

The Neuroscience of Consciousness

First published Tue Oct 9, 2018

Conscious experience in humans depends on brain activity, so neuroscience will contribute to explaining consciousness. What would it be for neuroscience to explain consciousness? How much progress has neuroscience made in doing so? What challenges does it face? How can it meet those challenges? What is the philosophical significance of its findings? This entry addresses these and related questions.

To bridge the gulf between brain and consciousness, we need neural data, computational and psychological models, and philosophical analysis to identify principles to connect brain activity to conscious experience in an illuminating way. This entry will focus on identifying such principles without shying away from the neural details. The notion of neuroscientific explanation here conceives of it as providing informative answers to concrete questions that can be addressed by neuroscientific approaches. Accordingly, the theories and data to be considered will be organized around constructing answers to two questions (see [section 1.4](#) for more precise formulations):

- **Generic Consciousness:** How might neural properties explain when a state is conscious rather than not?
- **Specific Consciousness:** How might neural properties explain what the content of a conscious state is?

A challenge for an objective science of consciousness is to dissect an essentially subjective phenomenon. As investigators cannot experience another subject's conscious states, they rely on the subject's observable behavior to track consciousness. Priority is given to a subject's introspective reports as these express the subject's take on her experience. Introspection thus provides a fundamental way, perhaps *the* fundamental way, to track consciousness. That said, consciousness pervasively influences human behavior, so other forms of behavior beyond introspective reports provide a window on consciousness. How to leverage disparate behavioral evidence is a central issue.

The term “neuroscience” covers those scientific fields whose explanations advert to the properties of neurons, populations of neurons, or larger parts of the nervous system.^[1] This includes, but is not limited to, psychologists' use of various neuroimaging methods to monitor the activity of tens of millions of neurons, computational theorists' modelling of biological and artificial neural networks, neuroscientists' use of electrodes inserted into brain tissue to record neural activity from individual or populations of neurons, and clinicians' study of patients with altered conscious experiences in light of damage to brain areas.

Given the breadth of neuroscience so conceived, an overview of sufficient depth must restrict breadth. On the neuroscience side, this review focuses on the central nervous system and the electrical properties of neurons, particularly in the cerebral cortex. On the side of consciousness, it focuses on perceptual consciousness, with emphasis on vision. This is not because visual consciousness is more important than other forms of consciousness. Rather, the level of detail in empirical work on vision often speaks more comprehensively to the issues that we shall confront.

That said, there are many forms of consciousness that we will not discuss. Some are covered in other entries such as split brain phenomena (see the entry on [the unity of consciousness](#),

section 4.1.1), animal consciousness (see the entry on [animal consciousness](#)), and neural correlates of the will and agency (see the entry on [agency](#), section 5). In addition, this entry will not discuss the neuroscience of consciousness in audition, olfaction or gustation; disturbed consciousness in mental disorders such as schizophrenia; conscious aspects of pleasure, pain and the emotions; the phenomenology of thought; the neural basis of dreams; and modulations of consciousness during sleep and anesthesia among other issues. These are important topics, and the principles and approaches highlighted in this discussion will apply to many of these domains.

- [1. Fundamentals](#)
 - [1.1 A Map of the Brain](#)
 - [1.2 Neurons and Brain](#)
 - [1.3 Access Consciousness and Phenomenal Consciousness](#)
 - [1.4 Generic and Specific Consciousness](#)
 - [1.5 The Hard Problem](#)
 - [1.6 Neural Correlates of Consciousness](#)
- [2. Methods for Tracking Consciousness](#)
 - [2.1 Introspection and Report](#)
 - [2.2 Access as a Condition on Phenomenal Consciousness](#)
 - [2.3 No Report Paradigms](#)
 - [2.4 Confidence and Metacognitive Approaches](#)
 - [2.5 The Intentional Action Inference](#)
 - [2.6 Vegetative State and the Intentional Action Inference](#)
- [3. Neurobiological Theories of Consciousness](#)
 - [3.1 The Global Neuronal Workspace](#)
 - [3.2 Recurrent Processing Theory](#)
 - [3.3 Higher-Order Theory](#)
 - [3.4 Information Integration Theory](#)
 - [3.5 Frontal or Posterior?](#)
- [4. Neuroscience of Generic Consciousness: Unconscious Vision as Case Study](#)
 - [4.1 Unconscious Vision and the Two Visual Streams](#)
 - [4.2 Blindsight](#)
 - [4.3 Unconscious Vision and the Intentional Action Inference](#)
- [5. Specific Consciousness](#)
 - [5.1 Neural Representationalism](#)
 - [5.2 The Contrast Strategy: Binocular Rivalry](#)
 - [5.3 Neural Stimulation](#)
 - [5.3.1 Visual Motion Perception](#)
 - [5.3.2 Tactile Vibration](#)
 - [5.3.3 Visual Word Form](#)
 - [5.4 Neural Representation and Probabilistic Coding](#)
- [6. The Future](#)
- [Bibliography](#)
- [Academic Tools](#)
- [Other Internet Resources](#)
- [Related Entries](#)

1. Fundamentals

1.1 A Map of the Brain

It will be helpful to grasp the basic anatomy of the brain. A central distinction concerns the difference between the cerebral *cortex* and the *subcortex*. The cortex is divided into two hemispheres, left and right, each of which can be divided into four lobes: frontal, parietal, temporal and occipital.



Figure 1: The Cerebral Cortex and Salient Areas

Figure Legend: The four lobes of the primate brain, shown for the left hemisphere. Some areas of interest are highlighted. Abbreviations: PFC: prefrontal cortex; IT: inferotemporal cortex; S1: primary somatosensory cortex; IPL and SPL: Inferior and Superior Parietal Lobule; MST: medial superior temporal visual area; MT: middle temporal visual area (also called V5 in humans); V1: primary visual cortex; V2-V6 consist of additional visual areas.

The discussion that follows will highlight specific areas of cortex including the *prefrontal cortex* that will figure in discussions of confidence ([section 2.2](#)), the global neuronal workspace ([section 3.1](#)) and higher order theories ([section 3.3](#)); the *dorsal visual stream* that projects into parietal cortex and the *ventral visual stream* that projects into temporal cortex including visual areas specialized for processing places, faces, and word forms (see sections [2.6](#) on places, [4.1](#) on visual agnosia and [5.3.3](#) on seeing words); *primary somatosensory cortex* S1 (see section [5.3.2](#) on tactile sensation); and early visual areas in the occipital cortex including the *primary visual area*, V1 (see sections [4.2](#) on blindsight and [5.2](#) on binocular rivalry) and a motion sensitive area V5, also known as the *middle temporal area* (MT; [section 5.3.1](#) on seeing motion). Beneath the cortex is the subcortex, divided into the forebrain, midbrain, and hindbrain, which covers many regions although our discussion will largely touch on the *superior colliculus* and the *thalamus*, two areas that play an important role in visual processing.

1.2 Neurons and Brain

A neuroscientific explanation of consciousness adduces properties of the brain, typically the brain's electrical properties. A salient phenomenon is neural signaling through *action potentials* or *spikes*. A spike is a large change in electrical potential across a neuron's cellular membrane which can be transmitted between neurons that form a neural circuit. For a sensory neuron, the spikes it generates are tied to its *receptive field*. For example, in a visual neuron, its receptive field is understood in spatial terms and corresponds to that area of external space where an appropriate stimulus triggers the neuron to spike. Given this correlation between stimulus and spikes, the latter carries information about the former. Information processing in sensory systems involves processing of information regarding stimuli within receptive fields.

Which electrical property provides the most fruitful explanatory basis for understanding consciousness remains an open question. For example, when looking at a single neuron, neuroscientists are not interested in spikes per se but the spike rate generated by a neuron per unit time. Yet spike rate is one among many potentially relevant neural properties. Consider the blood oxygen level dependent signal (BOLD) measure in functional magnetic resonance imaging (fMRI). The BOLD signal is a measure of changes in blood flow in the brain when neural tissue is active and is postulated to be a function of electrical properties at a different

part of a neuron than that part tied to spikes. Specifically, given a *synapse* which is the connection between two neurons to form a basic circuit motif, spikes are tied to the *presynaptic* side while the BOLD signal is thought to be a function of electrical changes on the *postsynaptic* side (signal flow is from pre to post). Furthermore, neuroscientists are typically not interested in the response of a single neuron but rather that of a population of neurons, of whole brain regions, and/or their interactions. Higher order properties of brain regions include the *local field potential* generated by populations of neurons and correlated activity such as *synchrony* between activity in different areas of the brain (neural oscillations were postulated to be central to consciousness by Crick & Koch 1990).

The number of neural properties potentially relevant to explaining mental phenomena is dizzying. This review focuses on the facts that neural sensory systems carry information about the subject's environment and that neural information processing can be tied to a notion of *neural representation*. How precisely to understand neural representation is itself a vexed question (Cao 2012, 2014; Shea 2014), but we will deploy a simple assumption with respect to spikes which can be reconfigured for other properties: where a sensory neuron generates spikes when a stimulus is placed in its receptive field, the spikes carry information about the stimulus (strictly speaking, about a random variable). *Information* as used in neuroscience is typically not a semantic notion, but bearing in mind that caveat, it will simplify matters to speak of a sensory neuron's activity as *representing* the relevant aspect of the stimulus that drives the neuron's response (e.g., direction of motion or intensity of a sound).^[2] This way of speaking is imprecise, so we shall return to neural representation in the final section when discussing how neural representations might explain conscious contents.

1.3 Access Consciousness and Phenomenal Consciousness

An important distinction separates *access consciousness* from *phenomenal consciousness* (Block 1995). "Phenomenal consciousness" refers to those properties of experience that correspond to *what it is like* for a subject to have those experiences (Nagel 1974 and the entry on [qualia](#)). These features are apparent to the subject from the inside, so tracking them arguably depends on one's having the relevant experience. For example, one understands what it is like to see red only if one has visual experiences of the relevant type (Jackson 1982).

As noted earlier, introspection is the first source of evidence about consciousness. Introspective reports bridge the subjective and objective. They serve as a behavioral measure that expresses the subject's own take on what it is like for her in having an experience. While there have been recent concerns about the reliability or empirical usefulness of introspection (Schwitzgebel 2011; Irvine 2012a), there are plausibly many contexts where introspection is reliable (Spener 2015; see Irvine 2012b for an extended discussion of introspection in consciousness science; for philosophical theories, see Smithies & Stoljar 2012).

Introspective reports demonstrate that the subject can access the targeted conscious state. That is, the state is *access-conscious*: it is accessible for use in reasoning, report, and the control of action. Talk of access-consciousness must keep track of the distinction between *actual* access versus *accessibility*. When one reports on one's conscious state, one accesses the state. Thus, access consciousness provides much of the evidence for empirical theories of consciousness. Still, it seems plausible that a state can be conscious even if one does not access it in report so long as that state is *accessible*. One *can* report it. Access-consciousness

is usually defined in terms of this dispositional notion of accessibility.

We must also consider the type of access/accessibility. Block's original characterization of access-consciousness emphasized accessibility in terms of the *rational* control of behavior, so we can summarize his account as follows:

A representation is access-conscious if it is poised for free use in reasoning and for direct "rational" control of action and speech.

Rational access contrasts with a broader conception of *intentional* access that takes a mental state to be access-conscious if it can inform goal-directed or intentional behavior including behavior that is not rational or done for a reason. This broader notion allows for additional measurable behaviors as relevant in assessing phenomenal consciousness especially in non-linguistic animals. So, if access provides us with evidence for phenomenal consciousness, this can be (a) through introspective reports; (b) through rational behavior, (c) through intentional behavior including nonrational behavior. Indeed, in certain contexts, reflexive behavior provides measures of consciousness ([section 2.3](#)).

1.4 Generic and Specific Consciousness

Explanations answer specific questions. Two questions regarding phenomenal consciousness frame this entry: Generic and Specific. The first focuses on a mental state's *being conscious* in general as opposed to not being conscious. Call this property *generic consciousness*, a property shared by specific conscious states such as seeing a red rose, feeling a touch, or being angry. Thus:

Generic Consciousness: What conditions/states N of nervous systems are necessary and (or) sufficient for a mental state, M , to be conscious as opposed to not?

If there is such an N , then the presence of N entails that an associated mental state M is conscious and (or) its absence entails that M is unconscious.

A second focus will be on the content of consciousness, say that associated with a perceptual experience's being of some perceptible X . This yields a question about specific contents of consciousness such as experiencing the motion of an object (see [section 5.3.1](#)) or a vibration on one's finger (see [section 5.3.2](#)):

Specific Consciousness: What neural states or properties are necessary and/or sufficient for a conscious perceptual state to have content X rather than Y ?

Expanding a bit, perceptual states have intentional content and specifying that content is one way of describing what that state is like. In introspectively accessing her conscious states, a subject reports what her experience is like by reporting what she experiences. Thus, the subject can report seeing an object moving, changing color, or being of a certain kind (e.g., a mug) and thus specify the content of the perceptual state. Discussion of specific consciousness will focus on perceptual states described as consciously perceiving X where X can be a particular such as a face, a property such as the frequency of a vibration or a proposition, say seeing that an object moves in a certain direction.

Many philosophers take perceiving X to be perceptually representing X . Intentional content on this reading is a semantic notion, and this suggests a linking principle tying conscious content to the brain: Perceptually representing X is based on neural representations of X . The

“based on” locution hedges on the precise relation between neural contents and conscious contents, but a simple relation is identity: neural content is perceptual content.^[3] This principle expresses a type of *first-order representationalism* about phenomenal content, a topic we return to in [section 5](#); see also the entry on [representational theories of consciousness](#). The principle explains specific consciousness by appeal to neural representational content.

Posing a clear question involves grasping its possible answers and in science, this is informed by identifying experiments that can provide evidence for such answers. The emphasis on necessary and sufficient conditions in our two questions indicates how to empirically test specific proposals. To test sufficiency, one would aim to produce or modulate a certain neural state and then demonstrate that consciousness of a certain form arises. To test necessity, one would eliminate a certain neural state and demonstrate that consciousness is abolished. Notice that such tests go beyond mere correlation between neural states and conscious states (see [section 1.6](#) on neural correlates and sections [2.2](#), [4](#) and [5](#) for tests of necessity and sufficiency).

In many experimental contexts, the underlying idea is *causal* necessity and sufficiency. However, if

$A=B$

, then A 's presence is also necessary and sufficient for B 's presence since they are identical. Thus, a brain lesion that eliminates N and thereby eliminates conscious state S might do so either because N is causally necessary for S or because

$N=S$

. An intermediate relation is that N *constitutes* or *grounds* S which does not imply that

$N=S$

(see the entry on [metaphysical grounding](#)). Whichever option holds for S , the first step is to find N , a neural correlate of consciousness ([section 1.6](#)).

In what follows, to explain generic consciousness, various global properties of neural systems will be considered ([section 3](#)) as well as specific anatomical regions that are tied to conscious versus unconscious vision as a case study ([section 4](#)). For specific consciousness, fine-grained manipulations of neural representations will be examined that plausibly shift and modulate the contents of perceptual experience ([section 5](#)).

1.5 The Hard Problem

David Chalmers presents the hard problem as follows:

It is undeniable that some organisms are subjects of experience. But the question of how it is that these systems are subjects of experience is perplexing. Why is it that when our cognitive systems engage in visual and auditory information-processing, we have visual or auditory experience: the quality of deep blue, the sensation of middle C? How can we explain why there is something it is like to entertain a mental image, or to experience an emotion? It is widely agreed that experience arises from a physical basis, but we have no good explanation of why and how it so arises. Why should physical processing give rise to a rich inner life at all? It seems objectively unreasonable that it should, and yet it does. If any problem qualifies

as *the* problem of consciousness, it is this one. (Chalmers 1995: 212)

The Hard Problem can be specified in terms of generic and specific consciousness (Chalmers 1996). In both cases, Chalmers argues that there is an inherent limitation to empirical explanations of phenomenal consciousness in that empirical explanations will be fundamentally either structural or functional, yet phenomenal consciousness is not reducible to either. This means that there will be something that is left out in empirical explanations of consciousness, a missing ingredient (see also the *explanatory gap* [Levine 1983]).

There are different responses to the hard problem. One response is to sharpen the explanatory targets of neuroscience by focusing on what Chalmers calls *structural features* of phenomenal consciousness, such as the spatial structure of visual experience, or on the contents of phenomenal consciousness. When we assess explanations of specific contents of consciousness, these focus on the neural representations that fix conscious contents. These explanations leave open exactly what the secret ingredient is that shifts a state with that content from unconsciousness to consciousness. On ingredients explaining generic consciousness, a variety of options have been proposed (see [section 3](#)), but it is unclear whether these answer the Hard Problem, especially if any answer to that the Problem has a necessary condition that the explanation must conceptually close off certain possibilities, say the possibility that the ingredient could be added yet consciousness not ignite as in a zombie, a creature without phenomenal consciousness (see the entry on [zombies](#)). Indeed, some philosophers deny the hard problem (see Dennett 2018 for a recent statement). Patricia Churchland urges: “Learn the science, do the science, and see what happens” (Churchland 1996: 408).

Perhaps the most common attitude for neuroscientists is to set the hard problem aside. Instead of explaining the existence of consciousness in the biological world, they set themselves to explaining generic consciousness by identifying neural properties that can turn consciousness on and off and explaining specific consciousness by identifying the neural representational basis of conscious contents.

1.6 Neural Correlates of Consciousness

Modern neuroscience of consciousness has attempted to explain consciousness by focusing on *neural correlates of consciousness* or NCCs (Crick & Koch 1998, 2003). Identifying correlates is an important first step in understanding consciousness, but it is an early step. After all, correlates are not necessarily explanatory in the sense of answering specific questions posed by neuroscience. That one does not want a mere correlate was recognized by Chalmers who defined an NCC as follows:

An NCC is a minimal neural system N such that there is a mapping from states of N to states of consciousness, where a given state of N is sufficient under conditions C , for the corresponding state of consciousness. (Chalmers 2000: 31)

Similarly, Christof Koch and others speak of “the minimal neural mechanisms jointly sufficient for any one specific conscious experience” (Koch et al. 2016: 307). One wants a minimal neural system since, crudely put, the brain is sufficient for consciousness but to point this out is hardly to explain consciousness even if it provides an answer to questions about sufficiency. There is, of course, much more to be said that is informative even if one does not drill down to a “minimal” neural system which is tricky to define or operationalize (see Chalmers 2000 for discussion; for criticisms of the NCC approach, see Noë & Thompson

2004; for criticisms of Chalmers' definition, see Fink 2016).

The emphasis on sufficiency goes beyond mere correlation, as neuroscientists aim to answer more than the question: What is a neural correlate for conscious phenomenon *C*? For example, Chalmers' and Koch's emphases on sufficiency indicate that they aim to answer the question: What neural phenomenon is sufficient for consciousness? Perhaps more specifically: What neural phenomenon is causally sufficient for consciousness? Accordingly, talk of "correlate" is unfortunate since sufficiency implies correlation but not vice versa. For example, correlation does not imply causal sufficiency, so not every correlate will be explanatory in the sense of answering Chalmers' and Koch's question. After all, assume that the NCC is type identical to a conscious state. Then many neural states will correlate with the conscious state: (1) the NCC's typical effects, (2) its typical causes, and (3) states that are necessary for the NCCs obtaining (e.g., the presence of sufficient oxygen). Thus, some correlated effects will not be explanatory. For example, citing the effects of consciousness will not provide causally sufficient conditions for consciousness.

While many theorists are focused on explanatory correlates, it is not clear that the field has always grasped this, something recent theorists have been at pains to emphasize (Graaf, Hsieh, & Sack 2012; Aru et al. 2012; Koch et al. 2016). In other contexts, neuroscientists speak of the neural *basis* of a phenomenon where the basis does not simply correlate with the phenomenon but also explains and possibly grounds it. However, talk of correlates is entrenched in the neuroscience of consciousness, so one must remember that the goal is to find the subset of neural correlates that are explanatory, in answering concrete questions. Reference to neural correlates in this entry will always mean neural explanatory correlate of consciousness (on occasion, I will speak of these as the neural *basis* of consciousness). That is, our two questions about specific and generic consciousness focus the discussion on neuroscientific theories and data that contribute to explaining them. This project allows that there are limits to neural explanations of consciousness, precisely because of the explanatory gap (Levine 1983).

2. Methods for Tracking Consciousness

Since studying consciousness requires that scientists track its presence, it will be important to examine various methods used in neuroscience to isolate and probe conscious states.

2.1 Introspection and Report

Scientists primarily study phenomenal consciousness through subjective reports. We can treat reports in neuroscience as conceptual in that they express how the subject recognizes things to be, whether regarding what they perceive (perceptual or observational reports, as in psychophysics) or regarding what mental states they are in (introspective reports). A report's conceptual content can be conveyed in words or through another overt behavior whose significance is fixed within an experimental context (e.g., pressing a button to indicate that a stimulus is present or that one sees it). Subjective reports of conscious states draw on distinctively first-personal access to that state. The subject *introspects*.

Introspection raises questions that science has only recently begun to address systematically in large part because of longstanding suspicion regarding introspective methods. Early modern psychology relied on introspection to parse mental processes but ultimately abandoned it due to worries about introspection's reliability (Feest 2012; Spener

forthcoming). Introspection was judged to be an unreliable method for addressing questions about mental processing. To address these worries, we must understand how introspection works, but unlike many other psychological capacities, we lack detailed models of introspection of consciousness (Feest 2014; for theories of introspecting propositional attitudes, see Nichols & Stich 2003; Goldman 2006; Heal 1996; Carruthers 2011). This makes it difficult to address long-standing worries about introspective reliability regarding consciousness.

In science, questions raised about the reliability of a method are answered by calibrating and testing the method. This calibration has not been done with respect to the type of introspection commonly practiced by philosophers. Such introspection has revealed many phenomenal features that are targets of active investigation such as the phenomenology of mineness (Ehrsson 2009), sense of agency (Bayne 2011; Vignemont & Fournier 2004; Marcel 2003; Horgan, Tienson, & Graham 2003); transparency (Harman 1990; Tye 1992), self-consciousness (Kriegel 2003: 122), cognitive phenomenology (Bayne & Montague 2011); phenomenal unity (Bayne & Chalmers 2003) among others. A scientist might worry that philosophical introspection merely recycles rejected methods of a century ago, indeed without the stringent controls or training imposed by earlier psychologists. How can we ascertain and ensure the reliability of introspection in the empirical study of consciousness?

One way to address the issue is to connect introspection to attention. Philosophical conceptions of introspective attention construe it as capable of directly focusing on phenomenal properties and experiences. As this idea is fleshed out, however, it is clearly not a form of attention studied by cognitive science, for the posited direct introspective attention is neither perceptual attention nor what psychologists call internal attention (e.g., the retrieval of thought contents as in memory recollection; (Chun, Golomb, & Turk-Browne 2011)). Calibrating introspection as it is used in the science of consciousness would benefit from concrete models of introspection, models we lack (see Spener 2015, for a general form of calibration).

One philosophical tradition links introspection to perceptual attention, and this allows construction of concrete models informed by science. The intuitive idea is expressed in Harman's observation:

Look at a tree and try to turn your attention to intrinsic features of your visual experience. I predict you will find that the only features there to turn your attention to will be features of the presented tree, including relational features of the tree "from here". (Harman 1990: 39)

This is related to a proposal inspired by Gareth Evans (1982): in introspecting perceptual states, say judging that one sees an object, one draws on the same perceptual capacities used to answer the question whether the object is present. In introspection, one then appends a further concept of "seeing" to one's perceptual report.^[4] Thus, instead of simply reporting that a red stimulus is present, one reports that one sees the red stimulus. Paradoxically, introspection relies on externally directed perceptual attention, but as noted earlier, identifying what one perceives is a way of characterizing what one's perception is like, so this "outward" perspective can provide information about the inner. Further, the advantage of this proposal is that questions of reliability come down to questions of the reliability of psychological capacities that can be empirically assessed, say perceptual, attentional and conceptual reliability. For example, Peters and Lau (2015) showed that accuracy in judgments about the visibility of a stimulus, the introspective measure, coincided with accuracy in judgments about stimulus properties, the objective measure (see also Rausch &

Zehetleitner 2016).

Introspection can be reliable. Successful clinical practice relies on accurate introspection as when dealing with pain or correcting blurry vision in optometry. The success of medical interventions suggests that patient reports of these phenomenal states are reliable. Further, in many of the examples to be discussed, the perceptual attention-based account provides a plausible cognitive model of introspection. Subjects report on what they perceptually experience by attending to the object of their experience, and where perception and attention are reliable, a plausible hypothesis is that their introspective judgments will be reliable as well. Accordingly, I assume the reliability of introspection in the empirical studies to be discussed. Still, given that no scientist should assert the reliability of a method without calibration, introspection must be subject to the same standards. There is more work to be done.

2.2 Access as a Condition on Phenomenal Consciousness

Introspection illustrates a type of cognitive access, for a state that is introspected is access conscious. This raises a question that has epistemic implications: is access consciousness necessary for phenomenal consciousness? If it is not, then there can be phenomenal states that are not access conscious, so are in principle not reportable. That is, phenomenal consciousness can *overflow* access consciousness (Block 2007).

Access is tied to attention. For example, the Global Workspace theory of consciousness understands consciousness in terms of access ([section 3.1](#)) where the accessibility of a perceived object requires attention to the object (attention puts an object, namely a representation of it, in the global workspace). So, the necessity of attention for phenomenal consciousness is entailed by the necessity of access for phenomenal consciousness.^[5] In contrast, recurrent processing theory holds that there can be phenomenal states that are not accessible ([section 3.2](#)).

Many scientists of consciousness take there to be evidence for no phenomenal consciousness without access and little if no evidence of phenomenal consciousness outside of access. An important set of studies focuses on the thesis that attention is a necessary gate for phenomenal consciousness, where attention is tied to access. Call this the *gatekeeping* thesis. To assess that evidence, we must ask: what is attention? An uncontroversial conception of attention is that it is subject selection of a target to inform task performance (Wu 2014b). The experimental studies thought to support the necessity of attention for consciousness draw on this conception.^[6] For example, in *inattentional blindness* paradigms (Mack & Rock 1998), attention is directed by asking subjects to perform a task on target *T* while a surprising stimulus *S* is presented. This approach tests necessity by ensuring through task performance that the subject is not attending to *S*. One then measures whether the subject is aware of *S* by observing whether the subject reports it. If the subject does not report *S*, then the hypothesis is that failure of attention to *S* explains the failure of conscious awareness of *S* and hence the failure of report.

A well-known experiment asks subjects to attend to the number of passes of a ball thrown by players in white shirts while ignoring a second ball passed by players in black shirts (Simons & Chabris 1999). During the task, a person in a gorilla costume walks across the scene. Half of the subjects fail to notice and report the gorilla, this being construed as evidence for the absence of visual awareness of the gorilla. Hence, failure to attend to the gorilla is said to

render subjects phenomenally blind to it. Similar claims are made in *change blindness* where subjects fail to detect the difference between two similar pictures (Simons & Ambinder 2005), the *attentional blink* where subjects fail to detect a stimulus presented immediately after detecting a prior stimulus (Dux & Marois 2009; Martens & Wyble 2010), and *hemispatial neglect* where patients fail to report objects in a part of their visual field that they cannot attend to due to brain lesions.

The gatekeeping thesis holds that attention is necessary for consciousness, so that removing it from a target eliminates consciousness of it. Yet there is a flaw in the methodology. To report a stimulus, one must attend to it, i.e., select it for report. The experimental logic requires eliminating attention to a stimulus *S* to test if attention is a necessary condition for consciousness (e.g., eliminating attention to the gorilla by distracting the subject with the ball). Yet even if the subject were conscious of *S*, when attention to *S* is eliminated, one can *predict* that the subject will fail to act (report) on *S* since attention is necessary for report. The observed results are actually consistent with the subject being conscious of *S* without attending to it, and thus are neutral between overflow and gatekeeping. Instead, the experiments concern parameters for the capture of attention and not consciousness.

While those antagonistic to overflow have argued that it is not empirically testable (M.A. Cohen & Dennett 2011), gatekeeping might be equally untestable. After all, to test the necessity of attention for consciousness, we must eliminate attention to a target while gathering evidence for the absence of consciousness. Yet if gathering evidence for consciousness requires attention, then in fulfilling the conditions for testing the necessity of attention, we undercut the access needed to substantiate the absence of consciousness (Wu 2017; for a monograph length discussion of attention and consciousness, see Montemayor & Haladjian 2015). How then can we gather the required evidence to assess competing theories?

2.3 No Report Paradigms

One response to is to draw on *no-report paradigms* which measure reflexive behaviors correlated with conscious states to provide a window on the phenomenal that is independent of access (Lumer & Rees 1999; Tse et al. 2005). For example, Frässle et al. (2014) demonstrate that certain ocular reflexes are correlated with perceptual experience in binocular rivalry (Naber, Frässle, & Einhäuser 2011). They presented subjects either with stimuli moving in opposite directions or stimuli of different luminance values, one stimulus in each pair presented separately to each eye. This induces binocular rivalry, an alternation in which of the two stimuli is visually experienced (see [section 5.2](#)). Where the stimuli involved motion, subjects demonstrated *optokinetic nystagmus* where the eye slowly moves in the direction of the stimulus and then makes a fast, corrective saccade (ballistic eye movement) in the opposite direction. Frässle et al. observed that optokinetic nystagmus tracked the perceived direction of the stimulus as reported by the subject. Similarly, for stimuli of different luminance, the pupils would dilate, being wider for dimmer stimuli, and narrower for brighter stimuli, again correlating with subjective reports of the intensity of the stimulus.

No-report paradigms use reflexive responses to track the subject's perceptual experience in the absence of explicit (conceptualized) report. They seem to provide a way to track phenomenal consciousness even when access is eliminated. This would broaden the evidential basis for consciousness beyond introspection and indeed, beyond intentional behavior (our "broad" conception of access). Yet no-report paradigms do not circumvent introspection (Overgaard & Fazekas 2016). For example, optokinetic nystagmus's usefulness

depends on validating its correlation with alternating experience *given subjective reports*. Once it is validated, monitoring this reflex can provide a way to substitute for subjective reports within that paradigm. One cannot, however, simply extend the use of no-report paradigms outside the behavioral contexts within which the method is validated. With each new experimental context, we must revalidate the measure with introspective report.

Can we use no report paradigms to address whether access is necessary for phenomenal consciousness? A likely experiment would be one that validates no-report correlates for some conscious phenomenon P in a concrete experimental context C . With this validation in hand, one then eliminates accessibility and attention with respect to P in C . If the no-report correlate remains, would this clearly support overflow? Perhaps, though gatekeeping theorists likely will respond that the result does not rule out the possibility that phenomenal consciousness disappears with access consciousness *despite* the no-report correlate remaining. For example, the reflexive response and phenomenal consciousness might have a common cause that remains even if phenomenal consciousness is selectively eliminated by removing access.

2.4 Confidence and Metacognitive Approaches

Given worries about calibrating introspection, researchers have asked subjects to provide a different metacognitive assessment of conscious states via reports about *confidence* (Grimaldi, Lau, & Basso 2015; Pouget, Drugowitsch, & Kepecs 2016). A standard approach is to have subjects perform a task, say perceptual discrimination of a stimulus, and then indicate how confident they are that their perceptual judgment was accurate. This judgment about perception can be assessed for accuracy by comparing the metacognitive judgment with perceptual performance (for discussion of formal methods such as metacognitive signal detection theory, see Maniscalco & Lau 2012). Related paradigms include post-decision wagering where subjects place wagers on specific responses as a way of estimating their confidence (Persaud, McLeod, & Cowey 2007; but see Dienes & Seth 2010).

How is metacognitive assessment of performance tied to consciousness? The metacognitive judgment reflects introspective assessment of the quality of perceptual states and can provide information about the presence of consciousness. For example, Peters and Lau (2015) tested whether normal subjects have unconscious vision under conditions of visual masking, a method that seems to eliminate conscious vision to a stimulus but allows for accurate visually guided behavior to it (Breitmeyer & Ogmen 2006). They presented stimuli in two temporal “windows” under masking but where the stimulus was present in only one window and a “blank” present in the other. If subjects accurately respond to the stimulus but showed no difference in metacognitive confidence in respect of the quality of perception of the target versus of the blank, this would provide evidence of the absence of consciousness in vision (effectively, blindsight in normal subjects; [section 4.2](#)). Interestingly, Peters and Lau found no evidence for unconscious vision in their specific paradigm.

One concern with metacognitive approaches is that they also rely on introspection (Rosenthal 2018; see also Sandberg et al. 2010; Dienes & Seth 2010). If metacognition relies on introspection, does it not accrue all the disadvantages of the latter? One advantage of metacognition is that it allows for psychophysical analysis. There has also been work done on metacognition and its neural basis. Studies with non-human primates and rodents have begun to shed light on neural processing for metacognition (for a review, see Grimaldi, Lau, & Basso 2015; Pouget, Drugowitsch, & Kepecs 2016). From animal studies, one theory is that metacognitive information regarding perception is already present in perceptual areas that

guide observational judgments, and these studies implicate parietal cortex (Kiani & Shadlen 2009; Fetsch et al. 2014) or the superior colliculus (Kim & Basso 2008; but see Odegaard et al. 2018). Alternatively, information about confidence might be read out by other structures, say prefrontal cortex (see [section 3.3](#) on Higher-Order Theory; also the entry on [higher order theories of consciousness](#)).

2.5 The Intentional Action Inference

Metacognitive and introspective judgments result from intentional action, so why not look at intentional action, broadly construed, for evidence of consciousness? Often, when subjects perform perception guided actions, we infer that they are relevantly conscious. It would be odd if a person cooks dinner and then denies having seen any of the ingredients. That they did something intentionally provides evidence that they were consciously aware of what they acted on. An emphasis on intentional action embraces a broader evidential basis for consciousness. Consider the *Intentional Action Inference* to phenomenal consciousness:

If some subject acts intentionally, where her action is guided by a perceptual state, then the perceptual state is phenomenally conscious.

An epistemic version takes the action to provide good evidence that the state is conscious. Notice that introspection is typically an intentional action so it is covered by the inference. In this way, the Inference effectively levels the evidential playing field: introspective reports are simply one form among many types of intentional actions that provide evidence for consciousness. Those reports are not privileged.

The intentional action inference and no-report paradigms highlight the fact the science of consciousness has largely restricted its behavioral data to one type of intentional action, introspection. What is the basis of privileging one intentional action over others? Consider the calibration issue. For many types of intentional action deployed in experiments, scientists can calibrate performance by objective measures such as accuracy. This has not been done for introspection of consciousness, so scientists have privileged an uncalibrated measure over a calibrated one. This seems empirically ill-advised. On the flip side, one worry about the intentional action inference is that it ignores guidance by unconscious perceptual states (see sections [4](#) and [5.3.1](#)).

2.6 Vegetative State and the Intentional Action Inference

The Intentional Action Inference is operative when subjective reports are not available. For example, it is deployed in arguing that some patients diagnosed as being in the vegetative state are conscious (Shea & Bayne 2010; see also Drayson 2014).

A patient in the vegetative state appears at times to be wakeful, with cycles of eye closure and eye opening resembling those of sleep and waking. However, close observation reveals no sign of awareness or of a ‘functioning mind’: specifically, there is no evidence that the patient can perceive the environment or his/her own body, communicate with others, or form intentions. As a rule, the patient can breathe spontaneously and has a stable circulation. The state may be a transient stage in the recovery from coma or it may persist until death. (Working Party RCP 2003: 249)

Vegetative state patients are not clinically comatose but fall short of being in a “minimally conscious state”. Unlike vegetative state patients, minimally conscious state patients

seemingly perform intentional actions.

Recent work suggests that some patients diagnosed as in the vegetative state are conscious. Owen et al. (2006) used fMRI to demonstrate correlated activity in such patients in response to commands to deploy imagination. In an early study, a young female patient was scanned by fMRI while presented with three auditory commands: “imagine playing tennis”, “imagine visiting the rooms in your home”, “now just relax”. The commands were presented at the beginning of a thirty-second period, alternating between imagination and relax commands. The patient demonstrated similar activity when matched to control subjects performing the same task: sustained activation of the supplementary motor area (SMA) was observed during the motor imagery task while sustained activation of the parahippocampal gyrus including the parahippocampal place area (PPA) was observed during the spatial imagery task. Later work reproduced this result in other patients and in one patient, the tasks were used as a proxy for “yes”/ “no” responses to questions (Monti et al. 2010; for a review, see Fernández-Espejo & Owen 2013). Note that these tasks probe *specific* contents of consciousness by monitoring neural correlates of conscious imagery.

Several authors (Greenberg 2007; Nachev & Husain 2007) have countered that the observed activity was an automatic, non-intentional response to the command sentences, specifically to the words “tennis” and “house”. In normal subjects, reading action words is known to activate sensorimotor areas (Pulvermüller 2005). Owen et al. (2007), responded that the sustained activity over thirty-seconds made an automatic response less likely than an intentional response. One way to rule out automaticity is to provide the patient with different sentences such as “do *not* imagine playing tennis” or “*Sharlene* was playing tennis”. Owen et al. (2007) demonstrated that presenting “*Sharlene* was playing tennis” to a normal subject did not induce the same activity as when the subject obeyed the command “imagine playing tennis”, but the result was not duplicated in patients.

Owen et al. draw on a neural correlate of imagination, a mental action. Arguing that the neural correlate provides evidence of the patient’s executing an intentional action, they invoke a version of the Intentional Action Inference to argue that performance provides evidence for specific consciousness tied to the information carried in the brain areas activated. Of note, experiments stimulating the parahippocampal place area induces seeming hallucinations of places (Mégevand et al. 2014).^[7]

3. Neurobiological Theories of Consciousness

Recall that the Generic Consciousness question asks:

What conditions/states N of nervous systems are necessary and/or sufficient for a mental state, M , to be conscious as opposed to not?

Victor Lamme notes:

Deciding whether there is phenomenality in a mental representation implies putting a boundary—drawing a line—between different types of representations... We have to start from the intuition that consciousness (in the phenomenal sense) exists, and is a mental function in its own right. That intuition immediately implies that there is also *unconscious* information processing. (Lamme 2010: 208)

It is uncontroversial that there is unconscious information processing, say processing

occurring in a computer. What Lamme means is that there are conscious and unconscious mental states (representations). For example, there might be visual states of seeing X that are conscious or not ([section 4](#)).

In what follows, the theories discussed provide higher level neural properties that are necessary and/or sufficient for generic consciousness of a given state. To provide a gloss on the hypotheses: For the Global Neuronal Workspace, entry into the neural workspace is necessary and sufficient for a state or content to be consciousness. For Recurrent Processing Theory, a type of recurrent processing in sensory areas is necessary and sufficient for perceptual consciousness, so entry into the Workspace is not necessary. For Higher-Order Theories, the presence of a higher-order state tied to prefrontal areas is necessary and sufficient for phenomenal experience, so recurrent processing in sensory areas is not necessary nor is entry into the workspace. For Information Integration Theories, a type of integration of information is necessary and sufficient for a state to be conscious.

3.1 The Global Neuronal Workspace

One explanation of generic consciousness invokes the *global neuronal workspace*. Bernard Baars first proposed the global workspace theory as a cognitive/computational model (Baars 1988), but we will focus on the neural version of Stanislas Dehaene and colleagues: a state is conscious when and only when it (or its content) is present in the global neuronal workspace making the state (content) globally *accessible* to multiple systems including long-term memory, motor, evaluational, attentional and perceptual systems (Dehaene, Kerszberg, & Changeux 1998; Dehaene & Naccache 2001; Dehaene et al. 2006). Notice that the previous characterization does not commit to whether it is phenomenal or access consciousness that is being defined.

Access should be understood as a relational notion:

A system X accesses content from system Y iff X uses that content in its computations/processing.

The *accessibility* of information is then defined as its potential access by other systems. Dehaene (Dehaene et al. 2006) introduces a threefold distinction: (1) neural states that carry information that is not accessible (*subliminal* information); (2) states that carry information that is accessible but not accessed (not in the workspace; *preconscious* information); and (3) states whose information is accessed by the workspace (*conscious* information) and is globally accessible to other systems. So, a necessary and sufficient condition for a state's being conscious rather than not is the access of a state or content by the workspace, making that state or content accessible to other systems. Hence, only states in (3) are conscious.



Figure 2. The Global Neuronal Workspace

Figure Legend: The top figure provides a neural architecture for the workspace, indicating the systems that can be involved. The lower figure sets the architecture within the six layers of the cortex spanning frontal and sensory areas, with emphasis on neurons in layers 2 and 3. Figure reproduced from Dehaene, Kerszberg, and Changeux 1998. Copyright (1998) National Academy of Sciences.

The global neuronal workspace theory ties access to brain architecture. It postulates a cortical

structure that involves workspace neurons with long-range connections linking systems: perceptual, mnemonic, attentional, evaluational and motoric.

What is the global workspace in neural terms? Long-range workspace neurons within different systems can constitute the workspace, but they should not necessarily be identified with *the* workspace. A subset of workspace neurons becomes the workspace when they exemplify certain neural properties. What determines which workspace neurons constitute the workspace at a given time is the activity of those neurons given the subject's current state. The workspace then is not a rigid neural structure but a rapidly changing neural network, typically only a proper subset of all workspace neurons.

Consider then a neural population that carries content p and is constituted by workspace neurons. In virtue of being workspace neurons, the content p is accessible to other systems, but it does not yet follow that the neurons then constitute the global workspace. A further requirement is that workspace neurons are (1) put into an active state that must be sustained so that (2) the activation generates a *recurrent* activity between workspace systems. Only when these systems are recurrently activated are they, along with the units that access the information they carry, constituents of the workspace. This activity accounts for the idea of global *broadcast* in that workspace contents are accessible to further systems. Broadcasting explains the idea of consciousness as *for the subject*: globally broadcasted content is accessible for the subject's use in informing behavior.

The global neuronal workspace theory provides an account of *access* consciousness but what of phenomenal consciousness? The theory predicts widespread activation of a cortical workspace network as correlated with phenomenal conscious experience, and proponents often appeal to imaging results that reveal widespread activation when consciousness is reported (Dehaene & Changeux 2011). There is, however, a potential confound. We track phenomenal consciousness by access in introspective report, so widespread activity during reports of conscious experience correlates with both access and phenomenal consciousness. Correlation cannot tell us whether the observed activity is the basis of phenomenal consciousness or of access consciousness in report (Block 2007). This remains a live question for as discussed in [section 2.2](#), we do not have empirical evidence that overflow is false.

To eliminate the confound, experimenters ensure that performance does not differ between conditions where consciousness is present and where it is not. Where this was controlled, widespread activation was not clearly observed (Lau & Passingham 2006). Still, the absence of observed activity by an imaging technique does not imply the absence of actual activity for the activity might be beyond the limits of detection of that technique. Further, there is a general concern about the significance of null results given that neuroscience studies focused on prefrontal cortex are typically underpowered (for discussion, see Odegaard, Knight, & Lau 2017).

3.2 Recurrent Processing Theory

A different explanation ties perceptual consciousness to processing independent of the workspace, with focus on *recurrent* activity in sensory areas. This approach emphasizes properties of first-order neural representation as explaining consciousness. Victor Lamme (2006, 2010) argues that recurrent processing is necessary and sufficient for consciousness. Recurrent processing occurs where sensory systems are highly interconnected and involve feedforward and feedback connections. For example, forward connections from primary visual area V1, the first cortical visual area, carry information to higher-level processing

areas, and the initial registration of visual information involves a forward sweep of processing. At the same time, there are many feedback connections linking visual areas (Felleman & Van Essen 1991), and later in processing, these connections are activated yielding dynamic activity within the visual system.

Lamme identifies four stages of normal visual processing:

- Stage 1: Superficial feedforward processing: visual signals are processed locally within the visual system.
- Stage 2: Deep feedforward processing: visual signals have travelled further forward in the processing hierarchy where they can influence action.
- Stage 3: Superficial recurrent processing: information has traveled back into earlier visual areas, leading to local, recurrent processing.
- Stage 4: Widespread recurrent processing: information activates widespread areas (and as such is consistent with global workspace access).

Lamme holds that recurrent processing in Stage 3 is necessary and sufficient for consciousness. Thus, what it is for a visual state to be conscious is for a certain recurrent processing state to hold of the relevant visual circuitry. This identifies the crucial difference between the global neuronal workspace and recurrent processing theory: the former holds that recurrent processing at Stage 4 is necessary for consciousness while the latter holds that recurrent processing at Stage 3 is sufficient. Thus, recurrent processing theory affirms phenomenal consciousness without access by the global neuronal workspace. In that sense, it is an overflow theory (see [section 2.2](#)).

Why think that Stage 3 processing is sufficient for consciousness? Given that Stage 3 processing is not accessible to introspective report, we lack introspective evidence for sufficiency. Lamme appeals to experiments with brief presentation of stimuli such as letters where subjects are said to report seeing more than they can identify in report (Lamme 2010). For example, in George Sperling's partial report paradigm (Sperling 1960), subjects are briefly presented with an array of 12 letters (e.g., in 300 ms presentations) but are typically able to report only three to four letters even as they claim to see more letters (but see Phillips 2011). It is not clear that this is strong motivation for recurrent processing, since the very fact that subjects can report seeing more letters shows that they have some access to them, just not access to letter identity.

Lamme also presents what he calls *neuroscience arguments*. This strategy compares two neural networks, one taken to be sufficient for consciousness, say the processing at Stage 4 as per Global Workspace theories, and one where sufficiency is in dispute, say recurrent activity in Stage 3. Lamme argues that certain features found in Stage 4 are also found in Stage 3 and given this similarity, it is reasonable to hold that Stage 3 processing suffices for consciousness. For example, both stages exhibit recurrent processing. Global neuronal workspace theorists can allow that recurrent processing in stage 3 is correlated, even necessary, but deny that this activity is explanatory in the relevant sense of identifying sufficient conditions for consciousness.

It is worth reemphasizing the empirical challenge in testing whether access is necessary for phenomenal consciousness (sections [2.1–3](#)). The two theories return different answers, one requiring access, the other denying it. As we saw, the methodological challenge in testing for the presence of phenomenal consciousness independently of access remains a hurdle for both theories.

3.3 Higher-Order Theory

A long-standing approach to conscious states holds that one is in a conscious state if and only if one relevantly represents oneself as being in such a state. For example, one is in a conscious visual state of seeing a moving object if and only if one suitably represents oneself being in that visual state. This higher-order state, in representing the first-order state that represents the world, results in the first order state's being conscious as opposed to not. The intuitive rationale for such theories is that if one were in a visual state but in no way aware of that state, then the visual state would not be conscious. Thus, to be in a conscious state, one must be aware of it, i.e., represent it (see the entry on [higher order theories of consciousness](#); Rosenthal 2002). Higher-order theories merge with empirical work by tying high-order representations with activity in prefrontal cortex which is taken to be the neural substrate of the required higher-order representations. On certain higher-order theories, one can be in a conscious visual state even if there is no visual system activity, so long as one represents oneself as being in that state.

The focus on prefrontal cortex allows for empirical tests of the higher-order theory as against other accounts (Lau & Rosenthal 2011). For example, on the higher-order theory, lesions to prefrontal cortex should affect consciousness (Kozuch 2014), testing the necessity of prefrontal cortex for consciousness. Against higher-order theories, some reports claim that patients with prefrontal cortex surgically removed maintain preserved perceptual consciousness (Boly et al. 2017). This would lend support to recurrent processing theories that hold that prefrontal cortical activity is not necessary for consciousness. It is not clear, however, that the interventions succeeded in removing *all* of prefrontal cortex, leaving perhaps sufficient frontal areas needed to sustain consciousness (Odegaard, Knight, & Lau 2017). Bilateral suppression of prefrontal activity using transcranial magnetic stimulation also seems to selectively impair visibility as evidenced by metacognitive report (Rounis et al. 2010). Furthermore, certain syndromes and experimental manipulations suggest consciousness in the absence of appropriate sensory processing as predicted by some higher-order accounts (Lau & Brown forthcoming), a claim that coheres with the theory's sufficiency claims.

3.4 Information Integration Theory

Information Integration Theory of Consciousness (IIT) draws on the notion of *integrated information*, symbolized by Φ , as a way to explain generic consciousness (Tononi 2004, 2008). IIT defines integrated information in terms of the effective information carried by the parts of the system in light of its causal profile. For example, we can focus on a part of the whole circuit, say two connected nodes, and compute the effective information that can be carried by this microcircuit. The system carries *integrated information* if the effective informational content of the whole is greater than the sum of the informational content of the parts. If there is no partitioning where the summed informational content of the parts equals the whole, then the system as a whole carries integrated information and it has a positive value for Φ . Intuitively, the interaction of the parts adds more to the system than the parts do alone.

IIT holds that a non-zero value for Φ implies that a neural system is conscious, with more consciousness going with greater values for Φ . For example, Tononi has argued that the human cerebellum has a low value for Φ despite there being four to five times the number of neurons in the cerebellum versus in human cortex. On IIT, what matters is the presence of

appropriate connections and not the number of neurons.

A potential problem for IIT is that it treats many things to be conscious which are *prima facie* not (in [Other Internet Resources](#), see Aaronson 2014a; for striking counterexamples and Aaronson 2014b with a response from Tononi).^[8] That said, the idea of integrated information might be useful for neuroscience, but we must show that invoking Φ can explain generic consciousness.^[9]

3.5 Frontal or Posterior?

In recent years, one way to frame the debate between theories of generic consciousness is whether the “front” or the “back” of the brain is crucial. Using this rough distinction allows us to draw the following contrasts: Recurrent processing theories focus on sensory areas (in vision, the “back” of the brain) such that where processing achieves a certain recurrent state, the relevant contents are conscious even if no higher-order thought is formed or no content enters the global workspace. Similarly, proponents of IIT have recently emphasized a “posterior hot zone” covering parietal and occipital areas (Boly et al. 2017) as a neural correlate for consciousness, as they speculate that this zone may have the highest value for Φ . For certain higher-order thought theories, having a higher-order state, supported by prefrontal cortex, without corresponding sensory states can suffice for conscious states. In this case, the front of the brain would be sufficient for consciousness. Finally, the global neuronal workspace, drawing on workspace neurons that are present across brain areas to form the workspace, might be taken to straddle the difference, depending on the type of conscious state involved. They require entry into the global workspace such that neither sensory activity nor a higher order thought on its own is sufficient, i.e., neither just the front nor the back of the brain.

The point of talking coarsely of brain anatomy in this way is to highlight the neural focus of each theory and thus, of targets of manipulation as we aim for explanatory neural correlates in terms of what is necessary and/or sufficient for generic consciousness. What is clear is that once theories make concrete predictions of brain areas involved in generic consciousness, neuroscience can test them.

4. Neuroscience of Generic Consciousness: Unconscious Vision as Case Study

Since generic consciousness is a matter of a state’s being conscious or not, we can examine work on specific types of mental state that shift between being conscious or not and isolate neural substrates. Work on unconscious vision provides an informative example. In the past decades, scientists have argued for unconscious seeing and investigated its brain basis especially in *neuropsychology*, the study of subjects with brain damage. Interestingly, if there is unconscious seeing, then the intentional action inference must be restricted in scope since some intentional behaviors might be guided by unconscious perception ([section 2.5](#)). That is, the existence of unconscious perception blocks a direct inference from perceptually guided intentional behavior to perceptual consciousness. The case study of unconscious vision promises to illuminate more specific studies of generic consciousness along with having repercussions for how we attribute conscious states.

4.1 Unconscious Vision and the Two Visual Streams

Since the groundbreaking work of Leslie Ungerleider and Mortimer Mishkin (1982), scientists divide primate cortical vision into two streams: dorsal and ventral (for further dissection, see Kravitz et al. 2011). The dorsal stream projects into the parietal lobe while the ventral stream projects into the temporal lobe (see [Figure 1](#)). Controversy surrounds the functions of the streams. Ungerleider and Mishkin originally argued that the streams were functionally divided in terms of *what* and *where*: the ventral stream for categorical perception and the dorsal stream for spatial perception. David Milner and Melvyn Goodale (1995) have argued that the dorsal stream is for *action* and the ventral stream for “*perception*”, namely for guiding thought, memory and complex action planning (see Goodale & Milner 2004 for an engaging overview). There continues to be debate surrounding the Milner and Goodale account (Schenk and McIntosh 2010) but it has strongly influenced philosophers of mind.

Substantial motivation for Milner and Goodale’s division draws on lesion studies in humans. Lesions to the dorsal stream do not seem to affect conscious vision in that subjects are able to provide accurate reports of what they see (but see Wu 2014a). Rather, dorsal lesions can affect visual-guidance of action with *optic ataxia* being a common result. Optic ataxic subjects perform inaccurate motor actions. For example, they grope for objects, yet they can accurately report the object’s features (for reviews, see Andersen et al. 2014; Pisella et al. 2009; Rossetti, Pisella, & Vighetto 2003). Lesions in the ventral stream disrupt normal conscious vision, yielding visual agnosia, an inability to see visual form or to visually categorize objects (Farah 2004).

Dorsal stream processing is said to be unconscious. If the dorsal stream is critical in the visual guidance of many motor actions such as reaching and grasping, then those actions would be guided by unconscious visual states. The visual agnostic patient DF provides critical support for this claim.^[10] Due to carbon monoxide poisoning, DF suffered focal lesions largely in the ventral stream spanning the *lateral occipital complex* that is associated with processing of visual form (high resolution imaging also reveals small lesions in the parietal lobe, [James et al. 2003]). Like other visual agnostics with similar lesions, DF is at chance in reporting aspects of form, say the orientation of a line or the shape of objects. Nevertheless, she retains color and texture vision. Strikingly, DF can generate accurate visually guided action, say the manipulation of objects along specific parameters: putting an object through a slot or reaching for and grasping round stones in a way sensitive to their center of mass. Simultaneously, DF denies seeing the relevant features and, if asked to verbally report them, she is at chance. In this dissociation, DF’s verbal reports give evidence that she does not visually experience the features to which her motor actions remain sensitive.

What is uncontroversial is that there is a division in explanatory neural correlates of visually guided behavior with the dorsal stream weighted towards the visual guidance of motor movements and the ventral stream weighted towards the visual guidance of conceptual behavior such as report and reasoning (see [section 5.3.3](#) on manipulation of seeing words via ventral stream stimulation). A substantial further inference is that consciousness is segregated away from the dorsal stream to the ventral stream. How strong is this inference?

Recall the intentional action inference. In performing the slot task, DF is doing something intentionally and in a visually guided way. For control subjects performing the task, we conclude that this visually guided behavior is guided by conscious vision. Indeed, a folk-psychological assumption might be that consciousness informs mundane action (Clark 2001;

for a different perspective see Wallhagen 2007). Since DF shows similar performance on the same task, why not conclude that she is also visually conscious? Presumably, one hesitates because DF's introspective reports clash with the intentional action inference. DF denies seeing features she is visually sensitive to in action. Should introspection then trump intentional action in attributing consciousness?

Two issues are worth considering. The first is that introspective reports involve a specific type of intentional action guided by the experience at issue. One type of intentional behavior is being prioritized over another in adjudicating whether a subject is conscious. What is the empirical justification for this prioritization? The second issue is that DF is possibly unique among visual agnosics. It is a substantial inference to move from DF to a general claim about the dorsal stream being unconscious in neurotypical individuals (see Mole 2009 for arguments that consciousness does not divide between the streams and Wu 2013 for an argument for unconscious visually guided action in normal subjects). What this shows is that the methodological decisions that we make regarding how we track consciousness are substantial in theorizing about the neural bases of conscious and unconscious vision.

4.2 Blindsight

A second neuropsychological phenomenon also highlighting putative unconscious vision is *blindsight* which results from lesions in primary visual cortex (V1) typically leading to blindness over the part of visual space contralateral to the sight of the lesion (Weiskrantz 1986). For example, left hemisphere V1 deals with right visual space, so lesions in left V1 lead to deficits in seeing the right side of space. Subjects then report that they cannot see a visual stimulus in the affected visual space. Strikingly, these clinically blind subjects can draw on information from the “unseen” stimulus to visually inform behavior regarding it, often in striking ways. For example, a blindsight patient with bilateral damage to V1 (i.e., in both hemispheres) who is blind across the visual field can walk down a hallway around obstacles he reports being unable to see (Gelder et al. 2008). Blindsight patients see in the sense of visually discriminating the stimulus to act on it yet deny that they see it. The contrast between behavior and report leads to the paradoxical term, “blindsight”. Like DF, blindsighters show a dissociation between certain actions and report, but unlike DF, they do not spontaneously respond to relevant features but must be encouraged to generate behaviors towards them.^[11]

The neuroanatomical basis of blindsight capacities remains unclear. Certainly, the loss of V1 deprives later cortical visual areas of a normal source of visual information. Still, there are other ways that information from the eye bypasses V1 to provide inputs to later visual areas. Alternative pathways include the superior colliculus (SC), the lateral geniculate nucleus (LGN) in the thalamus, and the pulvinar as likely sources.



Figure 3: Subcortical Pathways and their Connection to Cortical Vision (from Urbanski, Coubard, & Bourlon 2014)

Figure Legend: The front of the head is to the left, the back of the head is to the right. One should imagine that the blue-linked regions are above the orange-linked regions, cortex above subcortex. V4 is assigned to the base of the ventral stream; V5, called area MT in nonhuman primates, is assigned to the base of the dorsal stream.

The latter two have direct *extrastriate* projections (projections to visual areas in the occipital lobe outside of V1) while the superior colliculus synapses onto neurons in the LGN and pulvinar which then connect to extrastriate areas (Figure 3). Which of these provide for the basis for blindsight remains an open question though all pathways might play some role (Covey 2010; Leopold 2012). If blindsight involves nonphenomenal, unconscious vision, then these pathways would be a substrate for it, and a functioning V1 might be necessary for normal conscious vision.

Campion et al. (1983) raised an important alternative explanation: blindsight subjects in fact have severely degraded conscious vision but merely report on them with low confidence. In their reports, blindsight subjects feel like they are guessing about stimuli they can objectively discriminate. Campion et al. drew on *signal detection theory*, which emphasizes two determinants of detection behavior: *perceptual sensitivity* and *response criterion*. A subject's ability to visually detect a signal will depend partly on how well her visual system extracts the signal from noise (sensitivity) but also on the criterion that is set as a threshold for response. Consider trying to detect something moving in the brush at twilight versus at noon. In the latter, the signal will be greatly separated from noise (the object will be easier to detect) while in the former, the signal will not be (the object will be harder to detect). Yet in either case, one might operate with a conservative response criterion, say because one is afraid to be wrong. Thus, even if the signal is detectable, one might still opt not to report on it given a *conservative bias* (criterion), say if one is in the twilight scenario and would be ridiculed for "false alarms", i.e., claiming the object to be present when it is not.

Campion et al. hypothesized that blindsight patients are conscious in that they are aware of visual signal where discriminability is low (cf. the twilight condition). Further, blindsight patients are more conservative in their response so will be apt to report the absence of a signal by saying that they do not see the relevant stimulus even though the signal is there, and they can detect it, as verified by their above chance visually guided behavior. This possibility was explicitly tested by Azzopardi and Covey (1997) with the well-studied blindsight patient, GY. They compared blindsight performance with normal subjects at threshold vision using signal detection measures and found that with respect to motion stimuli, the difference between discrimination and detection used to argue for blindsight can be explained by changes in response criterion, as Campion et al. hypothesized. That is, GY's claim that he does not see the stimulus is due to a conservative criterion and not to a detection incapacity. Interestingly, for static stimuli, his response criterion did not change but his sensitivity did, as if he was tapping into two different visual processing mechanisms in each task (for an alternative explanation based on shifting response criterion, see Ko & Lau 2012).

In introspecting, what concepts are available to subjects will determine their sensitivity in report. In many studies with blindsight, subjects are given a binary option: do you see the stimulus or do you not see it? The concern is that the *do not see* option would cover cases of degraded consciousness that subjects might be unwilling to classify as seeing due to a conservative response criterion. So, what if subjects are given more options for report? Ramsøy and Overgaard (2004; see also Overgaard et al. 2006) provided subjects with four categories for introspective report: *no experience*; *brief glimpse*; *almost clear experience*; *clear experience*. Using this *perceptual awareness scale*, they found that subjects' objective performance tracked their introspective reports where performance was at chance when subjects reported no visual experience. As visibility increased, so did performance. When the scale was used with a blindsight patient (Overgaard et al. 2008), no above chance performance was detected when the subject reported no visual experience (see also Mazzi, Bagattini, & Savazzi 2016 for further evidence). A live alternative hypothesis is that

blindsight does not present a case of unconscious vision, but of degraded conscious vision with a conservative response bias that affects introspection. At the very least, the issue depends on how introspection is deployed, a topic that deserves further attention (see Phillips 2016 for further discussion of blindsight).

4.3 Unconscious Vision and the Intentional Action Inference

Blindsight and DF show that damage to specific regions of the brain disrupts normal visual processing, yet subjects can access visual information in preserved visual circuits to inform behavior despite failing to report on the relevant visual contents. The received view is that these subjects demonstrate unconscious vision. One implication is that the normal processing in the ventral stream, tied to normal V1 activity, plays a necessary role in normal conscious vision. Another is that dorsal stream processing or visual stream processing that bypasses V1 via subcortical processing yields only unconscious visual states. This points to a set of networks that begin to provide an answer to what makes visual states conscious or not. An important further step will be to integrate these results with the general theories noted earlier ([section 3](#)).

Still, the complexities of the empirical data bring us back to methodological issues about tracking consciousness and the following question: What behavioral data should form the basis of attributions of phenomenal consciousness? The intentional action inference is used in a variety of cases to attribute conscious states, yet the results of the previous sections counsel us to be wary of applying that inference widely. After all, some intentional behavior might be unconsciously guided.

In the case of DF, we noted that unlike many other visual agnosics, she can direct motor actions towards stimuli that she cannot explicitly report and which she denies seeing. In her case, we prioritize introspective reports over intentional action as evidence for unconscious vision. Yet, one might take a broader view that vision for action is always conscious and that what DF vividly illustrates is that some visual contents (dorsal stream) are tied directly to performance of intentional motor behavior and are not directly available to conceptual capacities deployed in report. In contrast, other aspects of conscious vision, supported by the ventral stream, are directly available to guide reports. This functional divergence is explained by the anatomical division in cortical visual processing.

For some time now, these striking cases have been taken as clear cases of unconscious vision and if this hypothesis is correct, the work has begun to identify visual areas critical for creating seeing, sometimes conscious and sometimes not. The neuroanatomy demonstrates that visually-guided behavior has a complex neural basis involving cortical and subcortical structures that demonstrate a substantial level of specialization. Understanding consciousness and unconsciousness in vision will need to be sensitive to the complexities of the underlying neural substrate.

5. Specific Consciousness

We turn to experimental work on specific consciousness:

Specific Consciousness: What neural states or properties are necessary and/or sufficient for a conscious perceptual state to have content *X* rather than *Y*?

In this section, we examine attempts to address claims about necessity and sufficiency by manipulation of the contents of consciousness through direct modulation of neural representational content.

5.1 Neural Representationalism

In thinking about neural explanations of specific consciousness, namely the contents of consciousness, we will provisionally assume a type of first-order representationalism about phenomenal content, namely that such content supervenes on neural content (see the entry on [representational theories of consciousness](#)). One strong position would be that phenomenal content is *identical* to appropriate neural content. A weaker correlation claim affirms only supervenience: no change in phenomenal content without a change in neural content. This neural representationalism allows us to link phenomenal properties to the brain via linking neural contents to perceptual contents.

5.2 The Contrast Strategy: Binocular Rivalry

A common approach, the *contrast strategy*, enjoins experimentalists to identify relevant correlates for some phenomenon *P* by contrasting cases where *P* is present from cases where *P* is not. Work on binocular rivalry illustrates this strategy (among many reviews, see Tong, Meng, & Blake 2006; Blake, Brascamp, & Heeger 2014). When each eye receives a different image simultaneously, the subject does not see both, say one stimulus overlapping the other. Rather, visual experience alternates between them. Call this *phenomenal alternation*. An initial restatement of our question about specific consciousness in respect of binocular rivalry is:

Specific Rivalry: What neural property is necessary and/or sufficient for phenomenal alternation in binocular rivalry in condition *C*?

That is, empirical theories aim to explain how visual content alternates in binocular rivalry. ^[12] Notice that this is a question about specific rather than generic consciousness, as the contrast is not between a state's being conscious versus not but about the contrast between two conscious states with different contents.

Neural explanations of binocular rivalry concern competition at some level of visual processing: (a) "interocular" competition between *monocular* neurons early in the visual system, namely visual neurons that receive input from only one eye or (b) competition between *binocular* neurons later in the visual system, namely neurons that receive input from both eyes. The winner of competition fixes which stimulus the subject experiences at a given time. Some of the earliest electrophysiological studies (Leopold & Logothetis 1996; Logothetis, Leopold, & Sheinberg 1996) on awake behaving monkeys supported later binocular processing as the neural basis of binocular rivalry. Processing in later (inferotemporal cortex, IT; see [figure 1](#)) rather than earlier visual areas (V1 or V2) were observed to be best correlated to the monkey's reported perception based on the monkey's stimuli-specific response. In contrast, imaging studies in humans suggested that neural activity in V1 did correlate with alternation. For example, Polonsky et al. used fMRI to demonstrate that V1 activity to competing stimuli tracked perception (Polonsky et al. 2000; but see Maier et al. 2008).

Recent accounts have taken binocular rivalry as resulting from processes at multiple levels

(Wilson 2003; Freeman 2005; Tong, Meng, & Blake 2006). For example, when the two competing stimuli have parts that can be fused into a coherent stimulus, as when half of a picture is presented to each eye, the subject can perceive the fusion, integrating content from each eye (Kovács et al. 1996; Ngo et al. 2000). This suggests that binocular rivalry can be sensitive to global properties of the stimulus (see Baker & Graf 2009). What unifies the mechanisms, perhaps, is the function of resolving a conflict generated by the stimuli.

Assume that some neural process R resolves interocular competition: when R resolves competition between stimuli X and Y in favor of X , then the subject is phenomenally conscious of X rather than Y and vice versa. Notice that R has the same “gating” function for any stimuli X and Y that are subject to binocular rivalry. So, while the presence of R can explain why the subject is having one conscious visual experience rather than another, R is not tied to a specific content. This suggests that in answering the question about rivalry, we will at best be identifying a necessary but not sufficient condition for a conscious visual state having a content X . R is a general gate for consciousness (cf. attention in global workspace theory).

A narrower explanation of specific consciousness would identify the specific neural representations that explain a conscious state’s having the specific content X (rather than Y). By the representationalism assumption, this will involve identifying neural representations with the same content, X . Focusing on a gate in explaining alternation in rivalry stops short of identifying those representations. Still, binocular rivalry can provide a useful method for isolating neural populations that carry relevant content. In principle, for any stimulus type of interest, X (e.g., faces, words, etc.), so long as X is subject to binocular rivalry, we can use rivalry paradigms to isolate brain areas that carry the relevant information that correlate with the subject’s perceiving X . That would allow us to identify potential candidates for the neural basis of conscious content.

5.3 Neural Stimulation

There are limited opportunities to manipulate human brain activity in a targeted way. Recent use of transcranial magnetic stimulation to activate or suppress neural activity has provided illumination, but such interventions are coarse-grained. Ultimately, to locate an explanatory correlate for specific conscious contents, we will need more fine-grained interventions in brain tissue. In humans, such opportunities are generally confined to manipulation before surgical interventions, say for brain tumors or epilepsy (see [section 5.3.3](#) for work with epilepsy patients).

In the middle of the last century, neurosurgeon Wilder Penfield and colleagues performed a set of direct electrical microstimulations during preoperative procedures (Penfield & Perot 1963), and in certain cases induced hallucinations by stimulating primary sensory cortices such as V1 or S1 (see [figure 1](#)). This provided evidence that endogenous activity could be causally sufficient for phenomenal experiences. Penfield’s interventions, however, were not based on fine-grained targeting of specific neural representations. As Cohen and Newsome note,

Penfield’s approach failed to generate substantial new insights into the neural basis of perception and cognition...because the gross electrical activation elicited by surface electrodes could not be related mechanistically to the information being processed within the excited neural tissue. (Cohen & Newsome 2004: 1)

A different approach begins with a more detailed understanding of underlying neural representations tied to different brain regions. For example, the fusiform face (FFA) area appears to be necessary for normal human face experience in that lesions in FFA lead to *prosopagnosia*, the inability to see faces even if one can see their parts. FFA is part of a larger network that is important in visual processing of faces (Behrmann & Plaut 2013). Recently, microstimulation of FFA in an awake human epilepsy patient induced visual distortions of actual faces as opposed to other objects (Parvizi et al. 2012). Alterations of visual experience were also reported during microstimulating the parahippocampal place (PPA) area in an awake pre-operative epileptic patient that induced visual hallucinations of scenes (Mégevand et al. 2014). PPA is the same area that showed activation in vegetative state patients when they putatively imagined walking around their home ([section 2.5](#)).



Figure 4. Ventral Stream Areas

A view from the bottom of cortex with location of areas FFA, PPA and LO identified. Occipital cortex is on the bottom. LO is lesioned in the visual agnostic patient, DF (see [section 4.1](#)). This figure is modified from figure 1 of Behrmann and Plaut 2013, kindly provided by Marlene Behrmann and used with her permission.

It is worth noting that many neuroscientists of vision take themselves to be investigating seeing in the ordinary sense, one that implies consciousness, but very few of them would characterize their work as about consciousness. That said, their work is of direct relevance to our understanding of specific consciousness even if it is not always characterized as such.

An important approach in visual neuroscience was articulated by A.J. Parker and William Newsome in “Sense and the Single Neuron” (1998) via “principles” to connect electrophysiological data about information processing to perception (for a recent discussion, see Ruff & Cohen 2014). To probe the neural basis of perception, neuroscientists need to explanatorily link neural data to the subject’s perception that guides behavior. The experimenter must ensure that recorded neural content correlates with perceptual content and not just response. Further, manipulation of the neurons carrying information should affect perception: inducing appropriate neural activity should shift perceptual response while abolishing or reducing that activity should eliminate or reduce perceptual response as measured in behavior. These proposals address concerns about necessity and sufficiency.

The intentional action inference is applicable (or at least its evidential version):

If some subject acts intentionally, where her action is guided by a perceptual state, then that state is phenomenally conscious.

We will consider the strength of this inference in three cases. The first case, visual motion perception, introduces the principles that guide the manipulation of neural content while the second case concerns tactile experience of vibration. These cases involve experiments with non-human primates, so we lack introspective reports. The final case concerns direct manipulation of the human brain along with introspective reports.

These experiments involve microstimulation of small populations of neurons that are targeted precisely because of their informational content. Microstimulation involves injecting a small current from the tip of an electrode inserted into brain tissue that directly stimulates nearby neurons or, through synaptic connections to other neurons, indirectly activates more distant

neurons (see Histed, Ni, & Maunsell 2013 for a review). It is assumed that neurons *tuned* in similar ways, that is neurons that respond to similar stimuli, tend to be interconnected, so microstimulation is taken to largely drive similarly tuned neurons.

5.3.1 Visual Motion Perception

We begin with visual motion perception in primates. Since the principles introduced here are central to much perceptual neuroscience and provide the basis for probing the link between neural representations and perceptual content, I examine it carefully. The salient question will be whether conscious experience is changed by the manipulations.

The work we shall discuss was done in awake behaving macaque monkeys. Visual area MT in the monkey brain (called V5 in humans) plays an important role in the visual experience of motion. MT is taken to lie in the dorsal visual stream ([figure 1](#)). Lesions that disrupt MT are known to cause *akinetopsia*, the inability to see motion. One patient with an MT (V5) lesion reported the following phenomenology: “people were suddenly here or there but I have not seen them moving” (Zihl, Von Cramon, & Mai 1983: 315). MT processing looks to be necessary for normal visual motion experience. Furthermore, MT neurons represent (carry information regarding) the direction of motion of visible stimuli: MT neurons are *tuned* for motion in specific directions with the highest firing rate for a specific direction of motion (for other functions and responses of MT, see Born & Bradley 2005). By placing motion stimuli in a neuron’s receptive field, scientists can map its tuning:



Figure 5. MT Neuron Tuning Curve

Figure Legend: Tuning of a neuron in MT showing a peak response in spiking rate at 0 degrees of motion. The dashed curve is generated when the animal is attending to the motion stimulus while it is in the receptive field (we shall not discuss the neural basis of attention, but see Wu 2014b, chap. 2, for a summary of the neuroscience of attention). The solid curve shows MT response when the animal is not attending to the motion stimulus in the receptive field. Figure from Lee & Maunsell 2009.

What is plotted is the activity of an MT neuron, in spikes per second, to a specific type of motion stimulus placed within its receptive field. How to relate a tuning curve to a determinate content is complicated. Since the neural response is not simply to one stimulus value, it is not obvious that the neuron should be taken to represent 0 degrees of motion, namely the value at its peak response. Indeed, theorists have noted that the tuning curve looks like a probability density function, and many now take neurons to have *probabilistic* content ([section 5.4](#)).

Experimenters have trained macaque monkeys to perform discrimination tasks reporting direction of motion. Typically, the monkey maintains fixation while the moving stimulus is placed within the receptive field of the recorded neuron. The monkey reports the direction of the stimulus by moving its eyes to a target that stands for either leftward or rightward motion (other behavioral reports can be generated such as moving a joystick). Provisionally, we apply the intentional action inference, so we assume that such reports are guided by conscious visual experience of the stimuli. Thus, *changes in behavior will be evidence for changes in conscious content*.

Early work suggested that the activity of a single neuron provides a strong correlate of the

animal's visually guided performance. This can be seen by plotting both the animal and the neuron's performance across different stimulus values. In these experiments, the value concerns the percent *coherence* of motion of a set of dots defined as the number of dots moving in the same direction (0% coherence being random motion; 100% being all dots moving in the same direction). In the first case, we construct a *psychometric curve* that plots the animal's percent correct reports relative to percent coherence of motion of the stimuli. As one might expect, percent correct reports drop as coherence drops, and the inflection point reflects where the subject is equally likely to indicate left or right motion. We can do the same for the neural activity of the neuron across the same stimulus values, a *neurometric curve*.

The experimentalist's window onto conscious experience is through behavior, the assumption being that report about motion correlates with perceptual experience. Correlation is assessed by asking the following question: would an ideal observer, using the activity of the neuron in question, be able to *predict* the animal's visually guided performance? Essentially, do the psychometric and neurometric curves overlap? Strikingly, yes. MT neurons were observed to predict the animal's behavior (Britten et al. 1992).



Figure 6

Figure Legend: Psychometric and Neurometric curves for a single MT neuron during performance of a motion direction detection task. Percent correct performance is plotted on the y-axis while percent motion coherence is plotted in a log scale on the x-axis. Figure modified from Ruff & Cohen 2014 and kindly provided by Doug Ruff.

This shows that the activity of a single MT neuron provides a neural correlate of the animal's visual discrimination of motion. Note that this is just a neural correlate of behavior. No one suggested that this neuron was causally sufficient for the behavior or for perception. Later results have suggested that individual neurons are not quite as sensitive as Britten suggested, but that small groups of MT neurons are sufficient to predict behavior (Cohen & Newsome 2009).

Earlier, we worried about mere correlates. To get causal or explanatory purchase, the content of the MT neurons correlated to the animal's behavior must be shown to contribute to perceptual guidance. This predicts that if we manipulate the content of the neurons, i.e., manipulate neural representations, then we should manipulate the content of the animal's visual experience of motion as reflected by predicted changes in behavior. This would be to test sufficiency with respect to specific consciousness.

Newsome and colleagues demonstrated that microstimulation of MT neurons shifted the animal's performance in predictable ways. Assume that neural population P , by encoding information about stimulus motion, can inform the subject's report of motion direction. This information is accessible for the control of behavior. Activation of P by microstimulation should shift behavior in a motion selective way correlated with the direction that P is tuned to (represents). Metaphorically, if a downstream control system is sensitive to the response of P when it generates behavior, then if we change P in a specific way, say amplifying its signal, we should change behavior in a way biased by P 's content. This was first demonstrated by Salzman et al. (1990). They inserted electrodes into MT and identified neurons tuned to a particular orientation. During a motion discrimination task, microstimulation of neurons with that tuning led to a shift in the psychometric curve as if that neuron was given more weight in

driving behavior.

In conditions of microstimulation relative to its absence, the monkey was more likely to report that there was motion in the stimulated neuron's preferred direction. In the original experiment, the psychophysical effect of microstimulation was equivalent to the addition of 7-20% coherence in the stimulus with respect to the neuron's preferred direction, depending on the experimental conditions. Further, as a test of necessity, a selective lesion of MT disrupted motion discrimination though the animals were able to recover some function suggesting that other visual information streams could be tapped so as to support performance (Newsome & Paré 1988).

Adopting the intentional action inference, one can conclude that the microstimulation shifted perceptual content (or again, that we have good evidence for this shift). That said, given our discussion of unconscious vision ([section 4](#)), another possibility is that MT microstimulation only changes unconscious visual representations. Newsome himself asked:

What is the conscious experience that accompanies the stimulation and the monkey's decision? Even if you knew everything about how the neurons encode and transmit information, you may not know what the monkey experiences when we stimulate his MT. (Singer 2006)

Clearly, having the monkey provide an introspective report would add evidential weight, but obtaining such reports from non-linguistic creatures is difficult. How can we get the animal to turn attention *inward* to their perceptual states in an experimental context?^[13]

5.3.2 Tactile Vibration

What of microstimulation in the absence of a stimulus? Might we induce *hallucinations* as Penfield did in his patients? Rather than the work from the Newsome group that modulated ongoing perceptual processing, the issue here is to create an internal signal that mimics perception. Romo et al. (1998) demonstrated that monkeys can carry out sensory tasks via activation triggered by microstimulation. The monkeys' task was to discriminate the frequency of two sequential "flutters" on their fingertips, that is, mechanical vibrations on the skin at specific frequencies. In an experimental trial, an initial *sample* flutter was presented for 500 ms and after a gap of 1-3 seconds, a second *test* flutter of either higher or lower frequency was presented. The animal reported whether the second test frequency was higher or lower than the sample.

The experimenters examined whether direct microstimulation in the absence of a stimulus could tap into the same neural representations that guided the animal's report. They isolated neurons in primary somatosensory cortex responsive to vibration frequency on the fingers (S1, the somatosensory homunculus discovered by Penfield [Penfield & Boldrey 1937]; see [figure 1](#)). The investigators then stimulated the same neurons in S1 in the absence of the test flutter, so used stimulation as a substitute for an actual vibration. Thus, the animal had to make a comparison between the frequency of a mechanical sample to either a subsequent (1) real mechanical test vibration (i.e., the good case with an actual stimulus) or (2) to a microstimulation test stimulus (i.e., the "hallucinatory" case where direct activation of the S1 neurons occurred in the absence of a stimulus). Romo et al. demonstrated that discrimination performance based either on mechanical stimulation or microstimulation was *equivalent*. In other words, the animals could match either mechanical or microstimulation to a remembered

mechanical sample.^[14]

In subsequent work (Romo et al. 2000), the investigators inverted the experiment, using the microstimulation as the *sample*. In this case, the animals had to remember the information conveyed by the microstimulation (effectively, a hallucination) and then compare it to either a subsequent (a) mechanically generated stimulation on the finger (actual test stimulus) or (b) a microstimulation of S1 as test (i.e., no stimulus). In both cases, performance was similar to earlier results. The striking finding is that behavior could be driven entirely by microstimulation. At least for the tactile stimulations at issue, the animal might have been in the *Matrix*!

One might think that the intentional action inference is stronger in this paradigm, given the elegant flipping of stimuli in Romo et al. 2000. Still, the authors comment:

This study, therefore, has directly established a strong link between neural activity and perception. However, we do not know yet whether microstimulation of the QA circuit in S1 elicits a subjective flutter sensation in the fingertips. This can only be explored by microstimulating S1 in an attending human observer. (Romo et al. 2000: 277)^[15]

Like Newsome, the authors reach for introspection. Yet they might undersell their result, for it seems that the animals are having a tactile hallucination: (a) Penfield showed that stimulation of primary sensory cortices like S1 induces hallucinations in humans; (b) action is engaged not at low stimulation of S1 in monkeys but only at higher level stimulation; (c) at that point, when the stimulation grabs their attention, the monkeys do what they were trained to do, namely discriminate stimuli, either with (d) just mechanical stimulation (normal experience), or (e) with a mix of mechanical and microstimulation or with just microstimulation; (f) given the behavioral equivalence of these three cases, one might then argue that if performance in the mechanical stimulation cases involves conscious tactile experience, then that same experience is involved in the other cases.

5.3.3 Visual Word Form

In humans, language is lateralized to the left hemisphere, and in visual word recognition, the left midfusiform gyrus (lmFG; sometimes referred to as the *visual word form area*, VWFA) is important for normal processing of visual word forms during reading. For example, damage to lmFG affects reading in adults (Gaillard et al. 2006; Behrmann & Shallice 1995) while learning novel words sharpens representations therein (Glezer et al. 2015).

Nevertheless, it is possible that lmFG is tuned to general visual form and is not specific to visual word forms. This is a common point of contention in addressing the function of various areas in the human visual system, notably the fusiform face area (FFA): is it a content specific area or is it more a general visual expertise area (Kanwisher 2000; Tarr & Gauthier 2000)?

In a recent study, Hirshorn et al. (2016) used microstimulation to disrupt processing in lmFG in human epileptic patients with preoperatively implanted electrodes spanning that area. These electrodes are used to map sites that will be surgically removed to relieve intractable seizures. Subjects read words or letters during actual or sham microstimulation. Crucially, stimulation in lmFG selectively disrupted word and letter reading but *not* general form perception. During lmFG stimulation, one subject when presented with “illegal” reported not seeing the word (see [Movie S1](#) in Other Internet Resources). Rather, she reported thinking of

different words (still, she did not report *seeing* different words). With “message” she reported thinking that an “n” was present. In a second patient, the identification of letters was completely disrupted (see [Movie S2](#) in Other Internet Resources). The patient reported seeing an “A” when presented with an “X”, and then an “F” and “H” when presented with “C”. The results suggest that normal word reading requires ImFG processing to parse linguistic forms. A plausible hypothesis is that microstimulation disrupted visual experience of specific types of stimuli, a test of necessity. That said, the first patient’s introspective reports of thinking rather than seeing words complicates matters.

Techniques for decoding information processing (using machine learning) suggests that processing in ImFG becomes more finely tuned to word form. Initially, ImFG represents a more *gist*-like representation but then develops more precise representations that individuate words of similar form. These converging results provide evidence that the areas stimulated carry information about word form such that in disrupting that activity, word perception was selectively disrupted.

Taken together, the three cases provide examples of detailed manipulations in different sensory modalities, animals, and contents that test for causal sufficiency and necessity across different levels of the sensory processing hierarchy, from early levels (e.g., S1) to mid-levels (MT) and finally to higher levels (ImFG or FFA). Working backwards, the experimental strategy is as follows: Given a perceptual experience with a specific content, one identifies neural correlates that carry equivalent content, say one’s seeing motion, feeling vibration, or parsing words linked to neural correlates processing information about motion, vibrational frequency and word forms. Next, fine-grained manipulations of the neural content are then correlated with related effects on perceptual experience. One issue that remains open is whether in tapping into neural processing by microstimulation, one has simply identified an earlier causal node in the neural processes that generate perceptual experience, there being more informative neural correlates later in the causal pathway. An important question is how one might identify the neural basis of the experience as opposed to its cause.

A couple of salient methodological challenges stand in the way of explaining specific consciousness. The first is that much of the detailed work will for the foreseeable future be done on non-human animals where introspective report is not easily available and where the intentional action inference will be essential. To strengthen that inference, we will need more detailed models that make plausible that conscious experience figures in the generation of the observed behavior. There is an experiment-theory circle that we must break into: we need a theory that supports the role of consciousness in behavior but the theory itself will be supported by behavioral data. How will we break into this cycle? The intentional action inference will be an important means of tracking consciousness, and the limits of its applicability must be investigated. The second challenge is that there is a need to individuate different kinds of explanatory correlates of consciousness, for some will be causes (upstream of conscious states), some will be enabling conditions, and some will be constituents of the state itself. Dividing these cases involves not just gathering data, but having a clear conceptual framework to draw distinctions in a principled way. Clearly joint philosophical and experimental work is needed.

5.4 Neural Representation and Probabilistic Coding

In invoking neural content, the assumption is that neural content mirrors perceptual content, so that if one is experiencing dots moving in a certain direction, there is a neural representation with the same content. This is a simplification and does not cohere with a

common current approach to neural content that takes it to be *probabilistic*. Consider again the tuning curve from an MT neuron, M ([figure 5](#)).

If asked to assign a determinate content to M , one might choose the value that corresponds to its peak response, here, 0° .^[16] This would be the most natural option if the tuning curve were essentially a sharp line at 0° , i.e., were the neuron only active for its preferred stimulus but otherwise not. Of course, that is not the neuron's response profile, so what is the content of the neural representation?

One approach that converges on a determinate content considers the activity of a neural population. To evaluate more information, the brain might integrate MT response by giving each MT neuron,

M

n

, one vote weighted according to the *strength* of its response. Thus, the tuning curve represents how strongly the neuron votes for its preferred value (the value at the peak). The votes are tallied by a downstream system, and the result can be represented as a *population vector* whose direction is understood to be the direction represented by the neural population. This specific approach was taken by Georgopoulos and colleagues to decode the direction of bodily movement from the activity of a population of motor neurons (Georgopoulos et al. 1982; for discussion of different coding approaches, see Pouget, Dayan, & Zemel 2003).

In recent perceptual neuroscience, an alternative picture of neural representational content often tied to *Bayesian approaches* to perceptual computation has gained traction (for accessible discussions, see Colombo & Seriès 2012; Rescorla 2015). On Bayesian models, extracting information from populations of (say) MT neurons does not yield a specific value of motion direction but rather a *probability density function*, across the space of possible motion directions (for a philosophical discussion of neural probabilistic codes, see Shea 2014). A key idea is not the generation of specific values as what neurons represent, say 0 degree motion as proposed earlier, but rather the conceptualization of the population response as reflecting uncertainty inherent in neural activity given noise.

If we plot the activity of all MT neurons responding to a specific motion stimulus, one hypothesis is that the population response codes the *likelihood*,

$P(r|s)$

, namely the conditional probability that one has the observed MT response r given the stimulus s . A Bayesian approach to neural population codes then understands neural processing to involve computation of the *posterior probability*,

$P(s|r)$

from the likelihood and prior knowledge of the probability of the stimulus

$P(s)$

in accordance with Bayes Theorem (this is normalized so that the probabilities sum to one). The details of Bayesian computation need not concern us since our main concern is with the possibility of neural content as probabilistic, something that seems counterintuitive relative to the approach illustrated by Georgopoulos and colleagues.

Knill and Pouget contrast the two approaches:

This is the basic premise on which Bayesian theories of cortical processing will succeed or fail—that the brain represents information probabilistically, by coding and computing with probability density functions or approximations to probability density functions...The opposing view is that neural representations are deterministic and discrete, which might be intuitive but also misleading. This intuition might be due to the apparent ‘oneness’ of our perceptual world and the need to ‘collapse’ perceptual representations into discrete actions, such as decisions or motor behaviors. (Knill & Pouget 2004)

How might a probabilistic account of neural representation affect our thinking about phenomenal consciousness via neural representationalism? Consider this possibility: What if probabilistic content is pervasive? Pouget, Dayan and Zemel note:

decoding is not an essential neurobiological operation because there is almost never a reason to decode the stimulus explicitly. Rather, the population code is used to support computations involving s , whose outputs are represented in the form of yet more population codes over the same or different collections of neurons. (2003: 385)

Put another way, the determinacy is apparent only at the output stage, the goal of processing. In the case of motor action, neural content is probabilistic until the actual movement when a determinate path is implemented in a specific movement trajectory.

Yet perceptual content does not *seem* probabilistic. This emphasizes a *prima facie* disconnect between current theories of neural content and those of phenomenal content. The linking principles we have deployed assume a specific view of neural content that might not cohere with current approaches to neural coding, leaving us with the challenge of explanatorily linking probabilistic content at the neural level with more determinate, nonprobabilistic content at the phenomenal level. One option is to find nonprobabilistic content at the neural level (e.g., as in the population vector approach). The other is to find probabilistic content at the phenomenal level (for related ideas, see Morrison 2017 and response by Denison 2017; also Block 2018). Either way, explanations of specific content will need to deal with this *prima facie* disconnect between phenomenal content as revealed by introspection and current theories of neural content.

6. The Future

Talk of the neuroscience of consciousness has, thus far, focused on the neural correlates of consciousness. Not all neural correlates are explanatory, so finding correlates is a first step in the neuroscience of consciousness. The next step involves manipulation of relevant correlates to test claims about sufficiency and necessity, as isolated in our two questions:

Generic Consciousness: What conditions/states N of nervous systems are necessary and (or) sufficient for a mental state, M , to be conscious as opposed to not?

Specific Consciousness: What neural states or properties are necessary and/or sufficient for a conscious perceptual state to have content X rather than Y ?

A productive neuroscience of consciousness requires that we understand the relevant neural properties at the right level of analysis. For generic consciousness, this will involve manipulation of relevant properties in a way that can avoid the access/phenomenal confound,

and recent work focuses on pitting the many theories we have considered against each other. For specific consciousness, the critical issue will be to understand neural representational content and to find ways to link experimentally and explanatorily neural content to phenomenal content. We have tools to manipulate neural contents to affect phenomenal content, and in doing so, we can begin to uncover the neural basis of conscious contents. There is much interesting work yet to be done, philosophically and empirically, and we can look forward to a productive interdisciplinary research program.

Bibliography

- Aglioti, Salvatore, Joseph F.X. DeSouza, and Melvyn A. Goodale, 1995, “Size-Contrast Illusions Deceive the Eye but Not the Hand”, *Current Biology*, 5(6): 679–685. doi:10.1016/S0960-9822(95)00133-3
- Andersen, Richard A., Kristen N. Andersen, Eun Jung Hwang, and Markus Hauschild, 2014, “Optic Ataxia: From Balint’s Syndrome to the Parietal Reach Region”, *Neuron*, 81(5): 967–83. doi:10.1016/j.neuron.2014.02.025
- Arnold, Derek Henry, 2011a, “I Agree: Binocular Rivalry Stimuli Are Common but Rivalry Is Not”, *Frontiers in Human Neuroscience*, 5(157). doi:10.3389/fnhum.2011.00157
- —, 2011b, “Why Is Binocular Rivalry Uncommon? Discrepant Monocular Images in the Real World”, *Frontiers in Human Neuroscience*, 5(116). doi:10.3389/fnhum.2011.00116
- Aru, Jaan, Talis Bachmann, Wolf Singer, and Lucia Melloni, 2012, “Distilling the Neural Correlates of Consciousness”, *Neuroscience and Biobehavioral Reviews*, 36(2): 737–746. doi:10.1016/j.neubiorev.2011.12.003
- Azzopardi, Paul and Alan Cowey, 1997, “Is Blindsight like Normal, near-Threshold Vision?” *Proceedings of the National Academy of Sciences* 94(25): 14190–14194. doi:10.1073/pnas.94.25.14190
- Baars, Bernard J., 1988, *A Cognitive Theory of Consciousness*, Cambridge: Cambridge University Press.
- Baker, Daniel H. and Erich W. Graf, 2009, “Natural Images Dominate in Binocular Rivalry”, *Proceedings of the National Academy of Sciences of the United States of America*, 106(13): 5436–5441. doi:10.1073/pnas.0812860106
- Bayne, Tim, 2011, “The Sense of Agency”, in *The Senses: Classic and Contemporary Philosophical Perspectives*, Fiona Macpherson (ed.), Oxford; New York: Oxford University Press, chapter 18.
- —, 2018, “On the Axiomatic Foundations of the Integrated Information Theory of Consciousness”, *Neuroscience of Consciousness*, 2018 (1). doi:10.1093/nc/niy007
- Bayne, Tim and Michelle Montague, 2011, *Cognitive Phenomenology*, Oxford: Oxford University Press. doi:10.1093/acprof:oso/9780199579938.001.0001
- Bayne, Tim and David J. Chalmers, 2003, “What Is the Unity of Consciousness?”, in *The Unity of Consciousness: Binding, Integration, and Dissociation*, Axel Cleeremans (ed.), Oxford: Oxford University Press, 23–58. doi:10.1093/acprof:oso/9780198508571.003.0002
- Behrmann, Marlene and David C. Plaut, 2013, “Distributed Circuits, Not Circumscribed Centers, Mediate Visual Recognition”, *Trends in Cognitive Sciences*, 17(5): 210–219. doi:10.1016/j.tics.2013.03.007
- Behrmann, Marlene and Tim Shallice, 1995, “Pure Alexia: A Nonspatial Visual Disorder Affecting Letter Activation”, *Cognitive Neuropsychology*, 12(4): 409–454. doi:10.1080/02643299508252004

- Blake, Randolph, Jan Brascamp, and David J. Heeger, 2014, “Can Binocular Rivalry Reveal Neural Correlates of Consciousness?” *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1641): 20130211. doi:10.1098/rstb.2013.0211
- Block, Ned, 1995, “On a Confusion about a Function of Consciousness”, *Behavioral and Brain Sciences* 18: 227–47. doi:10.1017/S0140525X00038188
- —, 2007, “Consciousness, Accessibility, and the Mesh between Psychology and Neuroscience”, *The Behavioral and Brain Sciences*, 30(5–6): 481–499. doi:10.1017/S0140525X07002786
- —, 2018, “If Perception Is Probabilistic, Why Does It Not Seem Probabilistic?” *Philosophical Transactions of the Royal Society B: Biological Sciences*, 373(1755): 20170341. doi:10.1098/rstb.2017.0341
- Boly, Melanie, Marcello Massimini, Naotsugu Tsuchiya, Bradley R. Postle, Christof Koch, and Giulio Tononi, 2017, “Are the Neural Correlates of Consciousness in the Front or in the Back of the Cerebral Cortex? Clinical and Neuroimaging Evidence”, *Journal of Neuroscience*, 37(40): 9603–9613. doi:10.1523/JNEUROSCI.3218-16.2017
- Born, Richard T. and David C. Bradley, 2005, “Structure and Function of Visual Area MT”, *Annual Review of Neuroscience*, 28: 157–189. doi:10.1146/annurev.neuro.26.041002.131052
- Breitmeyer, Bruno and Haluk Ogmen, 2006, *Visual Masking: Time Slices Through Conscious and Unconscious Vision*, Oxford: Oxford University Press. doi:10.1093/acprof:oso/9780198530671.001.0001
- Britten, Kenneth H., Michael N. Shadlen, William T. Newsome, and J. Anthony Movshon, 1992, “The Analysis of Visual Motion: A Comparison of Neuronal and Psychophysical Performance”, *Journal of Neuroscience* 12(12): 4745–4765. doi:10.1523/JNEUROSCI.12-12-04745.1992
- Champion, John, Richard Latto, and Y. M. Smith, 1983, “Is Blind-Sight an Effect of Scattered Light, Spared Cortex, and Near-Threshold Vision”, *Behavioral and Brain Sciences*, 6(03): 423–448. doi:10.1017/S0140525X00016861
- Cao, Rosa, 2012, “A Teleosemantic Approach to Information in the Brain”, *Biology and Philosophy*, 27(1): 49–71. doi:10.1007/s10539-011-9292-0
- —, 2014, “Signaling in the Brain: In Search of Functional Units”, *Philosophy of Science*, 81(5): 891–901. doi:10.1086/677688
- Carruthers, Peter, 2011, *The Opacity of Mind: An Integrative Theory of Self-Knowledge*, Oxford: Oxford University Press. doi:10.1093/acprof:oso/9780199596195.001.0001
- Chalmers, David J, 1995, “Facing up to the Problem of Consciousness”, *Journal of Consciousness Studies*, 2(3): 200–219.
- —, 1996, *The Conscious Mind: In Search of a Fundamental Theory*, New York: Oxford University Press.
- —, 2000, “What Is a Neural Correlate of Consciousness”, *Neural Correlates of Consciousness: Empirical and Conceptual Questions*, Thomas Metzinger (ed.), Cambridge, MA: MIT Press, 17–40.
- Chun, Marvin M., Julie D. Golomb, and Nicholas B. Turk-Browne, 2011, “A Taxonomy of External and Internal Attention”, *Annual Review of Psychology* 62: 73–101. doi:10.1146/annurev.psych.093008.100427
- Churchland, Patricia S., 1996, “The Hornswoggle Problem”, *Journal of Consciousness Studies*, 3(5–6): 402–408.
- Clark, Andy, 2001, “Visual Experience and Motor Action: Are the Bonds Too Tight?” *Philosophical Review*, 110(4): 495–520. doi:10.2307/3182592

- Cohen, Marlene R. and William T. Newsome, 2004, “What Electrical Microstimulation Has Revealed about the Neural Basis of Cognition”, *Current Opinion in Neurobiology*, 14(2): 169–177. doi:10.1016/j.conb.2004.03.016
- —, 2009, “Estimates of the Contribution of Single Neurons to Perception Depend on Timescale and Noise Correlation”, *Journal of Neuroscience*, 29(20): 6635–6648. doi:10.1523/JNEUROSCI.5179-08.2009
- Cohen, Michael A. and Daniel C. Dennett, 2011, “Consciousness Cannot Be Separated from Function”, *Trends in Cognitive Sciences*, 15(8): 358–364. doi:10.1016/j.tics.2011.06.008
- Colombo, Matteo and Peggy Seriès, 2012, “Bayes in the Brain—On Bayesian Modelling in Neuroscience”, *The British Journal for the Philosophy of Science*, 63(3): 697–723. doi:10.1093/bjps/axr043
- Cowey, Alan, 2010, “The Blindsight Saga”, *Experimental Brain Research*, 200(1): 3–24. doi:10.1007/s00221-009-1914-2
- Crick, Francis and Christof Koch, 1990, “Toward a Neurobiological Theory of Consciousness”, *Seminars in the Neurosciences*, 2: 263–275.
- —, 1998, “Consciousness and Neuroscience”, *Cerebral Cortex*, 8(2): 97–107. doi:10.1093/cercor/8.2.97
- —, 2003, “A Framework for Consciousness”, *Nature Neuroscience*, 6(2): 119–126. doi:10.1038/nn0203-119
- deCharms, R. Christopher and Anthony Zador, 2000, “Neural Representation and the Cortical Code”, *Annual Review of Neuroscience*, 23: 613–647. doi:10.1146/annurev.neuro.23.1.613
- Dehaene, Stanislas and Jean-Pierre Changeux, 2011, “Experimental and Theoretical Approaches to Conscious Processing”, *Neuron*, 70(2): 200–227. doi:10.1016/j.neuron.2011.03.018
- Dehaene, Stanislas, Jean-Pierre Changeux, Lionel Naccache, Jérôme Sackur, and Claire Sergent, 2006, “Conscious, Preconscious, and Subliminal Processing: A Testable Taxonomy”, *Trends in Cognitive Sciences*, 10(5): 204–211. doi:10.1016/j.tics.2006.03.007
- Dehaene, Stanislas, Michel Kerszberg, and Jean-Pierre Changeux, 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. doi:10.1073/pnas.95.24.14529
- Dehaene, Stanislas and Lionel Naccache, 2001, “Towards a Cognitive Neuroscience of Consciousness: Basic Evidence and a Workspace Framework”, *Cognition* 79(1): 1–37. doi:10.1016/S0010-0277(00)00123-2
- Denison, Rachel N., 2017, “Precision, Not Confidence, Describes the Uncertainty of Perceptual Experience: Comment on John Morrison’s ‘Perceptual Confidence.’” *Analytic Philosophy*, 58(1): 58–70. doi:10.1111/phib.12092
- Dennett, Daniel C., 2018, “Facing up to the Hard Question of Consciousness”, *Philosophical Transactions of the Royal Society B: Biological Sciences*, 373(1755): 20170342. doi:10.1098/rstb.2017.0342
- Dienes, Zoltán and Anil Seth, 2010, “Gambling on the Unconscious: A Comparison of Wagering and Confidence Ratings as Measures of Awareness in an Artificial Grammar Task”, *Consciousness and Cognition*, 19(2): 674–681. doi:10.1016/j.concog.2009.09.009
- —, 2010, “Measuring Any Conscious Content versus Measuring the Relevant Conscious Content: Comment on Sandberg et al”, *Consciousness and Cognition*, 19(4): 1079–1080. doi:10.1016/j.concog.2010.03.009

- Drayson, Zoe, 2014, “Intentional Action and the Post-Coma Patient”, *Topoi*, 33(1): 23–31. doi:10.1007/s11245-013-9185-8
- Dretske, Fred, 1981, *Knowledge and the Flow of Information*, Cambridge: MIT Press.
- Dux, Paul E. and René Marois, 2009, “The Attentional Blink: A Review of Data and Theory”, *Attention, Perception, & Psychophysics*, 71(8): 1683–1700. doi:10.3758/APP.71.8.1683
- Ehrsson, H. Henrik, 2009, “Rubber Hand Illusion”, in *The Oxford Companion to Consciousness*, Tim Bayne, Axel Cleeremans, and Patrick Wilken (eds), Oxford: Oxford University Press, 531–573.
- Evans, Gareth, 1982, *The Varieties of Reference*, Oxford: Oxford University Press.
- Farah, Martha J., 2004, *Visual Agnosia: Disorders of Object Recognition and What They Tell Us About Normal Vision*, second edition, Cambridge, MA: MIT Press.
- Feest, Uljana, 2012, “Introspection as a Method and Introspection as a Feature of Consciousness”, *Inquiry*, 55(1): 1–16. doi:10.1080/0020174X.2012.643619
- —, 2014, “Phenomenal Experiences, First-Person Methods, and the Artificiality of Experimental Data”, *Philosophy of Science*, 81(5): 927–939. doi:10.1086/677689
- Felleman, Daniel J. and David C. Van Essen, 1991, “Distributed Hierarchical Processing in the Primate Cerebral Cortex”, *Cerebral Cortex*, 1(1): 1–47.
- Fernández-Espejo, Davinia and Adrian M. Owen, 2013, “Detecting Awareness after Severe Brain Injury”, *Nature Reviews Neuroscience*, 14(11): 801–809. doi:10.1038/nrn3608
- Fetsch, Christopher R., Roozbeh Kiani, William T. Newsome, and Michael N. Shadlen, 2014, “Effects of Cortical Microstimulation on Confidence in a Perceptual Decision”, *Neuron*, 83(4): 797–804. doi:10.1016/j.neuron.2014.07.011
- Fink, Sascha Benjamin, 2016, “A Deeper Look at the ‘Neural Correlate of Consciousness’”, *Frontiers in Psychology*, 7(1044). doi:10.3389/fpsyg.2016.01044
- Franz, Volker H., 2001, “Action Does Not Resist Visual Illusion”, *Trends in Cognitive Science*, 5(11): 457–459. doi:10.1016/S1364-6613(00)01772-1
- Franz, Volker H. and Karl R. Gegenfurtner, 2008, “Grasping Visual Illusions: Consistent Data and No Dissociation”, *Cognitive Neuropsychology*, 25(7–8): 920–950. doi:10.1080/02643290701862449
- Frässle, Stefan, Jens Sommer, Andreas Jansen, Marnix Naber, and Wolfgang Einhäuser, 2014, “Binocular Rivalry: Frontal Activity Relates to Introspection and Action but Not to Perception”, *Journal of Neuroscience*, 34(5): 1738–1747. doi:10.1523/JNEUROSCI.4403-13.2014
- Freeman, Alan W., 2005, “Multistage Model for Binocular Rivalry”, *Journal of Neurophysiology*, 94(6): 4412–4420. doi:10.1152/jn.00557.2005
- Gaillard, Raphaël, Lionel Naccache, Philippe Pinel, Stéphane Clémenceau, Emmanuelle Volle, Dominique Hasboun, Sophie Dupont, et al, 2006, “Direct Intracranial, fMRI, and Lesion Evidence for the Causal Role of Left Inferotemporal Cortex in Reading”, *Neuron*, 50(2): 191–204. doi:10.1016/j.neuron.2006.03.031
- Gelder, Beatrice de, Marco Tamietto, Geert van Boxtel, Rainer Goebel, Arash Sahraie, Jan van den Stock, Bernard M. C. Stienen, Lawrence Weiskrantz, and Alan Pegna, 2008, “Intact Navigation Skills after Bilateral Loss of Striate Cortex”, *Current Biology*, 18(24): R1128–R1129. doi:10.1016/j.cub.2008.11.002
- Georgopoulos, Apostolos P., John F. Kalaska, Roberto Caminiti, and Joe T. Massey, 1982, “On the Relations between the Direction of Two-Dimensional Arm Movements and Cell Discharge in Primate Motor Cortex”, *Journal of Neuroscience*, 2(11): 1527–1537. doi:10.1523/JNEUROSCI.02-11-01527.1982
- Glezer, Laurie S., Judy Kim, Josh Rule, Xiong Jiang, and Maximilian Riesenhuber,

- 2015, “Adding Words to the Brain’s Visual Dictionary: Novel Word Learning Selectively Sharpens Orthographic Representations in the VWFA”, *Journal of Neuroscience*, 35(12): 4965–4972. doi:10.1523/JNEUROSCI.4031-14.2015
- Goldman, Alvin I., 2006, *Simulating Minds: The Philosophy, Psychology and Neuroscience of Mindreading*, New York: Oxford University Press. doi:10.1093/0195138929.001.0001
 - Goodale, Melvyn A. and A. David Milner, 2004, *Sight Unseen: An Exploration of Conscious and Unconscious Vision*, Oxford: Oxford University Press. doi:10.1093/acprof:oso/9780199596966.001.0001
 - Graaf, Tom A. de, Po-Jang Hsieh, and Alexander T. Sack, 2012, “The ‘Correlates’ in Neural Correlates of Consciousness”, *Neuroscience & Biobehavioral Reviews*, 36(1): 191–197. doi:10.1016/j.neubiorev.2011.05.012
 - Greenberg, Daniel L., 2007, “Comment on ‘Detecting Awareness in the Vegetative State.’” *Science*, 315(5816): 1221. doi:10.1126/science.1135284
 - Grimaldi, Piercesare, Hakwan Lau, and Michele A. Basso, 2015, “There Are Things That We Know That We Know, and There Are Things That We Do Not Know We Do Not Know: Confidence in Decision-Making”, *Neuroscience & Biobehavioral Reviews*, 55(August): 88–97. doi:10.1016/j.neubiorev.2015.04.006
 - Haffenden, Angela M. and Melvyn A. Goodale, 1998, “The Effect of Pictorial Illusion on Prehension and Perception”, *Journal of Cognitive Neuroscience*, 10(1): 122–136. doi:10.1162/089892998563824
 - Haffenden, Angela M., Karen C. Schiff, and Melvyn A. Goodale, 2001, “The Dissociation between Perception and Action in the Ebbinghaus Illusion”, *Current Biology*, 11(3): 177–181. doi:10.1016/S0960-9822(01)00023-9
 - Harman, Gilbert 1990, “The Intrinsic Quality of Experience”, *Philosophical Perspectives*, issue: *Action Theory and Philosophy of Mind*, 4: 31–52. doi:10.2307/2214186
 - Heal, Jane, 1996, “Simulation, Theory and Content”, in *Theories of Theories of Mind*, Peter Carruthers and Peter K. Smith (eds.), Cambridge: Cambridge University Press, 75–89. doi:10.1017/CBO9780511597985.006
 - Hirshorn, Elizabeth A., Yuanning Li, Michael J. Ward, R. Mark Richardson, Julie A. Fiez, and Avniel Singh Ghuman, 2016, “Decoding and Disrupting Left Midfusiform Gyrus Activity during Word Reading”, *Proceedings of the National Academy of Sciences*, 113(29): 8162–8167. doi:10.1073/pnas.1604126113
 - Histed, Mark H., Amy M. Ni, and John H. R. Maunsell, 2013, “Insights into Cortical Mechanisms of Behavior from Microstimulation Experiments”, *Progress in Neurobiology*, Special Issue: Conversion of Sensory Signals into Perceptions, Memories and Decisions, 103(April): 115–130. doi:10.1016/j.pneurobio.2012.01.006
 - Horgan, Terence, John Tienson, and George Graham, 2003, “The Phenomenology of First Person Agency”, In *Physicalism and Mental Causation: The Metaphysics of Mind and Action*, Walter, Sven and Heinz-Dieter Heckmann (eds.), Exeter: Imprint Academic, 323–341.
 - Irvine, Elizabeth, 2012a, “Old Problems with New Measures in the Science of Consciousness”, *The British Journal for the Philosophy of Science*, 63(3): 627–648. doi:10.1093/bjps/axs019
 - —, 2012b, *Consciousness as a Scientific Concept: A Philosophy of Science Perspective: A Philosophy of Science Perspective*, Dordrecht: Springer Netherlands. doi:10.1007/978-94-007-5173-6
 - Jackson, Frank, 1982, “Epiphenomenal Qualia”, *Philosophical Quarterly* 32(127): 127–136. doi:10.2307/2960077

- James, Thomas W., Jody Culham, G. Keith Humphrey, A. David Milner, and Melvyn A. Goodale, 2003, “Ventral Occipital Lesions Impair Object Recognition but Not Object-Directed Grasping: An fMRI Study”, *Brain*, 126(11): 2463–2475. doi:10.1093/brain/awg248
- Kanwisher, Nancy, 2000, “Domain Specificity in Face Perception”, *Nature Neuroscience*, 3(8): 759–63. doi:10.1038/77664
- Kiani, Roozbeh and Michael N. Shadlen, 2009, “Representation of Confidence Associated with a Decision by Neurons in the Parietal Cortex”, *Science*, 324(5928): 759–64. doi:10.1126/science.1169405
- Kim, Byoungsoon and Michele A. Basso, 2008, “Saccade Target Selection in the Superior Colliculus: A Signal Detection Theory Approach”, *Journal of Neuroscience*, 28(12): 2991–3007. doi:10.1523/JNEUROSCI.5424-07.2008
- King, Sheila M., Paul Azzopardi, Alan Cowey, John Oxbury, and Susan Oxbury, 1996, “The Role of Light Scatter in the Residual Visual Sensitivity of Patients with Complete Cerebral Hemispherectomy”, *Visual Neuroscience*, 13(1): 1–13. doi:10.1017/S0952523800007082
- Klein, Colin, 2017, “Consciousness, Intention, and Command-Following in the Vegetative State”, *The British Journal for the Philosophy of Science*, 68(1): 27–54. doi:10.1093/bjps/axv012
- Knill, David C. and Alexandre Pouget, 2004, “The Bayesian Brain: The Role of Uncertainty in Neural Coding and Computation”, *Trends in Neurosciences*, 27(12): 712–719. doi:10.1016/j.tins.2004.10.007
- Ko, Yoshiaki and Hakwan Lau, 2012, “A Detection Theoretic Explanation of Blindsight Suggests a Link between Conscious Perception and Metacognition”, *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 367(1594): 1401–1411. doi:10.1098/rstb.2011.0380
- Koch, Christof, Marcello Massimini, Melanie Boly, and Giulio Tononi, 2016, “Neural Correlates of Consciousness: Progress and Problems”, *Nature Reviews Neuroscience*, 17(5): 307–321. doi:10.1038/nrn.2016.22
- Kovács, Iilona, Thomas V. Pappathomas, Ming Yang, and Ákos Fehér, 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): 15508–15511. doi:10.1073/pnas.93.26.15508
- Kozuch, Benjamin, 2014, “Prefrontal Lesion Evidence against Higher-Order Theories of Consciousness”, *Philosophical Studies*, 167(3): 721–746. doi:10.1007/s11098-013-0123-9
- Kravitz, Dwight J., Kadharbatcha S. Saleem, Chris I. Baker, and Mortimer Mishkin, 2011, “A New Neural Framework for Visuospatial Processing”, *Nature Reviews Neuroscience*, 12(4): 217–230. doi:10.1038/nrn3008
- Kriegel, Uriah, 2003, “Consciousness as Intransitive Self-Consciousness: Two Views and an Argument”, *Canadian Journal of Philosophy*, 33(1): 103–132. doi:10.1080/00455091.2003.10716537
- Lamme, Victor A. F., 2006, “Towards a True Neural Stance on Consciousness”, *Trends in Cognitive Sciences*, 10(11): 494–501. doi:10.1016/j.tics.2006.09.001
- —, 2010, “How Neuroscience Will Change Our View on Consciousness”, *Cognitive Neuroscience*, 1(3): 204–220. doi:10.1080/17588921003731586
- Lau, Hakwan and Richard Brown, forthcoming, “The Emperor’s New Phenomenology? The Empirical Case for Conscious Experience without First-Order Representations”, in *Blockheads! Essays on Ned Block’s Philosophy of Mind and Consciousness*, Adam Pautz and Daniel Stoljar (eds), Cambridge, MA: MIT Press.

- Lau, Hakwan C. and Richard E. Passingham, 2006, “Relative Blindsight in Normal Observers and the Neural Correlate of Visual Consciousness”, *Proceedings of the National Academy of Sciences*, 103(49): 18763–18768. doi:10.1073/pnas.0607716103
- Lau, Hakwan and David Rosenthal, 2011, “Empirical Support for Higher-Order Theories of Conscious Awareness”, *Trends in Cognitive Sciences*, 15(8): 365–373. doi:10.1016/j.tics.2011.05.009
- Lee, Joonyeol and John H. R. Maunsell, 2009, “A Normalization Model of Attentional Modulation of Single Unit Responses”, *PLoS ONE*, 4(2): e4651. doi:10.1371/journal.pone.0004651
- Leopold, David A., 2012, “Primary Visual Cortex: Awareness and Blindsight”, *Annual Review of Neuroscience*, 35: 91–109. doi:10.1146/annurev-neuro-062111-150356
- Leopold, David A. and Nikos K. Logothetis, 1996, “Activity Changes in Early Visual Cortex Reflect Monkeys’ Percepts during Binocular Rivalry”, *Nature*, 379(6565): 549–553. doi:10.1038/379549a0
- Levine, Joseph, 1983, “Materialism and Qualia: The Explanatory Gap”, *Pacific Philosophical Quarterly* 64(4): 354–361. doi:10.1111/j.1468-0114.1983.tb00207.x
- Li, Fei Fei, Rufin VanRullen, Christof Koch, and Pietro Perona, 2002, “Rapid Natural Scene Categorization in the near Absence of Attention”, *Proceedings of the National Academy of Sciences of the United States of America*, 99(14): 9596–9601. doi:10.1073/pnas.092277599
- Logothetis, Nikos K., David A. Leopold, and David L. Sheinberg, 1996, “What Is Rivalling during Binocular Rivalry?” *Nature*, 380(6575): 621–624. doi:10.1038/380621a0
- Lumer, Erik D. and Geraint Rees, 1999, “Covariation of Activity in Visual and Prefrontal Cortex Associated with Subjective Visual Perception”, *Proceedings of the National Academy of Sciences*, 96(4): 1669–1673. doi:10.1073/pnas.96.4.1669
- Mack, Arien and Irvin Rock, 1998, *Inattentional Blindness*, Cambridge, MA: MIT Press.
- Maier, Alexander, Melanie Wilke, Christopher Aura, Charles Zhu, Frank Q. Ye, and David A. Leopold, 2008, “Divergence of fMRI and Neural Signals in V1 during Perceptual Suppression in the Awake Monkey”, *Nature Neuroscience*, 11(10): 1193–1200. doi:10.1038/nn.2173
- Maniscalco, Brian and Hakwan Lau, 2012, “A Signal Detection Theoretic Approach for Estimating Metacognitive Sensitivity from Confidence Ratings”, *Consciousness and Cognition*, 21(1): 422–430. doi:10.1016/j.concog.2011.09.021
- Marcel, Anthony J., 2003, “The Sense of Agency: Awareness and Ownership of Action”, in *Agency and Self-Awareness: Issues in Philosophy and Psychology*, Johannes Roessler and Naomi Eilan (eds), Oxford: Oxford University Press, 48–93.
- Martens, Sander and Brad Wyble, 2010, “The Attentional Blink: Past, Present, and Future of a Blind Spot in Perceptual Awareness”, *Neuroscience & Biobehavioral Reviews*, 34(6): 947–957. doi:10.1016/j.neubiorev.2009.12.005
- Mazzi, Chiara, Chiara Bagattini, and Silvia Savazzi, 2016, “Blind-Sight vs. Degraded-Sight: Different Measures Tell a Different Story”, *Frontiers in Psychology*, 7(901). doi:10.3389/fpsyg.2016.00901
- Mégevand, Pierre, David M. Groppe, Matthew S. Goldfinger, Sean T. Hwang, Peter B. Kingsley, Ido Davidesco, and Ashesh D. Mehta, 2014, “Seeing Scenes: Topographic Visual Hallucinations Evoked by Direct Electrical Stimulation of the Parahippocampal Place Area”, *Journal of Neuroscience*, 34(16): 5399–5405. doi:10.1523/JNEUROSCI.5202-13.2014

- Milner, A. David and Melvyn A. Goodale, 1995, *The Visual Brain in Action*, Oxford: Oxford University Press. doi:10.1093/acprof:oso/9780198524724.001.0001
- Mole, Christopher, 2009, “Illusions, Demonstratives and the Zombie Action Hypothesis”, *Mind*, 118(472): 995–1011. doi:10.1093/mind/fzp109
- Montemayor, Carlos and Harry Haroutioun Haladjian, 2015, *Consciousness, Attention, and Conscious Attention*, Cambridge, MA: MIT Press.
- Monti, Martin M., Audrey Vanhaudenhuyse, Martin R. Coleman, Melanie Boly, John D. Pickard, Luaba Tshibanda, Adrian M. Owen, and Steven Laureys, 2010, “Willful Modulation of Brain Activity in Disorders of Consciousness”, *New England Journal of Medicine*, 362(7): 579–589. doi:10.1056/NEJMoa0905370
- Morrison, John, 2017, “Perceptual Confidence and Categorization”, *Analytic Philosophy*, 58(1): 71–85. doi:10.1111/phib.12094
- Naber, Marnix, Stefan Frässle, and Wolfgang Einhäuser, 2011, “Perceptual Rivalry: Reflexes Reveal the Gradual Nature of Visual Awareness”, *PLoS ONE*, 6 (6): e20910. doi:10.1371/journal.pone.0020910
- Nachev, Parashkev and Masud Husain, 2007, “Comment on ‘Detecting Awareness in the Vegetative State.’” *Science*, 315(5816): 1221. doi:10.1126/science.1135096
- Nagel, Thomas, 1974, “What Is It like to Be a Bat?” *Philosophical Review*, 83 (October): 435–450. doi:10.2307/2183914
- Newsome, William T. and Edmond B. Paré, 1988, “A Selective Impairment of Motion Perception Following Lesions of the Middle Temporal Visual Area (MT)”, *Journal of Neuroscience*, 8(6): 2201–2211. doi:10.1523/JNEUROSCI.08-06-02201.1988
- Ngo, Trung T., Steven M. Miller, Guang B. Liu, and John D. Pettigrew, 2000, “Binocular Rivalry and Perceptual Coherence”, *Current Biology*, 10(4): R134–R136. doi:10.1016/S0960-9822(00)00399-7
- Nichols, Shaun and Stephen P. Stich, 2003, *Mindreading: An Integrated Account of Pretence, Self-Awareness, and Understanding Other Minds*, Oxford: Clarendon. doi:10.1093/0198236107.001.0001
- Noë, Alva and Evan Thompson, 2004, “Are There Neural Correlates of Consciousness?” *Journal of Consciousness Studies*, 11(1): 3–28.
- Odegaard, Brian, Piercesare Grimaldi, Seong Hah Cho, Megan A. K. Peters, Hakwan Lau, and Michele A. Basso, 2018, “Superior Colliculus Neuronal Ensemble Activity Signals Optimal Rather than Subjective Confidence”, *Proceedings of the National Academy of Sciences of the United States of America*, 115(7): E1588–E1597. doi:10.1073/pnas.1711628115
- Odegaard, Brian, Robert T. Knight, and Hakwan Lau, 2017, “Should a Few Null Findings Falsify Prefrontal Theories of Conscious Perception?” *Journal of Neuroscience*, 37(40): 9593–9602. doi:10.1523/JNEUROSCI.3217-16.2017
- O’Shea, Robert Paul, 2011, “Binocular Rivalry Stimuli Are Common but Rivalry Is Not”, *Frontiers in Human Neuroscience*, 5(148). doi:10.3389/fnhum.2011.00148
- Overgaard, Morten and Peter Fazekas, 2016, “Can No-Report Paradigms Extract True Correlates of Consciousness?” *Trends in Cognitive Sciences*, 20(4): 241–242. doi:10.1016/j.tics.2016.01.004
- Overgaard, Morten, Katrin Fehel, Kim Mouridsen, Bo Bergholt, and Axel Cleeremans, 2008, “Seeing without Seeing? Degraded Conscious Vision in a Blindsight Patient”, *PloS ONE*, 3(8): e3028. doi:10.1371/journal.pone.0003028
- Overgaard, Morten, Julian Rote, Kim Mouridsen, and Thomas Zoëga Ramsøy, 2006, “Is Conscious Perception Gradual or Dichotomous? A Comparison of Report Methodologies during a Visual Task”, *Consciousness and Cognition*, Special Issue on

Introspection, 15(4): 700–708. doi:10.1016/j.concog.2006.04.002

- Owen, Adrian M., Martin R. Coleman, Melanie Boly, Matthew H. Davis, Steven Laureys, Dietsje Jolles, and John D. Pickard, 2007, “Response to Comments on ‘Detecting Awareness in the Vegetative State.’” *Science*, 315(5816): 1221. doi:10.1126/science.1135583
- Owen, Adrian M., Martin R. Coleman, Melanie Boly, Matthew H. Davis, Steven Laureys, and John D. Pickard, 2006, “Detecting Awareness in the Vegetative State”, *Science*, 313(5792): 1402. doi:10.1126/science.1130197
- Parker, A. J., and W. T. Newsome, 1998, “Sense and the Single Neuron: Probing the Physiology of Perception”, *Annual Review of Neuroscience*, 21: 227–277. doi:10.1146/annurev.neuro.21.1.227
- Parvizi, Josef, Corentin Jacques, Brett L. Foster, Nathan Withoft, Vinitha Rangarajan, Kevin S. Weiner, and Kalanit Grill-Spector, 2012, “Electrical Stimulation of Human Fusiform Face-Selective Regions Distorts Face Perception”, *Journal of Neuroscience*, 32(43): 14915–14920. doi:10.1523/JNEUROSCI.2609-12.2012
- Penfield, Wilder and Edwin Boldrey, 1937, “Somatic Motor and Sensory Representation in the Cerebral Cortex of Man as Studied by Electrical Stimulation”, *Brain*, 60(4): 389–443. doi:10.1093/brain/60.4.389
- Penfield, Wilder and Phanor Perot, 1963, “The Brain’s Record of Auditory and Visual Experience”, *Brain*, 86(4): 595–696. doi:10.1093/brain/86.4.595
- Peng, Yueqing, Sarah Gillis-Smith, Hao Jin, Dimitri Tränkner, Nicholas J. P. Ryba, and Charles S. Zuker, 2015, “Sweet and Bitter Taste in the Brain of Awake Behaving Animals”, *Nature*, 527(7579): 512–515. doi:10.1038/nature15763
- Persaud, Navindra, Peter McLeod, and Alan Cowey, 2007, “Post-Decision Wagering Objectively Measures Awareness”, *Nature Neuroscience*, 10(2): 257–261. doi:10.1038/nn1840
- Peters, Megan A.K. and Hakwan Lau, 2015, “Human Observers Have Optimal Introspective Access to Perceptual Processes Even for Visually Masked Stimuli”, *ELife*, 4 (October): e09651. doi:10.7554/eLife.09651
- Phillips, Ian B., 2011, “Perception and Iconic Memory: What Sperling Doesn’t Show”, *Mind and Language*, 26(4): 381–411. doi:10.1111/j.1468-0017.2011.01422.x
- —, 2016, “Consciousness and Criterion: On Block’s Case for Unconscious Seeing”, *Philosophy and Phenomenological Research*, 93(2): 419–451. doi:10.1111/phpr.12224
- Pisella, Laure, Lauren Sergio, Annabelle Blangero, Héloïse Torchin, Alain Vighetto, and Yves Rossetti, 2009, “Optic Ataxia and the Function of the Dorsal Stream: Contributions to Perception and Action”, *Neuropsychologia*, 47(14): 3033–3044. doi:10.1016/j.neuropsychologia.2009.06.020
- Polonsky, Alex, Randolph Blake, Jochen Braun, and David J. Heeger, 2000, “Neuronal Activity in Human Primary Visual Cortex Correlates with Perception during Binocular Rivalry”, *Nature Neuroscience*, 3(11): 1153–1159. doi:10.1038/80676
- Pouget, Alexandre, Peter Dayan, and Richard S. Zemel, 2003, “Inference and Computation with Population Codes”, *Annual Review of Neuroscience*, 26: 381–410. doi:10.1146/annurev.neuro.26.041002.131112
- Pouget, Alexandre, Jan Drugowitsch, and Adam Kepecs, 2016, “Confidence and Certainty: Distinct Probabilistic Quantities for Different Goals”, *Nature Neuroscience*, 19(3): 366–374. doi:10.1038/nn.4240
- Prinz, Jesse, 2012, *The Conscious Brain*, Oxford: Oxford University Press. doi:10.1093/acprof:oso/9780195314595.001.0001

- Pulvermüller, Friedemann, 2005, “Brain Mechanisms Linking Language and Action”, *Nature Reviews: Neuroscience*, 6(7): 576–582. doi:10.1038/nrn1706
- Ramsøy, Thomas Zoëga, and Morten Overgaard, 2004, “Introspection and Subliminal Perception”, *Phenomenology and the Cognitive Sciences*, 3(1): 1–23. doi:10.1023/B:PHEN.0000041900.30172.e8
- Rausch, Manuel and Michael Zehetleitner, 2016, “Visibility Is Not Equivalent to Confidence in a Low Contrast Orientation Discrimination Task”, *Frontiers in Psychology*, 7 (591). doi:10.3389/fpsyg.2016.00591
- Rescorla, Michael, 2015, “Bayesian Perceptual Psychology”, in *The Oxford Handbook of Philosophy of Perception*, Mohan Matthen (ed.), Oxford: Oxford University Press, 694–716.
- Romo, Ranulfo, Adrián Hernández, Antonio Zainos, Carlos D. Brody, and Luis Lemus, 2000, “Sensing without Touching: Psychophysical Performance Based on Cortical Microstimulation”, *Neuron*, 26(1): 273–278. doi:10.1016/S0896-6273(00)81156-3
- Romo, Ranulfo, Adrián Hernández, Antonio Zainos, and Emilio Salinas, 1998, “Somatosensory Discrimination Based on Cortical Microstimulation”, *Nature*, 392(6674): 387–390. doi:10.1038/32891
- Rosenthal, David M., 2002, “Explaining Consciousness”, In *Philosophy of Mind: Classical and Contemporary Readings*, David J. Chalmers (ed.), Oxford: Oxford University Press, 109–131.
- —, 2018, “Consciousness and Confidence”, *Neuropsychologia*, February. doi:10.1016/j.neuropsychologia.2018.01.018
- Rossetti, Yves, Laure Pisella, and Alain Vighetto, 2003, “Optic Ataxia Revisited: Visually Guided Action versus Immediate Visuomotor Control”, *Experimental Brain Research*, 153(2): 171–179. doi:10.1007/s00221-003-1590-6
- Rounis, Elisabeth, Brian Maniscalco, John C. Rothwell, Richard E. Passingham, and Hakwan Lau, 2010, “Theta-Burst Transcranial Magnetic Stimulation to the Prefrontal Cortex Impairs Metacognitive Visual Awareness”, *Cognitive Neuroscience*, 1(3): 165–175. doi:10.1080/17588921003632529
- Ruff, Douglas A. and Marlene Cohen, 2014, “Relating the Activity of Sensory Neurons to Perception”, in *The Cognitive Neurosciences*, fifth edition, Michael S. Gazzaniga and George R. Mangun (eds), Cambridge, MA: MIT Press, 349–362.
- Salzman, C. Daniel, Kenneth H. Britten, and William T. Newsome, 1990, “Cortical Microstimulation Influences Perceptual Judgements of Motion Direction”, *Nature*, 346(6280): 174–177. doi:10.1038/346174a0
- Sandberg, Kristian, Bert Timmermans, Morten Overgaard, and Axel Cleeremans, 2010, “Measuring Consciousness: Is One Measure Better than the Other?” *Consciousness and Cognition*, 19(4): 1069–1078. doi:10.1016/j.concog.2009.12.013
- Schenk, Thomas and Robert D McIntosh, 2010, “Do We Have Independent Visual Streams for Perception and Action”, *Cognitive Neuroscience*, 1(1): 52–78. doi:10.1080/17588920903388950
- Schwitzgebel, Eric, 2011, *Perplexities of Consciousness*, Cambridge, MA: MIT Press.
- Shafto, Juliet P., and Michael A. Pitts, 2015, “Neural Signatures of Conscious Face Perception in an Inattentive Blindness Paradigm”, *Journal of Neuroscience*, 35(31): 10940–10948. doi:10.1523/JNEUROSCI.0145-15.2015
- Shannon, Claude Elwood, 1949, *The Mathematical Theory of Communication*, Urbana, IL: University of Illinois Press.
- Shea, Nicholas, 2014, “Neural Signalling of Probabilistic Vectors”, *Philosophy of Science*, 81(5): 902–913. doi:10.1086/678354

- —, forthcoming, *Representation in Cognitive Science*, Oxford, New York: Oxford University Press.
- Shea, Nicholas and Tim Bayne, 2010, “The Vegetative State and the Science of Consciousness”, *The British Journal for the Philosophy of Science*, 61(3): 459–484. doi:10.1093/bjps/axp046
- Simons, Daniel J. and Christopher F. Chabris, 1999, “Gorillas in Our Midst: Sustained Inattentive Blindness for Dynamic Events”, *Perception*, 28(9): 1059–1074. doi:10.1068/p281059
- Simons, Daniel J. and Michael S. Ambinder, 2005, “Change Blindness Theory and Consequences”, *Current Directions in Psychological Science*, 14(1): 44–48. doi:10.1111/j.0963-7214.2005.00332.x
- Singer, Emily, 2006, “Big Brain Thinking”, *MIT Technology Review*, February 10, URL = <<https://www.technologyreview.com/s/405296/big-brain-thinking/>>
- Smeets, Jeroen B. J., and Eli Brenner, 2006, “10 Years of Illusions”, *Journal of Experimental Psychology. Human Perception and Performance*, 32(6): 1501–1504. doi:10.1037/0096-1523.32.6.1501
- Smithies, Declan and Daniel Stoljar, 2012, *Introspection and Consciousness*, Oxford: Oxford University Press. doi:10.1093/acprof:oso/9780199744794.001.0001
- Spener, Maja, forthcoming, “Introspecting in the Twentieth Century”, in *Philosophy of Mind in the Twentieth and Twenty-First Centuries*, Amy Kind (ed.), New York: Routledge.
- —, 2015, “Calibrating Introspection”, *Philosophical Issues*, 25(1): 300–321. doi:10.1111/phils.12062
- Sperling, George, 1960, “The Information Available in Brief Visual Presentations”, *Psychological Monographs: General and Applied*, 74(11): 1–29. doi:10.1037/h0093759
- Stoerig, Petra, Martin Hübner, and Ernst Pöppel, 1985, “Signal Detection Analysis of Residual Vision in a Field Defect Due to a Post-Geniculate Lesion”, *Neuropsychologia*, 23(5): 589–599. doi:10.1016/0028-3932(85)90061-2
- Tarr, Michael J., and Isabel Gauthier, 2000, “FFA: A Flexible Fusiform Area for Subordinate-Level Visual Processing Automated by Expertise”, *Nature Neuroscience*, 3(8): 764–769. doi:10.1038/77666
- Tong, Frank, Ming Meng, and Randolph Blake, 2006, “Neural Bases of Binocular Rivalry”, *Trends in Cognitive Sciences*, 10(11): 502–511, doi:10.1016/j.tics.2006.09.003
- Tononi, Giulio, 2004, “An Information Integration Theory of Consciousness”, *BMC Neuroscience*, 5(November): 42. doi:10.1186/1471-2202-5-42
- —, 2008, “Consciousness as Integrated Information: A Provisional Manifesto”, *The Biological Bulletin*, 215(3): 216–242. doi:10.2307/25470707
- Tse, Peter U., Susana Martinez-Conde, Alexander A. Schlegel, and Stephen L. Macknik, 2005, “Visibility, Visual Awareness, and Visual Masking of Simple Unattended Targets Are Confined to Areas in the Occipital Cortex beyond Human V1/V2”, *Proceedings of the National Academy of Sciences of the United States of America*, 102(47): 17178–17183. doi:10.1073/pnas.0508010102
- Tye, Michael, 1992, “Visual Qualia and Visual Content”, in *The Contents of Experience: Essays on Perception*, Tim Crane (ed.), Cambridge: Cambridge University Press, 158–176. doi:10.1017/CBO9780511554582.008
- Ungerleider, Leslie G. and Mortimer Mishkin, 1982, “Two Cortical Systems”, in *Analysis of Visual Behavior*, David J. Ingle, Melvyn A. Goodale, and Richard J.W. Mansfield (eds), Cambridge, MA: MIT Press, 549–586.

- Urbanski, Marika, Olivier A. Coubard, and Clémence Bourlon, 2014, “Visualizing the Blind Brain: Brain Imaging of Visual Field Defects from Early Recovery to Rehabilitation Techniques”, *Frontiers in Integrative Neuroscience*, 8(74). doi:10.3389/fnint.2014.00074
- Vignemont, F. de and P. Fournieret, 2004, “The Sense of Agency: A Philosophical and Empirical Review of the ‘Who’ System”, *Consciousness and Cognition*, 13(1): 1–19. doi:10.1016/S1053-8100(03)00022-9
- Wallhagen, Morgan, 2007, “Consciousness and Action: Does Cognitive Science Support (Mild) Epiphenomenalism?” *The British Journal for the Philosophy of Science*, 58(3): 539–561. doi:10.1093/bjps/axm023
- Weiskrantz, L., 1986, *Blindsight: A Case Study and Implications*, Oxford: Clarendon Press. doi:10.1093/acprof:oso/9780198521921.001.0001
- Wilson, Hugh R., 2003, “Computational Evidence for a Rivalry Hierarchy in Vision”, *Proceedings of the National Academy of Sciences*, 100(24): 14499–14503. doi:10.1073/pnas.2333622100
- [Working Party RCP] Working Party of the Royal College of Physicians, 2003, “The Vegetative State: Guidance on Diagnosis and Management”, *Clinical Medicine*, 3(3): 249–254. doi:10.7861/clinmedicine.3-3-249
- Wu, Wayne, 2013, “The Case for Zombie Agency”, *Mind*, 122(485): 217–230. doi:10.1093/mind/fzt030
- —, 2014a, “Against Division: Consciousness, Information and the Visual Streams”, *Mind and Language*, 29(4): 383–406. doi:10.1111/mila.12056
- —, 2014b, *Attention*, Abingdon, UK: Routledge.
- —, 2017, “Attention and Perception: A Necessary Connection?” in *Current Controversies in Philosophy of Perception*, Bence Nanay (ed.), New York: Routledge, 148–162.
- Zihl, J., D. Von Cramon, and N. Mai, 1983, “Selective Disturbance of Movement Vision after Bilateral Brain Damage”, *Brain*, 106(2): 313–340. doi:10.1093/brain/106.2.313

Academic Tools



[How to cite this entry.](#)



[Preview the PDF version of this entry](#) at the [Friends of the SEP Society](#).



[Look up topics and thinkers related to this entry](#) at the Internet Philosophy Ontology Project (InPhO).



[Enhanced bibliography for this entry](#) at [PhilPapers](#), with links to its database.

Other Internet Resources

- [Movie S1](#), from Hirshorn et al. 2016: electrical stimulation session with P2. This movie shows P2’s word-naming ability completely disrupted during high stimulation, but no errors during low stimulation.
- [Movie S2](#), from Hirshorn et al. 2016: electrical stimulation session with P1. This

movie shows P1 misnaming letters under high stimulation, but no errors during low stimulation.

- Aaronson, Scott, 2014a, “Why I Am Not An Integrated Information Theorist (or, The Unconscious Expander)”, Blog *Shtetl-Optimized*, May 21. URL = <<https://www.scottaaronson.com/blog/?p=1799>>. Accessed May 26, 2016.
- Aaronson, Scott, 2014b, “Giulio Tononi and Me: A Phi-Nal Exchange”, Blog *Shtetl-Optimized*, May 30. URL = <<https://www.scottaaronson.com/blog/?p=1823>>. Accessed May 21, 2018.
- Lau, Hakwan, 2017, “In Consciousness We Trust: How to Make IIT (& Other Theories of Consciousness) More Respectable”, in blog *Consciousness We Trust*, August 10, URL = <<http://inconsciousnesswetrust.blogspot.com/2017/08/how-to-make-iit-and-other-theories-of.html>>.

Related Entries

[agency](#) | [consciousness: higher-order theories](#) | [consciousness: representational theories of](#) | [consciousness: unity of](#) | [grounding, metaphysical](#) | [information](#) | [mental representation](#) | [qualia](#) | [zombies](#)

Acknowledgments

Thanks to Hakwan Lau and Susanna Siegel and especially Dave Chalmers who refereed the article. Special thanks to Mark Sprevak and David Barak for organizing discussion groups on the entry at the University of Edinburgh and at Columbia University respectively and for their feedback. Thanks to Jorge Morales and Doug Ruff for extensive feedback on central sections. Among many others, thanks for comments to: Jake Berger, Ned Block, Richard Brown, Alessandra Buccella, Denis Buehler, Tony Cheng, Mazviita Chirimuuta, Andy Clark, Sam Clarke, Carrie Figdor, Cressida Gaukroger, Michelle Liu, Chris Mole, John Morrison, Will Nalls, David Papineau, David Rosenthal, Ian Phillips, Adina Roskies, Forrest Schreick, Nick Shea, and Cecily Whiteley.

Copyright © 2018 by
Wayne Wu <waynewu@andrew.cmu.edu>

[Open access to the SEP is made possible by a world-wide funding initiative.](#)
[Please Read How You Can Help Keep the Encyclopedia Free](#)

Browse

- [Table of Contents](#)
- [What's New](#)
- [Random Entry](#)
- [Chronological](#)
- [Archives](#)

About

- [Editorial Information](#)
- [About the SEP](#)
- [Editorial Board](#)
- [How to Cite the SEP](#)
- [Special Characters](#)
- [Advanced Tools](#)

- [Contact](#)

Support SEP

- [Support the SEP](#)
- [PDFs for SEP Friends](#)
- [Make a Donation](#)
- [SEPIA for Libraries](#)

Mirror Sites

View this site from another server:

[USA \(Main Site\) Philosophy, Stanford University](#)

- [Australia Library, University of Sydney](#)
- [Netherlands ILLC, University of Amsterdam](#)

The Stanford Encyclopedia of Philosophy is [copyright © 2021](#) by [The Metaphysics Research Lab](#), Department of Philosophy, Stanford University

Library of Congress Catalog Data: ISSN 1095-5054