli (i.e. words and meaningful images)

2. Heightened spatiotemporal precision in visual representation

3. integration of visual information over larger spatiotemporal windows (the latter of which also seems to enable integrating increasingly complex contextual information)

Note that these functions are not necessarily captured by the leading monolithic theories like GWT and IIT. GWT for example, which assumes that *the* function of consciousness is to access and broadcast information via the global workspace, is ill equipped to explain exactly how consciousness contributes functionally to aspects of visual perception that are not accessed for global broadcasting. In other words, none of the functions specified here are necessarily related to the activity of the global workspace; that is, the global broadcasting of informationally-rich representations is neither necessary nor sufficient for realizing these psychological functions. The same is true of IIT: the mere integration of information is neither necessary nor sufficient for conscious visual processing. Both theories employ graded psychological constructs, namely integration and access, that can occur unconsciously at least to some theoretically significant degree. In this way, integration and access, understood as monolithic constructs, are psychological functions that conceptually dissociate from consciousness; that is, at least until we can discover the different qualifying conditions that link them to different conscious processes (e.g. perhaps integration at reliable spatiotemporal thresholds).

The functional pluralism that this project is grounded on acknowledges that these functions, while likely sufficient for conscious experience, are likely not necessary. That is, it is plausible that while these functions require consciousness, many conscious experiences are not marked by these domain-specific functions. This makes intuitive sense: the way that experience enhances visual processing is likely not to be the same as the way it contributes to emotional processing, or to the processing of social bias, for example, where the same dimensions of representational complexity are not applicable. The intuition is strengthened when one considers the vast range of different kinds of specialized perceptual and cognitive mechanisms that emerge across the animal kingdom. Thus, the account on offer here points to rich avenues for future research. Specifically, researchers ought to shift their focus to domain-specific analyses of the functions of consciousness by performing similar comparisons between specific functional tasks carried out in other psychological domains. The end result of this inquiry should allow us to compile a range of different local, domain-specific, non-dissociable markers that can guide ongoing research in the philosophy and science of consciousness.

6. Conclusions

Several lines of research support the idea that although there are some fairly sophisticated unconscious visual processes, they are importantly limited in functional capacity. A careful analysis of the functional differences when awareness alone is manipulated in visual processing tasks reveals a cluster of capacities that consciousness likely facilitates. These include increased capacities for semantically interpreting visual stimuli, heightened spatiotemporal precision in visual representation, and integration of visual information over larger spatiotemporal windows. I urge that this sort of domain-specific functional analysis should precede overarching theories of the nature of consciousness. Moreover, the functions that consciousness contributes to vision are likely to be different from the functions it contributes to other psychological processes across the natural world, and perhaps in artificial systems. Further domain-specific analyses are required in order to determine whether 'consciousness' remains as a unified construct, or whether functional and structural differences prompt conceptual distinctions among phenomena previously unified under this conceptual category.

Chapter 3: Functions of Consciousness in Emotional Processing

Abstract

In this chapter, I defend the claim that conscious experience facilitates a variety of different functional capacities specific to the domain of emotional processing. The account is rooted in a general theoretical framework—what I'm calling a Valence Theory of Emotion—according to which emotional processing consists of valenced representational content, and the activity of the set of mechanisms responsible for generating, maintaining and modulating it. Evidence for genuine unconscious emotional processing is first reviewed, drawing on a) experimental research employing the tools of vision science (i.e. masking and suppression of emotionally relevant stimuli), and b) theoretical and clinical research on emotional disorder (i.e. generalized anxiety). After carefully comparing the functional capacities of the relevant unconscious and conscious emotional processes, I argue that conscious experience facilitates a cluster of functions in the domain of emotional processing, including: i) conceptualization of, ii) inhibition of, and iii) flexible response to valenced representational content.

1. Experiencing Emotion

1.1 Philosophical and Scientific Background

There is a deep and varied intellectual history devoted to understanding the nature and function of emotional processing as a feature of psychological systems. The prevalence and significance of emotion throughout the life of an organism continues to inspire rigorous philosophical and scientific investigation, often driven by the desire to integrate the phenomenon into broader theoretical and empirical models of the mind. Of particular relevance to the present analysis, much effort has been devoted to trying to understand how different sorts of conscious and unconscious processes contribute functionally to different emotional episodes. Despite all of this, there still is much disagreement about some important foundational issues, including what the proper target of emotion research is in the first place.

The first obstacle to a systematic account of the functions of consciousness in emotional processing, therefore, is to conceptually isolate the target of inquiry as clearly as possible.

In this section, I briefly sketch a working theory of emotion which I take to consolidate important insights from different prominent theoretical and empirical research programs. This will consequently become the conceptual backdrop for the subsequent comparative analysis aimed at isolating the functional contributions that consciousness makes (i.e. the Functional Contributions of Consciousness, or FCCs) within the domain of emotional processing. In section 2, I look closely at experimental paradigms that employ the tools of vision science to investigate emotional psychological processes, as well as clinical research on anxiety disorder, in order to illustrate the functional capacities and limitations of emotional processing in the absence of awareness. In section 3, I draw on these same bodies of research in order to identify psychological functions involved in emotion processing that are unique to consciousness, which include the conceptualization of, inhibition of, and flexible response to valenced representational content. In section 4, I consider neuroscientific research on emotion that provides supplemental evidence of consciousness' functions in this domain. Finally, in section 5, I briefly summarize and discuss some taxonomical results of my analysis.

Upon initial reflection, one might find intuitive reasons to doubt that 'emotion' constitutes a unified category or object of inquiry at all, given that it seems to involve a nebulous assortment of psychological and physiological phenomena (Griffiths, 2004). Existing attempts to carve out the relevant conceptual terrain are indeed largely unsatisfying, primarily because they reliably exclude some important feature of emotional processing. And yet, being able to clearly articulate what constitutes an emotion seems to be crucial for our ability to proceed with viable theoretical and experimental inquiry.

An important milestone in the modern study of emotion occurred in the early days of psychological science. William James' (1884) contention that emotion can be exhaustively explained in terms of sensory-perceptual systems and their 'felt' qualities inspired a tradition of theorists who assume emotional processing to be the product of "lower level" psychological and physiological mechanisms (e.g. Ekman 1977, Zajonc 1984, Stocker 1996). According to these *perceptual-affective* models, emotion consists of determined physiological responses (e.g. fixed patterns of facial expression, heart rate increase, and skin conductance response) to the perception of certain categories of environmental stimuli (e.g. threatening predators, significant conspecifics, tragic events). The implication is that emotional episodes occur to some degree beyond the volitional control of the subject. James also emphasized the *felt* qualities of these perceptually driven biological disturbances, and ultimately insisted that “our feeling of the same changes as they occur *is* the emotion” (James 1884, p. 189). Rooted in Darwinian evolutionary biology (1872), perceptual-affective accounts also typically characterize emotional processing as phylogenetically and/or ontogenetically primitive in some sense. These basic theoretical commitments seem to be largely compatible with common pre-theoretical assumptions about the nature and function of emotion.

Modern formulations of this general framework can be found both in philosophical models of emotion (e.g. Prinz 2004), and in the methodology of affective neuroscience developed by scientists like Damasio (1984; 2004). According to Damasio’s influential account, there are signature patterns of neural and physiological responses to perceived stimuli that constitute instances of different types of emotion. Damasio argues that those embodied processes can subsequently be represented within ‘second-order brain structures’, which reflects the felt aspects of emotion. However, the implication on this updated model is that *feeling* an emotion—or being consciously aware of it—requires the additional higher order representation of the emotional occurrence proper in the biological system. Consequently, this model implies that if, for example, certain ascending neural pathways are compromised pathologically, then emotional responses can in principle unfold entirely unconsciously (Damasio 2004).

Challenges to this theoretical camp are rooted in an opposing intuition: that some particular kind of “higher level” cognitive processing—like belief, appraisal, evaluation, categorization, or judgment—is both necessary and sufficient for emotion (e.g. Schachter and Singer 1962, Lazarus 1991, LeDoux 2017). Historically, Schachter and Singer (1962) offered a significant empirical challenge that continues to loom over perceptual-affective theories of emotion. Their experimental procedure involved injecting subjects with epinephrine to induce physiological arousal, and then manufacturing situations from which subjects derived cognitive rationalizations for why they felt the way they did. Subjects reliably reported having an emotional episode that was congruent with the evaluative interpretation provided by the experimental context (i.e. either joy or anger), despite experiencing the same physiological arousal. The study seemed to show, according to the authors, that “one labels, interprets, and identifies this stirred-up state in terms of the precipitating situation and one’s apperceptive mass” (Schachter & Singer 1962, p. 174).

This suggests that what truly characterizes and distinguishes the emotions is the cognitive or “top-down” evaluative appraisal of conceptual, perceptual and more broadly physiological (e.g. interoceptive) information. The idea that explicit, deliberate or voluntarily deployed higher level appraisal mechanisms are definitive of emotional processing reflects the common assumption that “to experience an emotion, people must *comprehend* that their well-being is implicated in a transaction, for better or worse” (Lazarus 1984, p.124, emphasis added). In general, these *cognitive theories* propose that emotional processing is either caused or constituted by some form of “thought-like” assessment by an organism about some feature of the environment that might positively or negatively impact their well-being, broadly understood. This proposal naturally fits with higher order theories of consciousness, where some explicit, cognitive processing (generally thought to be realized in prefrontal regions of the brain) is required to bring non-conscious (e.g. emotionally relevant) inputs together in conscious experience (Ledoux & Brown 2017).

Both approaches have explanatory merits, but both ultimately fall short of providing a comprehensive theoretical framework for studying emotion. On one hand, the assumption that emotional responses are nothing but lower level perceptual/affective processes cannot account for some phenomena that intuitively should be included in the scope of inquiry. “Long-standing” emotions for example (Solomon 2004)—say, the fact that I can be angry about something for five years without continuously experiencing a distinct physiological expression and without continuously perceiving the object of my anger—cast doubt on the idea that emotions can be exhaustively accounted for in terms of individual perceptual episodes and their resulting physiological expressions. Moreover, it is often argued that these models do not offer an adequate account of the intentional character of some emotional processes. That the object or content of an emotional state is central to its analysis has deep philosophical roots.[9] But it is difficult, some argue, to make sense of the content of more conceptually rich emotions, like existential dread for instance, on the perceptual-affective account (e.g. Green 1992, de Souza 2007, Reisenzein 2012). It is not clear, in the case of an emotional experience like existential dread what the emotionally relevant perceptual object could be. It also seems hard to imagine that each case of existential dread in the natural world has a common neurobiological and physiological signature pattern of response.

In contrast, what seems most puzzling about cognitive theories is the prevalence of situations in which avowed beliefs have no effect on emotional processing. Calhoun (1984) points out that emotional experiences are often inconsistent with our avowed and explicitly held beliefs about the world. For example, a spider might induce fear in a subject who knows enough about the spider's biology to justifiably and explicitly believe that it is harmless. Here, the emotional experience seems insulated from belief in such a way that suggests that they involve distinct processing networks. These "belief-emotion" conflicts pose a forceful challenge to cognitive theories: "the cognitivist must seriously entertain the apparent possibility of there being emotions that do not entail having the emotion-relevant beliefs" (Calhoun 1984, p. 322). In general, the "stubbornness" of emotions continues to be invoked as a challenge to cognitive theories (e.g. Pendoley forthcoming), the key criticism being that these accounts fail to account for the "deeply-rooted," subdoxastic, embodied nature of certain emotional responses. As a result, it has been argued that the state of emotions research resembles a cluster of distinct descriptions of various aspects of a multifaceted but unitary phenomenon, like the Indian parable of a group of blind people each describing different parts of an elephant (Russell & Barrett, 1999).

1.2 A Working Definition of Emotion

[9] For example, in Aristotle's Rhetoric (2004): "We will need with each of these emotions to investigate three particulars; in investigating anger, for instance, we will ask what the temperament is of angry people, with whom they most often become angry, and at what sorts of things."   .

Despite this theoretical divide, many have attempted to unify the domain of emotions research. It has been urged, for instance, that we ought to expand our understanding of the apparently necessary evaluative element of emotion to include both lower level perceptual-affective and higher level cognitive mechanisms. Interestingly, this kind of move is made on both sides of the debate: either to argue that some kind of cognitive evaluation is truly necessary and sufficient for emotion by showing that it can be manifested in lower level perceptual-affective mechanisms (e.g. Calhoun 1984, Solomon 2004) or to argue that lower level perceptual-affective mechanisms can indeed deliver the requisite sort of evaluative content thought to be in the jurisdiction of higher level cognitive processes like belief, judgment, or appraisal alone (e.g. Goldie 2000). It has also been argued, in an attempt to integrate insights from both cognitive and perceptual-affective theories, that emotions consist of hybrid composites of low and high level processing, such that each contributes something crucial to emotional episodes (e.g. Whiting 2011). This also seems to be the motivation behind "syndrome" theories of emotion (e.g. Averill 1985). These argumentative strategies ultimately represent the desire for a theoretical framework that is flexible enough to account for the wide range of psycho-physiological phenomena that seem to count as genuinely emotional.

Calhoun (1984), for example, thinks that we can explain belief-emotion conflicts by appealing to the notion that the same emotionally-relevant belief can be held *intellectually* but not *evidentially*, and vice versa. This kind of distinction suggests that emotional processing is varied and hierarchical, where networks of "evidential belief" or "seeings-as", rooted in lower level perceptual and affective processes, can be inconsistent with conceptually-richer beliefs that are held more explicitly. A similar move can be seen in Solomon's (2004) defense of his infamous slogan "emotion is judgment". He insists that judgement is a flexible construct that captures processes carried out by a wide variety of mechanisms, including our highest rational capacities and our lowest level bodily engagements with the world (i.e. "kinesthetic judgments"). Other thinkers similarly endorse a kind of emotional-representational "reflex" (Millikan 1996, Griffiths 2004), in which low level appraisal is construed as a "collapse of attitude", where "the 'affective computation' is simultaneously the belief that the world is a particular way and the intention to act in a particular way" (Griffiths 2004, p. 246). And Richard

Lazarus, who leans heavily on the assumption that cognitive appraisal is the key ingredient in emotional processing, also extends this to include both voluntary and automated responses, "conceptually" as well as "biologically" driven appraisal mechanisms, abstract-symbolic and more "elemental" processing, and processes occurring both consciously and unconsciously (Lazarus 1991). In contrast, some perceptual-affective theorists (e.g. Goldie 2000) have argued that low level perceptually driven evaluations better capture the world-directed content that forms the core of emotional processing, and upon which more elaborate cognitively driven evaluations will be based.

The possibility that genuine emotional processing occurs at different levels of the neurobiological and psychological hierarchy should inspire optimism about our ability to close the gap between cognitive and perceptual-affective theories of emotion by identifying a common element among them. But a common worry with the kinds of attempts at unification just surveyed is that terms such as 'belief' or 'judgment' might be illegitimately redefined so as to beg the question with regards to what counts as an emotional process. Griffiths (2004, p. 246) argues, for example, that "it is simply misleading to describe low level appraisal as evaluative judgment...instead, low level emotional appraisal seems to involve action-oriented representation." Similarly, in an ongoing debate about the nature of emotion, Zajonc (1984, p. 117) complains that:

> Lazarus's definitions of cognition and of cognitive appraisal also include forms of cognitive appraisal that cannot be observed, verified, or documented. Because the emotional reaction is *defined* as requiring cognitive appraisal as a crucial precondition, it must be present whether we have evidence of it or not…albeit at an unconscious level or in the form of the most primitive sensory registration.

It is important to recognize that some of the intuitions motivating these different theoretical camps are more convergent than they might first appear. Crucially, both perceptual-affective and cognitive accounts reflect the idea that emotional processing is essentially representational in nature (Charland, 1997). James' (1884) foundational model, for example, assumes that emotions are triggered by the perceptual representation of particular classes of objects and events. Solomon's flexible notion of *judgment*, for instance, is also driven by the assumption that “emotions are about the world” (2004, p. 77). In this vein, Nussbaum (2001) suggests that accounting for the diversity of emotional processing requires a flexible notion of *intentionality*. The point is that if both perceptual and cognitive processing are representational in nature, perhaps the construct of “emotion” spans this divide precisely because it picks out some general feature of representations, and is not limited to a particular processing domain or level. Indeed, it seems that the only potential candidate for being a common element among sensory-perceptual and cognitive aspects of emotional processing—a notion that is employed ubiquitously by philosophers, psychologists and neurobiologists—is their representational nature. Griffiths thinks “there are likely to be radical differences between the *representational* states involved in low-level and high-level appraisal” (2004, p. 246). And yet, the notion of representation is employed in the theoretical and empirical literature (and here by Griffiths himself) precisely as a way to capture something fundamental and ubiquitous about the capacities of information processing organisms.

I argue that the lack of consensus about the scope of this explanatory domain rests partly on a widespread category mistake about the kind of representational feature that emotions are. Specifically, while both perceptual and cognitive theories agree that emotion generally functions to encode evaluative significance, they mistakenly assume that certain kinds of representations themselves in some way constitute the emotion itself, such that the nature and function of emotion can be reduced to the nature and function of a subset of perceptual or cognitive representations. In contrast, the idea that emotional processing consists of *valenced representational content* seems to more accurately pick out the target representational feature, and provides the best candidate for being the common element among all the varied manifestations of emotion. This sort of definition seems compatible with a wide range of intuitions and theoretical frameworks across the field. On this view, emotions are evaluative information that is encoded into representations which are themselves either perceptual or cognitive (or possibly hybrid). Organisms represent the world in a variety of different ways, and representing the evaluative significance of different objects and events is central to all of these different sorts of representational processes at different levels of the psychological hierarchy. We can call this sort of general model a Valence Theory of Emotion (VTE): emotion just is valenced representational content, which encodes the value (very broadly construed) that represented objects and events (very broadly construed) have specifically for the evaluating subject. VTE consequently proposes quite a large domain of inquiry, and yet this reflects the intuition that emotional processing is valuable for a wide range of psychological operations (e.g. Etkin et al. 2015), presumably across organisms with very different information processing mechanisms. Indeed, as Lyon and Kuchling (2021) elegantly put it, "valence arguably is the fulcrum around which the dance of life revolves."

This theoretical proposal about the nature of emotion not only aims to capture the intuitions on both sides of the perceptual-affective/cognitive divide, but it also seems to capture the core principles of other widely influential theoretical and empirical research programs. For instance, pioneered largely by Barrett and colleagues (e.g. Barrett 2006, 2017, Russell 2003), the idea that the diverse range of emotional processing is comprised of some more basic psychological building blocks has continued to garner support. Barrett's work suggests that, instead of revealing robustly distinct emotion types with distinct processing signatures (cf. Ekman 1977), self-reports about occurrent emotion actually reveal that valence, or an evaluation of personal relevance and value resulting in various degrees of pleasure and displeasure, is the true "invariant core" of emotional life (Barrett 2006). On these views, valence is typically thought, along with intensity (i.e. arousal or level of activation[10]), to comprise something like an ever-present "core affect" that becomes the experiential background for the organism (Russell 2003). Further, these accounts typically appeal to some form of constructivism (i.e. psychological or social) in order to explain how such a primitive and ubiquitous representational encoding of valence can ultimately produce the full range of human emotional capacities.

This basic model also seems to be corroborated by neuroscientific considerations. The idea of a structurally and functionally isolated "limbic system" that is responsible for emotional processing has been cast into doubt for some time (e.g. Kotter & Meyer 1992, LeDoux 1996),

[10] My conjecture is that "intensity" should not be treated as an independent construct, as it simply picks out different degrees of positive or negative valence. However, this issue is beyond the scope of the present investigation. In what follows, although I'll only speak of "valence", I intend it to include different degrees or levels of activation, and in doing so remain neutral on whether or not we need to invoke "intensity" as an independent construct.

and the related idea that discrete emotions map cleanly onto neuroanatomy has also been cast into doubt by various research programs (e.g. Barrett & Wager 2006, Anderson 2014). Far from discovering identifiable patterns of activity among particular structures in the nervous system that are unique to distinct, linguistically-coded emotion categories, the neuroscientific study of emotional processing has revealed complex networks that draw on the fundamental causal properties of its component neural structures. In particular, the amygdala has emerged as a structure that is especially crucial, and perhaps necessary[11], for generating valence. The evidence is converging on the idea that the amygdala is one central neuroanatomical structure that supports the basic evaluation of the emotional relevance of information to the organism (Garavan et al. 2001, Glascher and Adolphs 2003, Sander et al. 2003, Bonnet et al. 2015). Neuroscientific research has even begun to isolate particular cellular pathways that correspond to the generation of positively and negatively valenced content. Beyeler et al. (2018), for example, mapped specific neural pathways in mice that originate in the basolateral amygdala (BLA), which appear to be responsible for independently encoding positive and negative valence. The BLA receives dense sensory input, and electrophysiological research suggests that this structure mediates associated learning of both fear and reward (Paton et al. 2006, Shabel and Janak 2009).

The VTE model ought to exhibit the kinds of complexities we should expect in the domain of emotional processing, but which can sometimes remain underappreciated. Simply put, representations and their consequences in the system can be staggeringly complex and multifaceted, and therefore so can the opportunities for encoding valence. This means that a single complex emotional episode, like becoming jealous, can involve dynamic, intricate systems of representation drawing on perception, cognition, memory, language and more. It is not always appreciated that the different ways those component representational elements become valenced need not be uniform, but rather might form a rich and nuanced tapestry of evaluative information. Consider an instance of jealousy, which, among other things, will involve both positive evaluations concerning some object or person of desire and negative

[11] Although the case has been made for valence processing in natural systems that lack complex neural structures like the amygdala (e.g. see Lyon & Kuchling 2021)

evaluations concerning the lack of fulfillment of that desire. Moreover, the cascading physiological, psychological and behavioural consequences of a single emotionally charged experience can ripple through the ongoing activity of the organism for decades, causing temporally extended physiological disturbances, changes in attentional allocation, influences on rationality and memory, self-regulating behavioural responses, and more. One advantage of VTE is that despite this complexity, we now have grounds for drawing a fairly clear distinction between processes that are genuinely components of emotion itself—namely, any representational process that takes on valenced content which implicates the evaluating subject—from those that are perhaps best understood as effects of the emotion itself—namely, physiological and behavioural responses to the presence of valenced content.

With this final point, a simple, and hopefully uncontroversial, working definition of emotion follows from VTE. The idea here is not to offer an innovative model that can challenge and replace existing frameworks, but to proceed with an account of emotion that captures the widest range of theoretical commitments possible. In what follows, I'll assume that emotion is valenced representational content in all its myriad forms, and that emotional processing functions to dynamically generate, maintain and modulate valenced representational content, encoding the evaluative significance of objects, events and ideas for the evaluating subject. This broad-stroke functional description obviously involves a wide range of nested subfunctions that apply to specific emotional contents; that is, particular emotions (e.g. joy) encode evaluative significance for an organism's particular values and contexts (e.g. things that bring joy like this should be pursued).

This was obviously a brief journey through a complex literature, but I take it to be a necessary one because there continues to be confusion surrounding the kind of things that emotions are. With a working definition in hand, we can now articulate the central question of this chapter more precisely: what functional capabilities, if any, does conscious experience contribute to the set of processes involved in generating, maintaining and modulating valenced representational content? On a broadly VTE model—that is, one that equates emotions with valenced content—there is already prima facie reason to believe that if there are indeed FCCs here, they will be different from those identified in the domain of vision, for example, as predicted by *functional pluralism*. This is because emotion involves a different aspect of representational processing than vision: visual representations are among the kinds of psychological phenomena that are themselves imbued with the evaluative content that characterizes emotion. We might therefore reasonably expect different functional capacities to be involved in the awareness of a particular kind of perceptual representation versus the awareness of that representation's evaluative content.

There are a variety of possible theoretical positions one could take on the relation between emotion and conscious experience. Indeed, there are still a number of prominent thinkers that equate emotion with experience. Higher order thought theories of emotion, for instance, assume that emotions are always conscious, and that the idea of an unconscious emotion is hopelessly confused (Ledoux & Brown 2017). This intuition is even shared by some who endorse the basic tenets of VTE (e.g. LeDoux 2017, Barrett 2017). One major difference between my formulation of VTE and others in this vein is their invocation of "core affect" and its definition in terms of conscious experience. According to Barrett (2017), for example, core affect necessarily refers to the ongoing *experience* of valence as an ever-present feature of an organism's conscious life. However, this construct seems to both confuse the nature of representation and neglect the functional capacities of unconscious processes. Something like core affect does seem utterly ubiquitous in the psychological life of an organism, given that valenced conscious representation is utterly ubiquitous in this way. When we consciously represent emotionally relevant information, we bring valenced content into awareness. But there seems to be no reason to begin with the assumption that processes trafficking in valenced content must always unfold in conscious experience. In fact, if valenced content is a property of most (maybe all) representations (Lebrecht et al. 2012), and some representations are unconsciously processed, then it follows that valenced content is also unconsciously processed. This point is reflected in the proposal of principled distinctions between "affect state" and "affective feelings" (Winkielman et al. 2005) or between "elicitation" and "experience" (Prinz 2005).

One lesson to be drawn here is that emotions must be understood as related to, but also in an important way conceptually distinct from, our ability to experience them. The next step in the current project, therefore, is to further motivate the idea that genuine emotional processing can be unconscious. Plausible evidence on this front should not only support the version of VTE on offer, but also licence functional comparisons with conscious emotional processes.

2. Unconscious Emotional Processing

The idea that a significant portion of emotional processing occurs outside of awareness was popularized with the rise of psychoanalysis, pioneered by psychologists like Freud and Jung. Despite criticisms of the Freudian picture of human psychology, many prominent philosophers and experimental scientists have continued to take this general idea seriously, and a diverse range of theoretical and empirical approaches aimed at understanding the murky realm of unconscious emotion have emerged over the last century**.** In what follows, I will survey a couple of distinct research programs—specifically a) research on unconscious perceptual (i.e. visual) processing of emotionally laden stimuli, and b) research on Generalized Anxiety Disorder and modern tools of psychotherapeutic intervention—in order to illustrate the functional capacities and limitations that unconscious mechanisms have for generating, maintaining and modulating valenced representational content.

2.1 Unconscious Perception of Emotionally Relevant Stimuli

One of the primary methods for studying unconscious emotional processing is by way of emotionally-laden perceptual, typically visual, stimuli (e.g. Garavan et al. 2001, Bonnet et al. 2015). The amygdala, thought to be a crucial neural structure for encoding emotional valence, has robust anatomical connections within the visual system (Diano et al. 2017), and has in fact been proposed to function as an intermediary step in early visual processing, providing a rapid evaluative assessment of incoming visual information (Sergerie 2008, Bonnet et al. 2015, Mendez-Bertolo et al. 2016, Diano et al. 2017, Ludwig 2020). Some have even argued that valence—or at least "micro-valence" characterized by very low levels of intensity or activation—is intrinsic to the very nature of all perceptual representation (Lebrecht et al. 2012). There is still much work to be done to map the complex relations between emotional and visual processing, however, and doing so is clearly beyond the scope of this paper. For the present purposes, the intimate interactivity between emotional and visual processing means that visual stimuli can be used to fairly reliably evoke emotional responses in experimental settings. Human faces are common visual stimuli used to induce emotional processing in the lab (e.g. Morris et al. 1998), but images containing other emotionally laden stimuli, like snakes and spiders, are also commonly employed (e.g. Ohman et al. 2001).

The reliance on visual stimuli to study emotional processing allows researchers to import sophisticated methods for rendering visual targets unconscious, namely masking and suppression techniques, that have already been widely and successfully employed in other experimental contexts. Emotion-inducing visual stimuli can be reliably masked or suppressed from awareness using a range of now-standard experimental designs, including everything from basic forward- and backward-masking techniques to more advanced methods like Continuous Flash Suppression (CFS). There are certainly legitimate worries about the validity of these sorts of experimental paradigms used to assess unconscious emotional processing, beyond those already inherent in masking and suppression research. For instance, it is controversial to assume that full-fledged emotional responses can occur in controlled laboratory settings at all. The psychological reaction to a visually presented stimulus (e.g. an angry face) in a controlled experimental setting is intuitively much different than what we'd expect to occur in real-life situations. Many researchers have pointed out the dynamic and multimodal nature of emotion-relevant perception, which is a significant challenge to replicate with static images (de Gelder 2005, Atkinson & Adolphs 2005). Some studies aim to remedy this by using classical conditioning techniques to reliably evoke responses to particular visual stimuli. For example, Morris et al. (1998) conditioned subjects to associate images of angry faces with bursts of white noise in order to reliably evoke emotional responses. But this remains a problem for all future research in this domain: figuring out how to evoke robust emotional responses in a laboratory setting while adhering to research ethics standards.

Despite these concerns, a wealth of research using various masking and suppression techniques has revealed that unconsciously processed emotionally laden visual stimuli reliably induce behavioural and physiological consequences indicative of genuine emotional processing.

These consequences include interference with or facilitation of performance (e.g. Hart et al. 2010, Diano et al. 2017), changes in perceptual processing such as delayed disengagement of attention (e.g. Georgiou et al. 2005), and physiological activity such as increased heart rate and skin conductance (e.g. Glascher & Adolphs 2003). In one study, Winkielman & Berridge (2004) used subliminally presented images of happy or angry faces to reliably and stably alter subject's preferences (i.e. willingness to pay) for a particular beverage, a strategy adopted by marketing campaigns around the world.[12] Moreover, there is evidence that patients with pathological conditions like 'blindsight' can encode the affective significance of visual stimuli despite severely degraded visual processing capacities (e.g. de Gelder 2005). In addition to this behavioural and physiological evidence, it has also been shown repeatedly via neuroimaging that amygdala reactivity indicative of learned associations of valence can be elicited even when visual stimuli remain unconscious (e.g. Morris et al. 1998, Gainotti 2012, Smith & Lane 2015).

It will be helpful to look at one particular experimental paradigm a little closer. On one application of Continuous Flash Suppression, stimuli that are "meaningful" to subjects (i.e. have a high degree of perceptual and cognitive salience) seem to "break through" induced suppression quicker than stimuli that are less meaningful, when the contrast between the binocularly presented target and suppression mask are slowly reversed (Stein & Sterzer 2014). This is taken as evidence of unconscious processing of the (e.g. semantic, contextual, evaluative, etc.) significance of suppressed stimuli. For example, Stein et al.'s (2012) experiment revealed that images of conspecifics—even atypical instances like headless human bodies—broke suppression much faster than when those same images are inverted, suggesting that only select visual information gains preferential access to awareness as a result of unconscious capacities for recognition and categorization. The theoretical interpretation here is that once a meaningful stimulus has been recognized and/or categorized unconsciously, its significance calls for the incoming signal to be quickly boosted, allowing conscious experience to be recruited for further processing.

[12] These results have been challenged on the grounds that the effect fails to replicate. One potential explanation of this failure is that subliminal priming (e.g. of a specific soft drink brand) only works if subjects also already have the relevant motivation (e.g. they are thirsty). For a brief review of this issue, see Smith & McCulloch (2012).

According to this method, given that emotionally relevant stimuli are inherently meaningful, they ought to break through suppression faster than emotionally neutral stimuli. The theoretical assumption is that because valence precisely encodes the positive or negative significance that objects and events have for the evaluating subject, it renders those with higher significance more psychologically salient, which ultimately facilitates enhanced perceptual processes like recognition and categorization (Zeelenberg et al. 2006). Several studies support this: stimuli with higher emotional relevance reliably break through suppression more rapidly than less emotionally relevant stimuli (e.g. Yang, Zald, & Blake, 2007; Sklar et al. 2012; Stein et al. 2014). Fear inducing stimuli especially tend to break suppression quickly, presumably as a result of the heightened value individuals place on potential threats. To lend further support to this idea, Capitao et al. (2014) showed that fear-inducing stimuli break suppression faster for individuals with higher baseline levels of anxiety, suggesting that because anxious individuals are biased towards fear inducing stimuli, they are better primed to efficiently detect and respond to them even at very early, preconscious processing stages. Moreover, in order to rule out the possibility that unconscious processing is merely acting on low level visual properties and not on emotional content itself, Viera et al. (2017) induced unconscious conditioned fear responses to otherwise neutral stimuli by pairing them with electric shocks, suggesting that unconscious processes are sensitive to emotional content itself and not merely learned associations with low level visual information. Gayet et al. (2016) ran a similar experiment, pairing otherwise neutral stimuli with electric shocks, and similarly found that stimuli that had become associated with the threat of shock broke suppression faster than stimuli that had not. Regardless of how best to interpret the results of CFS experiments in terms of the specific mechanisms at work, it is widely agreed that the phenomenon must be accounted for in some way in terms of unconscious processes that can recognize, categorize, and perform some preliminary processing of emotionally valenced representations.

The general consensus is that unconscious emotional responses are characterized by a rapid evaluation of incoming perceptual information that is mediated by a direct feedforward pathway from the retina to the amygdala (Ledoux 1996, Diano et al. 2017). Electrophysiological studies using human subjects during a presurgical evaluation period have placed amygdala reactivity to fearful faces in this subcortical processing route at 74 ms post-stimulus onset (Mendez-Bertolo et al. 2016), while more traditional masking paradigms and PET imaging have shown amygdala reactivity as early as 30 ms post-stimulus onset (Morris et al. 1998). Increased speed of amygdala reactivity during unconscious emotional evaluation has also been shown using binocular suppression experimental paradigms (Williams et al. 2004). These studies provide evidential support for the role of a subcortical route in facilitating the rapid evaluation of the emotional relevance of visual stimuli, even when they fail to reach conscious awareness.

It is also generally accepted that with the increases in speed of processing that this subcortical route enables, there is a corresponding decrease in the fidelity of information that is represented. Unconscious visual processing of emotionally relevant information certainly does reveal some category sensitivity. Whalen (1998) found different amygdala reactivity to fearful versus angry facial expressions, suggesting some ability to unconsciously distinguish between different categories of negative emotion. And some fMRI studies have revealed that amygdala reactivity to masked stimuli is in fact dependent upon categorical and contextual information about the stimuli. Fang et al. (2016), for example, found that masked, emotionally laden images of animals tend to elicit greater amygdala activation than emotionally laden images of inanimate objects, suggesting that amygdala reactivity discriminates between more or less ecologically relevant stimuli in the absence of awareness. Despite this basic category sensitivity, there is also mounting evidence for limitations in the recognitional and categorical resources of unconscious mechanisms that process emotional significance by way of visual perception. Celeghin et al. (2016), for instance, found that facial expressions that reflect more complex, socially constructed emotion categories, like guilt or arrogance, reliably fail to be processed unconsciously. More generally, visual scenes containing complex emotional information do not seem to elicit unconscious processing in patients with blindsight (de Gelder et al. 2002) or visual neglect (Grabowska et al. 2011).

This decrease in representational complexity is argued to make adaptive sense given the need to rapidly detect and respond to potentially threatening or beneficial information. The subcortical processing route to the amygdala is fed by the magnocellular pathway, which is biased towards low spatial frequency components of visual stimuli (Schiller & Malpeli 1977, Gainotti 2012). Several experiments provide evidence that unconscious emotional processing is sensitive to low but not high spatial frequency information (e.g. Mendez-Bertolo et al. 2016), in line with what has been discovered more generally about conscious versus unconscious visual processing. This rapid, feedforward, and coarse-grained subcortical processing allows the amygdala to perform its evaluations in response to minimal visual detail; only low spatial frequency features of stimuli (i.e. coarse in detail, global) are used to make an initial assessment of the emotional relevance of a particular stimulus to the organism. In evolutionary terms, it has been argued that there is adaptive value in being able to perform rapid assessments of the emotional significance of incoming information before it has the chance to be processed by higher cortical regions (Diano et al. 2017), in case an immediate behavioural response is required. This low spatial frequency processing appears to be sufficient for detection of threat and subsequent behavioural and physiological activation, presumably as a result of learned regularities in the visual system. The general implication here is that the visual system can facilitate the processing of emotional stimuli and generate genuine emotional episodes entirely in the absence of conscious experience, but only at a relatively low degree of representational complexity.

It is also important to note that there are neurobiological and psychological reasons for thinking that the same general story sketched here about unconscious visual processing of emotionally relevant stimuli is true of other sensory modalities. Consider verbal/auditory communication for example, one of the animal kingdom’s ancient methods for eliciting and sharing emotion. Minimal auditory information (e.g. a growl or a yelp) is needed to trigger a cascade of valence-driven emotional processes. It has been suggested that the emotional consequences of many specialized auditory signals (e.g. language) remain unconscious in humans, to the extent that the evaluative reaction and subsequent generation of a learned association of valence are not processed consciously (Owren et al. 2005). This point is supported by the neurobiological finding that the amygdala is similarly embedded in subcortical auditory (and other modality-specific) processing pathways (Diano et al. 2017). Taken together, there seems to be compelling evidence that perceptual processes facilitating emotional evaluation do not require (i.e. are not unique to, are insufficient for) conscious experience.

2.2 Emotional Disorder and Psychotherapy

Much important work has been done to understand and ultimately treat a range of emotional disorders by investigating the functional contributions of unconscious mechanisms. One significant improvement on the theoretical picture that emerged from the early days of Freudian psychoanalysis is the discovery that, far from operating entirely independently of and merely outputting in mysterious ways to conscious experiences of emotion, unconscious processes are intimately and intricately intertwined with conscious processes in the generation, maintenance, and modulation of valenced content (Ginot 2015). Generalized Anxiety Disorder provides a particularly clear case study for investigating some of the robust aspects of our emotional lives that escape conscious awareness. Research on anxiety and the kinds of psychological disorders it can fuel should also represent a more ecologically valid research program than those that rely on subliminal perceptual stimuli in the lab, because it investigates extremely common but complex emotional responses to extremely common but complex real-world stimuli.

At the heart of what we have come to understand as 'anxiety' in all its varied manifestations, is an evolutionarily ancient and biologically elegant mechanism: the fear or threat response system. The core operative mechanism responsible for this sort of fundamental negative evaluation is more or less preserved across an astounding range of organisms varying in complexity, maybe even in ants (see e.g. Normal et al. 2017). This phylogenetic ubiquity exhibits just how crucial of an ingredient threat response is in effectively representing the world, and ultimately surviving in it. Fear response systems serve to generate negatively valenced content in response to what the organism determines to be threatening, broadly speaking—including everything from predators to job interviews—leading to system-wide dysregulation "from which all people and animals wish to escape" (Panksepp & Biven, 2012, p. 36). Once the threat response system has determined that some aspect of the represented world is potentially harmful to the organism, these evaluations initiate dynamic, cascading processes that facilitate behavioural responses to the present threat, and can ultimately crystalize into learned associations that influence subsequent emotional evaluations and behavioural responses.

What is not always made explicit is the sheer complexity of the representational processing that comprises the threat response system and the resulting networks of emotional evaluation that it generates. On one hand, in adult human beings, the underlying evaluations that generate negative valence in threat response can be the result of learned associations that are surprisingly deeply rooted, not only evolutionarily but also developmentally speaking. Learned associations of valence begin to manifest extremely early in development, as infants "infuse" affective meaning into the world even before the capacity to construct and recall episodic memories come online (Mancia 2006). Early inputs to the fear response system that fuel these learned associations are also many and varied, including everything from overt threats to one's well-being, as seen in abusive households, to subtle signals and cues, such as a mother's neutral facial expression in response to their child's distress which signals that they are not available to help co-regulate the emotional response and ultimately alleviate the perceived threat (Ginot, 2015). Our psychological lives obviously become much more complex as we get older and become more engaged in the world, and hence so do the opportunities for additional, novel associations between stimuli and potential threat. Patterns of fear response also necessarily interact with each other in complex ways: one's fear of embarrassment might interact with their fear of novel environments, for instance, rendering them highly averse to entering new social situations. In addition, as is the case with emotional episodes more generally, a single instance of detected threat can initiate an astoundingly intricate cascade of physiological (e.g. heart rate increase, shortness of breath), psychological (e.g. reactivating old fears) and behavioural (e.g. coping strategies) consequences (Ginot 2015).

There is strong evidence, some of which was briefly reviewed in the previous section, that much of the processing involved in these kinds of fundamental emotional evaluations is unconscious. In particular, it has been reliably shown that conditioned fear responses—which involve the generation, maintenance, and modulation of valenced representational content that is characteristic of emotional processing—do not require conscious experience, suggesting that the detection of and response to threat is functionally distinct from our experience of fear (Morris et al., 1998; Ledoux, 2014). These learned associations driven by valenced content are also often not accessible to consciousness even upon effortful reflection, and yet they continue to exert far reaching influence on the psychological system. Often, subtle behavioural cues (like the mother's neutral facial expression) that generate threat response and carry enduring emotional significance are themselves "entirely out of awareness" (Ginot, 2015, p. 150), requiring significant effort to identify consciously. Moreover, many of the early fearful associations we make as infants, even if they were formed consciously, remain 'unresolved', in the sense that we fail to appropriately overcome the dysregulation caused by the emotional episode by effectively responding to the stimuli or coping with the triggered response. These unresolved evaluations of threat become buried under the constant bombardment of new inputs to the fear response system as our ability to represent and interact with the world becomes more psychologically sophisticated. And yet, this unconscious evaluative information often continues to exert significant influence on a range of psychological operations. The result is the strange but common situation in which, "we feel things, interpret events and act in certain ways, but we don't know why" (Ginot, 2015, p. 154).

The threat response system can malfunction in different ways and to different degrees, especially with regards to these failures to appropriately react to and eventually overcome or resolve the resulting dysregulation of the system (e.g. Hofmann 2007, Elman & Borsook 2018). This sort of emotional dysfunction is common in our species—consider how prevalent the irrational fear of public speaking is—and it will often escalate to the point where it is properly considered pathological. Indeed, there is evidence that anxiety disorders are the most common psychological disorder, affecting between one quarter to one third of individuals (e.g. Lepine 2002, Bandelow & Michaelis 2015). One major source of added complexity that triggers pathological malfunction occurs when the threat response itself (or its psychological and physiological manifestations) becomes a stimulus that provokes subsequent threat response, creating a vicious cycle that can eventually become extremely taxing on the system as a whole and can significantly impair normal functioning. Once again, the complex representational processes responsible for these malfunctions of the threat response system and the evaluative content it generates often occur in the absence of conscious awareness, and can require extensive therapeutic effort in order to identify and address (Ginot 2015). Anxiety disorder therefore provides a window into a sort of exaggerated set of unconscious emotional operations.

Generalized Anxiety Disorder (GAD), and the role of unconscious mechanisms in causing and sustaining it, is perhaps best illustrated through individual case studies. Ginot (2015) provides a colorful account of a patient, Ron, who presented typical psychological and behavioural symptoms of GAD. Ron first noticed in himself a general and diffuse sense of a lack of satisfaction with life, as well as decreased energy and feelings of ineffectiveness, both in his work and in his relationships. This general unhappiness and frustration, which began to consume every aspect of his life, compelled Ron to seek psychotherapy, aware that something in him needed to be addressed, but unaware of what it was. Ginot (2015, pp. 156-158) writes:

> What was notable and even startling about his initial presentation was his lack of awareness about his internal life, his clear patterns of complex and sophisticated avoidant behaviours both at work and in his love life—difficulties that quickly emerged in therapy…he was not aware of the extent that anxiety permeated much of his thinking and decisions…he knew something was wrong in the way he approached work and relationships, but was unconscious about the actual nature and extent of the amalgam of fear and defenses.

After engaging with psychotherapy, it became clear that "Ron was mired in a state of anxiety: a state of physical stress, fear and negative predictions" (Ginot, 2015, p. 56). This is a common state of affairs for patients with GAD: the evaluative content driving disordered physiological and behavioural responses remain inaccessible to the patient, even upon effortful conscious reflection. In other words, patients like Ron generate, maintain, and modulate valenced representational content in the absence of conscious experience. Typically a combination of pharmaceutical intervention and psychotherapy is necessary to treat the disorder in such a way as to mitigate the disruptive symptoms that result from a malfunctioning threat response system. In Ron's case, therapeutic treatment involved slowly unravelling a complex web of fear associations that had become buried in time, including a deeply rooted assumption that his father's frustration and abusive behaviour in his childhood was a result of his own failures and inadequacies.

It should be noted that probably all theories of emotion, even those that equate emotional valence with conscious experience (e.g. Barrett 2017), would agree that many of the upstream *causes* of a particular emotional episode are unconscious. Yet many theorists remain adamant that these upstream, lower level processes are not to be considered part of emotional processing proper; emotions are always consciously experienced. I conjecture that this is likely a symptom of the linguistic conflation of the terms "emotions" and "feelings", which seems to suggest that emotions are only what we actually experience. On HOT, for example, the assumption is that while there is a system of unconscious evaluative mechanisms that generate inputs to emotion systems, emotions themselves require consciousness (LeDoux & Brown 2017). This is based, in part on the idea that the self must be implicated in emotional evaluations, and that, according to HOT, consciousness functions to anchor unconsciously generated evaluations to the evaluating subject. In other words, an initial unconscious threat response becomes the subject of a higher order representation that includes the self, culminating in the experience of an emotion proper. One implication of this view is that we can never be mistaken about what emotion we are feeling (LeDoux & Brown 2017, p. 7), because emotion is characterized by what we feel.

However, the case of GAD exhibits just how representationally complex these unconscious processes can be, securing them as genuinely emotional processes according to VTE. The unconscious networks of valenced information that constitute generalized anxiety—and that are causally responsible for the physiological responses (e.g. tight chest, racing heart) that are felt by the anxious subject—include perceptions, thoughts, and memories that meet even highly restrictive criteria for being genuinely representational phenomena (e.g. Orlandi 2014). For example, the threat response system negatively valences perceptual encounters with *classes* of stimuli in the absence of awareness, which can be as concrete as "that scary dog" or abstract as "failure", securing the kind of abstraction thought by many to be necessary for representation. The stored memories constructed by the threat response system, as well as their lasting influence beyond the actual onset of the stimulus, also exhibit the sort of decoupling from proximal input that is a criterion for representation on many accounts. Unconsciously generated fear-infused representations also strongly influence behaviour at the system level, they are the result of extended periods of learning, and are—under the proper guidance of a psychotherapist—rationally responsive in the sense that they can be eventually mitigated or even eradicated through careful reconsideration of the patient's rational commitments and their logical structure. In order for this picture to hang together, we must necessarily assume that genuinely emotional processes—that is, processes that traffic in valenced representational content—can operate in the absence of conscious experience.

As already mentioned, note that within the HOT framework, it is generally agreed that first order representations of threat are indeed generated unconsciously, but consciousness is required for anchoring them to the subject and rendering them genuine emotions (Ledoux & Brown 2017). However, this position fails to acknowledge that these first-order representations must already involve the kind of "subject centredness" posited to be necessary for emotional evaluation. If these first order circuits signal threats to the well-being of the organism, they already encode the subject of evaluation, which seems sufficient for genuine emotion on all accounts. In short, if there is already an evaluation of significance to the organism in the first order representation, then there must also already be a representationally encoded subject of evaluation, which seems enough to secure genuine emotion regardless of whether these representations are conscious or not. Carruthers (2018) similarly argues that emotional evaluation and the goals and values it reflects must be understood as "subjective" in this way, even at the lowest levels of the processing hierarchy. These considerations ought to add to the conceivability of the idea that genuine emotional processes can occur outside of awareness.

The model on offer also explains why and how we are often "mistaken" about the emotional experiences we have, or at least unclear or confused about what we are emotionally experiencing and why. On one hand, it is quite common to not be able to accurately identify and articulate what we are feeling (i.e. physiologically) in the first place, because we simply remain unaware of the content of the evaluations driving those responses. There are cases, for instance, in which culturally propagated narratives negatively influence one's ability to identify one's own emotions, such as patriarchal and misogynistic constructions that cause women to misinterpret feelings of anger as more culturally encouraged feelings of sadness (Kring 2000). There are also simply many cases in which our current emotional state is so nuanced and multifaceted that it escapes clear identification based on introspection of our experience alone. On the other hand, in many cases in which we can identify the kind of emotional episode we are having, we remain unaware of *why* we are feeling it. This is the state of affairs for patients with GAD, where the emotional evaluations underlying sometimes intense episodes of fear are completely unknown. These fairly common phenomena seem to require positing complex unconscious processing of valence information.

This account of the role of unconscious processes in generating, maintaining and modulating valenced representational content also provides an interesting interpretation of the claim made by some philosophers that generalized anxiety is a rare but clear case of conscious experience that lacks intentional or representational content (e.g. Searle 2004). This assumption stems from the idea that experiences of anxiety are not typically accompanied by a clear sense of what the subject of the experience is anxious about. On the account just sketched, however, it should be clear that representational content responsible for the experience can certainly exist in the complex and often obscure realm of the unconscious, with its tangled webs of learned associations between classes of stimuli and positive or negative valence. Thus, it only seems as though generalized anxiety lacks clear representational content, because this content is so intricate and largely obscured from awareness. More work is needed to explore the mechanisms that are responsible for keeping this information from reaching consciousness, which sounds surprisingly similar to the original experimental motivations of Freudian psychoanalysis.

It should be noted that much of the preceding analysis might be relevant to emotional processing more generally. However, GAD simply allows us the opportunity to probe a specific psychological phenomenon that involves clearly identifiable contributions from unconscious mechanisms to emotional processing. Taken together, research on the unconscious perception of emotionally relevant stimuli and research on the mechanisms of threat response and generalized anxiety offer compelling evidence that valenced content is processed unconsciously, and that emotion can therefore in many ways be generated, maintained, and modulated without conscious experience.

3. Comparing Conscious Emotional Processing

We can now apply the general comparative method, using the functional capacities of these particular kinds of domain (i.e. emotion) specific unconscious processes as points of comparison in order to isolate any functional differences that result when consciousness is present. That is, if there is some psychological function involved in the generation, maintenance and modulation of valenced representational content that unconscious processes do not have the capacity to perform, then we can offer up that function as a candidate marker of experience in the domain of emotional processing. Emotional processing has deep phylogenetic and ontogenetic origins, and it has even been argued that emotional experiences are the earliest expression of consciousness in the natural world (e.g. Panksepp 2007). These markers could therefore be especially useful for ascribing consciousness in an extremely wide variety of biological organisms.

3.1. Conscious Perception of Emotionally Relevant Stimuli

Once again, because emotional valence is a property of different kinds of representations, we can leverage our ability to experimentally manipulate perceptual (e.g. visual) representation in order to understand some of the functional contributions that consciousness makes to emotional processing. The tools imported from vision science that are used to study the conscious and unconscious perceptual processing of emotional stimuli have been rigorously developed for decades now. As such, these experimental paradigms allow for uniquely precise functional comparisons between the conscious and unconscious processing of valenced representational content.

There are a few different functions that emerge from comparative analysis which are good candidates for being FCCs in the domain of emotion. First, being consciously aware of valenced (perceptual) content seems to facilitate increased capacities with regards to the recognition and categorization, and ultimately the conceptualization of emotionally valenced objects and events. Previous research on visual perception has revealed that conscious experience enhances the capacity to extract information from stimuli, specifically facilitating the visual system's ability to process category information that requires a certain degree of representational detail. Again, this is plausibly explained by the finding that unconscious processes detect emotionally relevant information via representations of low spatial frequency, whereas conscious processing seems to be needed for processing emotionally relevant information via representations of high spatial frequency (Gainotti 2012, Diano et al. 2017). The implication is that when a perceptual representation is deemed to be of high positive or negative value based on the rapid unconscious processing of low spatial frequency information about an object or event, consciousness can be recruited for further, more detailed processing of any significant informational elements (de Gelder 2005), much like attentional resources presumably continue to do downstream.

The idea that consciousness facilitates conceptualization in the domain of emotion is also reflected in the results of the breaking-CFS paradigm, where it is assumed that the most plausible explanation of why stimuli break suppression at different but predictable rates is that the resources of conscious experience are called upon when unconscious processes have recognized an emotionally relevant stimuli, but lack the resources for extracting more detailed information and further processing its informational content. The basic idea here is that because consciousness adds representational complexity to visual processing, consciousness ultimately increases the possible complexity of visually driven evaluations of valence. In this vein, Diano et al. (2017, p. 7) sum up their review as follows:

> Evidence therefore indicates that processing the emotional value of complex scenes, facial expressions of social emotions, or personal identity from faces depends critically on conscious visual perception and on the detailed processing of the high spatial frequency information that is characteristically performed by the cortical visual system in the ventral stream.

The representational resources that consciousness plausibly contributes to the perceptual processing of emotional stimuli seem to be enhanced further with the introduction of linguistic categories. Complex language use is widely taken to require conscious experience, and the conscious application of linguistic categories has been argued to have a modulating effect on basic perceptual processes more generally (Lupyan 2012). There is mounting evidence that this includes the modulation of perceptually processed emotional content, such that language actually structures and in some sense helps to constitute emotional perception (Lindquist et al. 2015). That language enhances an individual's "emotional granularity"—that is, the ability to make finer-grained distinctions among kinds of emotions (Wilson-Medenhall & Dunne 2021)—is supported by several lines of research. For instance, there is compelling evidence that while interference of linguistic resources impairs capacities for perceptually representing (e.g. recognizing or categorizing) valenced content, increasing the accessibility of linguistic resources enhances these capacities (Lindquist et al. 2015). This is exhibited clearly in individuals with semantic dementia, who reliably fail to sort images of different emotional expressions into groups beyond the broad categories of 'unpleasant' and 'pleasant' (Lindquist et al. 2014). Finally, cross-cultural differences have long been taken as evidence of the ways in which specific languages modulate emotion perception (Gendron et al. 2014). This general point about the role of language in emotion captures the intuitions of social constructivism, where more basic valencing processes are consciously elaborated upon and structured by cultural (e.g. linguistic) information (Hoemann et al. 2019, Barrett 2017). Future work ought to pursue the general hypothesis that all of this renders conscious emotion necessarily more vivid or intense than unconscious emotion, which would offer clues to the role that conscious emotion plays among the broader suite of psychological processes in which it is situated (e.g. perhaps more intense emotions are more attentionally salient or motivating).

Related to these differences in representational complexity are certain downstream psychological and behavioural consequences, which seem to reveal consciousness' role in the regulation of valenced representational (i.e. perceptual) content. This regulatory role involves at least two separable functions that might be considered FCCs in this psychological domain, each of which involves a kind of decoupling of valenced content from their preprogrammed responses.

On the one hand, emotional regulation can involve the inhibition of psychological, physiological, and behavioural consequences that typically follow from evaluations of positive and negative value. The consequences of unconsciously processed emotional information, while far reaching in the system, are taken to be "quantitatively and qualitatively different from those occurring during conscious emotion perception" (Diano et al. 2017, p. 8). Specifically, the unconscious perception of emotionally relevant stimuli sets off psychological, physiological and behavioural consequences that are more "characteristic" of typical emotional response, in the sense that they reflect deeply rooted associations between classes of stimuli with greater evolutionary and developmental significance. Evidence of this can be found in both the strength and speed of behavioural responses to unconsciously processed stimuli that are emotionally relevant (see, e.g. Williams et al. 2004, Tamietto et al. 2009). In contrast, these experiments suggest that consciousness is required for the inhibitory modulation of the kinds of "automatic" responses generated from unconscious mechanisms (e.g. Panksepp 2011). Fustos et al. (2013) argue, for example, that interoceptive awareness is required for the downregulation of an emotion's physiological consequences (e.g. heart racing), which is a crucial regulatory process.

Although it has been shown that some emotional regulation can occur unconsciously (e.g. Etkin et al. 2015, Koole & Rothermund 2011), these unconscious regulatory processes are constrained by the limited functionality of unconscious mechanisms. Essentially, "implicit" emotion regulation results from the same sorts of processes driving the initial formation of associative links between classes of stimuli and emotional responses. This is driven by controlled bottom-up processing, and can require quite extended learning in order to overcome or "rewire" associative links such that they no longer initiate automatic responses. In real-world settings, this sort of regulation might be outside the realm of possibility. In contrast, there seem to be different mechanisms at play when conscious experience is involved (e.g. Etkin et al. 2015). Here, emotional inhibition results from some sort of top-down modulation of an emotional episode, which might involve a range of complex strategies for overcoming an emotional reaction. One such strategy is the reappraisal of emotionally relevant information (Szczygiel et al. 2012), which has been shown to be a crucial regulatory strategy that can diminish the intensity of represented valence and its cascading consequences.

Finally, conscious experience also appears to facilitate more flexible and innovative behavioural response to valenced information. Emotions are widely assumed to be important for motivating behaviour (e.g. Helm 2010). In much the same way that top-down inhibition of emotion involves the attempt to decouple emotional valence from learned responses that have become relatively automated, the ability to react flexibly or innovatively to emotional content allows us to alter previously learned patterns of response. Flexible response to valenced representational content has been compellingly shown to require conscious experience. The idea is that the awareness of valenced content is required to motivate individuals to “overrule” patterns of response to the emotion-eliciting stimuli, facilitating flexibility, control, planning, and anticipation in terms of how the system responds to the dysregulation it produces (Diano et al. 2017). Lapate et al. (2016) showed, for example, that while unconsciously processed fearful faces can negatively bias subjects’ preferences for neutral objects as revealed by their behavioural responses, this negative bias was significantly reduced when subjects consciously perceived the emotionally relevant stimuli. It should be noted that overcoming this sort of biased response could involve either inhibition of the biased response or the flexible reaction to the initial biased response (or a combination of the two). Regardless of the operative mechanism, both seem to require conscious experience.

In sum, the same tools used to assess the functional differences between unconscious and conscious visual processing can be used to compare conscious and unconscious emotional processing. A comparative analysis reveals that conscious experience contributes a variety of functions in the domain of emotional processing, including increased capacities for the categorization of, inhibition of, and flexible response to valenced representational content. Given a sufficient body of reliable and replicable comparative experimental work, we might eventually be confident in arguing that these functions are sufficient for consciousness; that is, the presence of these domain-specific functions is enough to infer the presence of consciousness. These markers have the potential to be quite fruitful too, given the ubiquity of the visual processing of emotionally relevant stimuli across the natural world.

3.2. Consciousness and Emotional Disorder

The same FCCs that were isolated in the vision science paradigm also seem to emerge from research on emotional disorder and its therapeutic intervention. As VTE suggests, valence is a property of not just perceptual but of all kinds of representations, and so the tools of vision science only tell part of the story. Psychiatric disorders like GAD result from intricate, deeply rooted patterns of valence throughout the psychological hierarchy. Looking at the role that consciousness plays in the course of an anxiety disorder better reflects the functional complexity of real-world emotional processing. That consciousness' role in emotional processing appears to be preserved across different sorts of valenced representations is strong evidence that we have identified some candidate FCCs in this psychological domain.

First, there is evidence again that conscious experience facilitates increased capacities for categorizing and conceptualizing valenced information in the context of emotional disorder. Individuals with GAD and related anxiety disorders will often employ maladaptive strategies once they become aware that there is some robust source of emotional dysregulation occurring in the system. Such strategies can include worry and problem elaboration (i.e. the negative side of reappraisal), where individuals consciously generate representations of potential risks in response to, and in order to cope with, an initial threat (Stober 1998, Turk et al. 2005). These consciously deployed strategies are apt in some contexts, but can develop into a pattern of maladaptive response, where the anxious person excessively and obsessively produces highly detailed, abstract representations of perceived threat (Stober & Borkovec 2002). In this way, consciousness appears to play a crucial functional role in the construction of more elaborate evaluative representational content than can be generated by unconscious mechanisms alone, as seen quite clearly in these cases of pathology. In contrast, under the guidance of a psychotherapist, employing these conscious strategies like problem elaboration can nuance some of the valenced representations underlying an anxiety disorder in an attempt to diffuse the intensity of particularly salient evaluations of negative value (Ginot 2015). Finally, it is interesting to note that certain kinds of fear-infused memory might be so representationally complex as to be sufficient for consciousness (e.g. episodic memory with "autonoetic" content), but once formed, these sophisticated representations and their valenced content can become buried in the realm of more automated unconscious processes, and continue to influence the psychological system in myriad ways. Bringing these representations into consciousness is still a significant part of healing however, as it allows subjects to perform a sort of directed conceptualization that can diffuse their emotional impact.

Second, this research program also suggests that consciousness plays a crucial role in the regulation of emotional disorder, which relies heavily on strategies like cognitive reappraisal that involve the explicit (i.e. conscious) conceptual reformulation of an evaluative scenario (Jamieson et al. 2013, Campbell-Sills et al. 2014). Emotion regulation techniques like reappraisal can ultimately serve different specific functions in ongoing emotional processing, but in general it is assumed that consciousness contributes “the ability to reappraise one’s own emotional reactions” to the domain of emotional processing (Gross & Thompson 2007). On one hand, reappraisal might be deployed for the inhibition of “pre-programmed” psychological, physiological, and behavioural responses to emotionally relevant stimuli. Indeed much work has shown that explicit cognitive reappraisal is needed to overcome intense emotional episodes characteristic of GAD (e.g. Bebko et al., 2011, Hofmann, et al. 2009), by reducing automatically generated physiological and behavioural expressions of anxiety. In the context of psychotherapy, several different strategies appealing to conscious processing are employed in an attempt to downregulate the manifestations of emotional disorder. There is evidence that the conscious awareness of safety and lack of threat is likely required for inhibiting the threat response system (Ginot 2015, Panksepp and Biven, 2012). Finally, it has also been demonstrated that these conscious strategies directly reduce amygdala activation, which constitutes fairly reliable neurobiological evidence of emotional downregulation (e.g. Goldin et al., 2008; Ochsner et al., 2004).

Decoupling patterns of response to anxiety-inducing stimuli (perhaps by way of reappraisal) can also mean producing innovative behaviour in the face of the motivational pull of emotional content. Often, psychotherapists will suggest finding ways to shift behavioural patterns that accumulate in patients with generalized anxiety. The idea is to make the subject aware of the sorts of behavioural responses they have been employing but that are not working to reduce anxiety (e.g. obsessive thinking, unhealthy coping patterns like addiction), so that they can overwrite these patterns by (effortfully) developing new and more effective coping behaviours (e.g. deep relaxation, meditation, exercise). There is also a wealth of research supporting the effectiveness of consciously engaging in physical activity as a way to flexibly respond to the dysregulation caused by anxiety (e.g. Anderson & Shivakumar 2013). In a similar vein, Ginot (2015) argues that his clinical work with patients like Ron suggests that resolving conflicts between emotional evaluations and the behaviours they motivate also requires consciousness, which supports the idea that experience contributes the capacity for flexible response to valenced content. The idea that consciousness facilitates flexible behavioural response has deep roots, and is even assumed by some to be the primary biological function of consciousness (e.g. Earl 2014). However, this analysis reveals that, in the domain of emotional processing, flexible response is one FCC among many.

In sum, emotional disorders and their therapeutic interventions add further support to the candidate FCCs that emerged from the vision science paradigm; namely a) conceptualization, b) inhibition and c) flexible response. Indeed, the idea of bringing difficult emotional contents "to the surface" so that the resources of conscious experience can be brought to bear on them represents what I take to be a fairly widespread understanding of the nature of psychotherapy. With regards to Ginot's patient Ron, the conscious strategies employed in the context of therapy were crucial for his flourishing. Ginot (2015) relays how these strategies allow the patient to better understand, navigate and ideally overcome the learned associations and defensive reactions that operate unconsciously, and that sustain an emotional disorder like GAD. This research program therefore exhibits one valuable application of the overarching framework: isolating viable FCCs in the domain of emotional processing can help us better formulate strategies for emotional regulation, and ultimately for overcoming disruptive emotional disorders.

4. Structural Clues

We should also briefly consider any neuroscientific evidence concerning the relevant structural correlates when trying to formulate hypotheses about psychological function. Luckily, much work has been done to understand how emotion is realized in a variety of neurobiological networks, which, coupled with clues about the neural correlates of consciousness, can shed light on the functional contributions that consciousness makes in the domain of emotional processing.

There seems to be fairly widespread consensus that becoming conscious of one's emotions necessarily involves recurrent (i.e. feedback) processing from higher cortical structures back to significant subcortical structures like the amygdala. The amygdala has been shown to have similar patterns of activation in conscious and unconscious emotional processing (Phelps 2005), suggesting a preservation of the basic function of valence encoding across these conditions. However, mirroring debates in the domain of visual consciousness, one outstanding issue, is whether the recurrent processing proposed to be necessary for experiencing emotion originates in prefrontal, or in primarily sensory areas of the cortex. Higher order thought accounts of conscious emotion, for example, assume that structures that underwrite higher cognitive functions are the very same, and the only ones, that generate any conscious experience, including experiences of emotion (Ledoux & Brown 2017). Thus, this account predicts that recurrent processing from prefrontal structures to subcortical structures like the amygdala is necessary for consciousness. The brief survey that follows will help us get clearer on this issue, which should ultimately supplement the account on offer. However, the fact that the HOT model explicitly rejects pluralism about the structural and functional markers of consciousness should already caution us from adopting it.

4.1 Temporal Cortex and Enhanced Visual Processing

One network that has been implicated in conscious emotional processing involves recurrent connectivity from inferior or ventral regions of the temporal cortex to the amygdala (Pasley et al. 2004, Diano et al. 2017). Evidence for this kind of recurrent processing to the amygdala from temporal cortical regions has also been observed in primate models using florescent retrograde tracers (Stefanacci & Amaral 2000). The inferior temporal cortex (ITC) in particular is widely implicated in supporting more sophisticated visual processing (traditionally understood as part of the ventral visual stream), including object recognition and categorization at higher levels of abstraction (for a review of the role of the inferior temporal cortex in visual object categorization see Grill-Spector & Weiner 2014).

The structures that feedback to the amygdala in these core networks underlying emotional evaluation seem to lend support to our hypotheses concerning the different functional contributions conscious experience makes to valence encoding. On one hand, involvement of the ITC serves to recruit neural resources that are required for amplifying and stabilizing more fine-grained visual information extracted from a stimulus (Pasley et al. 2004). This is the most obvious contribution that the recurrent connectivity makes to processing in the amygdala, given that unconscious feedforward networks are selective for low-spatial frequency information only. When consciousness is established by way of feedback from ITC, higher-spatial frequency information can be extracted from emotionally relevant stimuli, which is consistent with the idea that conscious emotional processing provides a slower (given the more elaborate underlying neural architecture) but more detailed evaluation of visual stimuli in terms of their emotional significance or valence. In other words, feedback from ITC to the amygdala, which is highly associated with conscious experience, enhances an individual's ability to conceptually elaborate upon valenced perceptual content, ultimately achieving finer emotional granularity. Thus, this particular structural network speaks to the role consciousness plays in the categorization and conceptualization of valenced representational (i.e. perceptual) content.

4.2 Anterior Cingulate Cortex, Prefrontal Cortex and Behavioural Regulation

Other major structures that feed back to the amygdala in conscious emotional processing are the anterior cingulate cortex (ACC) and the dorsolateral prefrontal cortex (dlPFC). These are structures that are generally implicated in goal-directed action, cognitive control, and behavioural response selection. More specifically, the dlPFC seems to facilitate the implementation of executive control, while the ACC seems to facilitate performance monitoring, during the cognitive regulation of behavioural responses (MacDonald et al. 2000). The evidence is mounting that when the amygdala engages with the ACC and dlPFC via feedback connectivity, these structures can modulate the activity of subcortical areas and their relatively automatic evaluations of valence, as well as their resulting behavioural and physiological consequences (e.g. Stanley et al. 2008, Amodio 2014).

The anterior cingulate cortex and dorsolateral prefrontal cortex have previously been implicated in the regulation of spontaneously activated emotional evaluations. Enhanced BOLD responses in the amygdala correlating with a subliminally presented stimulus compared to a supraliminally presented stimulus suggest that the amygdala performs an automatic evaluative response that is suppressed or modulated when the subject is consciously aware of the stimulus and its emotional significance (Stanley et al. 2008). This speaks to consciousness' role specifically in the inhibition of emotional content and its cascading effects. The ACC is generally implicated in the automatic detection of conflict between "implicitly" held versus "explicitly" held attitudes (e.g. an unconsciously generated emotion and a conscious desire to overcome it), and the dlPFC is implicated in engaging cognitive control (either inhibition or flexible response) as a result of ACC activity. Given that the dlPFC is responsible for engaging cognitive control, and given its connectivity with motor systems, feedback originating in the dlPFC also likely supports the reformulation and execution of behavioural responses that are flexible; that is, not automatically activated based on previously adopted patterns of response. One interesting practical application of this research is that these regulatory effects seem to be further developed based on previous experience; for example, subjects with experience in interracial relationships show reduced emotional responses on the Implicit Association Test. This suggests that conscious strategies for emotional regulation might be employed more and be more effective in proportion to the richness and breadth of one's past experiences.

Thus, the amygdala, ACC and dlPFC are believed to constitute a network that is responsible for the conscious detection and regulation of automatic responses to evaluations of valence (Stanley et al. 2008), which again reflects the general point that recurrent networks that seem associated with consciousness are characterized in part by causal capacities for amplifying and stabilizing represented information. This neurobiological evidence also lends further support to the idea that being consciously aware of an emotionally laden stimulus facilitates the inhibition of and flexible response to that valenced information more specifically, in a way that unconscious emotional processing is incapable of. However, unlike HOT and related theories, it is important to appreciate the pluralistic nature of the structural and functional elements of consciousness in this domain and beyond. The neurobiological evidence briefly surveyed here reflects some of the wide variety of mechanisms underlying conscious emotional experience (some involving anterior neural regions and some not), which lends further support to a general pluralistic framework, and casts doubt on any monolithic theory of the function of consciousness.

5. Conclusions:

Functional Contributions of Consciousness in Emotional Processing

It is important to remember that functional pluralism and the search for domain-specific FCCs is still in its infancy, and so the conclusions drawn here are intended to be understood as preliminary guiding hypotheses rather than definitive claims. The more comparative research that we can draw on, the more confident we can be about which functions of emotional processing are unique to, or sufficient for, conscious experience. Still, we are in a position to offer up a few good candidate functions that seem to require consciousness in this psychological domain. There are three candidate FCCs that emerged from a preliminary look at the relevant data, each of which reflect different functional elements in the intricate networks of valenced representation that structure emotional systems:

1. Fine-grained evaluative contents (i.e. Conceptualization)

2. Down-regulation of evaluative content and its cascading effects (i.e. Inhibition)

3. Innovative response to dysregulation-causing evaluative content (i.e. Flexible Response)

Interestingly, in both experimental paradigms reviewed above, the same FCCs emerge. Since valenced content is the common denominator between them, it seems highly plausible that we have identified something important about how consciousness facilitates emotional processing more generally**.** Future research should be devoted to confirming that these hypothesized functions of consciousness are viable markers by performing rigorous comparisons with unconscious processing on carefully designed tasks.

More work also needs to be devoted to refining the constructs and taxonomical assumptions inherent in this analysis. Further philosophical analysis is required, for example, to determine whether these proposed FCCs are importantly different in terms of the underlying processes so as to necessitate “splitting” them into distinct psychological kinds. In contrast, there might be preliminary grounds for abstracting a unified account of emotional experience, centred around the notion of “regulatory functions” involved in the generation, maintenance and modulation of valenced representational content (this will be elaborated upon in the final chapter). Regardless of what the best way forward is, identifying strong candidates for being FCCs in the domain of emotion gives us important clues not only to what it is about experience such that it contributes what it does functionally to emotional processing, but also to the more general and elusive question of what consciousness is as a general feature of the psychological system.

Chapter 4: Functions of Consciousness and Biased Social Cognition

Abstract

This chapter evaluates a prevalent assumption concerning social cognition and bias; namely, that overcoming socially problematic (e.g. racist, sexist, ableist) biases requires conscious awareness. I briefly describe some of the specific psychological processes underlying social bias, and then survey the empirical literature in order to identify some bias intervention strategies that explicitly represent psychological functions that are likely unique to consciousness. As predicted by functional pluralism, a range of methods for overcoming social bias are indeed unique to consciousness, namely counterconditioning with vivid exemplars, perspective taking, and pre-emptive goal setting, while others are not, namely classical associative and evaluative counterconditioning. Although this is likely uncontroversial according to most existing theories, these conclusions contribute to the emerging functional profile of conscious experience. They also have practical implications with regards to formal and informal strategies for mitigating the social-psychological effects of systemic oppression, in that they reveal multiple points of possible intervention on representationally complex and deeply rooted biases.

1. Introduction:

The Philosophy and Science of Bias

The idea that human psychology is permeated with bias has drawn much attention from philosophers and scientists over the last century. That biases show up specifically in patterns of social cognition, such that they are both caused by and contribute to harmful social systems of oppression, has increasingly become the focus of theoretical and empirical inquiry. Given the complex contributions of a range of conscious and unconscious processes to biased social cognition and its intervention, this domain seems ripe for an analysis that aims to identify psychological capacities that are unique to conscious experience.

Several independent research programs have contributed to our understanding of the nature of humans' social biases. On one foundational line of inquiry, psychologists, behavioural economists and others interested in the status of human rationality sought to understand the extent to which human cognition relies on suboptimal heuristics in problem solving and decision making (e.g. Tversky & Kahneman 1974, Gigerenzer 2018). Much ink was spilled trying to identify and taxonomize different biased patterns of behavioural response, to isolate the operative psychological processes, and ultimately to evaluate their contributions to the success or failure of human intelligence (e.g. Gilovich et al. 2002, Gigerenzer & Brighton 2009). For instance, converging lines of experimental investigation suggest that individuals systematically interpret novel information according to previously held beliefs, expectations and hypotheses—the so-called “Myside” or “Confirmation” Bias (Nickerson 1998, Mercier 2017). These sorts of cognitive biases have been shown to be relatively stable across different contexts (e.g. Haselton & Nettle 2006, Stanovich et al. 2008), different sorts of tasks (e.g. Toplak et al. 2011), and different psychological domains (e.g. Rajsic et al. 2015), providing further evidence of their prevalence. Interestingly, it is still a matter of debate whether or not these biases should be construed as truly suboptimal information processing strategies according to normative standards of rationality. Contrary to the view that biases necessarily violate such standards, some have argued that the reliance on bias is often appropriate and even adaptive given how uncertain and unpredictable real-world contexts can be (Haselton et al. 2005, Gigerenzer & Brighton 2009). Rooted in the notion of ‘heuristic’, these research programs construe psychological bias as an automatically deployed step in information processing that is evolutionarily or developmentally acquired. The implication is that bias can render complex reasoning and decision-making tasks more efficient, although disagreement remains about the reliability or accuracy of different biased reasoning strategies (Tversky & Kahneman 1974, Gigerenzer & Brighton 2009). Regardless of whether or not bias necessarily violates normative standards of rationality, this line of inquiry emphasizes the systematicity (e.g. predictability, prevalence) thought to be central to the concept of bias in human psychology (Hahn and Harris 2014, Blanco 2017).

In another intellectual realm, feminist philosophers, critical-race theorists and others concerned with issues of social justice have a long tradition of drawing attention to the ways in which deeply rooted social systems of oppression infiltrate fundamental cognitive and perceptual processes (e.g. Bartky 1979, Sullivan and Tuana 2007, Anderson 2017). One central idea here is that an individual's particular position within oppressive social systems forms the psychological backdrop that they bring to bear on their ongoing cognitive and perceptual engagement with the world (Harraway 1988, Code 2006). These theoretical frameworks therefore call attention to the ways in which biased psychological processes actually emerge from and operate in existent social contexts, infusing existing research on bias with real-world content derived from the lived experiences of marginalized individuals. Take representativeness bias, for example, which has been described as the systematic tendency to weigh familiar objects and events as more probable, when their familiarity is irrelevant to the decision to be made (Gilovich & Griffin 2002). This mechanism is certainly prone to influence from social forces, and it is not hard to see how the systematic cultural misrepresentation of marginalized groups will produce widespread representativeness bias in individuals that reflects these social dynamics.[13] In general, a wide variety of harmful, socially propagated information (e.g. racism, sexism, ableism, etc.) inputs to individual psychological systems such that deeply rooted, biased assumptions about members of certain social groups are formed. In turn, these biased assumptions structure and ground the individual's psychological and behavioural tendencies, which ultimately functions to stabilize oppressive social systems. It has been argued that these social biases infiltrate processes across the psychological hierarchy, including everything from basic perceptual processing to abstract scientific theorizing (Peirce 1970, Mills 2007, Winston 2004). In addition, it is widely believed that these problematic social biases often manifest outside of conscious awareness and exert psychological influence even in those individuals who explicitly disavow them, which has significant theoretical and practical import (Lee 2017).

Finally, at the intersection of experimental psychology, cognitive science, and socially critical philosophy, research on the construct of implicit[14] bias represents an initial attempt to investigate different conscious and unconscious contributions to the psychology of oppression

[13] It should be emphasized that some biases that violate social and/or moral norms might not map cleanly onto biases that violate rational norms as discussed in the cognitive bias literature, suggesting that these research programs have importantly different explanatory targets. However, I think that an analysis of some of the intimate and intricate relationships between these research programs and the phenomena they study would be a fruitful avenue for future research.

[14] I will assume here that 'implicit' and 'unconscious' bias refer to the same psychological phenomena.

(e.g. Nosek et al. 2011, Brownstein 2018). The Implicit Association Test (Greenwald et al. 2009, Healy et al. 2014), for example, aims to systematically track how biases resulting from demographic membership unconsciously influence fundamental aspects of human information processing. There have certainly been challenges to the validity of the IAT as a direct measure of unconscious bias, although typically on the grounds that existing paradigms have merely failed to adequately operationalize the relation between the relevant psychological processes and observable behavioural measures (Forscher et. al 2017). Despite this challenge, it is generally agreed that social biases can be formed, maintained and deployed in the absence of conscious experience (Devine 1989, Amodio 2014). Much philosophical work has been done therefore to try to conceptualize the underlying psychological reality of implicit bias and its relationship to behaviour. While most theorists have assumed that implicit bias is characterized by basic associative learning mechanisms akin to those that drive classical conditioning (e.g. Brownstein 2015, Gendler 2011), others have argued that the fact that these biases are responsive to logical argumentation suggests instead that they are propositional in structure (e.g. Mandelbaum 2016). As discussed in the next section, there indeed seem to be a variety of psychological processes that are susceptible to social bias, including both those that are associative and those that are propositional in nature. In general, these research programs emphasize how deeply rooted social biases become, and how stably they persist and exert measurable influence on behaviour even in the face of genuine conscious disavowal. They also emphasize the need to carefully map the target psychological processes, so that claims about bias and bias-intervention are grounded in theoretically and empirically plausible accounts of the mind. Given the obvious goal of overcoming problematic social biases, this ought to motivate the search for intervention strategies that are sensitive to the underlying psychological dynamics.

The psychology of social bias is obviously a rich area of research, and so this chapter's aims are modest. I will simply offer some interpretation and evaluation of the assumption that conscious experience is required for overcoming oppressive social biases. In other words, I'll investigate the functional capacities that characterize some different bias intervention strategies that are unique to, or sufficient for, consciousness, and therefore that represent candidate functional markers of experience. In section 2, I'll say a bit more about the psychological processes underlying social bias, centred around an important distinction between "stereotype" and "prejudice". In section 3, I'll articulate different methods of intervention, and isolate those that are associated with consciousness. In section 4, I'll briefly discuss some of the theoretical implications this has for a general theory of the nature and functions of consciousness, as well as some of the practical implications it has for strategies aimed at mitigating the negative effects of social bias.

2. The Mechanisms of Bias

We can extract a general working definition of the target phenomenon by bringing together insights from the different research programs discussed above. Social biases are systematically activated psychological and behavioural sequences that reflect internalized assumptions about members of certain social groups, resulting from existing social systems of oppression. This abstract characterization, however, still does not do enough to specify the component psychological processes. In order to get clear on the functional role consciousness plays in bias intervention, we need a more precise understanding of the psychological manifestations of bias that we are trying to overcome.

Although perhaps not exhaustive, there is a good case to be made that social biases can be classified into two broad categories based on how they are realized psychologically and neurobiologically: prejudice and stereotype (Rudman et al. 2001, Amodio 2014). It should be noted that there are likely further taxonomical complexities within these categories, but this initial distinction can help us begin to get a sense of the kinds of psychological processes that become biased with problematic social information.

2.1. Prejudice

Social bias sometimes manifests as prejudice, which is characterized by learned emotion- or valence-driven responses to members of particular social groups (Cottrell & Neuberg 2005). These sorts of biases are largely generated through associative learning, where demographic membership becomes representationally linked with, and therefore comes to reliably invoke, negative emotional reactions like fear, hatred, and disgust, and their related behaviours (Amodio 2014). One example that has drawn a lot of attention from philosophers and scientists over the last decade is the biased perception of people of color as threatening by North American police officers. These kinds of culturally contingent associations are generated and reinforced in myriad ways, especially through forcefully propagated cultural narratives about oppressed individuals that remain unchallenged as a result of the lack of opportunity for positive mainstream representation of diverse cultures (e.g. Beeman & Narayan 2011). In general, categorization according to shared physical or cultural traits seems to be a fundamental aspect of human social cognition (Allport 1954), and yet our existing social categories have become systematically imbued with evaluative assumptions that reflect their oppressive origins.

Prejudice seems to be supported at least in part by neural structures that have long been implicated in emotionally driven learning more generally (Amodio 2014). The amygdala, which is widely believed to play a crucial role in the basic evaluations of positive and negative valence that get encoded in representations, seems to be particularly significant in producing prejudiced response to perceptual and conceptual information concerning demographic membership. Phelps et al. (2000), for example, used fMRI to investigate the role of the amygdala in prejudice, and found higher amygdala reactivity to Black versus White faces in individuals who had scored highly on independent measures of social bias. That some social biases manifest as learned emotional response to information that signals group membership means that some bias intervention strategies must target the relevant valence-generating processes.

2.2 Stereotype

Social bias sometimes manifests as stereotype, which is characterized by the conceptual linking of demographic membership with negative character traits and circumstantial attributes (Schneider 2005). These assumptions permeate conceptual rational thought, and are not necessarily accompanied by emotional valence as in prejudice. Stereotyping, while seemingly an appropriate cognitive strategy when employed in some contexts (Gendler 2011), can obviously negatively influence social cognition when it involves the systematic activation of conceptually linked attributes that reflect oppressive assumptions about marginalized groups (e.g. ‘person X’ + ‘criminal behaviour’). These are the sorts of biases thought to be tapped into by the IAT and related experimental paradigms, which purport to probe the biased selection and activation of stereotyped conceptual information for use in judgement and behaviour. However, this kind of problematic stereotyping emerges not just in simple conditioned associative relationships between particular elemental semantic contents as measured by the IAT, but also in more elaborate and temporally extended forms of reasoning and rational argumentation (Mandelbaum 2016).

The distinctness of stereotyping as a species of social bias is also reinforced by neurobiological considerations. The activation and deployment of stereotype is generally thought to be realized by temporal cortical structures and other regions that are implicated in conceptual representation and semantic memory (Wilson-Mendenhall et al. 2013, Binder 2016). More specifically, stereotypes that reflect oppressive assumptions about marginalized groups are thought to draw on conceptual associations realized in large part by neural activity in the anterior temporal lobe or ATL (Amodio 2014), which has been shown to selectively process conceptual representations of people and social groups (Olson et al. 2013). Although the processes underlying conceptual representation and semantic memory are likely distributed across the brain in a highly complex manner (Yee & Thompson-Schill 2016), the role of the ATL and related structures gives us a sense of the psychological processes underlying social stereotyping that are susceptible to bias. Once again, intervention strategies will need to be sensitive to these particular psychological dynamics, and target the relevant conceptual and semantic processes.

It should be noted that negative characteristics and negative emotional valence also typically become associatively linked with each other—prejudice and stereotyping networks are highly interactive after all (Sherman et al. 2005, Amodio 2014)—which reinforces and adds complexity to one’s bias against a particular social group. Presumably, emotional valence inputs crucial information that gets incorporated into stereotypical reasoning, while these conceptual structures proliferate and systematize prejudicial associations. This suggests further that many biases might be complex in terms of their processing dynamics, and therefore might afford multiple points of intervention.

3. Bias Intervention and Consciousness

Now that we have a clearer picture of some of the underlying psychological processes, we are in a better position to consider different possible interpretations of the hypothesis that conscious awareness is required for overcoming biased social cognition. This is a general assumption that seems to be widely held pretheoretically and in popular discussions about social bias. Calling attention to an individual's social biases—both "calling-out" which functions to expose some harm by bringing it to the perpetrator's awareness, and "calling-in" which functions to invite deeper reflection on the nature and cause of the harm in a more constructive way—has become perhaps one of the most commonly employed strategies aimed at mitigating oppressive thoughts and behaviours (e.g. Ortiz 2021). The idea seems to be that only when individuals become consciously aware that they harbour a certain bias can they take the steps necessary to break free of its systematic influence on their thought and behaviour.

The assumption that awareness is required for overcoming bias can also be seen in explicit theoretical claims and experimental paradigms (e.g. Izumi 2010, Devine et al. 2012, Lee 2017, Pope et al. 2018). Empirical research on diversity training, for example, suggests that awareness of one's particular biases is at least a necessary ingredient for "countering mental contamination" (Rudman et al. 2001, p. 866). Pope et al. (2018) also report statistical evidence that racial bias among professional basketball referees, for instance, was significantly reduced after an academic study that provided evidence of this bias gained widespread media attention. The statistical relationship between media coverage and bias reduction on this issue is taken as direct evidence of the crucial role that conscious experience plays in overcoming social bias. Again, the common claim here is typically that only once an individual becomes consciously aware of their bias—either through introspection, inference, or behavioural observation (Holroyd 2015)—are they adequately motivated to "break the habit", and ultimately to engage in both formal and informal anti-bias training (Devine & Monteith 1999).

In what follows, I'll discuss a few different specific strategies for effectively overcoming social bias[15], each representing a different iteration of this general hypothesis about the functional role of conscious experience. To be clear, there are two claims to be distinguished here: the claim that being consciously aware of the bias itself is required for overcoming it, and the claim that conscious experience provides processing resources required for overcoming certain biases, regardless of whether these biases are themselves conscious or not. It will likely be the case that conscious strategies for bias-intervention will also typically involve becoming conscious of the bias itself. But this need not be the case, as it is at least conceptually possible that we find a conscious strategy that works to overcome an unconscious bias. This point adds complexity to our understanding of the relationship between consciousness and social bias that unfortunately cannot be fully addressed here. The focus of inquiry going forward, therefore, will be restricted to the latter kind of claim: in what ways are different forms of conscious psychological processing required to overcome different kinds of prejudice and stereotype?

Still, I'll generally centre my discussion around biases that are held and that persist in the face of initial conscious disavowal on the part of the subject, as these are the sorts of biases that require the kinds of calculated and sustained intervention measures under consideration. Both prejudice and stereotype have long been thought to be capable of operating outside of conscious awareness and in the face of conscious disavowal in this way, (Devine 1989, Fazio et al. 1995), and so we will have to consider strategies for overcoming both kinds of deeply rooted bias. According to the functional pluralist framework, each strategy technically represents a candidate FCC, in that each describes a function that may or may not be unique to, or sufficient for, consciousness. The analysis that follows will likely not generate FCCs that are entirely surprising given what has emerged from the analysis of other domains, and given that many of the strategies surveyed explicitly draw on the resources of consciousness. However, each domain-specific FCC we can identify is an important contribution to the overall functional profile of conscious experience, and should ultimately contribute to our understanding of what consciousness is more generally as a feature of psychological systems.

[15] Drawn largely from a research contest that compared the efficacy of various bias-intervention strategies from leading research groups (Lai et al. 2014).

3.1 Counterconditioning

Both prejudice and stereotype can emerge from learned associations between members of certain social groups and either an emotional response or a stereotypical attribute, respectively. These associations are typically conditioned, in the classical sense that repeated exposure to a relationship between two phenomena can generate rigid processing patterns, such that exposure to one reliably, and in a sense automatically, activates the other (Maia 2009, Dang et al. 2015). As such, some bias intervention strategies take aim at these conditioned associations by way of counterconditioning.

Conditioning is widely agreed to be capable of unfolding unconsciously (i.e. learning conditioned associations without the awareness that one is learning them), especially when problematic associative relationships are so culturally ubiquitous that they blend into the very fabric of our social institutions (Coates 2011). Media portrayal provides one obvious illustration of the ubiquity of problematic associations individuals confront in their cultural landscapes. Islamophobia in North America, for example, is certainly at least buttressed by the systematic portrayal of individuals of Arab descent as villains in film and television (Saeed 2007, Jamal & Naber 2008), and this sort of systematic linking of group membership with negative characteristics (e.g. criminal activity) or negative emotional valence (e.g. fear/threat) typically goes unnoticed in the absence of effortful critical reflection (Beeman & Narayan 2011). Again, the associative interactions that form between negative characteristics and negative emotional valence add further complexity to the resulting biases.

Because at least some forms of conditioning can occur in the absence of conscious awareness, it follows that at least some forms of counterconditioning also do not require conscious awareness: the bias acquisition mechanisms and the bias intervention mechanisms must be the very same. In other words, the culturally ubiquitous inputs that contribute to unconsciously held social biases could in principle be different, such that an altogether different network of associations eventually replaces an existing one. This sort of counterconditioning does not require making either the biased stereotype or the nature of the bias intervention strategy conscious, rather subjects might remain totally unaware of the initial conditioned association, as well as the attempt to introduce the counterconditioning association (Pearson

2012). This conceptual point has not had significant uptake in the experimental sciences however, and thus a promising avenue for future research is to assess the efficacy of different kinds of unconscious counterconditioning specifically in terms of helping us overcoming social bias.

However, the fact that conditioning can unfold unconsciously has begun to be leveraged by experimental paradigms that seek to intervene on undesirable emotional responses more generally. Given that there is strong evidence that the same mechanisms are at play in both prejudice and emotional conditioning more broadly (Olsson and Phelps 2004), promising new strategies for counterconditioning emotional responses can be applied in the intervention on social bias. On one line of inquiry, Evaluative Conditioning represents a strategy for modulating valence-driven associations that does not seem to require subjects to be aware of the valence of the unconditioned stimulus or its relation to the conditioned stimulus (Waroquier et al. 2020). On a more elaborate (and indeed controversial) experimental paradigm, some researchers are beginning to use “decoded fMRI neurofeedback” (or DecNef) to countercondition undesired emotional responses (e.g. phobias). Here, positive (e.g. monetary) rewards become associated with participant-induced patterns of neural activity without subjects knowing the content or purpose of the procedure (Koizumi et al. 2016). Multi-voxel pattern analysis can map specific patterns of neural activation to mental representations of fear-inducing stimuli (e.g. colored disks that have been paired with mild electric shocks) with surprising accuracy. DecNef essentially involves motivating subjects via reward to induce a pattern of neural activity that reflects a fear-inducing stimulus without being aware that they are doing so. The goal is for that pattern of neural activity to eventually become associated with reward, drastically reducing phobic threat responses to subsequent presentations of the fear-inducing stimulus. These intervention strategies, while still very new, are no doubt applicable to the unconscious counterconditioning of social bias, when those biases are driven by learned emotional responses to members of a particular social group.

There are certainly forms of bias counterconditioning that cannot be carried out unconsciously, however. These are cases where it requires more extensive processing to make the counterconditioning association salient to the subject to the extent that it can influence subsequent psychological and behavioural events. With regards to stereotype, for example, counterconditioning sometimes requires the construction of “vivid” counterstereotypical scenarios (Lai et al. 2014), which is meant to explicitly signal a degree of representational complexity thought to be associated with consciousness alone. In one study, subjects heard elaborate stories in which they are the subject of a violent assault by a White man before being rescued by a Black man. The story includes intricate details and establishes personal relevance to the subject (e.g. “The last things you remember are the faint smells of alcohol and chewing tobacco and his wicked grin”), in order to maximize the vividness, and hence efficacy, of the counterstereotypical exemplar. Across a range of studies by the same team, this strategy significantly reduced implicit preferences that reflect oppressive assumptions about members of certain social groups (Lai et al 2014).

Another related intervention strategy that employs this more elaborate form of counterconditioning involves getting subjects to practice probes of implicit preference, like the IAT, with counterstereotypical exemplars (e.g. practice pairing Black with Good and White with Bad). Across different experimental trials, this strategy was shown to significantly reduce implicit preference for White over Black individuals. With regards to prejudice, counterconditioning can also require effortful practice overcoming previously formed associations between individuals from marginalized social groups and negative emotional valence. One lab implemented this sort of evaluative conditioning using an adapted “go/no-go association task”, getting subjects to practice responding only when stimuli pairs included a Black person and a positive word (Lai et al. 2014). This sort of “practice” breaking previously formed conditioned associations requires a certain amount of concentrated psychological effort, drawing on attentional and related executive functions that are highly associated with conscious experience. In each case discussed here (i.e. vivid counterstereotypical scenarios, practicing the IAT, go/no-go evaluative conditioning), there is no evidence that unconscious processes have the functional capacities necessary for concocting such elaborate counterconditioning scenarios or for facilitating such sustained effortful practice. These strategies require either the integration of a wide variety of perceptual and conceptual information or controlled and sustained behavioural discipline. Very few theories (if any) will deny that these sorts of higher level information processing functions require conscious experience.

There are more specific functional differences between these conscious and unconscious strategies for counterconditioning that are significant for the overall project of understanding the functional contributions of consciousness. Most significantly, there is mounting evidence that unconscious associative learning can only occur on standard delay conditioning paradigms, in which the conditioned stimulus remains present (Allen 2018). This sort of learning mechanism is certainly taken to be the "simplest", and therefore most ubiquitous, in the natural world (perhaps even manifesting in simple organisms like sea slugs that have less than 20,000 neurons, see Allen 2018). In contrast, trace conditioning, where associations are formed after the conditioned stimulus is no longer present and therefore must be held in working memory, is strongly correlated with consciousness. This is compatible with recently emerging theories that propose that a certain "distinctive kind of learning"; namely Unlimited Associative Learning which confers the systemic capacity for complex, open-ended networks of association, is an evolutionary marker of consciousness (e.g. Ginsburg & Jablonka 2019, Birch et al. 2020)—so long as this is taken to confer the capacity for trace conditioning in individual organisms (and, of course, it is acknowledged that this is one marker of consciousness among many). Other studies further support the idea that unconscious conditioning seems to require that stimulus pairings occur over shorter temporal intervals (Greenwald & De Houwer 2017), suggesting a similar functional limitation. Note that this mirrors consciousness' role in enhancing our ability to integrate information over larger spatiotemporal "windows" in visual perception (e.g. Faivre & Koch 2014), which might point to an important functional convergence across these domains.

It is also likely the case that more elaborate counterconditioning strategies will sometimes involve emotional reappraisal in the effort to override a prejudiced response. Recall that emotional reappraisal, one proposed mechanism potentially underlying the inhibition of and/or flexible response to valenced representational content, seems to require consciousness based on an analysis of the functional capacities of unconscious emotional processes. Evidence that some bias intervention strategies employ the same mechanism of emotional reappraisal would suggest another important convergence of function across these domains, which again would have important taxonomical implications. The point is that although we are at the preliminary stages of investigation into the functions of consciousness, it has already begun to reveal some common threads that will be crucial for drawing any plausible general taxonomical and metaphysical conclusions from this sort of careful, domain specific analysis (to be discussed in the final chapter).

To summarize, counterconditioning strategies that exploit the very same mechanisms for unconscious bias-intervention as those that are responsible for unconscious bias formation do not require consciousness. Strategies that require more elaborate construction of counterconditioning information seem to be unique to consciousness. The implication is that unconscious processing can be used to overcome bias only for simple conditioned associations, while conscious processing is required for constructing more representationally elaborate counterconditioning associations. While likely uncontroversial, these conclusions help to flesh out consciousness' functional profile.

3.2 Perspective Taking

Other forms of explicit thought can shift rational assumptions that reflect social bias without necessarily targeting the mechanisms of associative learning and conditioning. One bias intervention strategy that has attracted a lot of theoretical and empirical attention is "perspective taking". This approach is rooted in the intuitively plausible idea that there is value in attempting to empathetically adopt the perspective of individuals in different social circumstances, in order to motivate unbiased behaviour towards them. Presumably, being able to understand the experience of others puts us in a better position to understand the harms caused by socially biased behaviour, motivating us to mitigate those harms in others. It must be made clear, however, that perspective taking does not allow individuals in dominant social positions to truly experience oppression or even approximate the intricacies of living as an oppressed individual; this kind of understanding is likely out of reach from those dominant social positions (Bowman 2020). Still, engaging in empathy to the extent possible, such that one simulates to the best of their ability the experiences of another as if they were their own, appears to be a promising strategy for overcoming social bias.

Although similar in some ways, these intervention strategies can be treated distinctly from other forms of bias intervention that rely on more conceptually complex processing. Engaging in empathy is thought to involve an attempt at psychological “attunement” with others; a kind of moral attentiveness to the experiences of others that carries high motivational significance (Gruen 2015). There is evidence that cultivating empathy more generally, regardless of the particular perspective being adopted, can reduce the expression of particular biases, like racial prejudice (Pashak et al. 2018). However, empathetic perspective taking has now been developed and employed specifically as a tool for mitigating social bias.

As a bias intervention strategy, attempting to simulate the experience of others in empathetic perspective taking involves integrating debiasing information into more elaborate rational and moral psychological processes, and not merely conditioning a particular biased association in a particular context. In this way, perspective taking aims to restructure our general orientation to the world, and is thought to facilitate more temporally extended shifts in thought and behaviour. Although some forms of perspective taking and empathy training seem to be ineffective in altering implicit biases (e.g. game scenarios that require subjects to try to empathize with various hypothetical emotional reactions, see Lai et al. 2014), other techniques have shown promise (e.g. composing a narrative essay about a character from another demographic group, see Todd and Galinski 2012). Regardless of its efficacy, engaging in empathetic perspective taking represents a psychological phenomenon that is widely accepted to be unique to consciousness. Again, few (if any) theories would deny that the kind of conceptual resources required to engage in empathetic attempts at psychological attunement, including the ability to integrate complex socio-cultural information in simulations, require conscious experience. While likely uncontroversial, this kind of rare agreement among consciousness researchers ultimately reveals something stable in our theorizing about consciousness: at least some degree of conceptual intricacy cannot be achieved without conscious experience. This is a claim that seems to follow from many of our investigations into candidate FCCs, suggesting something important for our overall picture of what is it about consciousness such that it is involved in the particular psychological functions that it is (discussed in the final chapter).

3.3 Pre-emptive Goal Setting

Finally, another sort of bias intervention technique relies on explicit goal-setting and task orientation in order to steer the psychological system away from problematic social bias. The strategy here is to attempt to override the expression of bias in thought and action by having individuals pre-emptively and explicitly align themselves with anti-oppressive goals or motivational aims. There is experimental evidence in other contexts that an individual's explicit goals exert influence on their unconscious processing. Nakumara et al. (2007) for example, found that explicitly acknowledged task demands (i.e. either to read aloud or categorize a visually masked word) reliably changed the neural processing route for unconsciously perceived stimuli. It makes sense, then, that we can develop bias intervention strategies that capitalize on the influence of pre-emptive goal setting on unconscious processing.

One research paradigm, for instance, uses "implementation intentions", which are explicit plans for responding behaviourally to hypothetical situational cues (Stewart & Payne 2008). Implementation intentions specifically aim to pre-emptively strengthen associations between anti-oppressive goals and behavioural responses by rendering them more accessible in memory (Brandstätter, Lengfelder, & Gollwitzer 2001). In one series of studies, subjects were first informed about the nature of the IAT and the tendency for it to reveal implicit preferences, before being instructed to commit themselves to an implementation intention. This involved explaining to subjects the premise of the implementation intention strategy, and instructing them to say silently to themselves "I definitely want to respond to the Black face by thinking 'safe'…whenever I see a Black face on the screen, I will think 'safe'" before completing the IAT (Lai et al. 2014). The idea is to have subjects frame implementation intentions as hypothetical (i.e. if-then, or when-then) task instructions, thereby linking classes of environmental stimuli with specific behavioural responses. Explicit (i.e. verbal) commitment to the intention is thought to strengthen these desired stimuli-response associations, ultimately rendering them more psychologically accessible. Their results suggest that setting implementation intentions significantly decreases unconscious bias as measured by the IAT. This effect has since been replicated, and continues to show promise as a strategy for reducing both the activation and application of oppressive stereotypes (Rees et al. 2019).

Implementation intentions and related goal-centred strategies directly exploit the functional capacities of conscious processing. Built into the definition of the strategy, the very idea is to consciously commit oneself to hypothetical behavioural responses that reflect anti-oppressive goals in the attempt to restructure the unconscious activation and application of biased associations. In other words, there is no question that this is a bias intervention strategy that is unique to consciousness, and attempting to recreate it entirely unconsciously would necessarily change fundamental aspects of the intervention strategy. Implementation intentions therefore reveal how conscious processing can function to constrain, guide, and exert a kind of control over unconscious processing that is typically thought to be "automatic" (Brownstein 2015). Metaphorically speaking, consciousness functions here to "set the dials" for future episodes of unconscious processing. Some philosophers have even argued that this sense of "long term" conscious control over unconscious processing grounds the moral responsibility we have for our unconscious biases (e.g. Holroyd 2012). The efficacy of this intervention strategy ultimately exhibits why it is important to individually and collectively commit ourselves explicitly to anti-oppressive goals, and to try to cultivate a general "take" on the world, reflected in the overall state of the system, that is motivated by social justice. In terms of our project of distinguishing the functional capacities of conscious versus unconscious processes, that this bias intervention strategy requires the resources of conscious experience suggests that unconscious processing is limited in flexibility, specifically in terms of producing innovative (i.e. unbiased) psychological and behavioural responses to information about demographic membership.

4. Conclusions and Practical Implications

According to this preliminary survey, some bias intervention strategies clearly represent psychological functions that are unique to consciousness—namely, counterconditioning with representationally complex exemplars, perspective taking, and pre-emptive goal setting—while others do not—namely, classical associative and evaluative counterconditioning. It should be stressed that this is a relatively small review of a complex literature, and so it is best to think of these conclusions again as working hypotheses about the functions of consciousness that await further confirmation. That is, the following functions underlying some bias intervention strategies should be taken as candidate FCCs awaiting further experimental confirmation:

1. Counterconditioning with representationally complex exemplars

2. Perspective taking

3. Pre-emptive goal setting

This survey contributes to the overall picture that is emerging about the functional contributions of consciousness, and ultimately what it is about conscious experience such that it occupies the various functional roles that it does. Interestingly, each of these psychological functions involves a certain degree of representational complexity, which is a common theme in our search for FCCs. Certain intricacies were obviously outside of the scope of this brief review, including consideration of the ways that different (e.g. emotional and rational) conscious and unconscious processes interact to stabilize (Dang et al. 2015) but also help mitigate (Lee et al. 2017) social bias. However, there are still important implications to be drawn from this analysis, not only for our understanding of the nature and functions of consciousness, but also for the development of strategies for intervening on social bias.

It likely will not come as a surprise that many of the methods for overcoming social bias require conscious experience. One might think that being able to sustain social systems of oppression in general requires psychological resources that cannot be carried out unconsciously. Along these lines, Frith and Metzinger (2016) argue that consciousness is required for the integration of group preferences with our own self-models, enabling social emotions like regret that are crucial for moral cognition. However, given that many species with complex sociality are still denied ascriptions of consciousness on many leading theories, this is not a point we can take for granted at the moment. Moreover, given that social cognition is comprised of a variety of more fundamental psychological operations, there is value in isolating some of these particular processes in order to achieve a more precise understanding of the functional role that consciousness plays in this domain.

As such, there are some valuable insights here for our search for the functional contributions of consciousness. For instance, the bias intervention strategies reviewed above that are likely unique to consciousness all seem to require a kind or degree of informational integration that cannot be carried out unconsciously. IIT falls short as a general theory of the nature and functions of consciousness because unconscious processes often require explanation in terms of informational integration (Mudrik et al. 2011), and conscious experiences themselves need not always be maximally integrated (Brogaard 2020). However, this analysis points to a potential way forward for IIT. It seems that one productive avenue for further research is to distinguish the kinds or degrees of informational integration that are indeed truly unique to consciousness. We can then try to abstract from this functional profile a general theory about the nature of consciousness such that it is required for certain forms of informational integration but not others. The idea is to generate qualifying conditions on IIT's central identity claim "consciousness=integrated information", by appropriately restricting the scope of the latter class. The result should be a much more accurate and nuanced account of the relationship between consciousness and information integration. Furthermore, the fact that independent domain-specific analyses of visual perception and social bias both pointed to integration over certain spatiotemporal distances as a candidate FCC is perhaps evidence of a cross-domain functional feature of consciousness. Determining spatiotemporal thresholds for unconscious integration across different psychological domains therefore represents another important area for future research when it comes to devising bias intervention strategies.

There is also value here in seeing the general method in action. While the end result of the analysis suggests candidate FCCs that are not all that controversial, it is theoretically significant to demonstrate how a particular claim such as "overcoming bias requires consciousness" can be carefully interpreted in a variety of ways according to the functional pluralist model, and each interpretation can be evaluated independently as a claim about the functional role of consciousness. The wider variety of specific claims we subject to tests of necessity and sufficiency the better, as the result is either an important convergence of function that can ground more abstract and general claims about the nature and functions of consciousness, or an important divergence of function that fleshes out the overarching pluralism that initially motivates the search for FCCs.

This analysis also exhibits some of the practical applications of the overarching project, and helps to illustrate why it is important to link conscious experience to psychological functions. Bias training, comprised both of formal programs offered by organizations that are explicit attempts to eradicate social bias and informal psychological strategies that individuals can employ in daily life, can be highly volatile. It is widely documented, for example, that initially becoming conscious of one's bias can have a rebound or backlash effect in certain individuals for various social-psychological reasons (e.g. Macrae et al. 1994, Legault, Gutsell, & Inzlicht, 2011). One specific practical implication here, then, is that we might want to avoid conscious bias intervention strategies in some contexts, particularly in situations where backlash is expected, or in individuals who are, for whatever reason, not motivated to take the necessary steps to follow through with formal or informal anti-bias intervention training once they've become aware. As seen in the context of fear reduction more generally, people are averse to confronting stimuli that produce unwanted emotional consequences. This means that unconscious conditioning will be most effective in these contexts, perhaps both in laboratory settings and in our broader cultural contexts.

Another practical implication of this analysis is that because we have shown that a single psychological construct like "social bias" involves a dynamic assortment of psychological processes and their underlying neurobiological mechanisms, we have also simultaneously identified a variety of entry points where bias intervention might be effective. There is strong evidence that intervention is most effective when multiple strategies are employed that each target different psychological aspects of a particular bias (Lai et al. 2014). Moreover, on their own, none of these bias intervention strategies appear to result in significant lasting change in thought and behavior (Lai et al. 2016)), suggesting the need for a sustained and concerted effort in the attempt to eradicate social bias. Given this, a commitment to overcoming oppressive social structures must involve an assortment of strategies that target a range of different conscious and unconscious processes. This supports the theoretical assumption of functional pluralism: consciousness facilitates a range of different psychological functions, even within the domain of social cognition. Different strategies for overcoming bias draw on different functional contributions of consciousness, suggesting that consciousness is a multifunctional psychological property. Thus, while the psychological complexity of bias adds to its stability, it also renders it vulnerable to multi-faceted intervention strategies that exploit the different functional contributions that conscious experience makes in this domain. The hope is that continuing this sort of analysis should reveal a range of functional contributions that consciousness makes to the complex processes involved in biased social cognition, which will hopefully support the implementation of intervention strategies that are more targeted and multifaceted, and more effective as a result.

Chapter 5: Implications for the Philosophy and Science of Consciousness

Abstract

This brief final chapter brings together some of the key conclusions that can be drawn from the preceding analysis into the functional contributions of consciousness. In particular, I'll first consider some taxonomical implications of the functional pluralist project, and their consequences for future research on consciousness in the psychological and neural sciences. Next, I'll begin the process of abstracting a general philosophical account of the nature of consciousness from its emerging functional profile. In doing so, I sketch what I call a Local Workspace Theory of Consciousness, according to which local recurrent processing networks selectively amplify and stabilize information according to local processing needs, producing a range of local markers of consciousness. I conclude with what I take to be the best formulation of a general claim about the nature of consciousness as a psychological property that we have available to us at the moment; namely, that consciousness is at least in part characterized by a high degree of representational complexity afforded by the structural mechanisms that realize it and reflected in the psychological functions that it facilitates.

1. Drawing Conclusions

The final stage of this project involves bringing together the results of these three domain specific analyses in order to a) get a better sense of the specific kinds or types of functions that consciousness facilitates within and across psychological domains (discussed in section 1.1), and b) abstract from this functional profile any general features of conscious experience that help to explain why it facilitates the psychological functions that it does (discussed in section 1.2). In order to do so, I'll begin by drawing out some of the taxonomical conclusions that follow from my analysis, which will support and constrain my discussion of some of the metaphysical conclusions of the project.

1.1 Taxonomical Implications

It should be made explicit that the fact that this project proceeded by way of a domain specific analysis does not automatically commit functional pluralism to any particular philosophical positions regarding the nature of psychological domains themselves and how we ought to distinguish them. The categories of psychological phenomena explored here (i.e. vision, emotion, and biased social cognition) should be taken merely as guiding, and certainly revisable, taxonomical assumptions about the different kinds of conscious experiences that we need to account for in our explanations. Indeed, vision, emotion and bias are all intricately intertwined in many different complex psychological processes. We saw, for instance, that some visual representations carry emotional information that evokes biased response in certain social contexts (Amodio 2014). Teasing apart clear functional boundaries of each "domain" might seem like a hopeless enterprise in such cases. Indeed, it might turn out at the very least that upon closer scrutiny, these categories require substantial revision in order to maintain their usefulness for grounding important generalizations in the sciences. It might even eventually make sense to abandon domain talk altogether, and centre such analyses more precisely around the categories of psychological tasks being performed, reflecting instead a sort of "task ontology" (Burnston forthcoming). Still, in addition to being instrumentally useful for organizing scientific inquiry, there is a deeper reason that these particular domains emerge as relatively stable categorical distinctions in the cognitive sciences. Namely, these domains emerge on epistemological grounds because they reflect shared clusters of (e.g. functional) properties underlying the psychological phenomena that they delineate taxonomically. That is, there is clearly enough functionally (and structurally) similar about visual, emotional, and socially-biased psychological processes that warrants at least some degree of independent theoretical and experimental treatment.

It should be no surprise, then, that an exploration of the structures and functions associated with consciousness will in some ways remain divided along similar taxonomical boundaries. If some psychological functions reliably cluster in these domains, and consciousness contributes something significant to this set of functions, then it follows that the functions of consciousness will also plausibly cluster in these ways. For example, some of the functions underlying visual perception that consciousness contributes to, such as increased spatiotemporal resolution or capacities for semantic processing, cannot be coherently understood to operate in the inhibition of a biased emotional response; this is part of the motivation for adopting functional pluralism. Thus, we will get a natural categorical distinction here between FCCs that preserves existing taxonomical boundaries. We might predict that organisms with vastly different processing requirements altogether than humans (e.g. perhaps as a result of being endowed with utterly different perceptual systems, like echolocation or tentacles) will exhibit some FCCs that cluster into their relevant domain specific categories, as these different processing requirements involve different nested functional components.

It might also still turn out that within these commonly delineated domains we see important convergences of function, such that what seemed like distinct FCCs are actually best understood as different manifestations of one and the same function. For example, there might be sufficient grounds for eventually grouping FCCs like emotional inhibition and flexible emotional response under some broader category; either something like emotional regulation or some sort of “decoupling” operation that overrides unconsciously deployed patterns of response. More work needs to be done in terms of carefully spelling out the particular component functions at play in each case in order to make the comparative case for these sorts of taxonomic unifications. Achieving the most appropriate “grain” of functional explanation is an ongoing project, and I argue that progress here will require that we continue to survey and compare the functional capacities of conscious versus unconscious processing in as wide a range of particular tasks as possible. The better picture we have of the wide range of FCCs and their specific functional makeup, the better position we will be in to map the conceptual relationships among them, grouping them under more abstract categories as necessary for explanation.

Note, however, that there are also already some cases where these traditional taxonomical lines have become blurred in the search for FCCs. In other words, we already have evidence of important convergences of function not only within these commonly delineated domains, but across them as well. Most saliently, the increase in spatiotemporal resolution of one’s representational capacities, which repeatedly emerges as a stable functional marker of consciousness, operates in both paradigmatically visual (i.e. object recognition, motion tracking) and emotional (i.e. threat response) processing tasks. This is obviously because we study emotion by way of valenced visual perception, but the implication remains that this FCC is important both for visual perception in general and for the perception of emotionally relevant objects and events in particular, suggesting a singular FCC at work. It would be fruitful to explore the possibility of this kind of convergence across perceptual modalities in future research, as increases in representational resolution in particular seems to have plausible analogs in auditory perception, somatosensory perception, olfaction, gustatory perception, vestibular perception, and proprioception. Another obvious cross-domain convergence appears between FCCs underlying emotional regulation and those underlying certain strategies for bias intervention, when the target biases are driven by emotional content (i.e. prejudice) and hence the corresponding bias intervention strategy involves emotional regulation. For instance, pre-emptive goal setting might be best understood as involving the flexible response to emotional content (i.e. overriding negative emotional reactions to members of certain demographic categories and their resulting behaviours) suggesting the same basic functional mechanisms are at play in each case. The conceptualization of emotional content thought to require consciousness also probably reflects a more general FCC regarding the ability to extract more detailed (either emotionally relevant or irrelevant) information from the objects of perception. This FCC also likely functions in bias intervention strategies that require emotional reappraisal, when such appraisal involves conceptual nuancing negatively valenced representations in an attempt to mitigate their influence. Finally, increased capacities for spatiotemporal integration were linked with conscious experience both on paradigmatic visual processing tasks and in counterconditioning paradigms, suggesting that consciousness facilitates a more domain general capacity for informational integration over certain spatiotemporal thresholds.

Still there are already enough divergences to secure the general functional pluralist model on offer. We already saw this with spatiotemporal resolution, which is inapplicable in psychological tasks like pre-emptive goal setting that do not require fine grained perceptual discrimination, securing traditional taxonomical assumptions about the relevant psychological domains. Increased capacities for semantic processing of visual stimuli also seem to lack an analogue in cases of emotion regulation or bias intervention. Likewise, the conceptualization of emotional content lacks a clear analogue in many cases of bias intervention. Consider also the notion of “flexibility”, which has long been thought of as a function that is closely associated with consciousness (Earl 2014). Flexibility, understood as an FCC in emotional processing and likely the operative FCC in some bias intervention strategies, might taxonomically subsume a number of functions (e.g. perhaps a range of executive functions like attention, executive control, etc.) so long as they share the relevant causal properties. However, it is generally agreed that visual perception is not appropriately described as involving “flexible processing”, presumably because producing relatively stable representations of the visual world is extremely valuable for ongoing perceptual-motor processing. This suggests that flexibility is one FCC among many others, including various enhanced visual functions like semantic and spatiotemporal processing that aren’t appropriately described as involving flexibility.

All of this reflects the need for ongoing work on such taxonomical issues, as candidate FCCs might either a) cluster into the domain specific categories commonly delineated, b) cross these taxonomical boundaries and thus be domain general to some degree, or c) pick out even further fine-grained taxonomical distinctions within the existing domains of inquiry. It will be exciting to see where the results of continued careful analysis of FCCs takes us on these issues, but much more needs to be done before any sharp taxonomical boundaries are carved out. Thus, I urge that this is where the experimental science focus in the future; namely on employing comparative paradigms to test the association or dissociation of existing candidate FCCs, and ultimately on getting clearer on the underlying functional details such that a robust taxonomy of FCCs can be established.

However, as the functions of consciousness cluster into their appropriate taxonomical categories, it will inevitably be asked whether consciousness itself will remain as a unified construct after such an analysis matures, given the diversity of functions likely to be identifiable across the natural world. In other words, is there compelling grounds for unifying consciousness as a psychological construct? More generally, can a phenomenon with such diverse functions remain conceptually unified? I think this situation is ubiquitous. To employ a trivial analogy, although scissors may be realized by many different configurations of materials in the natural world, the two sharp blades characteristic of scissors have multifunctional applications (e.g.

cutting, stabbing, curling ribbons on presents). I do think that both the science and philosophy of consciousness (as well as their mutually informative relationships) have continued to vindicate the idea that we are in fact collectively investigating a single psychological property, albeit a property with multiple functional applications. It has always been relatively clear that the general object of inquiry here is a property of psychological systems captured by pretheoretical terms like ‘experience’ or ‘awareness’. Indeed, part of what makes consciousness so philosophically mysterious is the fact that we are simultaneously fully acquainted with what it is as an ever-present feature of the world, and yet baffled about how to make sense of it theoretically and experimentally. But at least on conceptual grounds, consciousness as a whole can still logically be lumped and not split taxonomically, even if the functions of consciousness might cluster into relatively discrete taxonomic categories. Of course, further domain-specific analyses are needed in order to either confirm this general picture of consciousness, or else force divisions among different conscious phenomena previously unified under a single construct based on theoretically significant differences in function.

The sort of domain-specific analysis offered here should ultimately deliver an importantly nuanced account of the functional characteristics of conscious psychological processing. In some sense, after this analysis we are also now a bit clearer with regards to what is being picked out as a target of inquiry in consciousness research; namely, what consciousness *is*, such that it has these functional characteristics. What more can we say about consciousness as a psychological property, such that it seems to support the range of different information processing functions that it does? In other words, why are the functions that we have identified uniquely associated with consciousness as a property of complex systems? Thus, these taxonomical issues naturally lead us directly to some crucial metaphysical considerations.

1.2 Metaphysical Implications

One of the central motivations behind this project is the desire to proceed with caution when it comes to devising wholesale theories about the nature of consciousness, until we have a more accurate picture of the range of different structural and functional properties associated with it in the natural world. In this vein, Miracchi (2019) suggests that research into the naturalistic bases of consciousness is in fact hindered when it is influenced by deeper metaphysical debates about the ultimate structure of the world. Saying what consciousness *is* as a general property of the natural world has been notoriously difficult, and it should be stressed that the preceding analysis should be compatible with an incredibly wide range of metaphysical commitments. That is, methodically tracking relationships of association and dissociation between conscious experience and certain structural-functional properties of the system could be meaningful even on an epiphenomenalist picture. But the story that seems most coherent here is the broadly naturalistic one, given that many of the FCCs and NCCs already discovered were revealed based largely on the causal manipulation of conscious experience as an experimental variable (e.g. CFS, pathology). This suggests, at the very least, that consciousness is a property importantly related to natural systems, bearing in mind that understanding the nature and relationships between different properties at different levels of organization in complex systems is no straightforward matter (Wimsatt 1994).

But more can be said beyond this reaffirmation of a broadly naturalistic framework, about the nature of consciousness as a property of psychological systems. After challenging monolithic theories of consciousness and its functions and instead starting from the assumption of functional pluralism, the picture that emerges might aptly be called something like a “Local Workspace Theory” of human consciousness:

> LWT: conscious experience is realized by local (versus global) recurrent processing circuits (between a currently unknown subset of neural structures) that selectively amplify and stabilize information in support of local (versus global) processing needs (e.g. mapping objects in a visual environment, regulating emotion, overcoming bias), which ultimately yields a variety of local (i.e. domain or task specific) markers that can be employed as proxies for experience in future psychological and neuroscientific research.

One interesting implication of this model is that the richness of conscious experience, both in terms of the quantity and quality of consciously represented information, likely reflects the consolidation of multiple local amplifications and stabilizations, which is compatible with some of the central principles of IIT (e.g. Tononi et al. 2016). That is, what we consciously experience at any one time is likely the result of a wide array of dissociable local processing networks, that become in some sense representationally unified. A similar "localist" approach is starting to surface in the search for NCCs as well, lending further credence to the overall framework. Malach (2021) in particular, has recently advocated for a localist perspective to the neuroscience of consciousness, according to which consciousness should be investigated by looking at the contributions of local neuronal relational structures in the cortex to the content of experience, rather than to global network activities that conscious processing only sometimes participates in. It should be noted that this model of consciousness is also entirely compatible with the idea that higher functions like access or attention play a further role in modulating information required for a given task, perhaps even similarly described as amplifying and/or stabilizing (see Fazekas and Nanay 2021). On this picture, the precise role of these kinds of executive functions is to selectively process information that has been amplified and stabilized by conscious (e.g. visual, emotional) processes, like a diner choosing the tastiest looking meal from a well curated menu.

This LWT formulation of the results reinforces the claim that the functional pluralist picture is compatible with consciousness being a unified construct. In essence, local processing networks and the different conscious psychological processes that they support share some structural-functional properties, beyond the working target of "experience", which only became salient after a careful analysis of consciousness' diverse manifestations. One such structural-functional property that has continued to be implicated across each specific analysis can be summed up as the unfortunately but necessarily vague *representational complexity* afforded by conscious processes and the localized recurrent processing networks that seem to be necessary for realizing them. This notion of representational complexity is necessarily vague precisely because the more distinct particular phenomena that reflect the general constituents of the construct, the less precise it can be in specifying the details of those particulars. But some form of representational complexity seems to be associated with consciousness' diverse functional capacities, such as spatiotemporal integration in vision and bias intervention strategies like pre-emptive goal setting. And yet, the functions that this property enables, and the kinds of structures that realize them, are causally distinct enough to suggest that the functional pluralist approach is the right way to proceed. These are to be taken as preliminary conclusions, and much more comparative analysis is required in order to assess the association between representational complexity and conscious experience. Once again it might turn out that only a certain kind or degree of representational complexity is necessary and/or sufficient for consciousness. More work is also needed in order to explain what it is about this structural-functional property such that consciousness is always recruited when representational demands reach a certain level of complexity. Once again, these metaphysical points ought to become clearer and clearer as we pursue FCCs in a functional pluralist framework.

One objection to this account is that no clear demarcating line has been offered that can distinguish conscious from unconscious processing in terms of functional capacity, and ultimately representational complexity. But this is to be expected given how dynamic and individualistic psychological systems are. Rather than being a disappointing conclusion, the fact that the conclusion here is stated in terms of a matter of degree (i.e. of representational complexity) tells us something crucial about the relationship between conscious and unconscious processing. Many of the functional capacities investigated in detail here—including access and integration more generally, as well as spatiotemporal and semantic functions in vision, and conceptual and regulatory functions in emotion more specifically—are notions that admit of degrees of informational or representational complexity in their own way. The upshot is that consciousness appears to enhance capacities that are present and can be carried out to some lesser degree unconsciously, suggesting the continuity of many psychological processes across the conscious/unconscious divide. It is also illuminating to recognize that the functional differences between conscious an unconscious processing admit of individual differences as the result of unique contextual factors. This suggests some adaptability in the functional capacities of conscious versus unconscious processing, which lends itself to a range of potentially illuminating experimental investigations.

Finally, it might be argued that the notion of representational complexity is merely synonymous with, or at least plays the same conceptual role as, information integration as defined by IIT. After all, it is also a notion that is taken to admit of degrees, and involves constructing more elaborate informational structures out of simpler ones. One thing to say here first is that information integration is identified with conscious experience according to IIT, and so this is a construct that, as it is currently wielded, cannot account for psychological continuity across the conscious and unconscious divide. Of course, IIT could simply acknowledge that unconscious processes are capable of (albeit less sophisticated) forms of information integration, although this would require the denial of some of the core tenets of the theory. But beyond the fact that we need a notion that is more widely assumed to capture something shared across conscious and unconscious phenomena, integration does not adequately capture all the functions picked out in our domain specific analyses, which is what we want our theoretical abstractions to do. In other words, some forms of representational complexity (e.g. spatiotemporal resolution) are not appropriately described as involving integration of information in any psychologically significant sense, and we ought to leave open the conceptual space for other dimensions of complexity not captured by information integration. It should to be made clear that *some* degree of information integration remains closely associated with consciousness, as *some* FCCs are characterized by this functional feature (e.g. spatiotemporal integration, constructing vivid counterstereotypes for bias intervention).

The resulting claim about representational complexity at least partially characterizing the nature of consciousness therefore might seem unsatisfyingly vague in its generality (although the same could be said for constructs like global access and information integration). But recall that it is grounded on a fairly detailed general account of the structural and functional properties shared across manifestations of conscious experience (i.e. Local Workspace Theory). I maintain that the more we focus on fleshing out the functional pluralist picture, the better we will be able to articulate these kinds of general claims about the nature of consciousness. Hopefully the fruitfulness of the overarching project is apparent. The search for FCCs (integrated with the search for NCCS) should continue to be extremely valuable for establishing viable markers that can be used as operationalizable proxies for consciousness in ongoing interdisciplinary research.

Chapter 1 References

Aizawa, K. (2017). Multiple Realization, Autonomy and Integration. In *Explanation and Integration in Mind and Brain Science*, ed. David M. Kaplan. Oxford University Press.

Ajina, S., & Bridge, H. (2016). Blindsight and Unconscious Vision: What They Teach Us about the Human Visual System. *The Neuroscientist: a review journal bringing neurobiology, neurology and psychiatry*, 23(5), 529–541.

Alexander, I., & Cowey, A. (2010). "Edges, Colour and Awareness in Blindsight," *Consciousness and Cognition* 19: 520-533.

Amodio, D. (2014). The Neuroscience of Prejudice and Stereotyping. *Nature Reviews Neuroscience*, Volume 15, 670–682.

Anderson, M. (2014). *After Phrenology: Neural Reuse and the Interactive Brain*. The MIT Press.

Arnold, D.H. (2005) Perceptual pairing of colour and motion. *Vision Res*. 45, 3015–3026.

Aru, J., Bachmann, T., Singer, W. & Melloni, L. (2012). Distilling the neural correlates of consciousness. *Neuroscience and Biobehavioral Reviews*, 36: 737-746.

Augusto, L. M. (2013). Unconscious Representations 1: Belying the Traditional Model of Human Cognition. *Axiomathes*, 23(4), 645-663.

Auksztulewicz, R., Spitzer, B. & Blankenburg, F. (2012). Recurrent Neural Processing and Somatosensory Awareness. *The Journal of Neuroscience*, 32 (3), 799-805.

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

Barack, David L. (2016). Cognitive Recycling. *British Journal for the Philosophy of Science*, 024.

Barrett, L. F. & Wager, T. D. (2006). The structure of emotion: evidence from neuroimaging studies. *Current Directions in Psychological Science*, Volume 15, Number 2, 79-83.

Beyeler, A. Chang, C-J., Silvestre, M. Leveque, C., Namburi, P., Wildes, C. P., & Tye, K. M. (2018). Organization of valence-encoding and projection-defined neurons in the basolateral amygdala. *Cell Reports*, 22: 905-918.

Birch, J. (2020). The Search for Invertebrate Consciousness. *Noûs*.

Block, N. (1995). On a confusion about a function of consciousness. *Brain and Behavioral Sciences* 18 (2):227-–247.

Block, N. (2010). Attention and Mental Paint. *Philosophical Issues*, 20: 23-63.

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*, 445-447.

Block, N. (2019a). What Is Wrong with the No-Report Paradigm and How to Fix It. *Trends in Cognitive Sciences*, 23 (12):1003-1013.

Block, N. (2019b). Empirical Science Meets Higher-Order Views of Consciousness: Reply to Hakwan Lau and Richard Brown. In *Blockheads!: essays on Ned Block's philosophy of mind and consciousness*, eds Adam Pautz and Daniel Stoljar, Cambridge, MA: MIT Press.

Block, N. & Philips, I. (2016). Unconscious Seeing. *Current Controversies in Philosophy of Perception*, Routledge.

Bourget, D. & Mendelovici, A. (2014). Tracking Representationalism. In Andrew Bailey (ed.), Philosophy of Mind: The Key Thinkers. Continuum. pp. 209-235.

Boyd, R. (1999). Homeostasis, species, and higher taxa. In R. A. Wilson (ed.), *Species: New Interdisciplinary Essays*. MIT Press. pp. 141-85.

Boly, M., Massimini, M., Tsuchiya, N., Postle, B. R., Koch, C., & Tononi, G. (2017). Are the neural correlates of consciousness in the front or in the back of the cerebral cortex? Clinical and neuroimaging evidence. *The Journal of Neuroscience*, 37(40): 9603-9613.

Boyle, A. (2020). The impure phenomenology of episodic memory. *Mind and Language* 35 (5):641-660.

Brogaard, B., Chomanski, B. & Gatzia, D.E. (2021). Consciousness and information integration. *Synthese* 198, 763–792.

Bronfman, Z. Z., Brezis, N., Jacobson, H., & Usher, M. (2014). We See More Than We Can Report: "Cost Free" Color Phenomenality Outside Focal Attention. *Psychological Science*, *25*(7), 1394–1403.

Brown, R., Lau, H., & LeDoux, J. E. (2019). Understanding the higher-order approach to consciousness. *Trends in Cognitive Sciences*, 23(9): 754-768.

Chalmers, D. J. (1995). Facing up to the problem of consciousness. *Journal of Consciousness Studies*, 2 (3):200-19.

Cleeremans, A., Achoui, D., Beauny, A., Keuninckx, L., Martin, J.-R., Munoz-Moldes, S., Vuillaume, L., & de Heering, A. (2020). Learning to be Conscious. *Trends in Cognitive Sciences*, 24(2): 112-123.

Cohen, M. A. & Dennett, D. C. (2011). Consciousness cannot be separated from function. *Trends in Cognitive Sciences*, Vol. 15, No. 8.

Crawford, J. D., Henriques, D. Y. P., & Medendorp, W. P. (2011). Three-dimensional Transformations for Goal-Directed Action. *Annual Review of Neuroscience*, 34:1, 309-331.

Cummins, R. (1975). Functional analysis. *Journal of Philosophy*, 72(20), 741–65.

Cummins, R. (1983). *The Nature of Psychological Explanation*. MIT Press

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

Dehaene, S., Changeux, J-P., Naccache, L., Sackur, J., & Sergent, C. (2006). Conscious, preconscious, and subliminal processing: a testable taxonomy. *Trends in Cognitive Sciences*, Vol. 10, No. 5, 204-211.

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

Dehaene, S. (2014). *Consciousness and the brain: Deciphering how the brain codes our thoughts*. Viking.

Diano, M., Celeghin, A., Bagnis, A., & Tamietto, M. (2017). Amygdala Response to Emotional Stimuli without Awareness: Facts and Interpretations. *Frontiers in Psychology*. Volume 7, 2029.

Douglas, R. J. & Martin, K. A. C. (2007). Recurrent neuronal circuits in the neocortex. *Current Biology*, 17 (13).

Dretske, F. (1988). *Explaining behavior: Reasons in a world of causes*. MIT Press.

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

Earl, B. (2014). The biological function of consciousness. *Frontiers in Psychology, 5,* Article 697.

Enns, J. T., & Di Lollo, V. (2000). What's new in visual masking? *Trends in Cognitive Sciences*, 4(9): 345-352.

Faivre, N., Mudrik, L., Schwartz, N., & Koch, C. (2014). Multisensory Integration in Complete Unawareness: Evidence From Audiovisual Congruency Priming. *Psychological Science*, *25*(11), 2006–2016.

Fodor, J. A. (1974). Special sciences (or: The disunity of science as a working hypothesis). *Synthese* 28 (2):97-115.

Frith, C. D. & Metzinger, T. (2016). What's the use of consciousness? How the stab of conscience made us really conscious. In: *The Pragmatic Turn: Toward Action-Oriented Views in Cognitive Science,* ed. A. K. Engel, K. J. Friston, and D. Kragic. Strüngmann Forum Reports, vol. 18, J. Lupp, series editor. Cambridge, MA: MIT Press.

Gennaro, R. J. (2018). Higher-Order Theories of Consciousness. *Internet Encyclopedia of Philosophy*.

Godfrey-Smith, P. (2017). The Evolution of Consciousness in Phylogenetic Context. *The Routledge Handbook of Philosophy of Other Minds*, ed. Kristin Andrews and Jacob Beck Routledge.

Griffiths, P. E. (1993). Functional analysis and proper functions. *British Journal for the Philosophy of Science*, 44 (3):409-422.

Gross, S., & Flombaum, J. (2017). Does perceptual consciousness overflow cognitive access? The challenge from probabilistic, hierarchical processes. *Mind & Language, 32*(3), 358–391.

Harman, G., 1990. The Intrinsic Quality of Experience, in *Tomberlin*; reprinted in Lycan & Prinz 2008.

Hassin, R. R., Ferguson, M. J., Shidlovski, D., & Gross, T. (2007). Subliminal exposure to national flags affects political thought and behaviour. *PNAS*, Vol. 104, No. 50, 19757-19761.

Healy, Graham F., Boran, L. & Smeaton, A.F. (2015). Neural Patterns of the Implicit Association Test. *Frontiers of Human Neuroscience*. Vol. 24.

Holcombe, A.O. (2009). Seeing slow and seeing fast: Two limits on perception. *Trends in Cognitive Science*, 13(5):216-21.

Husserl, E. (1931/1999). *Cartesian Meditations: An introduction to phenomenology*. Translated by Dorion Cairns. Kluwer Academic Publishers.

Khalidi, M. A. (2013). *Natural Categories and Human Kinds: Classification in the Natural and Social Sciences*. Cambridge: Cambridge University Press.

Khalidi, M. A. (2017). Crosscutting psycho-neural taxonomies: the case of episodic memory. *Philosophical Explorations*, Vol. 20, No. 2, 191-208.

Kiesel, A., Knude, W., & Hoffman, J. (2007). Mechanisms of subliminal response priming. *Advances in Cognitive Psychology*, Vol. 3, No. 1-2, 307-315.

Kim, J. (1992). Multiple Realization and the Metaphysics of Reduction. Philosophy and Phenomenological Research, 52(1), 1-26.

Koch, C. (2019). *The Feeling of Life Itself: Why Consciousness is Widespread but Can't Be Computed.* MIT Press, Cambridge.

Koch, C. & Tsuchiya, N. (2007). Attention and Consciousness: Two Distinct Brain Processes. *Trends in Cognitive Sciences*, 11(1): 16-22.

Koch, C., Massimini, M., Boly, M. Tononi, G. (2016). Neural correlates of consciousness: progress and problems. *Nature Reviews Neuroscience*, 17, pages 307–321.

Kouider, S., & Dehaene, S. (2007). Levels of processing during non-conscious perception: a critical review of visual masking. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *362*(1481), 857–875.

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

Kouider, S. & Faivre, N. (2017). Conscious and Unconscious Perception. In *The Blackwell Companion to Consciousness* (eds S. Schneider and M. Velmans).

Kriegel, U. (2002). PANIC Theory and the Prospects for a Representational Theory of Phenomenal Consciousness. *Philosophical Psychology*, 15: 55–64.

Lamme, Victor A. F. (2006). Towards a True Neural Stance on Consciousness, *Trends in Cognitive Sciences*, 10(11): 494–501.

Lamme, V. A. F., & Roelfsema, P. R. (2000). The Distinct Modes of Vision Offered by Feedforward and Recurrent Processing. Trends in Neurosciences, 23, 571-579.

Lamme, V. (2020). Visual Functions Generating Conscious Seeing. *Frontiers in psychology*, *11*, 83.

Lau, H. C., & Passingham, R. E. (2007). Unconscious activation of the cognitive control system in the human prefrontal cortex. *The Journal of neuroscience: the official journal of the Society for Neuroscience*, *27*(21), 5805–5811.

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

Lau, H. C. & Brown, R. (2019). The Emperor's New Phenomenology? The Empirical Case for Conscious Experiences without First-Order Representations. In *Blockheads!: essays on Ned Block's philosophy of mind and consciousness*, eds Adam Pautz and Daniel Stoljar, Cambridge, MA: MIT Press.

LeDoux, J. & Daw, N. D. (2018). Surviving threats: neural circuit and computational implications of a new taxonomy of defensive behaviour. *Nature Reviews Neuroscience*, 19: 269-282.

Lycan, W. G. (2001). The case for phenomenal externalism. *Philosophical Perspectives*, 15: 17-35.

Lycan, W. G. (2019). Block and the representational theory of sensory qualities. In *Blockheads!: essays on Ned Block's philosophy of mind and consciousness*, eds Adam Pautz and Daniel Stoljar, Cambridge, MA: MIT Press.

Mahon, B. Z. (2015). Missed Connections: A Connectivity-Constrained Account of the Representation and Organization of Object Concepts. In *The Conceptual Mind*, eds. Eric Margolis and Stephen Laurence, MIT Press.

Maley, C. J. & Piccinini, G. (2017). A Unified Mechanistic Account of Teleological Functions for Psychology and Neuroscience. In *Explanation and Integration in Mind and Brain Science*, ed. David M. Kaplan, Oxford University Press.

Mattiassi, A. D., Mele, S., Ticini, L. F. & Urgesi, C. (2014). Conscious and unconscious representations of observed action in the human motor system. *Journal of Cognitive Neuroscience*, 26(9), 2028-2041.

Mele, A. (2001). *Self-Deception Unmasked*. Princeton University Press.

Merleau-Ponty, M. (1962/2001). *The phenomenology of perception*. London: Routledge.

Millikan, Ruth G. (1984). *Language, Thought, and Other Biological Categories*. MIT Press.

Miracchi, L. (2017), Generative explanation in cognitive science and the hard problem of consciousness. *Philosophical Perspectives*, 31: 267-291.

Misselhorn, J., Schwab, B. C., Schneider, T. R., & Engel, A. K. (2019). Sychronization of sensory gamma oscillations promotes multisensory communication. *eNeuro* 6(5).

Mudrik, L., & Koch, C. (2013). Differential processing of invisible congruent and incongruent scenes: A case for unconscious integration. *Journal of Vision*. 13(13), 1-14.

Mudrik, L., Faivre, N., & Koch, C. (2014). Information integration without awareness. *Trends in cognitive sciences*. Volume 18, Issue 9, p. 488-496.

Neander, K. (1991). The teleological notion of 'function'. *Australasian Journal of Philosophy*, 69:4, 454-468.

Neri, P., & Levi, D. M. (2006). Spatial resolution for feature binding is impaired in peripheral and amblyopic vision. *Journal of Neurophysiology, 96*(1), 142–153.

Odegaard, B., Chang, M.Y., Lau, H.C., & Cheung, S. (2018). Inflation versus filling-in: why we feel we see more than we actually do in peripheral vision. *Philosophical Transactions of the Royal Society B: Biological Sciences, 373*.

Oizumi M., Albantakis L., Tononi G. (2014) From the Phenomenology to the Mechanisms of Consciousness: Integrated Information Theory 3.0. *PLOS Computational Biology*, 10(5).

Oswald, F. L., Mitchell, G., Blanton, H., Jaccard, J., & Tetlock, P. E. (2013). Predicting ethnic and racial discrimination: A meta-analysis of IAT criterion studies. *Journal of Personality and Social Psychology*, 105(2), 171–192.

Phillips, I. (2011). Perception and Iconic Memory: What Sperling Doesn't Show. *Mind and Language*. 26 (4):381-411.

Poldrack, R. A. (2020). The physics of representation. *Synthese*:1-19.

Poldrack, R. A., & Yarkoni, T. (2016). From brain maps to cognitive ontologies: informatics and the search for mental structure. *Annual review of psychology*, *67*, 587-612.

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

Putnam, H. (1975). The meaning of 'meaning'. *Minnesota Studies in the Philosophy of Science* 7:131-193.

Pynn, L. K. & DeSouza, J. F. X. (2013). The function of efference copy signals: Implications for symptoms of schizophrenia. *Vision Research*, 76(14), 124-133.

Quilty-Dunn, J. (2020). Is Iconic Memory Iconic? *Philosophy and Phenomenological Research* 101 (3):660-682.

Robinson, W. (2019). *Epiphenomenal mind: An integrated outlook on sensations, beliefs, and pleasure*. New York: Routledge.

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

Salomon, R., Noel, J-P., Lukowska, M., Faivre, N., Metzinger, T., Serino, A. & Blanke, O. (2017). Unconscious integration of multisensory bodily inputs in the peripersonal space shapes bodily self-consciousness. *Cognition*, 166: 174-183.

Sartre, J-P. (1940/2004). *The Imaginary: A phenomenological psychology of the imagination*. New York: Routledge.

Schwitzgebel, E. (2016). Phenomenal Consciousness, Defined and Defended as Innocently as I Can Manage. *Journal of Consciousness Studies* 23 (11-12): 224-235.

Seth, A. (2009). Functions of Consciousness. In Banks, P W. (ed.) (2009). *Encyclopedia of Consciousness: A - L*. Elsevier.

Shoemaker, S. (1994). Phenomenal Character. *Noûs*, 28: 21–38.

Siewert, C. (2013). Phenomenality and Self-Consciousness. In Uriah Kriegel (ed.), *Phenomenal Intentionality*. Oxford University Press.

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

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

Stoljar, D. (2019). In Praise of Poise. In *Blockheads!: essays on Ned Block's philosophy of mind and consciousness*, eds Adam Pautz and Daniel Stoljar, Cambridge, MA: MIT Press.

Storm, J. F., Boly, M., Casali, A. G., Massimini, M., Olcese, U., Pennartz, C. M. A., & Wilke, M. (2017). Consciousness Regained: Disentangling Mechanisms, Brain Systems, and Behavioral Responses. *The Journal of Neuroscience*, 37 (45): 10882-10893.

Strawson, G. (2011). Cognitive phenomenology: real life. In Tim Bayne & Michelle Montague (eds.), *Cognitive Phenomenology*. Oxford University Press. pp. 285--325.

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

Tononi, G. (2012a). The Integrated Information Theory of Consciousness: An Updated Account. *Archives Italiennes de Biologie,* 150: 290-326

Tononi, G. (2012b). *Phi: A Voyage from the Brain to the Soul*. New York: Pantheon Books.

Tononi, G., Boly, M., Massimini, M. & Koch, C. (2016). Integrated information theory: from consciousness to its physical substrate. *Nat Rev Neurosci* 17, 450–461.

Tononi, G. & Koch, C. (2015). Consciousness: here, there, and everywhere? *Phil. Trans. R. Soc. B* 370, 20140167.

Treisman, A. (2006). How the deployment of attention determines what we see. *Visual Cognition, 14*(4–8), 411–443.

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

Tyler, C. (2015). Peripheral color demo. *i-Perception* 6(6): 1-5.

Usher, M., Bronfman, Z. Z., Talmor, S., Jacobson, H., & Eitam, B. (2018). Consciousness without report: insights from summary statistics and inattention 'blindness'. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *373*(1755), 20170354.

van Gaal, Simon & Lamme, Victor A. F. (2012). Unconscious High-Level Information Processing: Implication for Neurobiological Theories of Consciousness. *The Neuroscientist*, 18(3) 287–301.

Wimsatt, W. C. (1994). The ontology of complex systems: levels of organization, perspectives, and causal thickets. *Canadian Journal of Philosophy* 20 (sup1): 207-274.

Woodward, J. (2017). Explanation in Neurobiology: an Interventionist Perspective. In *Explanation and Integration in Mind and Brain Science*, ed. David M. Kaplan, Oxford University Press.

Wu, W. (2017). Shaking up the Ground Floor: The Cognitive Penetration of Visual Attention. *The Journal of Philosophy*. Volume CXIV, No. 1, 5-32.

Wyart, V., & Tallon-Baudry, C. (2008). Neural dissociation between visual awareness and spatial attention. *The Journal of Neuroscience*, 28 10, 2667-79.

Zhou, L., Deng, C.L., Ooi, T.L., & He, Z.J. (2016). Attention modulates perception of visual space. *Nature Human Behaviour, 1*.

Chapter 2 References

Alexander, I., & Cowey, A. (2010). "Edges, Colour and Awareness in Blindsight," *Consciousness and Cognition* 19: 520-533.

Altman, D. G., and Bland, J. M. (1995). Absence of evidence is not evidence of absence. *BMJ* 311: 485.

Anderson, M. (2014). *After Phrenology: Neural Reuse and the Interactive Brain*. The MIT Press.

Armstrong, A.-M., & Dienes, Z. (2013). Subliminal understanding of negation: Unconscious control by subliminal processing of word pairs. *Consciousness and Cognition*, 22, 1022–1040.

Arzi, A. et al. (2012) Humans can learn new information during sleep. *Nat. Neurosci*. 15, 1460–1465

Axelrod, V., & Rees, G. (2014). Conscious awareness is required for holistic face processing. *Consciousness & Cognition,* 27: 233-245.

Axelrod, V., Bar, M., Rees, G., & Yovel, G. (2015). Neural correlates of subliminal language processing. *Cerebral Cortex*, 25(8), 2160–2169.

Bargh, J. A., and Morsella, E. (2008). The unconscious mind. *Perspect. Psychol. Sci*. 3, 73–79.

Barack, David L. (2019). Cognitive Recycling. *British Journal for the Philosophy of Science*. 70 (1):239-268.

Bednar, J. A., & Wilson, S. P. (2016). Cortical Maps. *The Neuroscientist*, Vol. 22, No. 6, pp. 604-617.

Berti, A. & Rizzolatti, G. (1992). Visual Processing without Awareness: Evidence from Unilateral Neglect. *Journal of Cognitive Neuroscience*, 4(4).

Binder, JR. (2016) In defense of abstract conceptual representations. *Psychon. Bull. Rev*. 23: 1096 – 1108.

Birch, J. (2020). The search for invertebrate consciousness. *Noûs*. 1– 21.

Bisiach, E. & Luzzatti, C. (1978). Unilateral neglect of representational space. *Cortex*, 14: 129-133.

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*, 445-447.

Block N., & Phillips I. (2016). Debate on Unconscious Perception. In: Nanay B (ed.), Current Controversies in Philosophy of Perception. New York: Routledge: Taylor & Francis Group.

Block, N. (2019). What Is Wrong with the No-Report Paradigm and How to Fix It. *Trends in Cognitive Sciences*, 23 (12):1003-1013.

Boly, M., Massimini, M., Tsuchiya, N., Postle, B. R., Koch, C., & Tononi, G. (2017). Are the Neural Correlates of Consciousness in the Front or in the Back of the Cerebral Cortex? Clinical and Neuroimaging Evidence. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *37*(40), 9603–9613.

Bonner, M. F., & Price, A. R. (2013). Where is the Anterior Temporal Lobe and What Does it Do? *The Journal of Neuroscience*, 33 (10): 4213-4215.

Bourget, D. (2010). Consciousness is underived intentionality. *Noûs*, 44(1): 32–58.

Born, R. T., & Bradley, D. C. (2005). Structure and function of visual area MT. *Annual Review of Neuroscience*, 28: 157-189.

Boyle, A. (2019) The impure phenomenology of episodic memory. *Mind & Language*. 1– 20.

Brascamp, J. et al. (2015) Negligible fronto-parietal BOLD activity accompanying unreportable switches in bistable perception. Nat. Neurosci. 18, 1672–1678.

Breitmeyer, B. G., & Öğmen, H. (2006). *Oxford psychology series. Visual masking: Time slices through conscious and unconscious vision (2nd ed.).* Oxford University Press.

Breitmeyer, B. (2014). Functional Hierarchy of Unconscious Object Processing. In B. Breitmeyer (Ed.), *The Visual (un)Conscious and its (dis)Contents*. Oxford: Oxford University Press.

Breitmeyer, B. G. (2015). Psychophysical "blinding" methods reveal a functional hierarchy of unconscious visual processing. *Consciousness and Cognition: An International Journal, 35,* 234–250.

Brickner, R.M. (1952) Brain of Patient A after bilateral frontal lobectomy: status of frontal-lobe problem. AMA Arch Neurol Psychiatry 68:293–313. CrossRef Medline

Brogaard, B. (2011). Are There Unconscious Perceptual Processes? *Consciousness and Cognition* 20 (2):449-63.

Brogaard, B., Chomanski, B., & Gatzia, D. E. (2020). Consciousness and information integration. *Synthese*.

Brown, R., Lau, H., & LeDoux, J. E. (2019). Understanding the higher-order approach to consciousness. *Trends in Cognitive Sciences*, 23(9): 754-768.

Buckner, R. L., & Krienen, F. M. (2013). The evolution of distributed association networks in the human brain. *Trends in Cognitive Sciences, 17*(12), 648–665.

Bullier J. (2001). Feedback connections and conscious vision. Trends in cognitive sciences, 5(9), 369–370.

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

Bussche, E.V., Segers, G., & Reynvoet, B. (2008). Conscious and unconscious proportion effects in masked priming. *Consciousness and Cognition, 17*, 1345-1358.

Cabeza, R., Stanley, M. L., & Moscovitch, M. (2018). Process-Specific Alliances (PSAs) in Cognitive Neuroscience. *Trends in cognitive sciences*, *22*(11), 996–1010.

Carrasco, M. (2011). Visual attention: The past 25 years. Vision research, 51(13), 1484-1525.

Carruthers, P. (2020). *Human and Animal Minds: The Consciousness Questions Laid to Rest*. Oxford University Press.

Chen, Lihong ; Qiao, Congying ; Wang, Ying & Jiang, Yi (2018). Subconscious processing reveals dissociable contextual modulations of visual size perception. *Cognition* 180:259-267.

Cohen, M. A. & Dennett, D. C. (2011). Consciousness cannot be separated from function. *Trends in Cognitive Sciences*, Vol. 15, No. 8.

Costello, P., Jiang, Y., Baartman, B., McGlennen, K., & He, S. (2009). Semantic and subword priming during binocular suppression. *Consciousness and cognition*, *18*(2), 375–382.

Cox, E. J., Sperandio, I., Laycock, R., Chouinard, P. A. (2018). Conscious awareness is required for the perceptual discrimination of the threatening animal stimuli: A visual masking and continuous flash suppression study. *Consciousness and Cognition*. 65: 280-292.

Crawford, J. D., Henriques, D. Y. P., & Medendorp, W. P. (2011). Three-dimensional Transformations for Goal-Directed Action. *Annual Review of Neuroscience*, 34:1, 309-331.

Crepaldi, D., Rastle, K., Coltheart, M., & Nickels, L. (2010). "Fell" primes "fall", but does "bell" prime "ball"? Masked priming with irregularly-inflected primes. *Journal of Memory and Language*. 63: 83-99.

Dehaene, S., Kerszberg, M., & Changeux, J. (1998). A Neuronal Model of a Global Workspace in Effortful Cognitive Tasks. *Proceedings of the National Academy of Sciences of the United States of America, 95*(24), 14529-14534.

Dehaene, S., Changeux, J-P., Naccache, L., Sackur, J., & Sergent, C. (2006). Conscious, preconscious, and subliminal processing: a testable taxonomy. *Trends in Cognitive Sciences*, Vol. 10, No. 5, 204-211.

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

Dehaene, S. (2014). *Consciousness and the brain: Deciphering how the brain codes our thoughts*. Viking.

Drhuv, N.T., & Carrandini, M. (2014). Cascaded effects of spatial adaptation in the early visual system. *Neuron*. 81, 529–535.

Diano, M., Celeghin, A., Bagnis, A., & Tamietto, M. (2017). Amygdala response to emotional stimuli without awareness: facts and interpretations. *Frontiers in Psychology*. Volume 7, Article 2029.

Doerig, A., Schurger, A., Hess, K., & Herzog, M. (2019). The unfolding argument: Why IIT and other causal structure theories cannot explain consciousness. *Consciousness and Cognition, 72*, 49-59.

Douglas, R. J. & Martin, K. A. C. (2007). Recurrent neuronal circuits in the neocortex. *Current Biology*, 17 (13).

Drinnenberg, A., Franke, F., Morikawa, R. K., Jüttner, J., Hillier, D., Hantz, P., Hierlemann, A., Azeredo da Silveira, R., & Roska, B. (2018). How Diverse Retinal Functions Arise from Feedback at the First Visual Synapse. *Neuron*, *99*(1), 117–134.

Elgendi, M., Kumar, P., Barbic, S., Howard, N., Abbott, D., & Cichocki, A. (2018). Subliminal priming—state of the art and future perspectives. *Behavioral Sciences*, *8*(6), 54.

Faivre, N. & Koch, C. (2014). Temporal structure coding with and without awareness. *Cognition*. 131 (3):404-414.

Fang, Z., Han, L., Chen, G., and Yang, J. (2016). Unconscious processing of negative animals and objects: role of the amygdala revealed by fMRI. *Frontiers in Human Neuroscience*, 10: 146.

Fahrenfort, J. J., Scholte, H. S., & Lamme, V. A. F. (2007). Masking Disrupts Reentrant Processing in Human Visual Cortex. *Journal of Cognitive Neuroscience*, 19 (9): 1488–1497

Federer, F., Ta'afua, S., Merlin, S., Angelucci, A. (2020). Stream-specific feedback Inputs to the primate primary visual cortex. bioRxiv 2020.03.04.977264; doi: https://doi.org/10.1101/2020.03.04.977264

Fisch, L., Privman, E., Ramot, M., Harel, M., Nir, Y., Kipervasser, S., Andelman, F., Neufeld, M. Y., Kramer, U., Fried, I., & Malach, R. (2009). Neural "ignition": enhanced activation linked to perceptual awareness in human ventral stream visual cortex. *Neuron*, *64*(4), 562–574.

Frässle, S., Sommer, J., Jansen, A., Naber, M., & Einhäuser, W. (2014). Binocular Rivalry: Frontal Activity Relates to Introspection and Action But Not to Perception. *The Journal of Neuroscience, 34*, 1738 - 1747.

Frith, C. D. and Metzinger, T. (2016). What's the use of consciousness? How the stab of conscience made us really conscious. In: *The Pragmatic Turn: Toward Action-Oriented Views in Cognitive Science,* ed. A. K. Engel, K. J. Friston, and D. Kragic. Strüngmann Forum Reports, vol. 18, J. Lupp, series editor. Cambridge, MA: MIT Press.

Gelbard-Sagiv, H., Faivre, N., Mudrik, L., & Koch C. (2016). Low-level awareness accompanies "unconscious" high-level processing during continuous flash suppression. *Journal of Vision*, 16 (1): 3, 1-16.

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

Goodale, M., & Milner, D. (2013). Sight unseen: An exploration of conscious and unconscious vision (2nd ed.). Oxford University Press.

Halligan, P. W., & Marshall, J. C. (1998). Neglect of awareness. *Consciousness and cognition*, *7*(3), 356–380.

Harris, J. J., Schwarzkopf, D. S., Song, C., Bahrami, B., & Rees, G. (2011). Contextual Illusions Reveal the Limit of Unconscious Visual Processing. *Psychological Science*, *22*(3), 399–405.

Hassin, R. R. (2013). Yes It Can: On the Functional Abilities of the Human Unconscious. *Perspectives on Psychological Science*, 8(2), 195–207.

Hesselmann, G., Darcy, N., Ludwig, K., Sterzer, P. (2016). Priming in a shape but not in a category task under continuous flash suppression. *Journal of Vision*. 16.

Hilgetag, C. C., Medalla, M., Beul, S. F., & Barbas, H. (2016). The primate connectome in context: Principles of connections of the cortical visual system. *Neuroimage*, 134: 685-702.

Hubel, D. H. & Wiesel, T. N. (1968) Receptive fields and functional architecture of monkey striate cortex. *J. Physiol*., 195: 215-243.

Hupé, J. M., James, A. C., Payne, B. R., Lomber, S. G., Girard, P., and Bullier, J. (1998). Cortical feedback improves discrimination between figure and background by V1, V2 and V3 neurons. *Nature* 394, 784–787. doi: 10.1038/29537

Hupé, J. M., James, A. C., Girard, P., Lomber, S. G., Payne, B. R., and Bullier, J. (2001). Feedback connections act on the early part of the responses in monkey visual cortex. *J. Neurophysiol.* 85, 134–145.

Jiang, Y., Costello, P., & He, S. (2007). Processing of Invisible Stimuli: Advantage of Upright Faces and Recognizable Words in Overcoming Interocular Suppression. *Psychological Science*, *18*(4), 349–355.

Kandel, E. R. (2012). *The age of insight: The quest to understand the unconscious in art, mind, and brain, from Vienna 1900 to the present. Random House.*

Koch, C., Massimini, M., Boly, M. Tononi, G. (2016). Neural correlates of consciousness: progress and problems. *Nature Reviews Neuroscience*, 17, pages 307–321.

Koivisto, M., Mäntylä, T., & Silvanto, J. (2010). The role of early visual cortex (V1/V2) in conscious and unconscious visual perception. *NeuroImage, 51*(2), 828–834.

Koivisto, M., Kastrati, G., & Revonsuo, A. (2014). Recurrent Processing Enhances Visual Awareness but Is Not Necessary for Fast Categorization of Natural Scenes. *Journal of Cognitive Neuroscience, 26*, 223-231.

Koivisto, M., & Rientamo, E. (2016). Unconscious vision spots the animal but not the dog: Masked priming of natural scenes. Consciousness and Cognition: An International Journal, 41, 10–23.

Koivisto, M., & Grassini, S. (2018). Unconscious response priming during continuous flash suppression. *PLoS ONE* 13(2): e0192201.

Kominsky, Jonathan F. & Scholl, Brian J. (2020). Retinotopic adaptation reveals distinct categories of causal perception. *Cognition* 203:104339.

Korisky, U., Hirschhorn, R. & Mudrik, L. (2019). “Real-life” continuous flash suppression (CFS)-CFS with real-world objects using augmented reality goggles. *Behav Res* 51, 2827–2839.

Kouider, S., & Dehaene, S. (2007). Levels of processing during non-conscious perception: a critical review of visual masking. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *362*(1481), 857–875. https://doi.org/10.1098/rstb.2007.2093

Kouider, S. & Faivre, N. (2017). Conscious and unconscious perception. In S. Schneider, M. Velmans (Eds.), *The Blackwell Companion to Consciousness* (2nd ed.), John Wiley & Sons Inc, Hoboken, NJ, pp. 551-561

Kovács, I.P., Papathomas, T.V., Yang, M., & Feher, A. (1996). When the brain changes its mind: interocular grouping during binocular rivalry. *Proceedings of the National Academy of Sciences of the United States of America, 93 26*, 15508-11.

Kreiman, G., & Serre, T. (2020). Beyond the feedforward sweep: feedback computations in the visual cortex. *Annals of the New York Academy of Sciences, 1464*.

Kriegel, U. (2013). *Phenomenal Intentionality*. Oxford: Oxford University Press.

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

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

Lamme, V. A. (2010). How neuroscience will change our view on consciousness. *Cognitive neuroscience*, *1*(3), 204–220.

Lamme, V. (2020). Visual Functions Generating Conscious Seeing. *Frontiers in psychology*, *11*, 83.

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

Li, K., & Malhotra, P. A. (2015). Spatial neglect. *Practical neurology*, *15*(5), 333–339. https://doi.org/10.1136/practneurol-2015-001115

Liang, H., Gong, X., Chen, M., Yan, Y., Li, W., & Gilbert, C. (2017). Interactions between feedback and lateral connections in the primary visual cortex. *Proceedings of the National Academy of Sciences of the United States of America, 114*(32), 8637-8642.

Locke, J. (1689). *An Essay Concerning Human Understanding*. Oxford University Press.

Ludwig, D. (2020). Social-eyes: Rich Perceptual Contents and Systemic Oppression. *Review of Philosophy and Psychology*, 11(1).

Lupyan, G., Rahman, R. A., Boroditsky, L., & Clark, A. (2020). Effects of language on visual perception. *Trends in Cognitive Sciences*, *24*(11), 930-944

Maniscalco, B., & Lau, H. (2012). A signal detection theoretic approach for estimating metacognitive sensitivity from confidence ratings. *Consciousness and cognition*, *21*(1), 422–430.

Marr, D. (1982). Vision: A computational investigation into the human representation and processing of visual information. San Francisco: W.H. Freeman.

McDowell, J. (1994). The Content of Perceptual Experience. *The Philosophical Quarterly*, 44(175): 190-205.

Mendez-Bertolo, C., Moratti, S., Toledano, R., Lopez-Sosa, F., Martinez- Alvarez, R., Mah, Y. H., et al. (2016). A fast pathway for fear in human amygdala. *Nat. Neurosci.* 19, 1041–1049.

Michel, M., & Morales, J. (2020). Minority reports: Consciousness and the prefrontal cortex. *Mind & Language*, *35*(4), 493-513.

Morris, J.S., Ohman, A. & Dolan, R.J. (1998). Conscious and Unconscious Emotional Learning in the Human Amygdala. *Nature*, Vol 393- 4: 467-470.

Movshon, A. J. & Simoncelli, E. P. (2014). Representation of Naturalistic Image Structure in the Primate Visual Cortex. *Cold Spring Harbour Symposia on Quantitative Biology*, Volume LXXIX.

Mudrik, L., Breska, A., Lamy, D., & Deouell, L. (2011). Integration Without Awareness: Expanding the Limits of Unconscious Processing. *Psychological Science, 22*(6), 764-770.

Mudrik, L., Faivre, N., & Koch, C. (2014). Information integration without awareness. *Trends in cognitive sciences*. Volume 18, Issue 9, p. 488-496.

Nadalini, A., Bottini, R., Casasanto, D., & Crepaldi, D. (2020). The limits of unconscious semantic processing. https://doi.org/10.31234/osf.io/6zf2n

Nakamura, K., Dehaene, S., Jobert, A., Le Bihan, D., & Kouider, S. (2007). Task-specific change of unconscious neural priming in the cerebral language network. *PNAS Proceedings of the National Academy of Sciences of the United States of America, 104*(49), 19643–19648.

Nakamura K., Oga T., Fukuyama H. (2012). Task-sensitivity of unconscious word processing in spatial neglect. Neuropsychologia. 50(7): 1570-1577.

Nakamura, K., Makuuchi, M., Oga, T., Mizuochi-Endo, T., Iwabuchi, T., Nakajima, Y. and Dehaene, S. (2018). Neural capacity limits during unconscious semantic processing. *European Journal of Neuroscience*. 47: 929-937.

Noy, N., Bickel, S., Zion-Golumbic, E., Harel, M., Golan, T., Davidesco, I., Schevon, C. A., McKhann, G. M., Goodman, R. R., Schroeder, C. E., Mehta, A. D., & Malach, R. (2015). Ignition's glow: Ultra-fast spread of global cortical activity accompanying local "ignitions" in visual cortex during conscious visual perception. *Consciousness and cognition*, *35*, 206–224.

Nurminen, L., Merlin, S., Bijanzadeh, M. *et al.* Top-down feedback controls spatial summation and response amplitude in primate visual cortex. *Nat Commun* 9, 2281 (2018).

O'Reilly, R. C., Wyatte, D., Herd, S., Mingus, B., & Jilk, D. J. (2013). Recurrent Processing during Object Recognition. *Frontiers in psychology*, *4*, 124.

Odegaard, B., Knight, R., & Lau, H. (2017). Should a Few Null Findings Falsify Prefrontal Theories of Conscious Perception? *The Journal of Neuroscience, 37*, 9593 - 9602.

Ogden, J. A. editor (2005). *Fractured Minds: A Case Study Approach to Clinical Neuroscience*. New York: Oxford University Press.

Oizumi M., Albantakis L., Tononi G. (2014) From the Phenomenology to the Mechanisms of Consciousness: Integrated Information Theory 3.0. *PLOS Computational Biology*, 10(5).

Oriet, C. and Brand, J. (2012) Size averaging of irrelevant stimuli cannot be prevented. *Vision Res*. 79, 8–16.

Orlandi, Nico. (2014). *The innocent eye: why vision is not a cognitive process*. Oxford University Press.

Ortells, J. J., Marí-Beffa, P., & Plaza-Ayllón, V. (2013). Unconscious congruency priming from unpracticed words is modulated by prime–target semantic relatedness. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*(2), 394–413.

Palmer, T. D., & Ramsey, A. K. (2012). The function of unconsciousness in multisensory integration. *Cognition*, *125*, 353–364.

Persaud, N., McLeod, P., and Cowey, A. (2007). Post-decision wagering objectively measures awareness. *Nat. Neurosci.* 10, 257–261.

Peters, M. A. K., Kentridge, R. W., Phillips, I., and Block, N. (2017). Does unconscious perception really exist? Continuing the ASSC20 debate. *Neuroscience of Consciousness.* 1–11.

Phillips, I. (2018). Unconscious Perception Reconsidered. *Analytic Philosophy* 4 (59):471-514.

Plass, J., Guzman-Martinez, E., Ortega, L., Grabowecky, M., & Suzuki, S. (2014). Lip Reading Without Awareness. Psychological Science, 25(9), 1835-1837.

Poldrack, R. A., & Yarkoni, T. (2016). From brain maps to cognitive ontologies: informatics and the search for mental structure. *Annual review of psychology*, *67*, 587-612.

Poldrack, R. A. (2020). The physics of representation. *Synthese*:1-19.

Prinz, J. (2011). The Sensory Basis of Cognitive Phenomenology. In Tim Bayne & Michelle Montague (eds.), *Cognitive Phenomenology*. Oxford University Press. pp. 174--196.

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

Prinz, J. (2015) Unconscious perception. In: Matthen M (ed.), *Oxford Handbook of Philosophy of Perception*. Oxford: Oxford University Press, 371–89.

Quilty-Dunn, J. (2019). Unconscious perception and phenomenal coherence. *Analysis*, Volume 79, Issue 3: 461–469

Reber, T.P. and Henke, K. (2012) Integrating unseen events over time. *Consciousness and Cognition*. 21, 953–960

Rosch, E. (1978). Principles of categorization. In *Cognition and categorization*, ed. B.E. Rosch and B.B. Lloyd, 28–49. Hillsdale: Erlbaum.

Rossetti, Y., Pisella, L., & McIntosh, R.D. (2017). Rise and fall of the two visual systems theory. *Annals of physical and rehabilitation medicine, 60 3*: 130-140.

Russell, B. (1912). *The Problems of Philosophy*, London: Williams and Norgate; New York: Henry Holt and Company.

Scholte, H. S., Jolij, J., Fahrenfort, J. J., & Lamme, V. A. F. (2008). Feedforward and recurrent processing in scene segmentation: Electroencephalography and functional magnetic resonance imaging. *Journal of Cognitive Neuroscience, 20*(11), 2097–2109.

Siegle, J. H., Jia, X., Durand, S., Gale, S., Bennett, C., Graddis, N., Heller, G., Ramirez, T. K., Choi, H., Luviano, J. A., Groblewski, P. A., Ahmed, R., Arkhipov, A., Bernard, A., Billeh, Y. N., Brown, D., Buice, M. A., Cain, N., Caldejon, S., Casal, L., … Koch, C. (2021). Survey of spiking in the mouse visual system reveals functional hierarchy. *Nature*, *592*(7852), 86–92.

Sikkens, T., Bosman, C. A., & Olcese, U. (2019). The Role of Top-Down Modulation in Shaping Sensory Processing Across Brain States: Implications for Consciousness. *Frontiers in Systems Neuroscience*, 13 (31): 1-15.

Sklar, A.Y. et al. (2012) Reading and doing arithmetic nonconsciously. Proc. Natl. Acad. Sci. U.S.A. 109, 19614–19619.

Song, C., & Yao, H. (2016). Unconscious processing of invisible visual stimuli. *Scientific Reports, 6*.

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

Sprenger, A., Kömpf, D., & Heide, W. (2002). Visual search in patients with left visual hemineglect. *Progress in brain research*, *140*, 395–416.

Stein, T., & Sterzer, P. (2014). Unconscious processing under interocular suppression: Getting the right measure. *Frontiers in Psychology, 5,* Article 387.

Stein, T., Utz, V., & van Opstal, F. (2020). Unconscious semantic priming from pictures under backward masking and continuous flash suppression. *Consciousness and Cognition*. 78.

Tamietto, M., & de Gelder, B. (2010). Neural bases of the non-conscious perception of emotional signals. *Nat Rev Neurosci* 11, 697–709.

Tapia, E., & Beck, D. M. (2014). Probing feedforward and feedback contributions to awareness with visual masking and transcranial magnetic stimulation. *Frontiers in psychology*, *5*, 1173.

Tononi, G., Boly, M., Massimini, M. & Koch, C. (2016). Integrated information theory: from consciousness to its physical substrate. *Nat Rev Neurosci* 17, 450–461.

Tsuchiya, N., & Koch, C. (2014). On the relationship between consciousness and attention. In M. S. Gazzaniga & G. R. Mangun (Eds.), The cognitive neurosciences (pp. 839–853). MIT Press.

Tye, M. (2000). *Consciousness, Color and Content*. Cambridge, MA: MIT Press.

Van Boxtel, J. J., Tsuchiya, N., & Koch, C. (2010). Consciousness and attention: on sufficiency and necessity. *Frontiers in Psychology*, *1*, 217.

Van den Bussche, E., Smets, K., Sasanguie, D. and Reynvoet, B. (2012). The power of unconscious semantic processing: The effect of semantic relatedness between prime and target on subliminal priming. *Psychologica Belgica*, 52(1), pp.59–70.

van Essen, D. C., Anderson, C. H., & Felleman, D. J. (1992). Information processing in the primate visual system: An integrated systems perspective. *Science*, 255: 419-23.

van Gaal, S., & Lamme, V. A. F. (2012). Unconscious High-Level Information Processing: Implication for Neurobiological Theories of Consciousness. *The Neuroscientist*, *18*(3), 287–301.

Venezia J.H., Matchin W. and Hickok G. (2015). Multisensory Integration and Audiovisual Speech Perception. In: Arthur W. Toga, editor. Brain Mapping: An Encyclopedic Reference, vol. 2, pp. 565-572. Academic Press: Elsevier.

Verdon, V., Schwartz, S., Lovblad, K.-O., Hauert, C.-A., & Vuilleumier, P. (2010). Neuroanatomy of hemispatial neglect and its functional components: A study using voxel-based lesion-symptom mapping. *Brain: A Journal of Neurology, 133*(3), 880–894.

Vetter, P., & Newen, A. (2014). Varieties of cognitive penetration in visual perception. Consciousness and Cognition, 27, 62–75.

Vuilleumier, P., & Landis, T. (1998). Illusory Contours and spatial neglect. *NeuroReport, 9*(11), 2481–2484.

Williams, M. A., Morris, A. P., McGlone, F., Abbott, D. F., and Mattingley, J. B. (2004). Amygdala responses to fearful and happy facial expressions under conditions of binocular suppression. *J. Neurosci.* 24, 2898–2904.

Wilson, H. R., & Wilkinson, F. (2015) From orientations to objects: Configural processing in the ventral stream. *Journal of Vision*. 15(7): 4.

Wu, W. (2014). Against Division: Consciousness, Information and the Visual Streams. *Mind and Language* 29 (4): 383-406.

Wu, W. (2017). Shaking up the Ground Floor: The Cognitive Penetration of Visual Attention. *The Journal of Philosophy*. Volume CXIV, No. 1, 5-32.

Wu, W. (2020). Is Vision for Action Unconscious? *Journal of Philosophy*. 117 (8):413-433.

Yang, E., Brascamp, J., Kang, M.-S., & Blake, R. (2014). On the use of continuous flash suppression for the study of visual processing outside of awareness. *Frontiers in Psychology, 5,* Article 724.

Yang, Y. H., and Yeh, S. L. (2010). Accessing the meaning of invisible words. *Consciousness and Cognition.* 20, 223–233.

Yokoyama, T., Noguchi, Y., & Kita, S. (2013). Unconscious processing of direct gaze: evidence from an ERP study. *Neuropsychologia*, *51*(7), 1161–1168.

Zabelina, D.L., Guzman-Martinez, E., Ortega, L. *et al.* (2013). Suppressed semantic information accelerates analytic problem solving. *Psychon Bull Rev* 20, 581–585

Chapter 3 References

Amodio, D. (2014). The Neuroscience of Prejudice and Stereotyping. Nature Reviews Neuroscience, Volume 15, 670–682.

Anderson, M. L. (2014). After phrenology: Neural reuse and the interactive brain. MIT Press.

Anderson, E., & Shivakumar, G. (2013). Effects of exercise and physical activity on anxiety. *Frontiers in psychiatry*, *4*, 27. https://doi.org/10.3389/fpsyt.2013.00027

Aristotle & Roberts, W. R. (2004). *Rhetoric*. Mineola, N.Y: Dover Publications.

Atkinson, A. P., & Adolphs, R. (2005). Visual Emotion Perception: Mechanisms and Processes. In L. F. Barrett, P. M. Niedenthal, & P. Winkielman (Eds.), Emotion and consciousness (pp. 150–182). The Guilford Press.

Averill, J. R. (1985). The Social Construction of Emotion With Special Reference to Love. *The Social Construction of the Person* (1985): 89–109.

Bandelow, B., & Michaelis, S. (2015). Epidemiology of anxiety disorders in the 21st century. Dialogues in clinical neuroscience, 17(3), 327–335.

Barrett, L. F. (2006). Valence is a basic building block of emotional life. *Journal of Research in Personality*, 40, 35-55.

Barrett, L. F. (2017). How emotions are made: The secret life of the brain. Houghton Mifflin Harcourt.

Barrett, L. F., & Wager, T. D. (2006). The Structure of Emotion: Evidence From Neuroimaging Studies. *Current Directions in Psychological Science*, 15(2), 79–83.

Bebko, G. M., Franconeri, S. L., Ochsner, K. N., & Chiao, J. Y. (2011). Look before you regulate: Differential perceptual strategies underlying expressive suppression and cognitive reappraisal. *Emotion*, 11(4), 732–742.

Beyeler, A., Chang, C. J., Silvestre, M., Lévêque, C., Namburi, P., Wildes, C. P., & Tye, K. M. (2018). Organization of Valence-Encoding and Projection-Defined Neurons in the Basolateral Amygdala. *Cell reports*, 22(4), 905–918.

Bonnet, L., Comte, A., Tatu, L., Millot, J-L., Moulin, T, & de Bustos, E. M. (2015). The role of the amygdala in the perception of positive emotions: an “intensity detector”. *Frontiers in Behavioral Neuroscience*. Volume 9, Article 178.

Calhoun, C. (1984). Cognitive Emotions? *What is an Emotion?* New York: Oxford University Press.

Campbell-Sills, L., Ellard, K. K., & Barlow, D. H. (2014). Emotion regulation in anxiety disorders.

Capitão, L. P., Underdown, S. J. V., Vile, S., Yang, E., Harmer, C. J., & Murphy, S. E. (2014). Anxiety increases breakthrough of threat stimuli in continuous flash suppression. *Emotion, 14*(6), 1027–1036.

Carruthers, P. (2018). Valence and value. *Philosophy and Phenomenological Research*, 97(3), 658–680.

Celeghin, A., Diano, M., Costa, T., Adenzato, M., Mosso, C. O., Weiskrantz, L., et al. (2016). Psychophysiological mechanisms guiding recognition of basic and complex emotions without visual cortex. *Neuropsychol. Trends* 20, 77.

Charland, L. (1997). Reconciling Cognitive and Perceptual Theories of Emotion: A Representational Proposal. *Philosophy of Science*, Vol. 64, No.4: 555-579.

Damasio, A. (1984). The Feeling of What Happens. *What is an Emotion?*, New York: Oxford University Press.

Damasio, A. (2004). William James and the Modern Neurobiology of Emotion. *Emotion, Evolution and Rationality*, ed. Dylan Evans and Pierre Cruse, New York: Oxford University Press.

Darwin, C. (1872). *The Expression of the Emotions in Man and Animals*, London: John Murray, Albemarle Street.

de Gelder, B. (2005). Nonconscious Emotions: New Findings and Perspectives on Nonconscious Facial Expression Recognition and Its Voice and Whole-Body Contexts. In L. F. Barrett, P. M. Niedenthal, & P. Winkielman (Eds.), *Emotion and consciousness* (pp. 123–149). The Guilford Press.

de Gelder, B., Pourtois, G., & Weiskrantz, L. (2002). Fear recognition in the voice is modulated by unconsciously recognized facial expressions but not by unconsciously recognized affective pictures. *Proc. Natl. Acad. Sci. U.S.A.* 99, 4121–4126.

de Sousa, R. (2007). Emotion. in E. Zalta (ed.), *The Stanford Encyclopedia of Philosophy.*

Diano, M., Celeghin, A., Bagnis, A., & Tamietto, M. (2017). Amygdala response to emotional stimuli without awareness: facts and interpretations. *Frontiers in Psychology*. Volume 7, Article 2029.

Earl B. (2014). The biological function of consciousness. Frontiers in psychology, 5, 697.

Ekman, P. (1977). Biological and Cultural Contributions to Body and Facial Movement. *The Anthropology of the Body*, Academic Press, London.

Elman, I., & Borsook, D. (2018). Threat Response System: Parallel Brain Processes in Pain vis-à-vis Fear and Anxiety. *Frontiers in psychiatry*, *9*, 29.

Etkin, A., Büchel, C., & Gross, J. J. (2015). The neural bases of emotion regulation. *Nature Reviews Neuroscience, 16*(11), 693–700.

Fang, Z., Li, H., Chen, G., & Yang, J. (2016). Unconscious processing of negative animals and objects: role of the amygdala revealed by fMRI. Frontiers in human neuroscience, 10, 146.

Füstös, J., Gramann, K., Herbert, B. M., & Pollatos, O. (2013). On the embodiment of emotion regulation: Interoceptive awareness facilitates reappraisal. *Social Cognitive and Affective Neuroscience, 8*(8), 911–917.

Gainotti, G. (2012). Unconscious processing of emotions and the right hemisphere. *Neuropsychologia*, 50: 205-218.

Garavan, H., Pendergrass, J. C., Ross, T. J., Stein, E. A., and Risinger, R. C. (2001). Amygdala response to both positively and negatively valenced stimuli. *Neuroreport*, Volume 12, Number 12.

Gayet, S., Paffen, C. L., Belopolsky, A. V., Theeuwes, J., & Van der Stigchel, S. (2016). Visual input signaling threat gains preferential access to awareness in a breaking continuous flash suppression paradigm. *Cognition*, *149*, 77–83.

Gendron, M., Roberson, D., van der Vyver, J. M., Barrett, L. F. (2014). Perceptions of emotion from facial expressions are not culturally universal: Evidence from a remote culture. Emotion, 14, 251–262.

Georgiou, G. A., Bleakley, C., Hayward, J., Russo, R., Dutton, K., Eltiti, S., et al. (2005). Focusing on fear: attentional disengagement from emotional faces. *Vis. Cogn.* 12, 145–158.

Ginot, E. (2015). *The Neuropsychology of the Unconscious—Integrating Brain and Mind in Psychotherapy*. Norton, New York.

Gläscher, J., & Adolphs, R. (2003). Processing of the arousal of subliminal and supraliminal emotional stimuli by the human amygdala. *The Journal of Neuroscience*, 23(32), 10274–10282.

Goldie, P. (2000). *The emotions: A philosophical exploration*. Oxford University Press.

Goldin, P. R., McRae, K., Ramel, W., & Gross, J. J. (2008). The neural bases of emotion regulation: reappraisal and suppression of negative emotion. *Biological psychiatry*, *63*(6), 577–586.

Grabowska, A., Marchewka, A., Seniów, J., Polanowska, K., Jednoróg, K., and Królicki, L. (2011). Emotionally negative stimuli can overcome attentional deficits in patients with visuo-spatial hemineglect. *Neuropsychologia* 49,3327–3337.

Green, O. H. (1992). The emotions: A philosophical theory. Dordrecht: Kluwer.

Griffiths, P. (2004). Is emotion a natural kind? In Robert C. Solomon (ed.), *Thinking About Feeling: Contemporary Philosophers on Emotions*. Oxford University Press.

Grill-Spector, K. & Weiner, K. S. (2014). The functional architecture of the ventral temporal cortex and its role in categorization. *Nature Reviews Neuroscience*, 15: 536-548.

Gross, J. J., & Thompson, R. A. (2007). Emotion Regulation: Conceptual Foundations. In J. J. Gross (Ed.), Handbook of emotion regulation (pp. 3–24). The Guilford Press.

Hart, S. J., Green, S. R., Casp, M., and Belger, A. (2010). Emotional priming effects during Stroop task performance. *Neuroimage* 49, 2662–2670.

Helm, B. W. (2010). Emotions and motivation: reconsidering neo-Jamesian accounts. *The Oxford Handbook of Philosophy of Emotion*, 303-323.

Hoemann, K., Xu, F., & Barrett, L. F. (2019). Emotion words, emotion concepts, and emotional development in children: A constructionist hypothesis. *Developmental Psychology, 55*(9), 1830–1849.

Hofmann, S. G. (2007). Cognitive factors that maintain social anxiety disorder: A comprehensive model and its treatment implications. *Cognitive behaviour therapy*, *36*(4), 193-209.

James, W. (1884). What is an Emotion? *Mind*, Vol. 9, No. 34.

Jamieson, J. P., Nock, M. K., & Mendes, W. B. (2013). Changing the Conceptualization of Stress in Social Anxiety Disorder: Affective and Physiological Consequences. *Clinical Psychological Science*, *1*(4), 363–374.

Koole, S. L., & Rothermund, K. (2011). "I feel better but I don't know why": the psychology of implicit emotion regulation. *Cognition & emotion*, *25*(3), 389–399.

Kötter, R., & Meyer, N. (1992). The limbic system: a review of its empirical foundation. Behavioural brain research, 52(2), 105-127.

Kring, A. M. (2000). Gender and anger. In A. H. Fischer (Ed.), *Gender and emotion: Social psychological perspectives* (pp. 211–231). Cambridge University Press.

Lapate, R. C., Rokers, B., Tromp, D. P. M., Orfali, N. S., Oler, J. A., Doran, S. T., ... & Davidson, R. J. (2016). Awareness of emotional stimuli determines the behavioral consequences of amygdala activation and amygdala-prefrontal connectivity. Scientific reports, 6(1), 1-16.

Lazarus, R. S. (1984). The Primacy of Cognition. *American Psychologist*, Vol 39, No. 2, 124-129.

Lazarus, R. S. (1991). Cognition and Motivation in Emotion. *American Psychologist*, Vol. 46, No. 4.

Lebrecht, S., Bar, M., Barrett, L. F., & Tarr, M. J. (2012). Micro-valences: perceiving affective valence in everyday objects. *Frontiers in psychology*, *3*, 107.

LeDoux, J. E. (1996). The emotional brain: The mysterious underpinnings of emotional life. Simon & Schuster.

LeDoux J. E. (2014). Coming to terms with fear. Proceedings of the National Academy of Sciences of the United States of America, 111(8), 2871–2878.

LeDoux J. E. (2017). Semantics, Surplus Meaning, and the Science of Fear. *Trends in cognitive sciences*, *21*(5), 303–306.

LeDoux, J. E., & Brown, R. (2017). A higher-order theory of emotional consciousness. *PNAS Proceedings of the National Academy of Sciences of the United States of America, 114*(10), e2016–e2025.

Lépine, J. P. (2002). The epidemiology of anxiety disorders: prevalence and societal costs. *Journal of Clinical Psychiatry*, 63, 4-8.

Lindquist, K. A., Gendron, M., Barrett, L. F., Dickerson, B. C. (2014). Emotion, but not affect perception, is impaired with semantic memory loss. *Emotion*, 14, 375–387.

Lindquist, K. A., Satpute, A. B., & Gendron, M. (2015). Does Language Do More Than Communicate Emotion? *Current Directions in Psychological Science*, *24*(2), 99–108.

Ludwig, D. (2020). Social-eyes: Rich perceptual contents and systemic oppression. *Review of Philosophy and Psychology*, 11(4), 939–954.

Lupyan, G. (2012). Linguistically modulated perception and cognition: the label-feedback hypothesis. *Frontiers in psychology*, 3, 54.

Lyon, P., & Kuchling, F. (2021). Valuing what happens: a biogenic approach to valence and (potentially) affect. *Philosophical Transactions of the Royal Society B*, *376*(1820), 20190752.

MacDonald, A. W., Cohen, J. D., Stenger, V. A., & Carter, C. S. (2000). Dissociating the role of the dorsolateral prefrontal and anterior cingulate cortex in cognitive control. *Science*, 288: 5472, 1835-1838.

Mancia, M. (2006). Implicit memory and early unrepressed unconscious: their role in the therapeutic process (how the neurosciences can contribute to psychoanalysis). The International journal of psycho-analysis, 87(Pt 1), 83–103.

Millikan, R. G. (1996). 'Pushmi-pullyu Representations'. In L. May & M. Friedman & A. Clark (Eds.), *Mind and Morals: Essays on cognitive science and ethics*. Cambridge, MA: MIT Press

Mudrik, L., Breska, A., Lamy, D., & Deouell, L. (2011). Integration Without Awareness: Expanding the Limits of Unconscious Processing. Psychological Science 22(6): 764-770.

Norman, V. C., Pamminger, T., & Hughes, W. (2017). The effects of disturbance threat on leaf-cutting ant colonies: a laboratory study. *Insectes sociaux*, *64*(1), 75–85. https://doi.org/10.1007/s00040-016-0513-z

Nussbaum, M. C. (2001). *Upheavals of Thought: The Intelligence of Emotions*, New York: Cambridge University Press.

Mendez-Bertolo, C., Moratti, S., Toledano, R., Lopez-Sosa, F., Martinez- Alvarez, R., Mah, Y. H., et al. (2016). A fast pathway for fear in human amygdala. *Nat. Neurosci.* 19, 1041–1049.

Morris, J.S., Ohman, A. & Dolan, R.J. (1998). Conscious and Unconscious Emotional Learning in the Human Amygdala. *Nature*, Vol 393- 4: 467-470.

Ohman, A., Flykt, A., & Esteves, F. (2001). Emotion drives attention: detecting the snake in the grass. J. Exp. Psychol. Gen. 130, 466–478.

Ochsner, K. N., Ray, R. D., Cooper, J. C., Robertson, E. R., Chopra, S., Gabrieli, J. D., & Gross, J. J. (2004). For better or for worse: neural systems supporting the cognitive down- and up-regulation of negative emotion. *NeuroImage*, *23*(2), 483–499.

Orlandi, N. (2014). The Innocent Eye: Why Vision is Not a Cognitive Process. OUP Usa.

Ortony, A. & Turner, T. J. (1990). What's Basic About Basic Emotions? *Psychological Review*, Vol. 97, No. 3.

Owren, M. J., Rendall, D., & Bachorowski, J.-A. (2005). Conscious and Unconscious Emotion in Nonlinguistic Vocal Communication. In L. F. Barrett, P. M. Niedenthal, & P. Winkielman (Eds.), *Emotion and consciousness* (pp. 185–204). The Guilford Press.

Panksepp, J. (2007). Affective consciousness. In M. Velmans & S. Schneider (Eds.), *The Blackwell companion to consciousness* (pp. 114–129). Blackwell Publishing.

Panksepp, J. (2011). Cross-species affective neuroscience decoding of the primal affective experiences of humans and related animals. *PLoS ONE* 6: e21236.

Panksepp, J., & Biven, L. (2012). *The archaeology of mind: Neuroevolutionary origins of human emotion.* W. W. Norton & Company.

Pasley, B. N., Mayes, L. C., & Schultz, R. T. (2004). Subcortical discrimination of unperceived objects during binocular rivalry. *Neuron* 42, 163–172.

Paton, J.J., Belova, M.A., Morrison, S.E., & Salzman, C.D. (2006). The primate amygdala represents the positive and negative value of visual stimuli during learning. Nature *439*, 865–870.

Pendoley, K. (Forthcoming). What Stubborn Emotions Can't Do for a Theory of Emotion.

Phelps, E. A. (2005). The Interaction of Emotion and Cognition: Insights from Studies of the Human Amygdala. In L. F. Barrett, P. M. Niedenthal, & P. Winkielman (Eds.), *Emotion and consciousness* (pp. 51–66). The Guilford Press.

Prinz, J. J. (2004). *Gut Reactions: A Perceptual Theory of the Emotions*. Oxford University Press.

Prinz, J. J. (2005). Emotions, Embodiment, and Awareness. In L. F. Barrett, P. M. Niedenthal, & P. Winkielman (Eds.), *Emotion and consciousness* (pp. 363–383). The Guilford Press.

Reisenzein, R. (2012). What is an Emotion in the Belief-Desire Theory of Emotion? In F. Paglieri, M. Tummolini, F. Falcone & M. Miceli (eds.), *The goals of cognition: Essays in honor of Cristiano Castelfranchi*. College Publications.

Russell, J. A. (2003). Core Affect and the Psychological Construction of Emotion. *Psychological Review*, Vol. 110, No. 1, 145-172.

Russell, J. A., & Barrett, L. F. (1999). Core affect, prototypical emotional episodes, & other things called emotion: Dissecting the elephant. *Journal of Personality and Social Psychology*, *76*, 805–819.

Sander, D., Grafman, J., & Zalla, T. (2003). The human amygdala: an evolved system for relevance detection. Reviews in the neurosciences, 14(4), 303–316.

Schachter, S. & Singer, J. E. (1962/1984). Cognitive, Social, and Physiological Determinants of Emotional State. *What is an Emotion?* New York: Oxford University Press.

Schiller, P. H., & Malpeli, J. G. (1977). Properties and tectal projections of monkey retinal ganglion cells. Journal of Neurophysiology, 40(2), 428–445.

Searle, J. R. (2004). *Mind: A Brief Introduction*. Oxford University Press.

Sergerie, K., Chochol, C., & Armony, J. L. (2008). The Role of the Amygdala in Emotional Processing: A Quantitative Meta-Analysis of Functional Neuroimaging Studies. *Neuroscience and Biobehavioral Reviews*, Vol. 32, 811-830.

Shabel, S.J., & Janak, P.H. (2009). Substantial similarity in amygdala neuronal activity during conditioned appetitive and aversive emotional arousal. Proc. Natl. Acad. Sci. USA *106*, 15031–15036.

Sklar, A. Y., Levy, N., Goldstein, A., Mandel, R., Maril, A., & Hassin, R. R. (2012). Reading and doing arithmetic nonconsciously. *Proc. Natl. Acad. Sci. U.S.A.* 109, 19614–19619.

Smith, R., & Lane, R. D. (2015). The neural basis of one's own conscious and unconscious emotional states. *Neuroscience and biobehavioral reviews*, 57, 1–29.

Smith, C. A. & Lazarus R. S. (1993). Appraisal Components, Core Relational Themes, and the Emotions. *Cognition and Emotion*, Vol. 7, 3/4.

Smith, P., & McCulloch, K. (2012). Subliminal perception. *Encyclopedia of human behavior 2nd edition. London: Elsevier*.

Solomon, R. C. (1984). Emotions and Choice. *What is an Emotion?*, New York: Oxford University Press.

Solomon, R. C. (2004). Emotions, thoughts, and feelings: Emotions as engagements with the world. In Thinking About Feeling: Contemporary Philosophers on Emotions. Oxford University Press. pp. 1-18.

Stanley, D., Phelps, E., & Banaji, M. (2008). The Neural Basis of Implicit Attitudes. Current Directions in Psychological Science. Volume 17, Issue 2, 164-170.

Stefanacci, L., & Amaral, D.G. (2000). Topographic organization of cortical inputs to the lateral nucleus of the macaque monkey amygdala: a retrograde tracing study. J. Comp. Neurol. 421, 52–79.

Stein, T., & Sterzer, P. (2014). Unconscious processing under interocular suppression: Getting the right measure. Frontiers in Psychology, 5, Article 387.

Stein, T., Sterzer, P., & Peelen, M. V. (2012). Privileged detection of conspecifics: evidence from inversion effects during continuous flash suppression. *Cognition*, *125*(1), 64–79.

Stein, T., Seymour, K., Hebart, M. N., & Sterzer, P. (2014). Rapid fear detection relies on high spatial frequencies. *Psychol. Sci*. 566–574.

Stöber, J. (1998). Worry, problem elaboration and suppression of imagery: The role of concreteness. *Behaviour Research and Therapy, 36*(7-8), 751–756.

Stöber, J., & Borkovec, T. D. (2002). Reduced concreteness of worry in generalized anxiety disorder: Findings from a therapy study. *Cognitive Therapy and Research, 26*(1), 89–96.

Stocker, M. (1996). The Irreducibility of Affect. *Valuing Emotions*, Cambridge University Press.

Szczygieł D., Buczny J., Bazińska R. (2012). Emotion regulation and emotional information processing: the moderating effect of emotional awareness. Pers. Ind. Differ. 52 433–437.

Tamietto, M., Castelli, L., Vighetti, S., Perozzo, P., Geminiani, G., Weiskrantz, L., et al. (2009). Unseen facial and bodily expressions trigger fast emotional reactions. *Proc. Natl. Acad. Sci. U.S.A.* 106, 17661–17666.

Turk, C. L., Heimberg, R. G., Luterek, J. A., Mennin, D. S., & Fresco, D. M. (2005). Emotion Dysregulation in Generalized Anxiety Disorder: A Comparison with Social Anxiety Disorder. *Cognitive Therapy and Research, 29*(1), 89–106.

Vieira, J. B., Wen, S., Oliver, L. D., & Mitchell, D. (2017). Enhanced conscious processing and blindsight-like detection of fear-conditioned stimuli under continuous flash suppression. *Experimental brain research*, *235*(11), 3333–3344.

Whalen, P. J. (1998). Fear, Vigilance, and Ambiguity: Initial Neuroimaging Studies of the Human Amygdala. Current Directions in Psychological Science, 7(6), 177–188.

Whiting, D. (2011). The Feeling Theory of Emotion and the Object-Directed Emotions. *European Journal of Philosophy,* 19 (2):281-303.

Williams, M. A., Morris, A. P., McGlone, F., Abbott, D. F., and Mattingley, J. B. (2004). Amygdala responses to fearful and happy facial expressions under conditions of binocular suppression. *J. Neurosci.* 24, 2898–2904.

Wilson-Mendenhall, C., & Dunne, J. (2021). Cultivating Emotional Granularity. *Frontiers in Psychology*, *12*.

Winkielman, P., Berridge, K. C., & Wilbarger, J. L. (2005). Emotion, Behavior, and Conscious Experience: Once More without Feeling. In L. F. Barrett, P. M. Niedenthal, & P. Winkielman (Eds.), *Emotion and consciousness* (pp. 335–362). The Guilford Press.

Winkielman, P., & Berridge, K. C. (2004). Unconscious emotion. Current Directions in Psychological Science, 13(3), 120–123.

Yang, E., Zald, D. H., & Blake, R. (2007). Fearful expressions gain preferential access to awareness during continuous flash suppression. *Emotion*, 7(4), 882–886.

Zajonc, R.B. (1984). On the Primacy of Affect. *American Psychologist*, Vol 39, No. 2, 117-123.

Zeelenberg, R., Wagenmakers, E.-J., & Rotteveel, M. (2006). The Impact of Emotion on Perception: Bias or Enhanced Processing? *Psychological Science*, *17*(4), 287–291.

Chapter 4 References

Allen, C. (2018). Associative learning. In K. Andrews & J. Beck (Eds.), *The Routledge handbook of philosophy of animal minds* (pp. 401–408). Routledge/Taylor & Francis Group.

Allport, G. W. (1954). *The Nature of Prejudice.* Addison- Wesley.

Anderson, E. (2017). Feminist epistemology and philosophy of science. The Stanford Encyclopedia of Philosophy, Edward N. Zalta (ed.)

Bartky, S. (1979). On psychological oppression. In Philosophy and Women, eds. Sharon Bishop and Marjorie Weinzweig. Wadsworth Publishing Company, 10.

Beeman, A., & Narayan, A. (2011). If You're White, You're Alright: The Reproduction Of Racial Hierarchies in Bollywood Films. In *Covert Racism* (pp. 155-173). Brill.

Binder JR. (2016) In defense of abstract conceptual representations. Psychon. Bull. Rev. 23, 1096 – 1108.

Birch, J., Ginsburg, S., & Jablonka, E. (2020). Unlimited Associative Learning and the origins of consciousness: a primer and some predictions. *Biology & philosophy*, *35*, 56.

Blanco F. (2017) Cognitive Bias. In: Vonk J., Shackelford T. (eds) Encyclopedia of Animal Cognition and Behavior. Springer, Cham.

Bowman, M. (2020). Privileged Ignorance, "World"-Traveling, and Epistemic Tourism. *Hypatia*, *35*(3), 475-489.

Brandstätter, V., Lengfelder, A., & Gollwitzer, P. M. (2001). Implemen- tation intentions and efficient action initiation. Journal of Personality and Social Psychology, 81, 946–960.

Coates, R. D. (2011). *Covert racism: Theories, institutions, and experiences*. Brill.

Brownstein, Michael (2015). Attributionism and Moral Responsibility for Implicit Bias. *Review of Philosophy and Psychology*, 7: 765-786.

Brownstein, Michael (2018). The Implicit Mind. Oxford University Press.

Cottrell, C. A. & Neuberg, S. L. (2005). Different emotional reactions to different groups: a sociofunctional threat- based approach to "prejudice". *J. Pers. Soc. Psychol.* 88, 770–789.

Dang, J., Xiao, S., & Mao, L. (2015). A new account of the conditioning bias to out-groups. *Frontiers in psychology*, *6*, 197. https://doi.org/10.3389/fpsyg.2015.00197

Devine, P. G. (1989). Stereotypes and prejudice: their automatic and controlled components. *J. Pers. Soc. Psychol.* 56, 5–18.

Devine, P. G., Forscher, P. S., Austin, A. J., & Cox, W. T. (2012). Long-term reduction in implicit race bias: A prejudice habit-breaking intervention. *Journal of experimental social psychology*, *48*(6), 1267–1278.

Fazio, R. H., Jackson, J. R., Dunton, B. C. & Williams, C. J. (1995). Variability in automatic activation as an unobtrusive measure of racial attitudes: a *bona fide* pipeline? *J. Pers. Soc. Psychol.* 69, 1013–1027.

Gigerenzer, G. (2018), "The Bias Bias in Behavioral Economics", Review of Behavioral Economics: Vol. 5: No. 3-4, pp 303-336.

Gigerenzer, G. & Brighton, H. (2009), Homo Heuristicus: Why Biased Minds Make Better Inferences. Topics in Cognitive Science, 1: 107-143.

Ginsburg, S., & Jablonka, E. (2019). *The evolution of the sensitive soul: Learning and the origins of consciousness.* (A. Zeligowski, Illustrator). The MIT Press.

Gendler, T. S. (2011). On the epistemic costs of implicit bias. Philosophical Studies, 156(1), 33.

Gilovich, T. & Griffin, D. (2002). Introduction-Heuristics and biases: Then and now. In Gilovich, T., Griffin, D., & Kahneman, D. (Eds.) *Heuristics and biases: The psychology of intuitive judgment*. Cambridge University Press.

Gilovich, T., Griffin, D., & Kahneman, D. (Eds.). (2002). Intro- Heuristics and biases: The psychology of intuitive judgment. Cambridge University Press.

Greenwald, A. G., & De Houwer, J. (2017). Unconscious conditioning: Demonstration of existence and difference from conscious conditioning. *Journal of Experimental Psychology: General, 146*(12), 1705–1721.

Gruen, L. (2015). Entangled Empathy: An Alternative Ethic for our Relationships with Animals. Brooklyn: Latern Press.

Haselton, M. G., & Nettle, D. (2006). The Paranoid Optimist: An Integrative Evolutionary Model of Cognitive Biases. Personality and Social Psychology Review, 10(1), 47–66.

Haselton, M.G., Nettle, D. and Murray, D.R. (2005). The Evolution of Cognitive Bias. In The Handbook of Evolutionary Psychology, D.M. Buss (Ed.).

Holroyd, J. (2012). *Responsibility for Implicit Bias*. Journal of Social Philosophy. Vol. 43 No. 3, 274–306.

Holroyd, J. (2015). Implicit bias, awareness and imperfect cognitions. *Consciousness and cognition*, *33*, 511-523.

Izumi, C.L. (2010). Implicit Bias and the Illusion of Mediator Neutrality. Washington University Journal of Law and Policy, 34, 71-155.

Jamal, A., & Naber, N. (Eds.). (2008). *Race and Arab Americans Before and After 9/11: From Invisible Citizens to Visible Subjects*. Syracuse University Press.

Koizumi, A., Amano, K., Cortese, A., Shibata, K., Yoshida, W., Seymore, B., Kawato, M., Lau, H. (2016). Fear reduction without fear through reinforcement of neural activity that bypasses conscious exposure. Nature Human Behaviour, Volume 1, Article 0006, pp. 1–7.

Lai, C. K., Marini, M., Lehr, S. A., Cerruti, C., Shin, J.-E. L., Joy-Gaba, J. A., Ho, A. K., Teachman, B. A., Wojcik, S. P., Koleva, S. P., Frazier, R. S., Heiphetz, L., Chen, E. E., Turner, R. N., Haidt, J., Kesebir, S., Hawkins, C. B., Schaefer, H. S., Rubichi, S., . . . Nosek, B. A. (2014). Reducing implicit racial preferences: I. A comparative investigation of 17 interventions. *Journal of Experimental Psychology: General, 143*(4), 1765–1785.

Legault, L., Gutsell, J. N., & Inzlicht, M. (2011). Ironic effects of anti-prejudice messages: How motivational interventions can reduce (but also increase) prejudice. Psychological Science, 22(12), 1472–1477.

Maia T. V. (2009). Fear conditioning and social groups: statistics, not genetics. *Cognitive science*, *33*(7), 1232–1251.

Macrae, C. N., Bodenhausen, G. V., Milne, A. B., & Jetten, J. (1994). Out of mind but back in sight: Stereotypes on the rebound. *Journal of Personality and Social Psychology, 67*(5), 808–817.

Mercier, H. (2017). Confirmation Bias—Myside Bias. In R. F. Pohl (Ed.), *Cognitive illusions: Intriguing phenomena in thinking, judgment and memory* (p. 99–114). Routledge/Taylor & Francis Group.

Mills, C. (2007). White ignorance. Sullivan, Shannon and Nancy Tuana, eds, 23. *Race and epistemologies of ignorance*. SUNY Press.

Nickerson, R. S. (1998). Confirmation Bias: A Ubiquitous Phenomenon in Many Guises. Review of General Psychology, 2(2), 175–220.

Olson, I. R., McCoy, D., Klobusicky, E., & Ross, L. A. (2013). Social cognition and the anterior temporal lobes: a review and theoretical framework. *Social cognitive and affective neuroscience*, *8*(2), 123-133.

Ortiz, S. M. (2021). Call-In, Call-Out, Care, and Cool Rationality: How Young Adults Respond to Racism and Sexism Online. Social Problems.

Pashak, T. J., Conley, M. A., Whitney, D. J., Oswald, S. R., Heckroth, S. G., & Schumacher, E. M. (2018). Empathy diminishes prejudice: active perspective-taking, regardless of target and mortality salience, decreases implicit racial Bias. *Psychology*, *9*(06), 1340.

Pearson, J. (2012). Associative learning: Pavlovian conditioning without awareness. Current Biology, 22(12), R495-R496.

Peirce, C. (1970). Offensive Mechanisms. Floyd B. Barbour (ed.) The Black Seventies. Porter Sargent Publisher p.: 267.

Pope, D. G., Price, J., & Wolfers, J. (2018). Awareness reduces racial bias. *Management Science*, *64*(11), 4988-4995.

Rajsic, J., Wilson, D. E., & Pratt, J. (2015). Confirmation bias in visual search. Journal of Experimental Psychology: Human Perception and Performance, 41(5), 1353–1364.

Rees, H. R., Rivers, A. M., & Sherman, J. W. (2019). Implementation Intentions Reduce Implicit Stereotype Activation and Application. *Personality and Social Psychology Bulletin*, *45*(1), 37–53.

Rudman, L. A., Ashmore, R. D., & Gary, M. L. (2001). " Unlearning" Automatic Biases: The Malleability of Implicit Prejudice and Stereotypes. *Journal of Personality and Social Psychology*, *81*(5), 856-868.

Saeed, A. (2007). Media, Racism and Islamophobia: The Representation of Islam and Muslims in the Media. *Sociology compass*, *1*(2), 443-462.

Schneider, D. J. (2005). *The psychology of stereotyping*. Guilford Press.

Sherman, J. W., Stroessner, S. J., Conrey, F. R., & Azam, O. A. (2005). Prejudice and Stereotype Maintenance Processes: Attention, Attribution, and Individuation. *Journal of Personality and Social Psychology, 89*(4), 607–622.

Stanovich, K. E., Toplak, M. E., & West, R. F. (2008). The development of rational thought: A taxonomy of heuristics and biases. In R. V. Kail (Ed.), Advances in child development and behavior: Vol. 36. Advances in child development and behavior (p. 251–285). Elsevier Academic Press.

Stewart, B. D., & Payne, B. K. (2008). Bringing automatic stereotyping under control: Implementation intentions as efficient means of thought control. Personality and Social Psychology Bulletin, 34, 1332–1345.

Sullivan, S., & N. Tuana. (2007). Race an epistemologies of ignorance. State University of New York Press.

Todd, A. R., Galinsky, A. D., & Bodenhausen, G. V. (2012). Perspective taking undermines stereotype maintenance processes: Evidence from social memory, behavior explanation, and information solicitation. Social Cognition, 30(1), 94-108.

Toplak, M. E., West, R. F., & Stanovich, K. E. (2011). The Cognitive Reflection Test as a predictor of performance on heuristics-and-biases tasks. *Memory & cognition*, *39*(7), 1275–1289.

Tversky, A. & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. *Science* 185 (4157):1124-1131.

Waroquier, L., Abadie, M., & Dienes, Z. (2020). Distinguishing the role of conscious and unconscious knowledge in evaluative conditioning. *Cognition*, *205*, 104460.

Wilson-Mendenhall, C. D., Simmons, W. K., Martin, A., & Barsalou, L. W. (2013). Contextual processing of abstract concepts reveals neural representations of non-linguistic semantic content. *Journal of Cognitive Neuroscience*, *25*, 920--935.

Winston, A. S. (2004). Defining difference: Race and racism in the history of psychology. American Psychological Association.

Yee, E., & Thompson-Schill, S. L. (2016). Putting concepts into context. *Psychonomic bulletin & review*, *23*(4), 1015–1027.

Zheng, R. (2016). Attributability, accountability and implicit attitudes. In Brownstein, & Saul (Eds.), Implicit bias and philosophy (Vol. 2) (pp. 62–89). Oxford, UK: Oxford University Press.

Chapter 5 References

Amodio, D. (2014). The Neuroscience of Prejudice and Stereotyping. *Nature Reviews Neuroscience*, Volume 15, 670–682.

Burnston, D. C. (forthcoming). Cognitive Ontologies, Task Ontologies, and Explanation in Cognitive Neuroscience. In *The Tools of Neuroscience Experiment* (pp. 259-283). Routledge.

Earl, B. (2014). The biological function of consciousness. *Frontiers in Psychology, 5,* Article 697.

Fazekas, Peter & Nanay, Bence (2021). Attention Is Amplification, Not Selection. *British Journal for the Philosophy of Science* 72 (1):299-324.

Malach, R. (2021). Local neuronal relational structures underlying the contents of human conscious experience. *Neuroscience of consciousness*, 2021(2), niab028.

Miracchi, L. (2019). None of These Problems Are That 'Hard'... or 'Easy': Making Progress on the Problems of Consciousness. *Journal of Consciousness Studies* 26 (9-10):160-172.

Tononi, G., Boly, M., Massimini, M. & Koch, C. (2016). Integrated information theory: from consciousness to its physical substrate. *Nat Rev Neurosci* 17, 450–461.

Wimsatt, W. C. (1994). The Ontology of Complex Systems: Levels of Organization, Perspectives, and Causal Thickets1. *Canadian Journal of Philosophy Supplementary Volume*, *20*, 207-274.