Note: This Target Article has been accepted for publication and has not yet been copyedited and proofread. The article may be cited as *Cognitive and Behavioral Neurology, in press*, but it must be made clear that it is not the final version.

Consciousness as a memory system

Andrew E. Budson, MD[1,2], Kenneth A. Richman, PhD[3], and Elizabeth A. Kensinger, PhD[4]

[1]Center for Translational Cognitive Neuroscience, Veterans Affairs Boston Healthcare System

[2]Alzheimer's Disease Research Center, Boston University

[3]Center for Health Humanities, MCPHS University

[4]Psychology and Neuroscience Department, Boston College

Date of Submission 4/21/2022; Resubmission 4/30/2022; Accepted 5/3/2022

Opinion paper

Running Head: Consciousness as a memory system

Author Note

This work was supported by NSF grant BCS-1823795 to Elizabeth A. Kensinger. We have no known conflict of interest to disclose. Correspondence concerning this article should be addressed to Andrew E. Budson, MD, VA Boston Healthcare System, 150 South Huntington Ave, 10B-67, Boston, MA 02130. 617-721-6236. Email: abudson@bu.edu.

Abstract

We suggest there is confusion between why consciousness developed and what additional functions, through continued evolution, it has co-opted. Consider episodic memory. If you believe it evolved solely to accurately represent past events, it seems like a terrible system—prone to forgetting and false memories. If you consider that it developed to flexibly and creatively combine and rearrange memories of prior events to plan for the future, then it is quite a good system. We argue that consciousness originally developed as part of the episodic memory system—quite likely the part needed to accomplish that flexible recombining of information. We posit further that it was subsequently co-opted to produce other functions that are not directly relevant to memory *per se*, such as problem solving, abstract thinking, and language. We suggest that this theory is compatible with many phenomena, such as the slow speed and after-the-fact order of consciousness, that cannot be explained well by other theories. We believe our explanation may have profound implications for understanding intentional action and consciousness in general. Moreover, we suggest that if we consider episodic memory along with its associated memory systems of sensory, working, and semantic memory, these systems as a whole ought to be considered together as the *conscious memory system* in that they, together, give rise to the phenomenon of consciousness. Lastly, we suggest that the cerebral cortex is the part of the brain that makes consciousness possible, and that every cortical region contributes to this conscious memory system.

Keywords: consciousness; episodic memory; cerebral cortex; neural correlates of consciousness

Questions

When you consider consciousness, a number of questions naturally arise. Why did consciousness develop? What is consciousness good for? If consciousness developed to help you plan and act for the future, why is consciousness so difficult to control? Why is mindfulness so hard? And for that matter, if your actions are under your conscious control, why is dieting (and resisting other urges) so hard for most people?

Why does it appear that you are an observer, peering out through your eyes at the world, while sitting in the proverbial Cartesian theater? Why do we speak, in William James's words, of a "stream of consciousness"? Can you perform complicated activities (such as driving) without being consciously aware of it?

Are animals conscious (and if so, which ones)? Are there developmental, neurologic, or psychiatric disorders that are actually disorders of consciousness?

There have, of course, been many answers to these questions over the last 2500 years. We hope to provide new insights to these and a number of related questions in this paper.

Definitions

Before we attempt to answer these questions, we should clarify what we mean when we use the word 'consciousness.'

For the most part, we mean what William James meant when he used the term: our own personal experience of perceiving, thinking, emoting, and acting (James, 1890). Self-consciousness, that is, being conscious of oneself as a thinking entity (à la Descartes), would certainly be included in what we mean, but only a small part of it.

Similarly, the various so-called "levels of consciousness" as measured by the Glasgow Coma Scale (Teasdale & Jennett, 1974), including conscious, confused, delirious, somnolent, obtunded, stuporous, and comatose (Posner et al., 2019), would be included in our use.

Any understanding of consciousness should also be consistent with the four basic properties that arise from studying its phenomenology: intentionality, unity, selectivity, and transience (Schacter et al., 2019). Consciousness is directed towards an object; it is about something (intentionality). We experience consciousness as unified, as one experience, rather than separate experiences of sights, sounds, smells, thoughts, feelings, etc. (unity). We have the capacity to be conscious of some things and not others (selectivity). Objects of consciousness are there transiently; the contents of consciousness have a tendency to change (transience).

Defining 'postdictive effects' will be important for understanding our theory. As counterintuitive as it sounds, in postdictive effects, a later stimulus can affect the perception of an earlier stimulus, or earlier and later stimuli can mutually affect each other (Herzog et al., 2020; Michel & Doerig, 2020; Sergent, 2018).

We wish to keep the commonsense use of the word 'conscious' when you say, "I was so wrapped up in my thoughts that, driving without being conscious of where I was going, I found myself sitting in the empty parking lot at work, despite intending to drive to the post office." We also want to retain the idea of having 'unconscious awareness,' such as when you see your friend, are sure that something is different about them, but it takes you a minute to figure out that they got a new haircut. We would like to separate this idea of 'unconscious awareness,' which may contain a conscious feeling (such as familiarity) short of a full conscious experience, from 'unconscious knowledge' that may influence our actions without any awareness (such as priming).

Block and others have promulgated the idea of two separate aspects of consciousness. Phenomenal consciousness is what it is like for a subject to have an experience. Access consciousness is when representations are made available to cognitive processing (Block, 2011). We will discuss this potentially useful distinction below.

Problems

James famously used the phrase "stream of consciousness" as a metaphor to describe the intuitive feeling that there is not only a "now" but also "downstream" events that have occurred in the past and "upstream" events that will come to pass (James, 1890). Why does consciousness feel this way when we know that the brain is actually processing massive amounts of information in parallel? And does consciousness flow linearly with time, as this metaphor implies?

In fact, postdictive and other order effects demonstrate that, at timescales less than 500 ms, consciousness does not flow linearly with time (Herzog et al., 2020; Michel & Doerig, 2020). Conscious awareness often occurs in the wrong order (after, rather than before or with, the perception, decision, or action) (Sergent, 2018). Conscious sensations are sometimes referred backwards in time (Hodinott-Hill et al., 2002; Libet et al., 1979). Consciousness is also too slow to guide many split-second decisions and actions that occur routinely when playing sports or musical instruments (Blackmore, 2017). Experiments performed with individuals who have experienced a brain injury have demonstrated that consciousness is not necessary to perform a number of activities that we usually think require conscious awareness (Weiskrantz et al., 1974). Lastly, mindfulness is hard, which suggests that controlling our conscious thoughts is not easy to do—something that is quite odd if the purpose of consciousness is to enable us to control our thoughts and actions. We will review these problems briefly here, along with how these problems lead us to the largest problem, that of the purpose of consciousness.

Order problems: Consciousness after the perception, decision, action

There are many examples in which the consciousness associated with a perception, decision, or action seems to occur only after the physiological perception, decision, or action has actually occurred. This order is incompatible with the idea that perceptions, decisions, and actions are only possible when conscious awareness and thought are present.

Tolling bells and cocktail parties

An oft mentioned example is from Exner, quoted by James in his *Principles of Psychology* (and often attributed to James): "Impressions to which we are inattentive leave so brief an image in the memory that it is usually overlooked. When deeply absorbed, we do not hear the clock strike. But our attention may awake after the striking has ceased, and we may then count off the strokes." (James, 1890.) The problem here is how we can become conscious of something after it has occurred. Were we conscious of it at the time it occurred, or not?

Another common example, which almost everyone has experienced at a cocktail party, is that you hear your name spoken, your attention is suddenly focused, and you can recall the earlier part of the sentence in which your name was mentioned (Blackmore, 2017). How is it that your consciousness can work "backwards" to perceive the earlier part of the sentence that you weren't paying attention to?

Block (2011) and others might argue that the simple answer to these two order problems is that you were phenomenally conscious of the sounds of the clock and the earlier part of sentence, but that your access to consciousness arose only later. Although attractive, we believe that this is not the best explanation for this phenomenon, as we will explain later.

Postdictive effects

Below we review several of the many experimentally-created postdictive effects; see Sergent (2018), Herzog et al. (2020), and Michel and Doerig (2020) for comprehensive reviews of these effects and their implications for theories of consciousness.

Cutaneous rabbits. In the cutaneous rabbit illusion, an individual holds out one arm while looking the other way. The experimenter then taps quickly at precisely equal intervals with equal pressure five times at the wrist, thrice near the elbow, and twice near the shoulder. This produces the odd sensation as if a little rabbit were running up the individual's arm—not three separate groups of taps (Geldard & Sherrick, 1972). There are Bayesian models that can closely replicate this illusion and thus in some sense explain it (Goldreich, 2007), but what isn't explained is how the brain knows where to put the intervening taps running up the forearm from the wrist to the elbow *before* the elbow taps have occurred? If we think about consciousness in the ordinary sense, it simply makes no sense. Blackmore states the problem very well when she says:

> If you stick to the natural idea that any tap (say the fourth one) must either have been conscious or unconscious (in the stream or not), then you get into a big muddle. For example, you might have to say that the third tap was consciously experienced at its correct place (i.e. on the wrist), but then later, after the sixth tap occurred, this memory

> was wiped out and replaced with the conscious experience of it happening half way between wrist and elbow. If you don't like this idea, you might prefer to say that consciousness was held up for some time—waiting for all the taps to come in before deciding where to place each one. In this case, the fourth tap remained unconscious until after the sixth tap occurred, and was then referred back in time so as to be put in its correct place in the stream of consciousness. (Blackmore, 2017)

Here consciousness simply fails to capture what is happening as it happens.

Color phi illusion. In his book, *Consciousness Explained* (Dennett, 1991) Dennett discusses extensively the color phi illusion. The phenomenon is that when a viewer watches a blue dot at the top of a frame, followed by a blank screen, and then a red dot at the bottom of the frame, all in quick succession, the viewer will report two odd things. First, they experience a sensation of motion, as if the first spot were moving downward. Second, abruptly and in the middle of its illusory path, it changes color (Kolers & von Grünau, 1976). The problem, of course, is that it makes no sense for subjects to consciously experience the downward motion or the color change before the second dot is consciously perceived. How (and why) do these phenomena happen?

Keuninckx and Cleeremans (2021) have recently suggested that the color phi illusion may simply be related to "Inherent dynamical and nonlinear sensory processing in the brain" and not related to consciousness, per se (Keuninckx & Cleeremans, 2021). This is an interesting idea that we discuss below in the context of our theory.

Color fusion effects. When a red disk is presented for 40 ms, the participant sees a red disk. Yet, when a red disk is presented for 40 ms, followed by a green disk presented for 40 ms in the same location, the percept is a single yellow disk (Pilz et al., 2013). Why are the colors fused together in our conscious perception? How does the later presentation of the green disk interfere with the prior perception of the red disk?

Illusory and invisible audiovisual rabbits. Postdictive effects can be crossmodal. In one experiment, three flashes are presented, each paired with a sound. When the sounds are repeated but the central flash is omitted, an illusory flash is perceived. Conversely, when all three flashes are present but the central sound is absent, the central flash is not perceived (Stiles et al., 2018). Why do these postdictive illusions happen?

Transcranial magnetic stimulation. In one ingenious experiment, transcranial magnetic stimulation (TMS) pulses to occipital cortex at various time intervals from 20 to 370 ms after stimuli were flashed altered how the stimuli were perceived. Thus, the TMS pulse itself acted as a postdictive stimulus, which had different effects depending on when TMS pulse was applied (Scharnowski et al., 2009).

Motor cortex first, conscious decision to move second

In one of the most convincing examples of this order problem, Dennett describes an experiment performed by the neurosurgeon W. Grey Walter and published in 1963. Patients who had electrodes implanted in their motor cortex were set up with a slide projector and told they could advance the slides by pressing the button on a controller. However, the controller was fake—it was not connected to the projector. What advanced the slides was actually a signal from the implanted electrodes. The patients experienced that the slide projector was *anticipating* their decisions. As Dennett described (Dennett, 1991), “They reported that just as they were ‘about to’ push the button, but before they had actually decided to do so, the projector would advance the slide—and they would find themselves pressing the button with the worry that it was going to advance the slide twice!" The commonsense view of consciousness tells us that the conscious decision to act precedes and causes the action itself. How do we explain this strange phenomenon in which motor actions occur prior to the conscious decisions to take these actions? How could the effect precede the cause?

Conscious sensations referred backwards in time

In addition to order problems, there are also situations where conscious sensations are referred backwards in time.

Chronostasis

The stopped-clock illusion, one example of the multisensory illusion of chronostasis (Hodinott-Hill et al., 2002). In this illusion, you make a saccade to a clock with a second hand. As in all saccades, your perception of the visual information during the saccade is masked to prevent you from experiencing motion blur. After the saccade is completed, you focus on the clock. Your experience is that the clock seems to be taking more than one second for the second hand to move. The explanation is that the sensory information of the image of the clock which your eyes receive after the fixation is projected backwards in time to fill in the time period when you were making the saccade (Thilo & Walsh, 2002). But how can a sensation be projected backwards in time?

Stimulation of the hand versus the somatosensory cortex

Following up on a prior experiment in which he produced conscious sensation by stimulating the cortex directly (Libet et al., 1964), Libet produced the conscious experience of a tingle in the subject's hand by either stimulating the back of the hand or by stimulating the contralateral somatosensory cortex of the brain directly (Libet et al., 1979). If we assume that the conscious experience is related to the somatosensory cortex receiving the stimulation, because the impulses initiated in the hand need to travel from the hand through the wrist, forearm (radius or ulnar nerve), arm (brachial nerve), shoulder (brachial plexus), neck (spinal nerves and cord), and head (including brainstem, internal capsule, and corona radiata), we would expect that stimulation of the cortex would be noticed more quickly than the stimulation of the hand. Two surprising results were found. First, in each case it took a long time—approximately 500 ms—from the stimulus onset until conscious experience. Second, as Libet writes: “after delayed neuronal adequacy is achieved, there is a subjective referral of the sensory experience backwards in time so as to coincide with this initial 'time-marker'.” (Libet et al., 1979)

What does it mean for a conscious experience to be referred “backwards in time”? Many scientists and philosophers have provided explanations for these results and Libet’s conclusions (Churchland, 1981; Dennett, 1991; Popper & Eccles, 1977). We have no evidence that any of them are wrong; we simply think our theory provides a more parsimonious explanation of these observations. As we will see, backwards referral does not pose a challenge for our theory.

Timing problems: Consciousness is too slow

The Libet experiment raises another problem of consciousness which Blackmore states explicitly: consciousness is too slow (Blackmore, 2017). Recall that Libet found it took approximately 500 ms from stimulus onset until the conscious experience occurred. Blackmore reminds us just how very long that amount of time is from a neurophysiological perspective, where impulses travel at speeds up to 100 m/s. If it takes 500 ms (long enough for an impulse to travel up to 50 meters) for conscious experience to occur, then it is too slow to be playing an active, controlling role in many activities, including playing sports and making music.

It is estimated that professional baseball players need to “decide” whether to swing at a pitch within 125 ms after it has left the pitcher’s hand, and the ball crosses the plate within 300 to 400 ms after it has left the pitcher’s hand (Science Non-Fiction, 2016). Ordinary reaction time measured by clicking a mouse in response to an auditory, tactile, or combined auditory/tactile stimulation produced speeds in one study ranging from about 210-320 ms for most people, but was as fast as 100-210 ms in trained musicians (Landry & Campoux, 2017). How can we be consciously in charge of our actions if our actions are occurring much faster than our conscious thoughts?

Lesion patients

Individuals with neurological disorders may also help us to understand consciousness. Some patients with brain lesions are unable to consciously perform a task but, when asked to do the task unconsciously—guess or just perform it without thinking about it—they are able to do it.

Visual apperceptive agnosia

For example, patient DF with visual apperceptive agnosia is unable to consciously perceive and report the size or shape of objects, yet she is able to correctly scale her grip to pick them up. She is also able to accurately perform actions, such as putting cards through slots of different angles, despite not being able to consciously perceive or report the angles (Ganel & Goodale, 2019). How is this performance without conscious awareness possible? Moreover, this dissociation between inaccurate visual perception and accurate grip scaling can be produced in individuals with normal vision as well (Aglioti et al., 1995; Chen et al., 2015); again, how is this possible if our actions are consciously controlled?

Blindsight

Blindsight causes similar problems (Kentridge et al., 2008; Poppel, et al., 1973; Weiskrantz et al., 1974). Individuals with blindsight are cortically blind at least in one hemifield, cannot consciously see objects in that hemifield, and yet are above chance when asked to guess, point, or otherwise act based upon a visual stimulus to their blind field. Explanations of the physiological phenomenon include preserved cortical islands; dissociation between a dorsal, action-related unconscious stream and a ventral, perception-related conscious stream (Brogaard, 2011); and visual information reaching subcortical structures such as the superior colliculi and lateral geniculate nuclei. These physiological explanations, however, still fail to explain the phenomena regardless of its biological underpinnings. How is it possible to accurately point at objects and perform similar tasks when such objects cannot be consciously visualized?

Mindfulness

Mindfulness is a problem because it is hard. Anyone who has tried to practice mindfulness knows that it is difficult. But if (as a commonsense view would have it) our consciousness[1] evolved to allow us to perform high-level abstract reasoning using language, logic, visuospatial abilities, or other cognitive capacities to allow us to carry out intentional

[1] This would be "access consciousness" if you like the distinction between phenomenal and access consciousness.

actions, then controlling our thoughts should be easy. So why is mindfulness so hard? Why is it so difficult for us to control our conscious thoughts?

Lack of apparent causal role for consciousness

That it is difficult to control our thoughts and patients with brain lesions and no relevant conscious perception can still perform tasks that we intuitively feel must require consciousness leads to the interesting—and for most people, uncomfortable—thought that perhaps consciousness is epiphenomenal. Many people have made this point, arguing that there could be philosophical zombies who act as if they are conscious but are not actually (e.g., Chalmers, 2010). Some biological traits will survive through natural selection only incidentally, because they arose in association with an adaptive trait (Gould & Lewontin, 1979). Is consciousness like that, merely epiphenomenal? Perhaps consciousness appears to be epiphenomenal only because we are looking in the wrong place to find its causal role. Could consciousness be epiphenomenal with respect to sensation, decision, and action, but play an important causal role in some other function?

How does consciousness contribute to evolutionary success?

To review, consciousness often occurs after the perception, decision, and action have taken place, in part because consciousness is too slow to participate in many real-time events. Consciousness fools us in many different ways, creating visual and tactile illusions. We know from patients with various brain lesions that a variety of judgments and actions can occur without conscious perception, suggesting that although consciousness is commonly present, it is not necessary to carry out at least some types of tasks. Lastly, consciousness is difficult to control—something that seems very odd if it evolved to allow us to carry out complex reasoning necessary for intentional action.

All these problems lead us to one of the most important questions regarding consciousness: What does consciousness “do”? How does consciousness contribute to the evolutionary success of human beings?

Consciousness as a memory system

The problems discussed above demonstrate just how many difficulties there are if we simply take at face value the idea that the role of consciousness is to enable our perceptions, decisions, and actions to occur. In this setting, we are ready to explain our theory that consciousness is, at its core, a memory system.

Consciousness is part of the episodic memory system

Tulving described episodic memory as the set of processes that allow us to mentally time-travel, and to re-experience a past moment in time (Tulving, 1985). To be able to do that, we must first take information that comes in through our sensory stores and our working memory and create a mental representation of a moment in time; this is the process of *encoding*. If we want that representation to be accessible later, we must store it in some durable form; this is the process of *consolidation*. And, when we want to later reflect on that moment in time, we must engage *retrieval* processes to do so.

We argue that one function of consciousness—and more importantly, what it initially developed to do—is to allow for each of these phases of episodic memory. Consciousness binds elements of an experience together, allowing for the creation of a memory trace that can include multisensory details. Over time, consciousness provides a medium in which these memory traces can be "replayed," a mechanism key to their successful storage.

This idea has been hinted at or suggested in some form for decades (Dafni-Merom & Aray, 2021). In 1985 Tulving wrote, "remembering is a conscious experience. To remember an event means to be consciously aware now of something that happened on an earlier occasion," and he further explains that different memory systems are characterized by different kinds of consciousness: procedural by anoetic (non-knowing) consciousness, semantic memory by noetic (knowing) consciousness, and episodic memory by autonoetic (self-knowing) consciousness (Tulving, 1985). In 1995 Moscovitch wrote, "Consciousness is an inherent property of the memory trace, being bound to it along with other aspects of the experienced event by the hippocampus and structures related to it. ... With respect to remembering, and perhaps with respect to no other function, consciousness is also an inherent property of the very object of our apprehension," (Moscovitch, 1995). However, as these examples show, it is normally explained as the other way around, that episodic memories bind conscious experiences together, not—as we are suggesting—that the conscious experience *is* the process of remembering, as we will explain in more detail below.

Cleeremans (2011) has also suggested that learning and memory are necessary for consciousness to develop in his "Radical Plasticity Thesis." In his account, the brain "learns" to be conscious by continuous attempts to predict the consequences of its actions on both itself and the outside world. This activity produces meta-representations that, when combined with the emotional value associated with them, produce conscious experience (Cleeremans, 2011).

Thus, although learning, memory, and plasticity are necessary for consciousness in his thesis, his views are clearly quite different from our hypothesis.

Episodic memory is now understood to have value not just for its ability to represent the past, but for its utility in allowing using past experiences to increase understanding of the present moment and to make predictions about future occurrences (Schacter et al., 2007; Suddendorf & Corballis, 2007). At retrieval, these traces of episodic events can be flexibly and creatively combined and rearranged in consciousness to allow anticipation, future planning, and intentional action. If we view consciousness as having developed as a critical part of the episodic memory system, we believe that all the questions that we asked can be answered, and all the problems that we raised can be solved—or they simply cease to look like problems.

To elaborate, our hypothesis rejects the idea that consciousness initially evolved in order to allow us to make sense of the world and act accordingly, and then, at some later point, episodic memory developed to store such conscious representations. Our hypothesis is that consciousness developed with the evolution of episodic memory simply—and powerfully—to enable the phenomena of remembering. We view the fact that our ability to imagine things in consciousness is constrained by and related to our episodic memory (James, 1890; Moulton & Kosslyn, 2009) as another piece of evidence supporting the idea that consciousness evolved as part of episodic memory.

Today, consciousness certainly participates in functions that we don't generally associate with episodic memory, such as problem solving, abstract reasoning, and language. But we argue that such functions developed later in evolutionary history, after consciousness was already functioning to furnish the content of episodic memory representations.

Consciousness, sensory memory, working memory, episodic memory, and semantic memory are part of the same system

At this point you might be thinking that we're simply confusing consciousness with working memory—the ability to keep information in mind and manipulate it. We respond that our theory is entirely consistent with the ideas put forth by Baddeley, Hitch, and others regarding the phonological loop, visuospatial sketchpad, central executive, and the episodic buffer (Repovs & Baddeley, 2006). Moreover, we would argue—along with many others—that although you can split semantic, episodic, and working memory into separate systems, in the healthy brain they function seamlessly together as a single system (Renoult et al., 2019; Repovs & Baddeley, 2006).

Information comes into working memory of which we are consciously aware either from our senses, via sensory memory, or from long-term memory stores. For example, let's say that you're walking in your neighborhood when you hear a bark. This auditory information is present in your sensory memory for mere seconds, but it is long enough for you to transfer it into your working memory. Once in your working memory, this auditory information acts as a cue that triggers the retrieval of an episodic memory from last week—when the dog that makes that particular bark chased you to the edge of its property! Now that the sensory information of the bark and last week's experience of being chased are together in your working memory, it isn't difficult to imagine the future: the dog is likely to chase you again. Without waiting for the dog to chase you, you quickly cross the street.

So we have sensory memory, episodic memory, and working memory helping us sense, remember, imagine, and act. But where does consciousness come in? Our argument is that you "hear" the bark when you are consciously aware of it. You consciously remember last week's chase when you retrieve the elements of the prior experience and build a conscious representation of that prior experience in working memory. Without effort, you consciously imagine the dog chasing you again. And—either consciously with deliberation or automatically and unconsciously without deliberation—you walk across the street. In this example, we view consciousness as necessary for the awareness, retrieval of episodic memories, and imagination of the future that may have led to your action of crossing the street. We view consciousness as an integral part of episodic memory, working memory, and sensory memory, and in that way we view all of these elements as part of the same memory system.

Now, you might be thinking that you don't really need consciousness for the action you performed of crossing the street. In fact, you might correctly note that you don't need episodic memory either. You just need operant conditioning to hear the bark and walk across the street. We agree—and it is one reason we choose this particular example. There are two points we would like to make here.

First, this is yet another example where, at least at a superficial level, consciousness does not seem to be necessary. So why is consciousness present? We now have an explanation that we didn't have when we first discussed this problem above: *Consciousness is necessary to remember this event*. Could consciousness be epiphenomenal in relation to the dog barking and your walking across the street? Yes. But it could not be epiphenomenal in relation to your forming and retrieving an episodic memory; we argue it is essential.

Second, we believe that your conscious episodic memory record of the dog barking is critically important in many situations. One is when the stimuli and setting are not similar enough to trigger the conditioning response. Another is when the conditioning response is triggered but the situation requires us to choose actions different from what we are conditioned to do. In both of these situations, the ability to consciously retrieve different memories and imagine different scenarios is critical for making decisions that could have evolutionary implications, as pointed out both by Schacter, Addis, and Buckner (2007) and Suddendorf and Corballis (2007). In fact, individuals with impaired episodic memory show impairment in their ability to learn from their prior errors despite showing near normal implicit memory (Baddeley & Wilson, 1994), again demonstrating that conscious episodic memory is critical for learning that leads to flexible future performance.

Now, it would certainly be useful to be able to generalize your experience with this particular dog to other dogs, so that, if you're taking a shortcut through a different neighbor's yard and their dog begins to bark, you'll know that you should move swiftly to get off their property even before the dog starts chasing you. This ability to form general facts from specific events stored in episodic memory is, of course, one way to define semantic memory. Here we would simply like to make the point (that many others have as well [e.g., Janssen et al., 2022; Renoult et al., 2019]) that although one can view semantic memory as a separate memory system, you can certainly think of it as being part of the same system as episodic memory.

Our view differs from Baars's global workspace theory

Perhaps at this point you are thinking we are merely espousing the global workspace theory of consciousness proposed by Baars (Baars, 2005), and expanded upon by others (e.g., Gaillard et al., 2009). To explain this theory, Baars uses a metaphor of a theater in which consciousness "resembles a bright spot on the stage of immediate memory, directed there by a spotlight of attention under executive guidance.” Although our theory is mostly consistent with global workspace theory, the major difference is that our theory adds the original purpose of consciousness: to allow us to store prior experiences in episodic memory.

Baars outlines six theoretical claims, as reproduced in Table 1.

Table 1. Theoretical claims: brain capacities enabled by conscious events (Baars, 2005)

1. Conscious perception enables access to widespread brain sources; unconscious sensory processing is much more limited
2. Conscious perception, inner speech, and visual imagery enable working memory functions; there is no evidence for unconscious access to working memory
3. Conscious events enable almost all kinds of learning: episodic and explicit learning, but also implicit and skill learning
4. Conscious perceptual feedback enables voluntary control over motor functions, and perhaps over any neuronal population and even single neurons
5. Conscious contents can evoke selective mechanisms (attention) and be evoked by it
6. Consciousness enables access to the ‘‘observing self’’ — executive interpreters, involving parietal and prefrontal cortex

Regarding claim 1, we certainly agree that conscious perception enables access to widespread brain sources, but we do not believe it has been proven or disproven that unconscious sensory processing is “much more limited.” Our suspicion is that it is not more limited than conscious perception.

We agree with claim 2; we believe that part of the definition of working memory includes that it is conscious.

Claim 3 also makes sense; episodic and other forms of explicit memory are conscious by definition. Many instances of skill learning and other forms of implicit memory are enabled or facilitated by conscious events, such as taking a tennis lesson or practicing the viola. There are, however, a number of instances of implicit learning, such as priming, in which the learning occurs without consciousness, but since the claim is "almost all," we can still accept this claim.

Claim 4 consists of two parts. We agree that conscious perceptual feedback occasionally enables voluntary control over motor functions (such as when you're working on improving your backhand or your vibrato), but note that, in our view, the vast majority of motor actions are unconscious and thus the perceptual feedback will also be unconscious. The second part of claim 4 relates to conscious perceptual feedback enabling voluntary control over "any neuronal population and even single neurons." On the one hand, if we are consciously changing our behavior there will necessarily be changes in the brain—this seems obvious unless one is a dualist. On the other hand, whether it is single neurons or assemblies of neurons that are changed is an empirical matter that future experiments can resolve. We will discuss our views on the relevant anatomy for consciousness below.

Claim 5 seems logical—attention can be directed by conscious thought but can also be captured by relevant stimuli (such as your name, your need to use the toilet, or a fire-alarm signal), which then becomes the contents of consciousness.

Like claim 4, claim 6 has phenomenological and brain components. The phenomenological claim, that through executive interpreters consciousness enables access to the "observing self" makes at least intuitive sense, although it seems to evoke a homunculus sitting in the Cartesian theater. Regarding the brain components, parietal and prefrontal cortex are certainly important, and we'll discuss them in more detail below.

Lastly, we will reiterate that, in addition to these differences, the major distinction between global workspace theory and our theory is that ours adds that the original purpose of consciousness was (and a major purpose still is) to facilitate the encoding, storage, retrieval, and flexible recombination of prior events using episodic memory and its related memory systems.

Attention is necessary but not sufficient for conscious awareness

Insofar as what you attend to determines which contents are engaged in working memory and thus have the potential for being stored as episodic memory, we believe that what you are conscious of (the object of consciousness) depends upon what you are attending to (the object of attention). Attention is therefore necessary for conscious awareness.

Attention is not, however, sufficient for consciousness, as has been demonstrated through many experiments that have manipulated both voluntary (endogenous) and external, reflexive (exogenous) attention producing a change in performance and/or reaction time without conscious perception (Kentridge et al., 2008; Hsieh et al., 2011; Zhang et al., 2012). For overviews see Breitmeyer (2015) and Breitmeyer et al. (2015).

There are also examples where attention to one stimulus can cause others to disappear or be absent from conscious perception. The Troxler fading effect shows that attention directed to one area of a visual scene can cause effects of either fading or intensifying to others, which may be due to microsaccades (Alexander et al., 2021). In motion-induced blindness, relevant targets may disappear when viewed against a moving background (Thomas et al., 2017). In change blindness (Andermane et al., 2019), inattentional blindness (Hutchinson et al., 2020), and the attention blink (Sy et al., 2021), even highly incongruent stimuli—such as someone in a gorilla costume walking through a basketball court (Simon, 2010)—can be invisible to conscious perception.

We believe that all of these phenomena are consistent (or, at least, not inconsistent) with our memory theory of consciousness. To reiterate, attention is necessary but not sufficient for stimuli to enter working memory. Therefore, stimuli which are unattended to will neither be consciously perceived nor remembered using working or episodic memory. If these unconsciously perceived stimuli are learned it occurs with unconscious memory processes, such as priming.

Our theory is consistent with the System 1 (unconscious) and System 2 (conscious) distinction

Our view of consciousness is fully consistent with the distinction Kahneman and Tversky made between the slow, effortful, logical, calculating, conscious System 2, versus the fast, automatic, stereotypic, unconscious System 1 (Kahneman, 2011; see also Carruthers, 2015). Our theory would simply add that conscious System 2 was made possible by the original purpose of consciousness, to be the contents of episodic memory.

From the contents of memory to problem solving and abstract reasoning

How did consciousness move from being solely the contents of episodic memory to being involved with problem solving, abstract reasoning, and the other processes made possible by System 2? We speculate that consciousness evolved and became involved in these other abilities due to consciousness being a key element of the episodic memory system's function of flexibly and creatively combining episodic memories to imagine the future.

We envision an early stage of this type of combination that would simply allow you to predict the future. For example, episodic memories of finding a delicious berry near a specific cave each autumn might combine with another episodic memory of being chased by a bear near that cave. The outcome is that you are able to predict that if you go to pick the berries you might end up being chased by a bear.

In addition to merely envisioning how the future might unfold, at some later point this conscious memory creative recombination process envisioned two or more possible future outcomes. In one future, when you go to pick the berries you're chased by the bear, while in another future, you're not chased.

Once consciousness is able to compare two possible futures, the problem solving comes in when you think about what you can do to help bring about the future you would like and avoid the future you don't. For example, memories may come to mind of ways to determine whether an animal is in their lair or not. Other memories may remind you that a bear can only chase one person at a time. Comparing these types of memory retrievals—that perhaps we may now refer to as thoughts—can allow a plan to develop. Problem solving in consciousness/working memory via episodic memory retrieval is now taking place.

Once problem solving is occurring, it is a small step to conscious abstract reasoning. As multiple events stored in episodic memory generalize into semantic memory, abstractions automatically occur. Episodic memories of individual dogs, bears, and rabbits allow for the general semantic memory categories of dog, bear, and rabbit to be formed. A more abstract semantic memory category of animal can then emerge, and be contrasted with the abstract semantic memory category of plant, etc.

Language

Language represents an interesting case. Much has been written about the development of language which will not be repeated here (e.g., Pinker, 2007). Succinctly, we believe that language developed from conjunction of consciousness and semantic memory.

One of the things that makes language an interesting case, however, is that although we can certainly speak with full conscious awareness and deliberation, it is our observation that we can also speak unconsciously, without thinking about it. We will return to this important concept in a later section. For now, we simply want to introduce the idea that just because a function developed with consciousness does not mean that it must only be present with consciousness.

Conscious perception as a memory

At this point in our dialogue, we imagine that you might be willing to accept our arguments that consciousness evolved as part of the episodic memory system but say, "So what? How does this explanation help us understand consciousness (or, for that matter, episodic memory)?"

If we believe that consciousness evolved as part of the episodic memory system, as a critical part of that system that allows us to store prior experiences in memory and retrieve them so that that the memories of these experiences can be flexibly and creatively combined to allow future planning and intentional action, there is no reason that consciousness needs to operate in real time. If consciousness is a system for memory encoding and retrieval—and not direct action—there is no reason that it cannot function properly with a small delay. We argue, in fact, that we do not consciously perceive events directly in real time. We perceive the world as a memory. In other words, technically, we are not directly consciously perceiving anything; we are actually experiencing a memory of a perception.

We suggest that we experience the world by remembering sensory memories. Moreover, most of the time we are not experiencing these “bottom-up” sensory memory processes by themselves. We experience sensory memory processes influenced by “top-down” episodic and/or semantic memory processes, such that the percept that is consciously perceived is a mashup between these bottom-up sensory memory processes and these top-down episodic and sematic memory processes.

Postdictive effects explained

Our theory can now explain postdictive effects. For example, Sergent and colleagues showed that not only does cuing a stimulus prior to its presentation improve conscious perception of the stimulus, but cuing *after* the stimulus presentation can as well (Sergent et al., 2013). Sergent and colleagues concluded that:

> the initial sensory processing associated with a stimulus can occur preconsciously, because its conscious or nonconscious fate can change drastically beyond this phase. Conscious perception would thus relate to secondary amplification of preconscious information held in sensory areas. …this secondary amplification does not have to be a direct consequence of the initial processing of the stimulus itself, but can be triggered by a subsequent and independent event. (Sergent et al., 2013)

Our memory theory of consciousness is both consistent with this explanation and can allow us to understand it better. Attention is drawn to the unconscious perception by the post-stimulus cue, and we then experience the conscious perception in the same way that we experience all conscious perceptions—by remembering it. We will review how our theory explains other postdictive effects in the next section.

The conscious memory of unconscious decisions and actions

If you've come along this far and are willing to consider that you—at least as a conscious self—don't perceive the world directly but rather remember it, we're ready to explain conscious decisions and actions. Our theory is simply that the "you" who decides and acts is unconscious. Your "conscious decisions" and "conscious actions" are actually memories of those unconscious decisions and actions. We believe that this explanation—that decisions and actions are fundamentally occurring through unconscious brain processes—is consistent with an evolutionary perspective that would argue there is no single conscious "decision-making system" in the brain. Instead, there are various processes that are unconsciously engaged to make specific decisions, such as when to eat, sleep, avoid, approach, grasp, release, etc. (Cisek, 2019).

We do think, however, that unconscious brain processes will sometimes engage the conscious memory system to facilitate optimal decision making and performance in certain situations. To explain these notions further we will provide an example using Kahneman and Tversky's System 1 and System 2 processing (Kahneman, 2011; see also Carruthers, 2015).

System 1 decisions and actions

Let's first consider System 1 decisions. These are the fast, automatic, unconscious decisions that require little or no thought or effort. Let's say you're working, perhaps engrossed in thought writing a paper on your computer. You suddenly decide, "I'd like a glass of water." Before you finish typing the paragraph you're working on and still thinking about the sentence that comes next, you rise, walk into the kitchen, open the cabinet door, reach inside, grasp a glass, pull it out, close the cabinet door, hold the glass under the faucet, turn the cold water on, watch the glass fill, turn the water off, raise the glass to your lips, tilt the end upwards, take a sip, and, while holding the glass steady so as not to spill, walk back to your computer, set the glass down, sit down, and get back to work.

We would argue that your decision to get a glass of water could well have been unconsciously initiated and acted out, as could every part of the sequence. In fact, we believe that almost all of the time when we walk or grasp an object it is completely unconscious brain processes that carries out these actions (consistent with Aglioti et al., 1995, Chen et al., 2015, and Cisek, 2019). Insofar as we are consciously aware of what we are doing, we suggest that this conscious awareness is a memory of this decision and action. On this account, consciousness is epiphenomenal with respect to the decisions and actions, but not epiphenomenal in general because it plays an important role via the episodic memory system.

System 2 decisions and actions

Now let's consider some analogous System 2 decisions and actions. Perhaps, instead of working on your computer, you're participating in a dangerous *Hunger Games*-like activity. You're thirsty, and you can see the cool water up ahead. But to reach it you need to either make your way across a bubbling lava field with floating rocks that you have to step on, or you can run across a grassy meadow filled with poisonous snakes and spiders. You carefully consider your options. In the end you decide that you'll have a better chance jumping from floating rock to floating rock across the lava. You carefully step on one rock, get your balance, and wait for another to float nearby. You time your jump perfectly and land in a crouch, distributing your weight. You continue this way until your reach the spring. You can see the glass you need, but to get it you need to carefully reach your hand through razor-sharp barbed wire. You contort and angle your hand and fingers to reach the glass, and delicately pull it back through. You fill your glass from the stream and drink down the precious liquid.

Here you have a series of decisions and actions that need to be thoughtfully and carefully carried out. Instead of acting automatically (perhaps while thinking about something else), each of these decisions and actions is consciously considered and fully attended to. However, we believe that the actual decisions and actions themselves are made and carried out by your unconscious self, and that you experience the conscious decision being made or action taking place only after the fact.

Unconscious perceivers, decision makers, and actors

Another way of saying this, is that it is the System 1 unconscious parts of your brain that actually perceive, make decisions, and act (consistent with Cisek, 2019). System 2 provides an additional layer of information that your unconscious brain can use (or not) to make decisions and act accordingly. System 2 uses consciousness and thus all of the explicit memory systems to review what you know about snakes, spiders, and lava flows (from semantic memory), how well you did the last time you had to jump from rock to rock (from episodic memory), and how, by counting in your head and watching the floating rock (using working memory), you might be able to time a jump perfectly.

The conscious memory system

Thus far we have sometimes been using the term 'episodic memory' in its standard definition (memory for prior events), and sometimes as shorthand for all of the explicit memory systems: working memory, episodic memory, and semantic memory. Because we believe that all of these explicit memory systems are truly part of one system—the explicit or *conscious memory system*—going forward we will use the terms episodic memory and episodic memory system just in their narrow senses, and use terms conscious memory and conscious memory system in this broader sense referring to all explicit memory systems.

Challenges for our theory

If we are correct that consciousness and explicit forms of memory are all part of the same system, there should not be examples of explicit memory that are unconscious. One could immediately argue that there cannot be examples of unconscious explicit memory, as that would be the same as saying that there are examples of unconscious conscious memory. Nonetheless, it has been shown through carefully conducted experiments that unconscious processes similar to episodic memory using similar anatomical networks have enabled participants to perform inferences that would usually require conscious awareness to perform (e.g., Schneider et al., 2021). In addition, other carefully conducted experiments have shown unconscious (i.e., guessing) above-chance performance on delayed-response working memory tasks that would usually require keeping information consciously in-mind (e.g., Trübutschek et al., 2017). Do these experimental results mean that our theory must be false? Although we readily admit that such experimental results are problematic for our theory and need to be explained, we would argue that they do not negate it, for the reasons explained below.

Unconscious episodic memory?

In the Schneider et al. (2021) study, participants were exposed to either weak masking of stimuli that allowed for conscious processing, or strong masking of stimuli that necessitated unconscious, subliminal processing. Conscious processing led to an improvement in reported accuracy as well as a reduction in reaction time. Unconscious processing led to a reduction in reaction time, but did not improve accuracy, which was at chance. Further, this reduction in reaction time for unconscious processing was only observed in “intuitive” decision makers who habitually would respond according to their instincts (using System 1), and not in “deliberative” decision makers who prefer relying on consciously accessible knowledge (using System 2). The fMRI experiments that showed the neuroanatomical correlates of this unconscious processing were conducted exclusively in intuitive decision makers.

Our first comment is that it is possible that even the strongly-masked stimuli were minimally or partially conscious, as it is difficult to exclude this possibility in experiments of this type (e.g., Holender, 1986; Timmermans & Cleeremans, 2015). Thus, one explanation of these results is that the strongly-masked stimuli engage the episodic memory system because they are minimally or partially conscious for some subjects.

Our second comment is that if we agree that that the strongly-masked stimuli are processed unconsciously, it is possible that they still do engage the episodic memory system, but only partially, and not strongly enough for a full, true, conscious episodic memory to be formed. Support for this view comes from the fact that these strongly-masked stimuli produced a change in reaction time, but not a change in accuracy. Thus, we would argue that although these strongly-masked stimuli unconsciously activated the episodic memory network and produced a change in reaction time, there was no change in accuracy, no true episodic memory formed, and therefore no requirement for consciousness.

Unconscious working memory?

In the Trübutschek et al. (2017) study, participants identified the location of visual stimuli after a delay. Stimuli were rated by participants on a 1 (unseen) to 4 (clearly seen) scale. They were instructed to guess at the location even if they were not able to see the stimulus. Behaviorally, participants performed greater than chance on both seen and unseen trials. Magnetoencephalographic (MEG) data was also obtained, in order to determine if the same or different neural mechanisms were used by participants when identifying the seen (correct) trials versus the unseen (correct) trials, with the hypothesis that if different neural mechanisms were engaged, accuracy on unseen (correct) trials were not simply due to participants misclassifying trials as unseen that were, in actuality, glimpsed. MEG data clearly showed two different neural mechanisms. The authors found that conscious and non-conscious working memory used different brain mechanisms, and that non-conscious working memory used an "activity-silent mechanism" based on slowly decaying calcium-mediated synaptic weights. They then postulated that perhaps this activity-silent mechanism underlies both conscious and non-conscious working memory, which they supported with modeling (but not empirical data).

Our comment on this study is simply that there are many examples of unconscious processing that leads to a future change in behavior or performance; priming and procedural memory being two. Thus, although we do not dispute the findings that there may be an activity-silent mechanism that supports unconscious processing of a spatial delayed response task, we would suggest that calling this processing "non-conscious working memory" may not be using the best nomenclature.

Answers and solutions

Let's now review the many previously inexplicable findings discussed earlier and consider how our theory can provide explanations for each, along with some additional inferences.

Order problems: Consciousness after the perception, decision, action

If the contents of our consciousness, i.e., what we are consciously aware of, is a memory of the perception, decision, and action, then there is no difficulty with consciousness occurring after the perception, decision, and action.

Tolling bells and cocktail parties

Thus, there is no difficulty with your being able to count off the strokes of the clock even though you didn't pay attention to them until the last chime; your conscious awareness is always a memory of the chimes. Similarly, it's not surprising that when you hear your name at a cocktail party that you can hear the earlier part of the sentence; you're remembering it, just as you do with all of your other experiences. This feature of our awareness just becomes more apparent to us when we reflect on certain types of situations.

Motor cortex first, conscious decision to move second

The experiment conducted by W. Grey Walter now makes sense as well. In the normal individual (without electrodes implanted into in their brain) who is controlling a slide show with a carousel and a push-button controller, their *experience* is that they make a conscious decision to advance the slide, consciously use their thumb to push the button, and then the slide advances. However, what *actually occurs* is that they make an unconscious decision to advance the slide—then consciously remember that unconscious decision, unconsciously use their thumb to push the button—then consciously remember that unconscious action, and then the slide advances. The conscious memory for the decision and action are “time-stamped” by the brain to occur not only in the proper order but also at the proper time such that it appears this conscious decision and conscious action coincided with the unconscious decision and unconscious action even though the conscious memory of these events was actually experienced after the events themselves. This is not strange or mysterious. It is the nature of memory that remembered events are referred to a previous time.

In the case of the subjects with electrodes implanted into their motor cortex, the subjects unconsciously decide to advance the slide—then consciously remember that decision. They unconsciously decide to use their thumb to push the button, which creates the motor cortex readiness potential that triggers the slide to advance, the slide begins to advance, they consciously remember the slide beginning to advance, then consciously remember pushing the button, generating this feeling of the slide projector anticipating their decisions.

There is also an experiment performed by Libet related to motor movement. Libet (1985) compared the time in which participants consciously decided to move their wrist (determined by participants noting the time they decided on a special clock), with the measured readiness potential of this voluntary action at the scalp. The readiness potential has been interpreted as representing the final stages of planning preparation for movement. What was startling was that Libet found that the readiness potential *preceded* the voluntary decision by about 350 ms. Libet concluded that the initiation of a spontaneous voluntary act “begins unconsciously” (Libet, 1985).

We have not introduced this important experiment previously because it is quite controversial for at least two reasons. The first is that not everyone who has tried to replicate it has been able to do so, although some have (e.g., Vinding et al., 2014), which should set that issue to rest. The second is that Schurger and colleagues have conducted a terrific set of related experiments and analyses that they suggest explains the readiness potential not as representing the final stages of motor preparation, but as spontaneous subthreshold fluctuations in neuronal activity (Schurger et al., 2012).

We do not disagree with the work done by Schurger and colleagues, as their explanation may be the correct one for this phenomenon. We would simply like to point out that our theory makes Libet's initial interpretation comprehensible—decisions and actions are initiated unconsciously and then we experience the conscious memory for those decisions and actions. If we believe Libet's interpretation and these motor results are generalizable, our conscious memory for our decisions and actions may occur 350 ms after the decisions and actions are unconsciously initiated.

The incredible slowness of consciousness

We can now understand why it doesn't matter that consciousness is too slow for real-time decisions and actions of athletes, musicians, and others who need to react quickly. All decisions and actions are occurring unconsciously. Our conscious memory for these decisions and actions occurs later.

Conscious sensations referred backwards in time

Stimulation of the hand versus the somatosensory cortex

Our memory theory of consciousness also helps us to understand some of Libet's other experimental results and how these results are informative regarding the timing of consciousness. Recall that Libet found that it took approximately 500 ms of cortical stimulation before conscious experience of a tingle occurred. If the stimulation was shorter than 500 ms, the subjects did not report feeling anything. He referred to this extended time as “the neuronal adequacy for consciousness.” He explained that we aren’t aware of this delay because events are referred backwards in time after neuronal adequacy has been achieved (Libet et al., 1979). The idea is that when you feel a touch on your arm, the impulses travel through your peripheral nerves, spinal cord, and brain until they reach the somatosensory cortex. If the stimulus activates your cortex for at least 500 ms, you consciously feel the touch and it is referred backwards in time such that you do not notice any delay in the conscious perception (Blackmore, 2017; Dennett, 1991; Libet et al., 1979).

The first point to make here is that our new understanding of consciousness—that we don’t perceive events directly but only remember them—makes it easy to understand this finding of Libet: Once the 500 ms threshold for conscious sensation is crossed you can simply remember the sensory memory of the touch. It’s not even surprising that it is referred backwards in time—again, it is fundamental to memory that it allows us to experience events that occurred earlier.

The second point that Libet’s experimental results could suggest is that 500 ms is the amount of time that our conscious perceptions are delayed. In other words, if this result from Libet is generalizable to other conscious sensory experiences, our conscious perception of the world may be delayed by half a second—although referred backwards in time so that we don’t notice this delay.

Chronostasis

The problem of the stopped clock illusion is that the final fixation of the clock is projected *backwards in time* to fill in the time period when you were making the saccade. This is no longer a problem because your visual perceptions are not directly experienced; they are consciously remembered. In other words, you are consciously perceiving a memory that is delayed approximately 500 ms. Thus, a top-down process in your visual system projects the final fixation backwards in time to fill in what would have been the missing perceptual experience while the saccade was taking place. Problem solved.

Postdictive effects

Because conscious perception is a memory, likely delayed by about 500 ms and referred backwards in time, we now have an easy and comprehensible explanation of postdictive effects.

Rabbits

In the case of the cutaneous rabbit, there are five taps at the wrist, three near the elbow, and two at the shoulder; yet you consciously experience intervening taps as well. Again, because sensations are consciously remembered (500 ms later) rather than being directly consciously experienced, it is not a problem for some top-down process to interpose, backwards in time, intervening taps between your wrist and elbow, and between your elbow and shoulder, such that you consciously perceive a little rabbit running up your arm. An analogous explanation can explain the illusory and invisible audiovisual rabbits.

Color fusion effects

Color fusion effects can be explained in a similar fashion. When a red disk is presented for 40 ms by itself, there is the obligatory delay and then you consciously perceive the red disk by remembering it. When the red disk is followed by the green disk in the same location, the colors are fused together in your sensory memory (because they are occurring in the same time window) and, after the 500 ms delay, you perceive the fused yellow image by remembering it.

TMS pulses

That a TMS pulse between 20 to 370 ms can induce postdictive effects should no longer be surprising. Because conscious perception is delayed, if there is disruption to the visual stream during the delay period, it can alter conscious perception of the prior stimulus just like a physical stimulus can.

Color phi illusion

Similarly, we can now explain why is it that when you watch the color phi illusion and see a blue dot at the top of a frame followed by a blank screen, and then a red dot at the bottom of the frame, you consciously experience the blue dot traveling down and changing color *before* you see the red dot. Again, a top-down process has interposed, backwards in time, the intervening dots and the color change into your conscious perception—easy to do because that conscious perception is actually a memory.

What about the suggestion by Keuninckx and Cleeremans (2021) that the color phi illusion may simply be related to "Inherent dynamical and nonlinear sensory processing in the brain" and not related to consciousness, per se (Keuninckx & Cleeremans, 2021)? We would argue that our work helps to elucidate one part of this phenomenon (how events are referred backwards in time), while theirs helps to elucidate another (why is there a sense of motion and color switch before the second stimulus is seen).

Creating uniformity: Remembering the gist

Our memory explanation of consciousness may also help to explain why we don't notice that our peripheral vision is greyscale or the blobs and stripes related our fixations and saccades with black areas in between. Although here we admit that we may be pushing the explanatory power of our theory past its limits, one possible explanation relates to the type of information that our conscious memory system remembers. In general, we tend to remember the general concept, idea, or gist of the information (Reyna & Brainerd, 1995). Because we now understand that we don't directly consciously perceive visual information—we remember it—we may be remembering the gist of the visual scene in the same way we might remember the gist of a conversation, movie, or a list of items. If consciousness does not have the function of capturing our occurrent sensory input, these features of our visual experiences are not problematic.

Our view of conscious perception is thus consistent with that espoused by Cohen et al. (2016). Using an understanding of perception that arises from the field of visual ensembles and summary statistics, they describe how observers can extract the gist of the scene with just a few fixations (Cohen et al., 2016).

Does conscious perception overflow working memory?

Block (2011) and others have suggested that phenomenally conscious perception has a rich content with a greater capacity that "overflows" the more limited access consciousness of perception. Evidence supporting this overflow idea comes from the Sperling paradigm, in which an array of letters (e.g., in a three by four grid) is briefly shown. Subjects generally report seeing all (or almost all) of the letters, yet they can report only three to four letters. Powerfully, however, when they are cued to report the letters from any row, they are able to recall three to four letters in that row (Sperling, 1960), suggesting that all of the twelve letters were potentially accessible. Our memory theory of consciousness would explain that the full array of items is present briefly in unconscious perceptual processes, although we only experience the conscious perception of the items that our attention is drawn to by the post-stimulus cue, as only after our attention is drawn to one row of letters are those items transferred into sensory memory and then into working memory.

This explanation is analogous to the one we discussed above for Sergent et al. (2013). Kentridge (2013) has noted that Sergent's result "does not necessarily invalidate the distinction between access and phenomenal consciousness, but it does lend weight to the alternative, and perhaps simpler, position that consciousness is just consciousness." We agree with Kentridge (2013) and Cohen et al. (2016) that although distinguishing between phenomenal and access consciousness appears to be a useful distinction, it leads to strange situations in which one can have a phenomenally conscious experience that one's conscious mind doesn't have access to. This could be rephrased to say that there can be phenomenally conscious experiences that are unconscious, an idea that doesn't make much sense to us. We believe that such distinctions have been postulated to solve some of the problems of consciousness, problems that we believe our memory theory of consciousness can explain without the need to invoke phenomenal and access consciousness.

Mindfulness

We can now understand why it is difficult to control our thoughts when we practice mindfulness. Our argument is that consciousness did not develop in order for us to perform high-level abstract reasoning using language, logic, visuospatial abilities, or other cognitive capacities to allow us to carry out intentional actions. Consciousness developed in order for us to remember events and information, as well as to creatively and flexibly recombine those events and information. We speculate that much of this remembering—and even the recombination of the remembered events and information—can occur without volitional control over consciousness. In other words, we may consciously perceive events and, at a later time, consciously imagine the recombination of elements of those events, even if the recombination is being directed by unconscious processes. As anyone knows who has practiced mindfulness, it can take great effort to sustain conscious attention to a single object—because, we argue, that is not what consciousness developed to do.

If we consider mindfulness from our new perspective, we might imagine the following processes occurring. A thought is generated unconsciously, perhaps of a meeting you are anticipating later today. Your attention is captured by this thought. Depending upon the goals of your mindfulness session, you may be content to simply be metacognitively aware of this thought, observing with some detachment the different emotions the thought produces (excitement that the meeting might go well as well as anxiety that it might go poorly). Or you may attempt to nudge your awareness from the meeting to your breath as you try to consciously attend to air going in and out of your nostrils.

The subjective experience of consciousness

One of the most exciting consequences of our memory theory of consciousness is how it might possibly explain certain aspects of the subjective experience of consciousness.

Stream of consciousness

You likely feel that James's metaphor of a stream of consciousness feels intuitively correct, with the momentary "now" where you are standing in the river, past events flowing progressively downstream, and future upstream events that are going to occur rushing toward you. Part of the power of this metaphor is that it is fairly linear. Yet, we know that the brain is processing a massive amount of information in parallel. Why do we experience events serially instead of in the parallel manner that the brain processes them? We would argue it is because it is a property of our conscious memory system to remember—and thus to consciously experience—events serially in time. Once we decouple real-time sensory input from consciousness, we no longer need the two to be processed by the brain in the same way.

We also believe that our understanding of consciousness explains, at least on one level, why events are bundled together over time such that they seem continuous. Again, we would argue that it is part of the neural architecture of our conscious memory system to store and retrieve events that are time-stamped in a certain order.

In the Cartesian theater

Do you have the intuitive experience that you, as a conscious self, are sitting inside your head—in the proverbial Cartesian theater—peering out at the world through your eyes, as if you were watching a movie? If so, we believe it is because you are not experiencing your perceptions directly; you are experiencing a memory of perception. Who is this "you"? Who is this homunculus sitting in the Cartesian theater? It is your conscious self remembering your perceptions, decisions, and actions. Why did it develop this way? We believe it developed this way to allow that flexible recombination of prior events and information through imagination, like a movie director moving scenes around on a storyboard.

Do we now have to deal with an infinite regress of homunculi sitting in Cartesian theaters? We would argue not. You have one conscious self. The conscious self is sitting in the theater, mostly passively, watching memories of experiences. The various parts of your unconscious self are processing information in parallel, sometimes paying attention to what is going on with the conscious self but mostly ignoring it. There is only one conscious homunculus. It stops right there.

Consistent with higher-order theories

Higher-order theories of consciousness claim that a mere first-order representation (being presented with a red stimulus) would only lead to consciousness if you are in some way aware of having that experience (being aware that you are seeing red). These theories also claim that conscious experiences involve some type of minimal inner awareness of one's ongoing mental functioning due to the first-order representation being monitored or represented by a relevant higher-order representation (Brown et al., 2019).

Although our memory theory of consciousness clearly differs from high-order theories of consciousness in many ways, we believe that our theory is consistent with the idea that you are not conscious of the first-order brain mechanisms that processes the presentation of a red stimulus in front of your eyes, and that you only become conscious of this stimulus when (sitting in the Cartesian theater) you experience the memory of its perception. Thus, we believe that our memory theory can explain this intuitive sense that we don't experience first-order representations directly, but only indirectly. Higher-order theories would say consciousness is experiencing higher-order representations of the first-order representations, whereas our memory theory would say consciousness is remembering the first-order representations.

Resonant metaphors

Our memory theory of consciousness helps explain why some metaphors resonate with us intuitively. Now that we understand why the Cartesian theater feels natural, we can understand why Plato's allegory of cave resonates with us. In a very real sense, we don't perceive the real world directly, we only perceive shadows or reflections of the real world through memory. Similarly, now that we understand how both the feeling of the theater and the serial appearance of reality are created by our conscious memory system, we can understand why the two metaphors of the movie *The Matrix* are so powerful: the thought that we are not truly experiencing the world directly, and that the world is actually made up of massive parallel processing streams of data.

Horse and rider

We would like to introduce one additional metaphor here that captures some of how we think about the conscious and unconscious self. Imagine your brain as a horse and rider together. Your unconscious, System 1 self is the horse, in control of the moment-to-moment journey you are taking. Just as you don't need to provide detailed instructions to the horse on how to cross a rocky field or jump over a short wall, neither do you need provide detailed instructions to your unconscious as to how to carry a cup full of hot coffee across the room and down a flight of stairs—you just need to look at the cup and your unconscious self does the rest. Your conscious, System 2 self is the human rider who is mostly just going for a ride. The rider can, of course, provide either moment-to-moment or more general, overall instructions to the horse, and the horse is usually happy to oblige.

How does this interaction between the conscious rider and the unconscious horse happen? Metaphorically, the rider says a few words, gives a gentle tug on the reins, or squeezes their legs to let the horse know which way they want to go. Approximately 500 ms later the rider can then sense whether the horse has, indeed, made the desired decision and moved in the desired direction. Of course, there are sometimes conflicts—such as when the horse wants to go down the easy path but the rider wants to travel up the mountain.

How is it that our System 2 rider is able to inform careful, considered decisions? Well, it’s sunny out there in the plains, and so the rider always wears a special pair of sunglasses. These sunglasses have a high-tech video screen built into them. In fact, these glasses have no lenses, only the built-in screen. Another downside or limitation is that there is a delay of approximately 500 ms. In other words, the rider is always perceiving the world by looking at the screen, which shows the visual world a little bit after the fact. But these special sunglasses allow a variety of amazing options: the rider can either (1) look out at the world (via sensory memory) 500 ms after it has passed by, (2) access prior autobiographical episodic memories using an active, creative, memory-building process, (3) access prior semantic information, (4) keep information on the screen so that it can be manipulated in working memory, or (5) use a combination of these features to flexibly, creatively imagine possible future outcomes. When faced with a difficult problem, the rider can use all these tools to come up with one or more possible solutions, which they then communicate to the horse, and the horse makes the actual decision—which may or may not be the same as what the rider suggested. Note that because the rider is always perceiving the world 500 ms after the fact, the rider depends upon their horse to make decisions without their input whenever quick, System 1 decisions are needed.

Oh, and there's just one other issue about these special sunglasses that needs to be mentioned. The images come from the horse. The horse is in control of the videos that the sunglasses are showing the rider. We can think of this part of the metaphor as explaining the 500 ms delay the rider experiences when perceiving the world. But the horse is also in charge of whether the rider views the world, prior autobiographical episodic memories, semantic information, working memory, or imagination. As always, the rider can communicate to the horse which images they wish to see, but the horse does not always comply, either because it cannot or because it wishes to show the rider another image. This is why mindfulness is so hard; the horse has a mind of its own.

To summarize this metaphor, we consider the rider (with their sunglasses) to be the conscious memory system—remembering rather than directly perceiving, deciding, and acting. We believe that this conscious memory system was always involved in providing information that could be used by the unconscious brain to make decisions informed by past events and information. Through continued evolution, we believe the conscious memory system developed additional capacities in humans as described above. By contrast, the horse represents all of the unconscious brain processes.

Limitation of these explanations of the subjective experience of consciousness

We wish to clearly state that we are well aware that our so-called 'explanations' of the subjective experience of consciousness do not even begin to get at the hard problem of consciousness—how a collection of neurons and supporting brain tissue produce subjective experience. We are, however, hopeful that our slightly increased understanding of the phenomenology of subjective experience of consciousness can help other researchers look in the right locations and do the right experiments to tackle the hard problem. Our suggestion to them is to focus on the conscious memory system.

Lesion Patients

Our memory theory of consciousness makes it easy to understand how individuals with visual apperceptive agnosia and individuals with blindsight can frequently respond correctly despite lacking conscious awareness of visual objects and other stimuli. Although the visual aspects of their conscious memory system are not functioning properly such that their sensory and working memory is impaired—meaning that perceptions do not enter consciousness—their unconscious self can still perceive stimuli and respond accordingly.

The causal role for consciousness and how consciousness contributes to evolutionary success

We have made our case that consciousness did not develop to play a direct causal role in decision and action, and instead developed as part of the conscious memory system. However, as implied by our horse and rider metaphor, we believe that, in modern human beings—and probably many other animals as well (as we will discuss below)—consciousness is essential to make good decisions and take proper actions. Consciousness enables System 2 (the rider) to use working memory as well as all prior autobiographical and semantic information to inform important decisions. Consciousness is thus tremendously important for evolutionary success—without consciousness, System 2 decisions could not be made. If we had no consciousness, we could still make decisions, but they would always be fast, System 1 (horse only) decisions. Consciousness allows us to make slow, carefully considered System 2 decisions. We will discuss these issues more below when we discuss the implications of our theory.

Consciousness is not epiphenomenal

You may now be thinking that all we are saying here is that conscious memory is important for evolution, and that the actual subjective experience of consciousness is epiphenomenal. Our belief is that subjective experience is an inherent property of the conscious memory system, and that to say you can have one without the other would be analogous to saying that you can have molecular motion without heat. Just as $½MV^2 = 3/2KT$, we believe that the use of the conscious memory system produces subjective experience. As mentioned above, we don't have the answer to this hard problem of consciousness, but we are hopeful that our theory will move the field toward finding that answer.

Neuroanatomical correlates and disorders of consciousness

As might be expected from our experience in philosophy, experimental psychology, cognitive neuroscience, and neurology, our theory of the neuroanatomical correlates of consciousness relates to brain regions and structures, and not to underlying cells, cellular assemblies, or neural oscillations (e.g., Lou et al., 2017). We understand that cellular and molecular microstructures may be crucial to gaining a full understanding of the neurophysiologic basis of consciousness, including the hard problem. Again, we hope that this discussion of what we consider to be key brain regions and structures will help others to dive deeper and achieve a more complete understanding. Because much of our discussion of the possible neuroanatomical correlates of consciousness is related to patients with various brain disorders, we will consider these two topics together.

Other theories of the neuroanatomical correlates of consciousness

There are currently four major theories that make predictions regarding the neural correlates of consciousness: recurrent processing theory (Lamme, 2015, 2018), global neuronal workspace theory (Mashour et al., 2020), integrated information theory (Tononi et al., 2016), and higher-order thought theory (Brown et al., 2019). Some of these theories specifically address the hard problem, whereas others, like ours, only try to point the way toward possible solutions. As pointed out by Yaron and colleagues, each of these four theories emphasizes largely different cortical regions as critical for consciousness (Yaron et al., 2021).

Recurrent processing theory

As introduced earlier, recurrent processing theory suggests that conscious processing depends on horizontal connections and recurrent loops between lower- and higher-brain regions that are extended in time and space and involve changes mediated by NMDA-dependent feedback activations (Lamme, 2015). Posterior cortical regions associated with visual processing of information are emphasized, with prefrontal cortex contributing but not essential for conscious processing (Lamme, 2018).

Global neuronal workspace theory

Arising out of the global workspace theory (Baars, 2005; reviewed above), global neuronal workspace theory proposes that conscious processes occur when information in specialized processors enters a large-scale reverberant brain-scale network of high-level cortical areas linked by long-distance re-entrant loops and becomes ignited, allowing global access by other specialized processors (Mashour et al., 2020). Parietal and prefrontal cortical areas are critical for routing information between other cortical processors. At the neuronal level, large pyramidal cells in cortical layer II/III as well as layer V play key roles.

Integrated information theory

Integrated information theory attempts to directly address the hard problem by starting from the essential phenomenal properties of experience and infers postulates about the characteristics that are required of its physical substrate (Tononi et al., 2016). It also provides a mathematical quantity of integrated information that provides a measure of the degree of consciousness of any system. The occipital and parietal lobes are considered to be critical and sufficient for conscious experience.

Higher-order theories

As introduced earlier, higher-order theories postulate that a first-order representation, such as awareness of a rose, would not be sufficient for conscious experience to arise. An organism must be in some way aware of itself as being in a first-order state in order to be conscious of it. Proponents of higher-order theories consider the prefrontal cortex to be important for conscious perception (Brown et al., 2019).

Neuroanatomical correlates of the conscious memory system

Having reviewed some of the leading theories that suggest which brain regions are important for consciousness, we will now state our hypothesis, which we will then work to support. Because we contend that the conscious memory system supports both consciousness and all forms of explicit memory, we argue that the neuroanatomical correlates of consciousness are the neuroanatomical structures involved in all forms of explicit memory.

Which structures are involved in explicit memory? The hippocampus should certainly be included along with related structures, such as neighboring entorhinal and perirhinal cortex, as well as Papez's circuit including the fornix, mamillary bodies, and anterior nucleus of the thalamus. We would also argue that the cerebral cortex is necessary. One might immediately think of the inferolateral temporal cortex for semantic memory, and then perhaps frontal and parietal cortex for working memory, including regions involved with the default mode network such as medial prefrontal cortex, posterior cingulate cortex, precuneus, and angular gyrus (e.g., Zheng et al., 2021). But we take the view espoused by Murray, Wise, Baldwin, and Graham in their book, *The Evolutionary Road to Human Memory*. They argue that the entire cerebral cortex is important for memory, writing:

> In our opinion, every cortical area contributes to memory, each in a specialized way. As our ancestors traveled along their evolutionary trajectories, cortical areas accumulated over time; and, in each instance, this happened for the same fundamental reason: to transcend problems and exploit opportunities that these animals faced in their time and place (Murray et al., 2020).

Their book provides a wonderful summary of the animal, human brain lesion, and neuroimaging studies that support the view that all cortical structures are not only involved with but are critical for a specific type of memory which is needed for a specific type of task, whether it be navigating the wilderness, distinguishing the sounds of prey and predators, or facial recognition (Murray et al., 2020).

We contend not only that every cortical region contributes to memory but that they each also contribute to a specific domain of conscious awareness. For example, we believe that the following speculations are likely correct: (1) Visual areas in occipital cortex are necessary for visual consciousness including visual imagery. (2) Auditory cortex in superior temporal cortex is necessary for auditory consciousness. (3) Parietal cortex (particularly of the nondominant hemisphere) is necessary for conscious awareness of space (particularly on the contralateral side). (4) Primary somatosensory cortex in the post-central gyrus of the parietal lobe is necessary for certain sub-domains of sensory consciousness such as graphesthesia (identification of numbers or letters written on the skin) and stereognosis (identification of objects by touch). (5) Primary motor cortex in the precentral gyrus of the frontal lobe is important for conscious awareness of the fine motor movements used to play musical instruments, thread needles, and perform other delicate tasks. (6) The frontal eye fields are important for conscious awareness of eye movements. (7) Broca's area is important for conscious awareness of your own speech, whether vocalized or just "in your head." (8) Insular cortex contributes to conscious awareness of your body. (9) Prefrontal cortex areas that facilitate complex thought, working memory, problem solving, and judgment are important for the conscious awareness that comes with these higher-level abilities.

The work of Gazzaniga and others with split-brain patients has demonstrated that when the corpus callosum is severed, the person may be left with two separate consciousnesses in one brain that, seemingly, work together just fine without any apparent functional difficulty or conflict (or, at least, no more internal conflict than any of us have with our unsplit brain from time-to-time) (Gazzaniga, 2015). We believe something analogous occurs not only with left and right hemispheres but with each region of the cortex.

We further suggest that each region of the cortex is autonomously conscious, and its island of consciousness is not dependent on any central executive or other region. That is, the minimally sufficient cortical region needed for conscious awareness may be any cortical region, whether it be a sensory region enabling conscious perception, a motor region enabling conscious movement, or an association region enabling some type of multimodal thought. We believe that this must be the case, as—as far as we are aware—there is no single cortical region (unilateral or bilateral) that, when removed, renders the individual unconscious. Thus, we contend that the visual conscious awareness in the occipital cortex is independent of the auditory consciousness in the superior temporal cortex and both are independent of the conscious awareness and guidance of motor movements occurring in frontal cortex.

What is the evidence for this hypothesis that any cortical region may be autonomously conscious? In large part it comes from the experience of working with several thousand patients who have strokes or neurodegenerative diseases affecting every part of their cortex, supported by the human patient literature. It is well known that damage to subcortical structures that are part of the reticular activating system (such as portions of the midbrain and thalamus) can lead to unconsciousness (e.g., Kinney et al., 1994), but—again—we know of no cortical regions that when damaged lead to unconsciousness. Indeed, it might be that some of these subcortical reticular activating system structures (such as the thalami) act as the hub, switching between different conscious cortical regions.

The literature supporting our view that (1) disruption of no cortical region (even wide-spread frontal or occipital/parietal regions) leads to unconsciousness but (2) virtually all cortical regions contribute to consciousness is well reviewed by the opposing 'front' versus 'back' cortical theories of consciousness (see Boly et al., 2017; Michel & Morales, 2020; Odegaard et al., 2017). We believe that our theory can reconcile those views espoused by the opposing front versus back camps.

For example, if posterior brain regions are critical for consciousness, as suggested by recurrent processing theory and integrated information theory, how is it that patients with posterior cortical atrophy or bilateral occipital and parietal strokes are not unconscious? As we describe below, they most certainly have deficits of conscious awareness, but we would never say that they are unconscious. Similarly, if prefrontal cortex is critical for consciousness, as suggested by global neuronal workspace theory and higher order theories, what about patients with behavioral variant frontotemporal dementia or bilateral frontal strokes? Again, as we describe below, there is no doubt that the consciousness in these patients is impaired, but we wouldn't say that they are the physical manifestation of unconscious philosophical zombies appearing to have a conscious inner world when they actually have none. We believe that they could consciously experience the beauty of a sunset as well as anyone else. Which group of patients is not conscious (in addition to those with reticular activating system damage)? We believe it is patients with either widespread or diffuse cortical dysfunction. These 'awake-but-not-conscious' patients are described as having an encephalopathy or delirium, as we will describe in more detail below.

Exactly how large would a cortical region need to be to support some form of independent consciousness? We don't know the answer, but we will suggest some methods to address that question in the **Future directions** section, below.

Note, however, that we are not saying that the only function of cortex is to provide conscious awareness for specific modalities. The cortex certainly performs much additional work that is solely involved in the unconscious processing of information.

We will also point out that saying the entire cortex is involved in consciousness is not the same as saying the entire brain is involved with consciousness. For those who think of the cortex as more or less synonymous with the brain, we remind you that although the cerebral cortex comprises 82% of the mass of the human brain is, it contains only 16 billion (19%) of the 86 billion neurons in the brain (Herculano-Houzel, 2009).

We now turn to a brief review of some of the major brain regions and structures involved in the conscious memory system, along with some of the relevant neurological disorders that impair consciousness in one way or another.

Hippocampus, related structures, and patients with amnesia

When considering the hippocampus, we typically also consider both neighboring medial temporal lobe structures such as entorhinal and perirhinal cortex, as well as anatomically connected structures such as the fornix, mamillary bodies, and anterior nucleus of the thalamus. Whether because of neurodegenerative disease, infection, inflammation, strokes, seizures, or surgery, the cognitive effects of damage to the hippocampus and related structures are well known. Disruption of episodic memory invariably occurs, leading to anterograde and some retrograde amnesia. Consciousness, at least in the ordinary sense of the term, does not appear to be disrupted by damage to the hippocampus and related structures. Individuals with such damage can certainly consciously experience many perceptions, decisions, and actions.

However, individuals with damage to the hippocampus and related structures show impairment in several aspects of cognition that we would argue is related to consciousness. First, individuals with hippocampal damage lose the ability to consciously perceive subtle differences in visual scenes, and individuals with neighboring perirhinal cortex damage lose the ability to consciously perceive subtle differences in faces (Mundy et al., 2013). Second, these individuals show reduced ability to imagine the future (Addis et al., 2009) and therefore to plan for it. Third, by losing the ability to form new episodic memories, these individuals are impaired in their ability to update their sense of self. We consider the ability to imagine the future and update one's sense of self to be important aspects of consciousness.

Lastly, we speculate that consciousness would almost certainly be impaired if the hippocampus and related structures were completely absent from birth. Although there are studies that show individuals with perinatal damage to some regions of the hippocampus develop normal semantic memory and have not been described as showing impairments in consciousness (Elward & Vargha-Khadem, 2018), we believe that consciousness would not be normal in individuals with *complete* absence or *complete* dysfunction of the hippocampus and related structures since birth.

Occipital Cortex, visual system, Anton syndrome, blindsight.

We described previously how individuals with damage to occipital cortex leading to visual apperceptive agnosia or hemifield blindness can sometimes perform tasks unconsciously that depend upon vision despite the fact that they cannot perform the task with conscious awareness. We believe that these cases support the idea that it is the visual cortex in the occipital lobes that provides conscious awareness of vision.

Patients with damage to their entire occipital cortex bilaterally leading to complete cortical blindness may deny that they are blind (Anton or Anton-Babinski syndrome), a form of 'anosognosia' or unawareness of their deficit (Das & Naqvi, 2021). We believe that this is one example of a general phenomenon whereby the brain region that generates a specific aspect of consciousness—in this case visual perception and visual imagery—is the same part responsible for the awareness of whether that aspect of consciousness is present, absent, or distorted. Our prediction would be that Anton syndrome would not be present if the damage to the visual system affected pathways prior to the occipital lobes (such as the eyes, optic nerves, tracts, radiations, or the lateral geniculate nuclei). Thus, we believe that the phenomenon of Anton syndrome supports the idea of visual consciousness being present in occipital cortex.

Parietal cortex, spatial neglect, and the "aha" moment of recollection

The parietal lobe plays an important role in attention to or awareness of one side of the world. Although both parietal lobes contribute to awareness of both left and right sides, damage to the parietal lobe of the language-dominant hemisphere (usually left), typically produces a mild and temporary loss of awareness (or 'neglect') of the contralateral right side, whereas damage to the nondominant (usually right) parietal lobe typically produces prominent and sometimes permanent neglect of the contralateral left side (Mesulam, 1999).

Neglect most commonly occurs for sensory stimuli that are localized in space, such as visual and tactile stimuli. Individuals with neglect may not notice food on the left half of their dinner plate, may find it difficult to pay attention to someone who is speaking to them on their left, and may not notice a touch to the left side of their body.

What is particularly striking about individuals with right parietal lobe damage and left neglect is the fact that they are typically completely unaware that the left side of the world is missing. As in Anton syndrome, individuals with neglect have anosognosia and are unaware of their deficit. We believe that parietal damage-induced neglect supports the idea of spatial awareness (consciousness) of one side of the world being present in parietal cortex, particularly of the nondominant hemisphere.

Other parietal lobe functions may also be relevant for consciousness. With the exception of retrosplenial cortex, damage to the parietal lobes is not known to impair episodic memory. Yet virtually every recognition memory task evaluated with functional MRI or event-related potentials produces activation of parietal cortex (Simons et al., 2008). How are we to reconcile this discrepancy? Our answer is that the so-called 'parietal old-new effect' (in which previously seen or 'old' items show greater parietal activation compared to novel or 'new' items) is part of the neural correlate of the conscious awareness that an item has been seen before. It's that "aha" moment when you consciously think, "Yes, I remember that" (Ally et al., 2008).

Frontal cortex, motor system and corticobasal syndrome

Regarding consciousness and frontal cortex, we begin by noting that neglect also occurs for movements and activities after supplementary motor cortex damage (Laplane & Degos, 1983). Patients might abandon washing, shaving, or brushing their hair on one side. Similarly, patients with damage to frontal eye fields typically show a form of neglect in that they experience difficulty moving their eyes voluntarily to the opposite side. We believe that these motor forms of neglect support the idea that the frontal lobes are important for the conscious control of movements and activities.

Some patients with brain lesions experience a bizarre delusion in which they do not recognize or believe that their paralyzed limb cannot move (anosognosia) or even that it is part of their body (somatoparaphrenia). Brain regions implicated in somatoparaphrenia include portions of the insula and middle and inferior frontal gyri (Gandola et al., 2012). Thus, conscious awareness of movement and even awareness that a body part is one's own may be related to proper cortical functioning.

Some patients with corticobasal syndrome experience alien limb phenomena, in which an affected limb appears to move on its own—without conscious control. We have seen patients in which the usually useless limb may rise and make simple movements. When the patient is performing tasks that would typically require two hands, such as tying shoelaces, the useless limb can sometimes be cajoled into helping, which it then may do easily and automatically (AEB observations). Although individuals with corticobasal syndrome show involvement in basal ganglia in addition to cortex, atrophy and/or hypometabolism of frontal and parietal cortex is prominent and generally detectable on brain imaging studies. This phenomenon supports the idea of the importance of the cortex in conscious motor control.

Patients with behavioral variant frontotemporal dementia show atrophy and/or hypometabolism of various regions of frontal cortex leading to behavioral problems. Many of these patients demonstrate utilization behavior, which occurs when individuals see a tool and automatically start to use it. A patient may pick up a pair of scissors and begin cutting or a pen and begin writing. We view utilization behavior as an example of damage to frontal cortex leading to impaired conscious control of behavior—but preserved unconscious actions. There are many other examples of damage to frontal lobes leading to loss of conscious control of behavior (e.g., Phineas Gage, Damasio et al., 1994), which we will not review here.

Apathy is a very common symptom in individuals with a variety of damage to prefrontal cortex. Here we would argue that some of the so-called higher frontal lobe functions, such as planning and abstract reasoning, are impaired because conscious awareness and control of such functions has been impaired. Not consciously realizing that one should be planning and acting leads to apathy.

Temporal lobes, auditory cortex, inferolateral temporal cortex, names, words, and meaning

Important aspects of the temporal lobe cortex include parts of the insula, auditory cortex, and the vast store of information that comprises semantic memory. Patients with damage to bilateral auditory cortex lose the conscious awareness of sounds but still may react to them, a phenomenon considered analogous to blindsight (Cavinato, et al., 2012). Animal studies have shown experimentally that damage to bilateral auditory cortex does not necessarily change behaviors in response to sounds (Floody et al., 2010). Thus, we have evidence that auditory cortex is involved in conscious awareness of sounds in a manner consistent with our theory.

Evidence both from imaging and individuals with brain lesions suggest that inferolateral temporal cortex is critical for conscious awareness of the names of people, animals, and tools (Damasio et al., 1996). Moreover, although unilateral, left-sided damage may impair access just to the names of such items, bilateral damage may lead to a complete loss of knowledge of animals, plants, and man-made objects. Individuals with neurodegenerative disease including semantic dementia and Alzheimer's disease frequently experience this loss of knowledge in the later stages—not knowing what a rabbit, pumpkin, or remote control is—which, we would argue, is the loss of a form of awareness of, or consciousness for, the items. In addition, even very early on in the course of their disease, once the concept of an item is completely lost from semantic memory, individuals with these semantic memory deficits are not consciously aware that anything is wrong—they just believe that they may never have encountered the missing items and therefore have no knowledge about them.

Properties of cortex that support consciousness

We hope that we have made our case that each cortical area contributes to specific conscious awareness related to the function of that cortical area, and that destruction of any cortical area will disrupt or abolish the domain of consciousness that was supported by that area, but not other domains of consciousness. Our theory that all cortical regions contribute to consciousness is consistent with a recent neuroimaging study which compared conscious versus unconscious object recognition and found widespread areas of cortical involvement (Levinson, et al., 2021). Our theory also provides one way to reconcile the different predictions made by the four major theories of consciousness regarding which cortical regions are most important (see Yaron et al., 2021, for review), arguing that they all are, but none is critical. Note that we are not implying that subcortical structures are not necessary for consciousness—they most certainly are—just that they are not sufficient for consciousness nor do subcortical structures contain the necessary neuronal architecture that allows for phenomenology or qualia of consciousness to occur.

Although we consider cerebral cortex architecture as the unit of neuronal assemblies that allows consciousness to occur, exactly how consciousness arises from cortex is unclear. We will leave it to others to determine whether consciousness is related to the spindle neurons (also known as von Economo neurons) found in layer V, thick-tufted pyramidal neurons in layer VB, large pyramidal cells in cortical layer II/III, other neuronal types, or none of these. It may be that assemblies of neurons are the 'unit' of consciousness and that looking at single cells to understand consciousness is similar to studying the nature of quarks inside the neutrons of iron atoms to understand why iron filings are attracted to magnets. Thus, it may be that concepts such as the perturbational complexity index will be key to understanding consciousness (Koch et al., 2016).

Supposing that assemblies of neurons in the cortex are the unit of consciousness, how many layers of cortex are required? Is only 6-layered cortex conscious, or is 5-, 4-, or 3-layered cortex sufficient? Can 3-layered allocortex produce some, ‘low-level’ perceptual or emotional consciousness whereas 6-layered neocortex is required for ‘higher-level’ self-consciousness and abstract reasoning? Because our theory is that consciousness developed as part of the conscious memory system, which includes hippocampally-based episodic memory, we speculate that some conscious awareness is present in even in allocortex.

Whatever the correct answer, we would argue that our lack of understanding of exactly how the cortex produces consciousness does not prevent us from using this theory of the involvement of the cortex in consciousness to make some speculations, which we will do now.

Implications

Our memory theory of consciousness with regions of the cerebral cortex as the fundamental units that allow for conscious awareness leads to a number of implications regarding which animals are conscious, which neurologic and psychiatric disorders may impair consciousness, and how the conscious and unconscious aspects of your mind work together day by day, minute by minute, and second by second. These implications, in turn, raise ethical implications, which we will also briefly discuss.

Animal minds

In the prior section we speculated that conscious awareness of basic perceptions and emotions may be present with 3- or 4-layered allocortex, whereas conscious abilities such as problem solving may require 6-layered neocortex.

Conscious mammals

Because all mammals have neocortex (Kaas, 2019), we argue that all mammals are conscious. Paralleling their differences in cortical structure, we believe the consciousness of mice differs in its complexity from that of a dog, which differs from that of a chimpanzee, which differs from that of a human. For example, Pine, Wise, and Murray (2021) review some of the changes in the neuroanatomy between nonhuman and human primates such as the expansion of homotypical association areas and hippocampus and how these expansions are related to "(i) a subjective sense of participating in and re-experiencing remembered events; and (ii) a limitless capacity to imagine details of future events." Although we certainly agree that humans developed these capacities to a greater extent than all other species, we also believe that all mammals will be able to have some conscious awareness of (memory for) perception, decisions, emotions, and actions. See Carruthers (2015) for a similar argument also based neuroanatomical homology.

Does this mean that all mammals have human-like conscious memory abilities to mentally time-travel? Suddendorf and Corballis (2007) argue there is little evidence to suggest that non-human mammals have developed this mental time-travelling ability, at least in the way that humans do, stating, "We maintain that the data so far continue to suggest that mental time travel is unique to humans." Although we do not contest their statement, we would still argue that non-human mammals possess some form of conscious memory that allows them to have some conscious awareness which provides them with evolutionary advantages, such as being able to determine temporal order (Eichenbaum et al., 2005).

What about the mirror self-recognition test? In this test, animals have a spot of color applied to their head when they are anesthetized or otherwise unaware of its application, in a location that they can only see in a mirror, such as their forehead. If the animal looks in the mirror and reaches for the spot or tries to rub the spot off, then we know the animal recognizes itself in the mirror. If not, we presume the animal doesn't realize that it is themselves in the mirror. Although a number of mammals have been shown to pass this test, including the four great apes, bottlenose dolphins, and Asian elephants (de Waal, 2019), most other mammals have not, including our feline and canine companions (although dogs can pass an odor version of the test [Horowitz, 2017]). We believe this test tells us something about visual perception and recognition of self-consciousness, but we would argue that just because an animal "fails" the mirror test doesn't mean that it has no conscious awareness of any type.

Consciousness in non-mammalian species

What about consciousness in other vertebrates such as birds, lizards, amphibians, and fish? Magpies have, in fact, been shown to pass the mirror self-recognition test (Prior et al., 2008) as have a species of fish (Kohda et al., 2019). Does that mean certain bird and fish species are self-conscious? de Waal has argued that passing the mirror test does indicate at least some rudimentary self-awareness (de Waal, 2019), while others suggest that one should consider all the data that shed light on the cognitive capacities of a species before drawing conclusions on its self-awareness (or lack thereof) (Vonk, 2020).

Our theory would be aligned with de Waal, that consciousness and self-awareness among species are on a continuum (de Waal, 2019). Based both on experimental work and their brain anatomy, we believe that there is evidence that most vertebrates have at least some rudimentary conscious memory system, since they have some form of a hippocampus (or a brain structure analogous to it) and some cortex (or analogous structure). It follows logically that those vertebrates that have the requisite anatomy for the conscious memory system would experience at least some conscious awareness, although it might be little more than perceptions and/or emotions. Some have, in fact, argued that at least one non-mammalian vertebrate species, California scrub-jays, can use their memory to spontaneously plan for the future without reference to their current motivation—something that had previously been thought to be a uniquely human ability (Raby et al., 2007; see also Carruthers, 2015).

Our theory is also consistent with many of the ideas put forth by Ginsburg and Jablonka (2021), who go even further than we do. They suggest that "unlimited associative learning" can be considered a marker of minimal consciousness. They note that such learning is present not only in almost all vertebrates, but also in octopods, squid, cuttlefish, honeybees, and cockroaches.

Ethical implications of conscious vertebrates

We are not vegetarians and do not want to imply that the inevitable ethical conclusions that stem from our theory means that we should all become vegetarians (or at least, non-vertebrate consumers) to avoid harming conscious animals. However, we would argue that we as societies and individuals should consider the ethical implications of consuming cows, pigs, chickens, and other vertebrates that we argue have forms of conscious awareness.

Disorders of consciousness

Now that we have a better understanding of both the phenomenology of consciousness and the neuroanatomical structures needed for it, we can speculate that a number of psychiatric, neurologic, and developmental brain disorders may be disorders of consciousness.

Strokes

We reviewed above how strokes that damage certain areas of the cerebral cortex result in specific impairments in consciousness related to the functions of those particular regions. Here we will simply add that strokes in the subcortical white matter of the corona radiata may also cause impairments in consciousness by disconnecting cortical regions from one another, leading to, for example, Wernicke aphasia (Mesulam et al., 2015) or alexia without agraphia (Geschwind, 1965). In brief, strokes that affect cortical and/or subcortical white matter frequently impair one or more domains of consciousness, the conscious memory system, and the ability to use previously learned information to problem solve and plan for the future.

Delirium

Patients with delirium (also known as an acute confusional state or encephalopathy) show an inability to carry out a coherent stream of thought or action (Lipowski, 1989). We suggest that these patients manifest a primary disorder of consciousness. As expected, they are also unable to properly store, retrieve, or use conscious memory for goal-directed activity.

Patients with delirium can (and often do) engage in a variety of activities including running, undressing, eating, and performing other complex activities (Lipowski, 1989). We suggest that such activity is performed without conscious control or awareness, and it is because of this lack of conscious awareness and control that the activities result in no episodic memories being formed and essentially no complex goals achieved. (The individual may undress because they are warm or drink because they are thirsty, but these types of basic goals are the only ones which can be achieved.)

Delirium may result from cortical dysfunction caused by metabolic disturbances (such as elevated calcium) or the inflammatory processes that accompany systemic infections (such as pneumonia) (Wilson et al., 2020). Delirium is more common when the cerebral cortex is already damaged by strokes or degenerative diseases (such as Alzheimer's) (Wilson et al., 2020).

Alzheimer's disease and other cortical dementias

Whether or not they start with involvement of subcortical structures (such as the substantia nigra of the midbrain in Parkinson's disease dementia), we contend that all cortical dementias eventually impair one or more domains of consciousness due to their involvement of cortical regions. Cortical dementias include Alzheimer's disease, Lewy body dementia (including Parkinson's disease dementia and dementia with Lewy bodies), progressive supranuclear palsy, corticobasal degeneration, and the frontotemporal lobar degenerations that may lead to behavioral variant frontotemporal dementia, primary progressive aphasia, or progressive amnestic dysfunction depending upon where in the cortex the pathology starts.

We believe that cortical dementias impair different domains of consciousness depending upon which cortical regions they disrupt, just as strokes do. In addition, the pathology in cortical dementias spreads throughout the brain, often affecting both adjacent cortical regions and regions that are neuronally connected. For example, Alzheimer's pathology spreads throughout the default mode network consisting of hippocampus, related medial temporal lobe structures, anterior and lateral temporal cortex, medial prefrontal cortex, and posterior cingulate, precuneus, and angular gyrus in parietal lobes (Buckner et al., 2008).

As cortical dementias spread and pathology involves more than one cortical region, we believe that more than one domain of consciousness becomes impaired. Once several cortical regions are involved, a phenomenon known as ‘sundowning’ becomes more likely to occur. Most cases of sundowning are a temporary and periodic state of delirium. Over time, most cortical dementias spread to most cortical areas of the brain, which we believe causes impairment in most domains of consciousness. Ultimately, the individual with cortical dementia spends more time in a delirious, confused—and we would argue unconscious—state, with few periods of conscious lucidity.

Migrainous phenomena

Individuals with classic migraines experience auras which are most commonly visual but may also cause numbness, speech or language difficulties, weakness, or confusion. Although migraine auras are not completely understood, it is clear that they are associated with cortical spreading depression (also known as spreading depolarization), in which there is a spreading of electrophysiological hyperactivity over the cortex at a velocity of several mm per min followed by spreading inhibition. Phenomenologically, the most common experience is that of positive visual phenomena (often described as scintillating zig-zag lines or fortifications) followed by a lack of vision or scotoma. Both the scintillations and the scotoma progress over minutes, and generally resolve in under an hour. Relevant here is the fact that cortical spreading depression causes disturbances of conscious perception and sometimes other conscious abilities.

Epilepsy

Similar to migraine auras, individuals with epilepsy may consciously experience an "aura" of positive phenomena, such as flashes of light or tingling and numbness which normally spreads over seconds. Following the positive phenomenon, a postictal loss of function may occur (sometimes called a Todd's paresis when related to motor function in a limb.) In addition to visual and tactile, auras may also be olfactory, gustatory, or cause simple or complex motor movements (such as lip smacking or bicycling movements). Seizure auras have been shown to be focal seizures which occur in some individuals prior to the seizure becoming general. Once the seizures become generalized, the individual becomes unconscious. Again, relevant here is that the temporary disruption of cortical activity may cause focal or general disturbances of consciousness, depending up on the extent of cortical disruption.

Dissociative disorders

Individuals with dissociative identity disorder have two or more distinct and relatively enduring personality states (American Psychiatry Association, 2013). This disorder generally develops in childhood due to prolonged and severe physical, sexual, and/or emotional abuse. We speculate that individuals with this disorder may have two or more separate consciousnesses, each with its own memories, which the other consciousnesses may or may not have access to. If we are correct, it may be an example in which psychological factors can cause a disorder of consciousness.

There is evidence from the literature, such as cases reported by James (1890), that not only can episodic memories be separated between identities but semantic memories as well. In one case he described, when an individual moved from one identity to another, her prior knowledge base was not accessible and she had to learn (or relearn, depending on your view) information which her other self knew.

Another related disorder is dissociative amnesia, in which an experienced or witnessed traumatic event—so emotionally laden that it would typically not be forgotten—cannot be remembered. We believe that this disorder may be considered a more focal case of the conscious memory system being fractured such that one or more explicit memories cannot be retrieved.

Depersonalization-derealization disorder includes symptoms of feelings of detachment from one's mental or bodily processes or one's surroundings. It is thought to be caused by early childhood emotional abuse and neglect. Triggers of depersonalization symptoms may include significant stress, panic attacks, and drug use. We speculate that, at least in some cases, this disorder may be a manifestation of a disruption of consciousness.

Schizophrenia

Schizophrenia consists of positive symptoms of psychosis, such as hallucinations, delusions, and disorganized thinking, and negative symptoms, such as social withdrawal, decreased emotional expression, and apathy. Although all modern theories of schizophrenia believe that it is a disorder of the brain resulting from interactions between individuals' genetic makeup and their environment, the specific biological pathophysiology has not been determined. We believe that it may represent a disorder of consciousness, with the hallucinations (usually of voices) being a symptom of disrupted and possibly fragmented consciousness (Tordjman et al., 2019). This idea is supported by the observation that individuals with schizophrenia show impaired conscious access but intact unconscious processing of perceptual stimuli (Berkovitch et al., 2017). Consistent with our memory theory of consciousness, memory deficits are common in individuals with this disorder (Avery et al., 2020).

Autism

The autism spectrum includes individuals of highly varied capacities, from those who are nonverbal and function at a cognitive level less than age two, to those who would have previously been diagnosed with Asperger syndrome and may show superior intellectual abilities. For the purposes of this paper, our brief discussion of autism will be restricted to those with high needs, focusing on those nonverbal individuals with significant impairment and intellectual abilities less than a typical two-year-old child.

Autism may be caused by several underlying brain differences, leading to several different phenotypes. In individuals with severe classic Kanner autism, there appear to be what we would consider severe deficits in the conscious memory system. Regarding memory, these individuals tend to learn in stereotyped ways, not in the typical episodic-memory-leading-to-semantic-memory manner. Education of these individuals therefore often involves operant conditioning (e.g., applied behavior analysis) to produce learning. Problem solving and responses to conflict and discomfort appear to us to predominantly use System 1 unconscious processes. There is little evidence (at least to the observer) for conscious, thoughtful, deliberative System 2 processing in some of these individuals. The unfortunate result for many is that they learn, via operant conditioning, maladaptive behaviors such as aggression and self-injury because those behaviors have previously provided them with escape from activities they did not like (Oliver & Richards, 2015).

We therefore speculate that individuals with severe classic Kanner autism have a disorder of the conscious memory system manifesting as impairments in both memory and consciousness. This speculation is supported by pathology studies showing that cerebral cortical architecture is disrupted in these individuals (Courchesne et al., 2011). Others have also speculated that autism may be related to deficits in consciousness (Tordjman et al., 2019). This speculation should not be misconstrued as suggesting that autistic people are somehow permanently unconscious or incapable of human experience.

There is also interesting anecdotal evidence that some of these individuals appear to be able to process information in parallel fashion more easily than typical individuals; for example, some desire to listen to music and watch an unrelated movie at the same time (AEB observation). This comfort with parallel processing may be related to a lack of serial one-thing-at-a-time thinking imposed by consciousness.

Lastly, we will note that just as no two neurotypical individuals are the same in their capacities, neither are two individuals with severe classic Kanner autism, such that this discussion may seem consistent with the capacities and limitations of some individuals and inconsistent with the capacities and limitations of others.

Most of our decisions and actions are unconscious, although we think they are otherwise

One of the most radical implications of our memory theory of consciousness is that although you generally have the subjective experience that most of your decisions and actions are consciously controlled and few occur automatically and unconsciously, we argue it is the other way around—few of your decisions and actions are consciously controlled and most occur automatically and unconsciously.

Unconscious routines

Let's take your morning routine, for example. We believe that you are acting primarily unconsciously as you get out of bed, walk to the bathroom, flip on the light, use the toilet, brush your teeth (including all the steps such as wet your toothbrush, put on toothpaste, move the brush across your teeth, rinse your toothbrush and your mouth, put the brush back), shower (including all the separate steps), dry yourself, dress, fix your hair, etc. It would only be if something were unusual—such as you needed to dress for a job interview or a snowstorm—that there would be much conscious deliberation. In fact, not only did you not make conscious decisions to guide your actions, but you likely were also only partially aware of your morning routine. There may have been moments when you attended to what you were doing, allowing you to form an episodic memory of those moments. At other times, however, your mind was free to use your conscious memory system to review what happened yesterday, reminisce about a prior event, or imagine what will likely happen in your 9 AM meeting. If these latter thoughts were occurring during the time that you were showering, instead of remembering the steps of your shower you will remember the topics you were consciously reviewing, reminiscing, and imagining. But even if you were perceptually aware of each step of your shower, that does not mean that you were consciously controlling those steps; your awareness simply allowed you to remember each step.

Conscious or unconscious decision?

You probably don't object to our claiming that you typically go through your morning routinely unconsciously, or that you would only remember whatever you were paying attention to at the time. But now you arrive at work, drink some coffee, and walk to your meeting. You're a little nervous because you need to explain to your boss why your team didn't meet its quarterly target, and you're going to have to explain that it's because your colleague Biff, also in the conference room, didn't do his part. To your surprise, as you launch into your explanation, you hear yourself taking responsibility for the team's shortfall, and not blaming Biff. As you're leaving the meeting, you consider what just happened. You end up consciously thinking that you said what you said because it probably wasn't really Biff's fault, mistakes can happen to anyone and, after all, you were leading this team, so you were responsible. You nod your head as you think, "Yes, that's what happened," and turn to other work.

However, we speculate that what actually occurred is that as you started to speak, you were also unconsciously processing in parallel the facial expressions, body language, and eye movements of those in the room—particularly of Biff and your boss. Your unconscious perceptions of his slightly red and determined face led you to unconsciously realize that Biff wasn't going to quietly accept that it was his fault; he would likely deny responsibility and blame you instead. Your boss's body language—mimicking Biff's posture and leaning toward him slightly, signifying that your boss was going to side with Biff—was also unconsciously processed. Your unconscious System 1 quickly made the decision to change tactics and say it was your fault—regardless of the truth or what your conscious System 2 had planned to say. (Remember that we believe unconscious brain processes make the final decisions—the rider can indicate to the horse where they wish to go but, in the end, the horse decides.)

This scenario raises the questions of (1) why you didn't realize what actually happened in the conference room and (2) why you came up with an alternative not-quite-right explanation. We argue that the answer to the first question is that you didn't realize what happened because you were not consciously aware of it at the time, and thus you could not consciously consider it then nor could you use your episodic memory (part of the conscious memory system) to remember it. (People can, of course, train themselves to notice subtle facial expressions, body language, and eye movements, which they could then be consciously aware of and remember.) The second question can best be answered by considering an experiment performed by Gazzaniga and his colleagues with one of his split-brain patients.

The conscious interpreter

A patient who had undergone a callosotomy to separate his left and right hemispheres to reduce the frequency of disabling seizures was shown separate images in his left and right visual fields. His task was to use each hand to point to a card that corresponded to the visual image he saw in each visual field. His right visual field, projecting to the verbal left hemisphere, was shown a chicken foot, and his right hand (controlled by the left hemisphere) pointed to a picture of a chicken. His left visual field, projecting to the nonverbal right hemisphere, was shown a snowy scene, and his left hand (controlled by the right hemisphere) pointed to a picture of a shovel. So far so good. But when he was asked why he was pointing to the shovel his response was, “You’re going to need a shovel to clean out the chicken coop.” From experiments like this one, Gazzaniga and colleagues came up with the explanation that our left hemisphere acts as an interpreter, explaining to others and our conscious selves why we do things (Gazzaniga, 2015).

It is not only patients with split brains, however, who may not be aware of why they act in a certain way. In their classic paper, “Telling more than we can know,” Nisbett and Wilson (1977) review evidence that any of us may be unaware of stimuli that influence our responses.

These ideas have been more recently articulated by Peter Carruthers (2011), who argues that we don't have any more access to most of our own thoughts than we do to anyone else's, and that we must use the same tools that we would use to discern how others think to discern our own thoughts.

These experiments by Gazzaniga (2015) and Nisbett and Wilson (1977) make it clear that what we consciously explain to others—and ourselves—may not always be the true explanation. Similarly, confabulation by individuals with memory impairment, such as those with Korsakoff's syndrome, may simply represent the conscious interpreter doing its best to make sense of the world with partial and outdated information.

To review, we believe that unconscious brain processes are ultimately responsible for your decisions and actions. If you consider your decisions and actions, most of the time you will have a ready explanation for them—and most of the time your explanation will be correct. But sometimes, as these examples demonstrate, the explanation you tell yourself will be incorrect. This phenomenon is likely one reason why people use the psychological defense mechanism of rationalization—in this case, because they are not consciously aware of the truth, only the narrative that their conscious mind comes up with.

Will power and difficulty controlling behavior

Perhaps some of the most obvious examples that you do not have direct conscious control over your decisions and actions relates to will power and the ability to control behavior. Have you ever tried to control your behavior and found it difficult to do so? Perhaps you promised yourself you're going to have just one spoonful of chocolate ice cream but, the next thing you know, the entire container is empty. While you were eating it, did you say to yourself, "I shouldn't be doing this. I should put the spoon down and put the container back in the freezer"? Did it sometimes appear that you were simply watching yourself eat the ice cream and you had little control over your actions?

We would argue that you now understand why it sometimes appears that you have little or no control over your own actions—because it is actually your unconscious brain processes that are in control. You now also understand why it appears that you are merely watching yourself perform actions that you don't consciously wish to do—because your conscious self is not participating directly, it watches memories of your actions (via your special video screen sunglasses as you are riding on your unconscious horse or via the Cartesian theater, depending on which metaphor you prefer). In fact, our memory theory of consciousness suggests that the sense *we are watching ourselves act* is a more accurate depiction of reality than the sense that *we are consciously engaged in deciding and action*.

How the conscious mind can influence decisions and actions

Note that just because we argue it is unconscious brain processes that ultimately decide and act, does not mean we believe that your conscious mind cannot influence those decisions and actions. Because conscious and unconscious processes are both occurring in the brain, it should not be surprising or mysterious that they interact, even if we don't understand the brain physiology underlying those processes and interactions. As Kahneman and Tversky described so eloquently in their work, we frequently have fast, unconscious System 1 "gut instinct" choices competing with slow, conscious System 2 logical choices. Luckily both for the individual and the community, the conscious System 2 logical choices often win out for important choices and the best decision is made. We would simply argue that, when that happens, the conscious System 2 'convinces' the unconscious System 1 to make the rational choice. The conscious memory system then remembers that the desired decision was made and thinks that it actually made the decision itself—even though it just remembered the decision that was actually made unconsciously. (The rider succeeds in cajoling the horse to go up the steeper path to avoid the dangers on what appears to be the easier route. The rider thinks they made the decision, but where they go is always up to the horse.)

Free will

We would like to make several points about free will. First, just because your decisions and actions are ultimately made unconsciously, doesn't mean that you don't have free will—or, at least, not any more than if you made your actions consciously. The implications regarding determinism are no different whether decisions are initiated consciously or unconsciously. Second, as we just discussed, your conscious mind can cajole your unconscious self into making the best decisions in particular instances and can change the tendencies of our unconscious self over time. Third, if major life decisions are made slowly, over minutes, hours, days, or longer, these important decisions will almost certainly have input from both your conscious mind as well as your unconscious brain processes. Lastly, our memory theory of consciousness makes no predictions about whether your decisions and actions are determined on a microscopic scale, so on that point we will not comment other than to say that we believe you must act as if you had free will.

Who are you?

One important issue that we have been skirting around is who, exactly, are you? Are you your conscious mind, your unconscious brain processes, both, or neither? The obvious answer is the correct one; you are both, of course. However, we would argue that the elements of your personality, your character—your innate tendencies of how you interact with the world—are determined by your unconscious brain processes, which we can call your unconscious self. Your conscious mind, on the other hand, is the part of you that does the conscious thinking, remembering, feeling, and calculating, and carries out other conscious processes. Note that your unconscious self certainly has different parts or aspects (Cisek, 2019). One part may crave chocolate ice cream while another part may wish to look svelte in the mirror.

Understandably, your conscious mind believes that you are primarily your conscious self, as your conscious mind spends most or all of each day alone with itself. From the perspective of everyone else interacting with you, however, you are primarily your unconscious self. Don't forget, we believe that most of your decisions and actions occur automatically without conscious input, and only few have an element of conscious control. This means that most of the time the world is interacting with your unconscious self. Thus, we experience self and other as a mix of conscious and unconscious perception and action. We almost certainly overestimate how much of the mix is conscious.

You can change who you are

Just because a large part of who you are is your unconscious self, does not mean that you cannot change and improve yourself over time. As educators and clinicians, we most certainly believe that people can grow, learn, and change for the better. For example, perhaps you have noticed a tendency toward selfishness in yourself. In contemplating this tendency along the lines we have discussed in this article, you realize that it stems from one aspect of your unconscious self. Your conscious self, as well as other aspects of your unconscious self, may feel embarrassed about this tendency. To create a change, the parts of your unconscious self that desire a change can use conscious System 2 processes to create a conscious plan that may include both large actions in the world (such as sharing your time or money with charitable organizations) and small, everyday acts of selflessness (such as being considerate to those around you and helping strangers in need). If these actions are rewarded by positive feelings, your nonconscious memory systems (such as operant conditioning) will reinforce the actions such that they become part of your character—that is, your unconscious self.

This is relevant to moral responsibility. If you can change your character and you find that your character tends toward bad behavior, then you are responsible for making yourself into a better person. For instance, a parent may have an obligation to develop the (unconscious) tendency to engage in caring behaviors toward their children (cf. Björnsson and Brülde 2017).

The memory theory of consciousness points us toward a mixed type of ethics that includes a healthy dose of virtue ethics. Most of our actions arise from our unconscious self such that the proper object of moral evaluation is character, in line with the virtue ethics familiar from Aristotle. Other actions will be guided by conscious deliberation. Each of these deliberate actions will be subject to moral evaluation individually, consistent with deontological ("rule-based" or "duty-bound") or consequentialist ("outcome-based") ethics.

You can control your behaviors and actions

In the same way that you can change who you are by consciously deciding to change your unconscious self, we firmly believe that you can control your actions. Perhaps this is obvious now that we have explained that "you" are both your conscious and unconscious selves. But even if we are just considering your conscious self, we believe that your conscious self can cajole and convince your unconscious self to make the decisions and take the actions you desire.

In fact, we hope that the viewpoint regarding consciousness expounded upon in this paper will help everyone (including educators and clinicians) understand that to change your actions, you need to change both your conscious and unconscious self. Furthermore, to change your unconscious self may require the use of nonconscious training such as described above, including operant conditioning, skill learning, and the like. For example, to train your unconscious self to put your keys in the same place each day so you won't spend hours hunting around the house for them, you initially need to set up conscious reminders until procedural memory takes over and your unconscious is trained (i.e., until it becomes a habit).

Returning our one-spoonful of chocolate ice cream example, for most people it isn't enough to simply consciously desire to stop after eating one spoonful using will power. Instead, you will need to use your conscious mind to set up systems to either reduce the temptation of eating the entire container (perhaps dividing the container into one-spoonful amounts each on a plastic spoon, with one in front and the others in the back of the freezer) or reward your unconscious self for stopping after one spoonful with something greater than the reward the chocolate ice cream itself would produce.

This also gives us a way to describe heroic actions. Some actions seem heroic to onlookers but come easily to the person because of training or experience. For instance, running into a burning building to save a child is "just part of the job" for a firefighter. Other actions can be called heroic because they require overriding our System 1, unconscious character. We know that this type of action is possible. Indeed, the underdog characters in popular movies are often portrayed as overcoming certain unconscious or characterological tendencies (such as trepidation or cowardice) in just this way. Neville Longbottom from the Harry Potter books is an example.

Consistent with the principles of cognitive behavioral therapy

Everything we are suggesting in the ways in which your conscious mind can influence and train your unconscious self is consistent with the principles of cognitive behavioral therapy (CBT). CBT explicitly recognizes that psychological problems are commonly based on both "unhelpful ways of [conscious] thinking" and "[unconscious] learned patterns of behavior" (APA, 2017, text in brackets added). Moreover, CBT works both to change conscious thinking patterns and to use role play to change patterns of behavior. We believe that CBT is a successful therapy because it directly addresses both the conscious mind and unconscious self, using techniques that each can understand.

Understanding mindfulness therapy

We can also understand mindfulness therapy in this context. By focusing attention on process that usually occur without our noticing, mindfulness can bring what is usually unconscious to the attention of the conscious self. That allows System 2 deliberate processing, deliberate evaluation, and deliberate changes in how we respond. To achieve the individual's goals for the therapy, the new responses can remain in the domain of the conscious self or be adopted by the unconscious System 1 self.

Ethical decisions and actions

It follows from our discussion that an ethical decision or action may stem entirely from our unconscious self, or from both our conscious and unconscious self. If you are in a horrible situation, you may jump in front of your child to save them from a bullet or other attack. Such actions would typically be automatic and too fast for conscious thought and deliberation. You likely make many small ethical decisions each week using your fast System 1 unconscious processing. At other times you engage in slow, deliberate, conscious System 2 thinking to determine the appropriate ethical course of action; perhaps stepping down from a committee in protest if you strongly disagree with their decision, despite the loss of prestige and revenue it would entail. Another example would be consciously choosing a less remunerative profession based on the amount of good it could produce in the world when certain aspects of your unconscious self would have chosen a more remunerative profession based on the amount of money or other comforts it could provide. Such consciously informed decisions would, of course, require your conscious mind to cajole and convince your unconscious self to make the actual decision.

Teaching ethical decisions and actions

Our hypothesis that all decisions and actions are ultimately made by the unconscious self has implications for the pedagogical methods that we use to teach ethics to our children and others in society. Teaching the theory behind ethics and ethical decisions and actions is helpful but not sufficient, as such theories will only inform their conscious minds and not their unconscious selves. Just as learning a new procedural skill first requires conscious memorization of a sequence and then practice of the actual motor/procedural skill so that the learning occurs in the unconscious self, so should ethical education involve not only conscious thinking and memorization, but also the visceral, emotional, empathetic learning that is acquired by doing. For example, children can participate in role playing various actions and situations to feel the emotional consequences that different actions engender.

All of this speaks to the importance of ethical approaches involving fictional and autobiographical narrative insofar as these approaches support integration of conscious rules with unconscious responses. It also suggests that we may want to examine the ethics curricula in professional training (such as in medical, nursing, education, and law schools) so that we avoid a one-and-done, ghettoized ethics course in favor of an ethics-across-the-curriculum model.

Understanding hurtful actions

If, as we argue, decisions and actions are made unconsciously, then we must work together as a society to use this understanding to reduce the occurrence of hurtful acts through pedagogical methods as described, such as role playing. We should not excuse those who commit hurtful acts, but we should also have greater sympathy for them as we understand that they may be acting on unconscious impulses. Nonetheless, to live together in society individuals need to learn to control their unconscious selves and thus their behavior through cognitive behavioral therapy and other ethical methods. (Although the horse determines where they will go, the rider needs to learn to tame and control their horse.)

Developing conscious control of the unconscious mind: The importance of mindfulness

Mindfulness has recently been touted as the key to everything in our society from avoiding burnout to achieving enlightenment. Initially we raised the issue of why mindfulness is hard and then we provided our explanation that mindfulness is hard because our conscious minds did not develop to be in control of our thoughts, rather, the unconscious parallel-processing brain developed our conscious minds simply as a necessary part of our conscious memory system. That is, we argue your conscious mind developed as a tool to be used by your unconscious brain to help it predict the future and make decisions accordingly. It is for this reason that we believe it is difficult for you to consciously control your conscious thoughts—because normally your unconscious brain is in control of your conscious thoughts!

If, however, you wish to achieve more control over your decisions and actions, it is helpful to have more control over your conscious thoughts. Once you have control over your conscious thoughts, you can use your thoughts to cajole and convince your unconscious self to make the decisions you consciously desire, whether that is eating only one spoonful of chocolate ice cream, leaving work problems at work when you come home to your family, or stopping the flow of thoughts that run through your mind when you are trying to sleep. Because mindfulness is one method of improving your ability to consciously control your thoughts, we believe that practicing mindfulness can help you control your decisions, behaviors, and actions as well.

Unlock the power of your unconscious self

Have you ever had the experience that someone asks you a question and, without thinking, you start to answer, only to be surprised at the words that leave your mouth? Have you stepped back and reflected, “Oh, I guess that is what I think about that…” We believe this situation is an example of how we may learn what our unconscious self thinks about an issue, and it is just one way that you can use your unconscious processes to your advantage.

Put your unconscious to work

You've probably had the experience of trying to come up with a name or an answer to a question and, after the conversation has finished, the name or answer then pops into your consciousness. You can actually task your unconscious to work on a problem, and your unconscious mind may (or may not) help you solve it. Simply say to yourself, "OK, I need to recall that piece of information," and then turn to other matters. You may find the information coming to you spontaneously later that day, often within minutes. Another method (and one that we use) is to review some of the work that you need to do the following day just prior to sleep. For example, prior to sleep, we will generally review the last section of writing that we have done as well as our notes or plan of what we will be writing the following day. Based on the work of Sanders, Beeman, Paller, and others (Paller et al., 2021; Sanders et al., 2019), we assume that this little bedtime preparation facilitates our unconscious brain processes while we are sleeping. In any event, when we sit down at our computer the next morning, the writing proceeds easily.

Bring unconscious processes to conscious awareness

Have you ever been in a meeting and thought to yourself, that person is scared, angry, sad, excited, or happy? We are all fairly good at detecting basic emotions in ourselves and others using either unconscious or conscious processes (Calvi et al, 2020; Ochsner et al., 2004; Smith & Lane, 2015). What about more complex states such as bluffing, lying, manipulating, seducing, protecting, defending, and supporting? These can be a bit more difficult to discern consciously, but you probably are aware of many of them some of the time at an unconscious level. You can train yourself to bring these states that other people are exhibiting to your consciousness by consciously analyzing their tone of voice, body language, eye contact, etc. But you can also tap into your unconscious perception by noticing how the person makes you feel. Do you feel sorry for them? Do you like or dislike them? Are you afraid of them? Do they make you angry? Do they make you smile—and if so, why? By analyzing your reactions to people and events you can bring at least some of your unconscious perceptions to consciousness.

Stereotype threat and other unconscious biases

One of the most important reasons to bring unconscious thoughts and emotions to conscious awareness is to reduce the effects of stereotype threat and other unconscious biases that we are all susceptible to. Stereotypes and biases have been shown to impact the careers of women and minorities (Grewal et al., 2013), in addition to having a detrimental effect on many other aspects of society. Education and training have been shown to increase awareness of these issues and provide concrete steps that individuals and institutions can follow to reduce the harmful effects of stereotypes and unconscious biases (Grewal et al., 2013; Rodriguez et al., 2021).

Conversations

Have you had the experience that it is sometimes hard to get a word in edgewise in a conversation—like you're half a second behind—and other times it is easy and effortless? We believe that, at least in part, it depends on whether you are responding consciously, with the built-in perceptual delay of about half a second, or unconsciously, which has no delay. If we are correct, the more consciously determined you are to make that point in the conversation, the more likely you are to have trouble doing it, because you will be experiencing the perceptual delay. You'll be more able to break into the conversation if you relax and let your reactions and responses be more spontaneous.

Unconscious talking?

Can your language be controlled by unconscious brain process? Or does speaking and writing always involve not only conscious awareness but conscious participation in such activities? We hinted at aspects of these questions previously, discussing how we can sometimes be surprised by the words that come out of our mouths. This example alone is perhaps sufficient to make the claim that unconscious brain processes can control our language—how else could our conscious mind be surprised at what we have just said?

Another example relates to when we are giving a lecture that we have given many times before. Here we have noticed that two related and interesting phenomena sometimes occur. The first is that we often enter a mode where we, as our conscious selves, can watch and listen to ourselves talk on what appears to be an automatic pilot—we at least do not feel as if we need to use any conscious effort to make the words come out of our mouths. It is as if not only are we in the Cartesian theater watching the world go by, we are also able to see ourselves on the stage! The second is that, during such a time when we can step back and watch ourselves lecture, we sometimes get distracted, start thinking of something completely different, and don't pay attention to ourselves lecturing. When our conscious awareness returns to our lecture, we find that we cannot remember what point we have just made, and whether we should be making the point that is on the slide on the screen or have already done that and should move on to the next slide.

Thus, speaking without conscious control or even awareness is not only possible but probably common. As with other perceptions, decisions, and actions, if we are not consciously aware of our speaking, we will not be able to remember what we have just said.

In the zone: Playing sports and musical instruments

As discussed earlier, reaction times are often too fast in sports and music to be consciously controlled. Thus, many athletes and musicians perform using their unconscious System 1 brain processes. In fact, many musicians and athletes perform much *better* when they are only using their unconscious processes to control their behavior. This is often referred to as being "in the zone," which the Oxford English Dictionary defines as, "A state of perfect concentration leading to optimum mental or physical performance." We suggest that being "in the zone" means you are allowing your unconscious brain processes full rein to make the right, often precise or creative choices much faster and more accurately than you would have been able to do with conscious control. This level of unconscious expertise is, of course, only possible when these unconscious brain processes have been trained through thousands of hours of practice. But once that training has taken place, you can let your unconscious brain processes loose! Conscious control is as likely to interfere with performance as it is to help.

Note, however, we are not saying that conscious processes have no role in sports or music. Some of the best athletes, such as the tennis champions Serena Williams and Roger Federer, are known to consciously modify their strategy when their current approach is not producing success. Similarly, some of the most creative music, such as that from Miles Davis and Lady Gaga, appears to result from the interaction of conscious and unconscious brain processes working together.

Art and insight

Many—perhaps most—creative endeavors from painters to novelists likely result from the combination of conscious and unconscious brain processes working together. Additionally, appreciation of beauty and other aesthetics in any sensory modality (vision, hearing, smell, taste, touch) may be primarily driven by unconscious properties, which may be why aesthetic appreciation is so difficult to describe using words. We suspect that creative insights in other fields, whether psychology, biology, chemistry, physics, mathematics, etc., often arise from unconscious processes in a brain that has been prepared by consciously struggling with the same or related problems for some time. As Louis Pasteur famously said, "*Dans les champs de l'observation le hasard ne favorise que les esprits prepares,*" which is often colloquially translated as, "Chance favors the prepared mind." Such scientific insights that likely arose through the combination of conscious and unconscious processes include those of Alice Ball, Rachel Carson, Marie Curie, Marie Daly, Jennifer Doudna, Albert Einstein, Jane Goodall, August Kekulé, Ada Lovelace, Isaac Newton, and many others. Some of the insights of these people might be because they are somehow able to see the world more the way it actually is, rather than the way the conscious mind suggests that it is. In other words, some of these individuals may have had more access to their unconscious brain processes.

Hume, for example, asserted that the self is "nothing but a bundle or collection of different perceptions" (Hume, 1978), and that we can have no idea of any subject to which these perceptions belong. It's simply that we "bundle" these experiences together that they appear continuous. It is possible that some of Hume's insights may be related to his glimpsing some of his unconscious brain processes.

Future directions

Here we outline some possible methods that can be used to test different aspects of our theory of consciousness. We will first discuss methods that may be able to prove or disprove the theory that consciousness developed as a memory system, and then methods that examine the idea that each region of the cerebral cortex is sufficient to make one aspect of consciousness possible.

Methods to evaluate consciousness as a memory system

We have argued that consciousness developed as part of a memory system and that all explicit forms of memory (sensory, working, episodic, and semantic) are part of that unified system. One line of investigation would therefore be to examine whether consciousness is truly bound up and an integral part of these forms of memory. Are there situations in which explicit memory can exist in the absence of conscious awareness? Conversely, can there be any type of conscious awareness without the involvement of one of these forms of explicit memory? Are there correlations (or not) between explicit memory performance and conscious perception abilities? Are there correlations between subjective measures of explicit memory and subjective measures of consciousness? Such experiments would provide important insights into this issue.

Experimental paradigms

Experiments similar to those performed by Schneider and colleagues (Schneider et al., 2021), which attempt to show engagement of the episodic memory system with and without conscious awareness by using different types of visual masking, may be one fruitful line of research. Other paradigms that contrast conscious versus unconscious processing of stimuli may also be useful (for reviews, see Breitmeyer, 2015 and Timmermans & Cleeremans, 2015).

The key consideration in all of these paradigms would be the relationship between conscious awareness of, and explicit memory for, the stimulus. These include paradigms that use change blindness (e.g., Simons, 2010), attentional blink (e.g., Sy et al., 2021), visual crowding, continuous flash suppression, perceptual fading, motion induced blindness, reversible figures, binocular rivalry, perceptual filling in (e.g., Davidson et al., 2020), and visual perception of degraded objects (e.g., Levinson et al., 2021). False memory paradigms may also be useful (Nichols & Loftus, 2019; Schacter & Dodson, 2001); these experiments often produce a vivid, confident, conscious recollection of an item or experience which never occurred (Sikora-Wachowicz et al., 2019). Visual imagery experiments, in which a conscious image is produced in the absence of a percept, may also be useful (e.g., Moulton & Kosslyn, 2009). Lastly, experimental paradigms involving sleep and memory may be particularly useful (e.g., Paller et al., 2021), as dreams appear to be one area in which consciousness and memory may not be co-occurring.

Subjective measures

Subjective measures have been used to fractionate memory into different components, such as the remember-know (Tulving, 1999) and recollection-familiarity distinction (Yonelinas, 1994). Subjective measures ask participants to use introspection (as in the case of remember-know) or to provide confidence judgments. Confidence judgements have been successfully used to create receiver operating characteristic (ROC) curves to estimate the degree to which subjects use recollection and familiarity during a task (Yonelinas, 1994). Process-dissociation procedures can also be used to separate components of memory (e.g., automatic versus intentional use, Jacoby, 1991).

Subjective measures have also been used to evaluate conscious awareness (for review see Timmermans & Cleeremans, 2015). These include the perceptual awareness scale (Ramsøy & Overgaard, 2004) in which subjects rate stimuli as glimpsed, almost clear, or clear; the continuous visual analog scale (Sergent & Dehaene, 2004) in which participants rate stimuli on a continuum from not seen to maximally visible; rule awareness scale (Wierzchoń et al. 2012) in which subjects rate how aware they are of a rule; post-decision wagering (Persaud et al., 2007) in subjects wager money on their response; confidence ratings (Dienes et al., 1995) in which subjects rate their conscious awareness; and feeling of warmth (Metcalfe, 1986) which is similar to confidence ratings but lends itself to a more intuitive response.

Methods to evaluate order and timing issues around consciousness

Regarding issues of the order and timing of consciousness, investigations of chronostasis (e.g., Melcher et al., 2020) and postdictive effects (for reviews, see Herzog et al., 2020; Michel & Doerig, 2020; and Sergent, 2018) would likely be informative. Such paradigms include the cutaneous rabbit illusion (Geldard & Sherrick, 1972), illusory and invisible audiovisual rabbits (Stiles et al., 2018), color fusion effects (Pilz et al., 2013), color phi illusion (Keuninckx & Cleeremans, 2021), sequential metacontrast paradigm (Otto et al., 2006), rapid serial visual presentation (Akyürek et al., 2016), and cues presented after a faint Gabor grating (Sergent et al., 2013). Key in these paradigms would be the relationship between explicit memory and the conscious postdictive experience.

Disrupting consciousness and memory

Another line of research that may be fruitful would be to disrupt either explicit memory or conscious awareness and evaluate whether the other is also disrupted. For example, will medications known to impair explicit memory (e.g., anticholinergics, benzodiazepines) impair consciousness as well? Will medications known to impair consciousness (e.g., sedatives, anesthetics) impair memory as well? Will a TMS pulse that disrupts explicit memory also disrupt at least one aspect of consciousness and vice-versa?

Neural signals of consciousness and memory

Although there is not currently a consensus as to which physiological markers correlate with conscious experience, one can still evaluate whether such markers correlate with both one domain of conscious awareness and explicit memory (rather than correlating with one but not the other). Putative markers of consciousness—such as electroencephalographic (EEG) activity and its derived components, event-related potentials (ERPs)—are discussed briefly in the next section.

Methods to evaluate the theory that the cerebral cortex is where consciousness occurs

We have argued that not only do all areas of cerebral cortex contribute to consciousness, but that each region of cortex may be autonomously conscious of its own particular aspect of consciousness. This theory is contrary to those of a number of other researchers, some of whom suggest that the frontal lobes are the minimally sufficient regions necessary for consciousness (Brown et al., 2019; Murray et al., 2020) and some of whom suggest the parietal and occipital cortex are the regions minimally sufficient for consciousness (Lamme, 2015; Lee et al., 2019; Tononi et al., 2016).

Methods to evaluate the anatomical aspect of our theory of consciousness and that of other theories can include: studies of patients with brain lesions or dementia, simulated lesions using TMS, brain imaging studies that provide anatomical localization using functional MRI (fMRI) or positron emission tomography (PET), brain imaging studies which provide information about the timing and physiology of brain processes using EEG and ERPs, derived EEG measures such as the bispectral index (BIS) used by anesthesiologists, and other derived measures such as the perturbation complexity index (PCI).

Different domains of consciousness may have different neural correlates

Because we believe that different regions of cortex can be autonomously conscious and that the consciousness awareness (the conscious phenomena or qualia) provided by each region will be different, there is no reason that the physiological markers of these different brain regions need to be the same. Therefore, from our perspective, the task is to look for physiologic markers of conscious versus unconscious processes for each cognitive domain (e.g., conscious visual awareness versus unconscious visual perception), rather than conscious versus unconscious processes of the brain as a whole. Even within a cognitive domain, different experimental paradigms may lead to different conclusions regarding the neural correlates of consciousness due to the specifics of the paradigm and how consciousness is measured and reported (e.g., Cohen et al., 2020).

If this idea is correct, one fruitful project would be to try to separate different domains of conscious awareness (such as visual, auditory, olfactory, gustatory, tactile, etc.) and perhaps also differences within a domain (such as visual perceptual fading, visual perceptual filling in, and visual perception of degraded objects, etc.) and then try to determine the neural correlate of consciousness for each. Some would argue that amodal thoughts (abstract thoughts completely separate from sensation) can also be conscious and should be added to this list whereas others would argue that is not possible (see Carruthers, 2015).

Fixed and temporary brain lesions

One straightforward way to evaluate whether our cortical hypothesis of consciousness is correct is to study patients with brain lesions that have resulted from strokes or surgical resections. The advantages of studying such patients include that their lesions are generally constant over time and, using structural MRI combined with measures of white matter integrity (such as diffusion tensor imaging), the anatomical extent of the lesions and their connections can be readily ascertained. Studies of patients who happen to have symmetric bilateral lesions may be particularly informative. Such patients can then be evaluated for different domains of consciousness via structured interviews, questionnaires, and experimental paradigms as described in earlier in this section.

A similar approach can use TMS to produce transient brain lesions in various cortical areas. Advantages include that lesions can be produced on demand and, when combined with structural MRI for localization, in the precise area you wish. Lesions can also be bilateral. Disadvantages include that the lesions are transient and that deep brain structures, such as thalamus, hippocampus, and insular cortex, are difficult to stimulate.

Another promising approach is to combine these methods, by evaluating a patient with a known, unilateral, fixed brain lesion and then to use TMS to temporarily inactivate the reciprocal cortex on the opposite hemisphere. In all these situations, one of the key questions is whether there are any bilateral cortical lesions that produce complete unconsciousness. Our theory would predict that bilateral lesions would impair specific domains of consciousness, but not other domains of consciousness. The hypotheses of several others predict that all domains of consciousness will be impaired when certain bilateral cortical regions are not functioning.

Dementia as bilateral and multifocal brain lesions

Although one does not frequently think of patients with dementia as helpful for addressing issues of focal brain dysfunction, there are several advantages of this approach. First, many dementias affect brain regions bilaterally. Although the impact of the disease is not generally identical between the two hemispheres, it is often roughly equivalent. For example, individuals with Alzheimer's disease may show roughly symmetric bilateral atrophy of hippocampi, anterior temporal lobes, and parietal lobes. Patients with posterior cortical atrophy may show roughly symmetric atrophy of parietal and occipital lobes. Some patients with behavioral variant frontotemporal dementia show roughly symmetric atrophy of frontal lobes.

But the real opportunity of studying patients with cortical dementias comes when you measure the precise degree of atrophy, by cortical thickness, for example, using structural MRI. When the degree of atrophy is measured along with consciousness-related variables including structured interviews, questionnaires, experimental paradigms, and potentially EEG, ERP, BIS, PCI, and other measures, you can ascertain which cortical areas are important for specific domains of consciousness. This approach has been successfully used in parcellating different aspects of memory (Wolk et al., 2011).

Functional MRI (fMRI) or positron emission tomography (PET)

fMRI and fluorodeoxyglucose (FDG) PET studies are the standard tools used to evaluate the function of various brain regions. Such tools have been used to evaluate consciousness and can yield valuable insights when combined with appropriate experimental paradigms. Paradigms that contrast conscious awareness and unconscious perception are key (e.g., Levinson, et al., 2021; for reviews see Mashour et al., 2020 and Michel & Morales, 2020). Similar paradigms can be used to evaluate other domains of consciousness by contrasting conscious awareness with unconscious perception in auditory, olfactory, tactile, and perhaps even gustatory modalities. New paradigms can be developed to evaluate conscious versus unconscious actions, decisions, and emotions. Taken together, such experiments will enable our hypothesis that all cortical regions participate in their own domain of consciousness to be tested.

Electroencephalography (EEG), derived EEG measures such as the bispectral index (BIS), event-related potentials (ERPs), and perturbation complexity index (PCI)

Although scalp EEG does not provide good spatial localization, it has the advantage of measuring the electrical activity of the cerebral cortex. Raw EEG data, such as proportion of different EEG rhythms, is already used to determine various levels of consciousness, related to both sleep as well as to stages of coma. Derived EEG measures, such as BIS, are used by anesthesiologists to measure the depth of anesthesia and reduce intraoperative awareness. Both raw and derived EEG measures can therefore provide a measurement of consciousness that may be informative regarding our and others' hypotheses about various aspects of consciousness.

ERPs are averages of the EEG signal time locked to specific auditory or visual stimuli. ERPs can provide information regarding the precise timing of events. ERPs have also been used to provide information regarding the cortical activity of a variety of different conscious and unconscious brain processes, including those related to attention, perception, working memory, episodic memory, and language. We believe that important insights can be gained when ERPs are measured in experimental paradigms that evaluate conscious versus unconscious brain processes.

Global neuronal workspace theory hypothesizes that consciousness can be measured by a late positive ERP component which occurs approximately 300 ms after a stimulus, which is often called a P300 or P3 (Mashour et al., 2020). This component is thought to be the same as the old-new parietal effect that in memory research is often termed the LPC. Using this ERP component to detect conscious awareness is consistent with studies of conscious awareness of remembering (Ally et al., 2008), and with the ~300 ms delay suggested by our memory theory of consciousness. Others, however, have suggested that the P3b tracks what observers are reporting—not what they are perceiving (Cohen et al., 2020). Regardless of the debate regarding this particular ERP component, we believe that ERPs will be helpful in the search for neural correlates of consciousness. Following our discussion above, it may be that different ERP components will reflect different domains of consciousness—one for conscious perception and one for the action of reporting.

PCI can be thought of as a special type of ERP, one that is triggered by TMS pulses rather than auditory or visual stimuli. Additionally, rather than simply averaging the resultant EEG signal to produce the ERP, the EEG data is evaluated regarding its integration (a measure of how wide the perturbation of the TMS pulses spread) and its information (a measure of how compressed its spatiotemporal pattern of activity can be). A single number is produced. The PCI has been shown to be able to reliably differentiate individuals who are awake from those whose consciousness is diminished due to sleep, anesthesia, or coma (Casali et al., 2013).

Evaluating MRI, FDG PET, EEG, BIS, ERP, and PCI in individuals who may exhibit impaired consciousness.

Lastly, if our hypotheses are correct that there are individuals who show impairments of consciousness (as described in the **Disorders of consciousness** section), the prediction naturally follows that measuring aspects of their consciousness using neuroimaging, neurophysiological, and computational derived techniques should yield differences between these individuals and those with normal consciousness.

One approach to answering the question of how large a cortical region would need to be to support independent consciousness would be to use structural MRI techniques to measure cortical areas in individuals who demonstrate phenomenologically impaired consciousness of a particular modality (such as visual consciousness) due to a brain lesion, degenerative disease, or other pathology. The relevant structural maps and total cortical volumes in these individuals could be compared to those of other individuals who also experienced damage to the relevant brain regions but have intact consciousness in that modality.

Conclusions and limitations.

Our central claim is that consciousness is essentially and originally part of explicit memory. We experience the world progressing serially because our conscious memory system creates a linear, coherent stream of experiences from our unconscious, parallel brain processes. We believe our memory theory of consciousness is useful (and perhaps correct) because it helps explain phenomena recognized as long-standing puzzles for previous theories, such as postdictive effects. We have shown how the memory theory of consciousness helps us understand clinical syndromes, experimental studies, and everyday experiences. We have also suggested that regions of the entire cerebral cortex are the functional units that makes consciousness possible.

We are hopeful that this paper provides the framework for a fruitful line of theoretical, observational, and experimental work that can prove or disprove each of the hypotheses that we have put forth. As is implicit in that statement, we are well aware that many—perhaps even most—of the hypotheses we are proposing may turn out to be incorrect. What we are confident of, however, is that research using the methodologies that we and others have outlined to test our hypotheses will move forward the cognitive neuropsychology, experimental psychology, and cognitive neuroscience of consciousness studies, bringing us closer to understanding the fundamental nature and anatomical basis of consciousness.

In addition to the possibility that some of our hypotheses are wrong, we must also acknowledge that our hypotheses only discuss several small aspects of consciousness and ignore many of the most important parts of any complete theory of consciousness, such as the so-called ‘hard problem’ (how a collection of biological material can produce the subjective experience that we call consciousness). Nonetheless, by careful observation and well-designed experiments, examining this memory theory of consciousness may move us toward a time in the future when such questions will seem quaint, similar to questions about what constitutes the ‘life force’ that living beings have or how light travels through the ether.

Acknowledgements: We thank Ken Paller and the many anonymous reviewers of prior versions of this manuscript for their insightful and helpful comments. For those interested in the origins of this memory theory of consciousness, please see the Supplementary Appendix.

References

Addis DR, Sacchetti DC, Ally BA, et al. 2009. Episodic simulation of future events is impaired in mild Alzheimer's disease. *Neuropsychologia*. 47:2660–2671. doi:10.1016/j.neuropsychologia.2009.05.018

Aglioti S, DeSouza JF, Goodale MA. 1995. Size-contrast illusions deceive the eye but not the hand. *Curr Biol*. 5:679-685. doi:10.1016/s0960-9822(95)00133-3

Akyürek EG, Wolff MJ. 2016. Extended temporal integration in rapid serial visual presentation: Attentional control at Lag 1 and beyond. *Acta Psychol*. 168:50–64 doi:10.1016/j.actpsy.2016.04.009

Alexander RG, Venkatakrishnan A, Chanovas J, et al. 2021. Microsaccades mediate perceptual alternations in Monet's "Impression, sunrise". *Sci Rep*. 11:3612. doi:10.1038/s41598-021-82222-3

Ally BA, Simons JS, McKeever JD, et al. 2008. Parietal contributions to recollection: Electrophysiological evidence from aging and patients with parietal lesions. *Neuropsychologia*. 46:1800–1812. doi:10.1016/j.neuropsychologia.2008.02.026

American Psychiatric Association. 2013. Diagnostic and Statistical Manual of Mental Disorders: DSM-5. 5th ed.

American Psychological Association. 2017. *What is cognitive behavioral therapy?* [online report]. Available at https://www.apa.org/ptsd-guideline/patients-and-families/cognitive-behavioral. Accessed July 5, 2021.

Andermane N, Bosten JM, Seth AK, et al. 2019. Individual differences in change blindness are predicted by the strength and stability of visual representations. *Neurosci Conscious*. 2019(1): niy010. doi:10.1093/nc/niy010

Avery SN, McHugo M, Armstrong K, et al. 2020. Habituation during encoding: A new approach to the evaluation of memory deficits in schizophrenia. *Schizophr Res*. 223:179–185. doi:10.1016/j.schres.2020.07.007

Baddeley A, Wilson BA. 1994. When implicit learning fails: Amnesia and the problem of error elimination. *Neuropsychologia*. 32:53-68. doi:10.1016/0028-3932(94)90068-x

Baars BJ. 2005. Global workspace theory of consciousness: Toward a cognitive neuroscience of human experience. *Prog Brain Res.* 150:45–53. doi:10.1016/S0079-6123(05)50004-9

Berkovitch L, Dehaene S, Gaillard R. 2017. Disruption of Conscious Access in Schizophrenia. *Trends Cogn Sci*. 21:878-892. doi:10.1016/j.tics.2017.08.006

Björnsson G, Brülde B. 2017. Normative responsibilities: Structure and sources. In: Hens K, Cutas D, Horstkötter D, eds. *Parental Responsibility in the Context of Neuroscience and Genetics.* Springer International Publishing; 69:13-33.

Blackmore S. 2017. *Consciousness: A Very Short Introduction.* 2nd ed. Oxford, UK: Oxford University Press.

Block N. 2011. Perceptual consciousness overflows cognitive access. *Trends Cogn Sci*. 15:567-575. doi:10.1016/j.tics.2011.11.001

Boly M, Massimini M, Tsuchiya N, et al. 2017. Are the neural correlates of consciousness in the front or in the back of the cerebral cortex? Clinical and neuroimaging evidence. *J. Neurosci*. 37:9603–9613. doi:10.1523/JNEUROSCI.3218-16.2017

Breitmeyer BG. 2015. Psychophysical "blinding" methods reveal a functional hierarchy of unconscious visual processing. *Conscious Cogn*. 35:234-250. doi:10.1016/j.concog.2015.01.012

Breitmeyer BG, Kiefer M, Niedeggen M. 2015. Exploring the visual (un)conscious. *Conscious Cogn*. 35:178-184. doi:10.1016/j.concog.2015.04.021

Brogaard B. 2011. Conscious vision for action versus unconscious vision for action?. *Cogn Sci*. 35(6):1076-1104. doi:10.1111/j.1551-6709.2011.01171.x

Brown R, Lau H, LeDoux JE. 2019. Understanding the higher-order approach to consciousness. *Trends Cogn Sci*. 23:754–768. doi:10.1016/j.tics.2019.06.009

Buckner RL, Andrews-Hanna JR, Schacter DL. 2008. The brain's default network: anatomy, function, and relevance to disease. *Ann N Y Acad Sci*. 1124:1-38. doi:10.1196/annals.1440.011

Calvi, E, Quassolo, U, Massaia, M, et al. 2020. The scent of emotions: A systematic review of human intra- and interspecific chemical communication of emotions. *Brain and Behav.* 10:e01585. doi:10.1002/brb3.1585

Carruthers P. 2015. *The Centered Mind: What the Science of Working Memory Shows Us About the Nature of Human Thought*. Oxford, UK: Oxford University Press.

Carruthers P. 2011. *The Opacity of Mind: An Integrative Theory of Self-Knowledge*. Oxford, UK: Oxford University Press.

Casali AG, Gosseries O, Rosanova M, et al. 2013. A theoretically based index of consciousness independent of sensory processing and behavior. *Sci Transl Med.* 5: 198ra105. doi:10.1126/scitranslmed.3006294

Cavinato M, Rigon J, Volpato C, et al. 2012. Preservation of auditory P300-like potentials in cortical deafness. *PLoS One*. 7:e29909. doi:10.1371/journal.pone.0029909

Chalmers DJ. 2010. *The Character of Consciousness.* Oxford, UK: Oxford University Press.

Chen J, Jayawardena S, Goodale MA. 2015. The effects of shape crowding on grasping. *J Vis*. 15:6. doi:10.1167/15.3.6

Churchland PS. 1981. The timing of sensations: Reply to Libet. *Philos Sci.* 48:492–497. doi:10.1086/289013

Cisek P. 2019. Resynthesizing behavior through phylogenetic refinement. *Atten Percept Psychophys*. 81:2265-2287. doi:10.3758/s13414-019-01760-1

Cleeremans A. 2011. The radical plasticity thesis: How the brain learns to be conscious. *Front Psychol*. 2:86. doi:10.3389/fpsyg.2011.00086

Cohen MA, Dennett DC, Kanwisher N. 2016. What is the bandwidth of perceptual experience?. *Trends Cogn Sci.* 20:324-335. doi:10.1016/j.tics.2016.03.006

Cohen MA, Ortego K, Kyroudis A, et al. 2020. Distinguishing the neural correlates of perceptual awareness and postperceptual processing. *J Neurosci*. 40:4925-4935. doi:10.1523/JNEUROSCI.0120-20.2020

Courchesne E, Mouton PR, Calhoun ME, et al. 2011. Neuron number and size in prefrontal cortex of children with autism. *JAMA*. 306, 2001–2010. doi:10.1001/jama.2011.1638

Dafni-Merom A, Arzy S. 2020. The radiation of autonoetic consciousness in cognitive neuroscience: A functional neuroanatomy perspective. *Neuropsychologia*. 143:107477. doi:10.1016/j.neuropsychologia.2020.107477

Damasio H, Grabowski T, Frank R, et al. 1994. The return of Phineas Gage: Clues about the brain from the skull of a famous patient. *Science*. 264:1102–1105. doi:10.1126/science.8178168

Damasio H, Grabowski TJ, Tranel D, et al. 1996. A neural basis for lexical retrieval. *Nature.* 380:499–505. doi:10.1038/380499a0

Das JM, Naqvi IA. 2021. Anton Syndrome. *StatPearls*. Treasure Island, Florida: StatPearls Publishing.

Davidson MJ, Graafsma IL, Tsuchiya N, et al. 2020 A multiple-response frequency-tagging paradigm measures graded changes in consciousness during perceptual filling-in. *Neurosci Conscious*. 2020(1): niaa002. doi:10.1093/nc/niaa002

de Waal FBM. 2019. Fish, mirrors, and a gradualist perspective on self-awareness. *PLoS Biol* 17:e3000112. doi:10.1371/journal.pbio.3000112

Dennett DC. 1991. *Consciousness Explained*. Boston, Massachusetts: Little, Brown and Company.

Dienes Z, Altmann GTM, Kwan L, et al. 1995. Unconscious knowledge of artificial grammars is applied strategically. *J. Exp. Psychol. Learn. Mem. Cog*. 21:1322–1338. doi:10.1037/0278-7393.21.5.1322

Eichenbaum H, Fortin NJ, Ergorul C, et al. 2005. Episodic recollection in animals: "If it walks like a duck and quacks like a duck. . ." *Learn. Motiv.* 36:190–207. doi:10.1016/j.lmot.2005.02.006

Elward RL, Vargha-Khadem F. 2018. Semantic memory in developmental amnesia. *Neurosci. Letts*. 680:23–30. doi.org/10.1016/j.neulet.2018.04.040

Floody OR, Ouda L, Porter BA, et al. 2010. Effects of damage to auditory cortex on the discrimination of speech sounds by rats. *Physiol. Behav.* 101:260–268. doi:10.1016/j.physbeh.2010.05.009

Gaillard R, Dehaene S, Adam C, et al. 2009. Converging intracranial markers of conscious access. *PLoS Biology*, 7:e61. doi:10.1371/journal.pbio.1000061

Gandola M, Invernizzi P, Sedda A, et al. 2012. An anatomical account of somatoparaphrenia. *Cortex*. 48:1165–1178. doi:10.1016/j.cortex.2011.06.012

Ganel T, Goodale MA. 2019. Still holding after all these years: An action-perception dissociation in patient DF. *Neuropsychologia*. 128:249–254. doi:10.1016/j.neuropsychologia.2017.09.016

Gazzaniga MS. 2015. *Tales From Both Sides of the Brain: A Life in Neuroscience*. New York, New York: Ecco/HarperCollins Publishers.

Geldard FA, Sherrick CE. 1972. The cutaneous "rabbit": A perceptual illusion". *Science.* 178:178–179. doi:10.1126/science.178.4057.178

Geschwind N. 1965. Disconnexion syndromes in animals and man. I. *Brain*. 88:237–294. doi:10.1093/brain/88.2.237

Ginsburg S, Jablonka E. 2021. Evolutionary transitions in learning and cognition. *Philos Trans R Soc Lond B Biol Sci*. 376:20190766. doi:10.1098/rstb.2019.0766

Goldreich, D. 2007. A Bayesian perceptual model replicates the cutaneous rabbit and other tactile spatiotemporal illusions. *PLoS One.* 2:e333. doi:10.1371/journal.pone.0000333

Gould SJ, Lewontin RC. 1979. The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme. *Proc R Soc Lond B Biol Sci*.;205(1161):581-598. doi:10.1098/rspb.1979.0086

Grewal D, Ku MC, Girod SC, et al. 2013. How to recognize and address unconscious bias. In: Roberts L. ed. *The Academic Medicine Handbook*. New York, New York: Springer.

Herculano-Houzel S. 2009. The human brain in numbers: A linearly scaled-up primate brain. *Front. Hum. Neurosci.* 3:31. doi:10.3389/neuro.09.031.2009

Herzog MH, Drissi-Daoudi L, Doerig A. 2020. All in good Time: Long-lasting postdictive effects reveal discrete perception. *Trends Cogn Sci*. 24:826-837. doi:10.1016/j.tics.2020.07.001

Hodinott-Hill, I, Thilo KV, Cowey A, et al. 2002. Auditory chronostasis: Hanging on the telephone. *Curr. Biol.* 12: 1779–1781. doi:10.1016/S0960-9822(02)01219-8

Holender D. 1986. Semantic activation without conscious identification in dichotic listening, parafoveal vision, and visual masking: a survey and appraisal. *Behav. Brain Sci*. 9:1–66. doi:10.1017/S0140525X00021269

Horowitz A. 2017. Smelling themselves: Dogs investigate their own odours longer when modified in an "olfactory mirror" test. *Behav Processes*. 143:17-24. doi:10.1016/j.beproc.2017.08.001

Hsieh PJ, Colas JT, Kanwisher N. 2011. Pop-out without awareness: Unseen feature singletons capture attention only when top-down attention is available. *Psychol Sci*. 22:1220-1226. doi:10.1177/0956797611419302

Hutchinson BT, Pammer K, Jack B. 2021. Pre-stimulus alpha predicts inattentional blindness. *Conscious Cogn*. 87:103034. doi:10.1016/j.concog.2020.103034

Hume D. 1978. *A Treatise of Human Nature (1739-40)*. Oxford, UK: Oxford University Press.

James, W. 1890. *The Principles of Psychology*. New York, New York: Henry Holt & Company.

Janssen J, Oudman E, Irish M, et al. 2022. Exploring episodic and semantic contributions to past and future thinking performance in Korsakoff's syndrome. *Mem Cognit*. 50:630-640. doi:10.3758/s13421-021-01262-2

Jacoby LL. 1991. A process-dissociation framework: Separating automatic from intentional uses of memory. *J. Mem Lang*. 30:513–541. doi/10.1016/0749-596X(91)90025-F

Kaas JH. 2019. The origin and evolution of neocortex: From early mammals to modern humans. *Prog Brain Res*. 250:61-81. doi:10.1016/bs.pbr.2019.03.017

Kahneman D. 2011. *Thinking, Fast and Slow.* New York, New York: Farrar, Straus & Giroux.

Kentridge RW. 2013. Visual attention: bringing the unseen past into view. *Curr Biol*. 23:R69-R71. doi:10.1016/j.cub.2012.11.056

Kentridge RW, Nijboer TCW, Heywood CA. 2008. Attended but unseen: Visual attention is not sufficient for visual awareness. *Neuropsychologia*, 46:864–869. doi:10.1016/j.neuropsychologia.2007.11.036

Keuninckx L, Cleeremans A. 2021. The color phi phenomenon: Not so special, after all?. *PLoS Comput Biol*. 17:e1009344. doi:10.1371/journal.pcbi.1009344

Kinney HC, Korein J, Panigrahy A, et al. 1994. Neuropathological findings in the brain of Karen Ann Quinlan. The role of the thalamus in the persistent vegetative state. *N Engl J Med*. 330:1469-1475. doi:10.1056/NEJM199405263302101

Koch C, Massimini M, Boly M, et al. 2016. Neural correlates of consciousness: progress and problems. *Nat. Rev. Neurosci.* 17:307–321. doi:10.1038/nrn.2016.22

Kohda M, Hotta T, Takeyama T, et al. 2019. If a fish can pass the mark test, what are the implications for consciousness and self-awareness testing in animals? *PLOS Biol.* 17:e3000021. doi:10.1371/journal.pbio.3000021

Kolers PA, von Grünau M. 1976. Shape and color in apparent motion. *Vision Res.* 16(4), 329–335. doi:10.1016/0042-6989(76)90192-9

Lamme V. 2015. The crack of dawn: perceptual functions and neural mechanisms that mark the transition from unconscious processing to conscious vision. *Open MIND*. 22(T). doi:10.15502/9783958570092

Lamme V. 2018. Challenges for theories of consciousness: seeing or knowing, the missing ingredient and how to deal with panpsychism. *Philos. Trans R. Soc. Lond., B, Biol Sci.* 373: 20170344. Doi:10.1098/rstb.2017.0344

Landry SP, Champoux F. 2017. Musicians react faster and are better multisensory integrators. *Brain Cogn*. 111:156-162. doi:10.1016/j.bandc.2016.12.001

Laplane D, Degos JD. 1983. Motor neglect. *J. Neurol. Neurosurg. Psychiatry*. 46:152–158. doi:10.1136/jnnp.46.2.152

Lee M, Baird B, Gosseries O, et al. 2019. Connectivity differences between consciousness and unconsciousness in non-rapid eye movement sleep: a TMS-EEG study. *Sci Rep*, 9:5175. doi:10.1038/s41598-019-41274-2

Levinson M, Podvalny E, Baete SH, et al. 2021. Cortical and subcortical signatures of conscious object recognition. *Nat. Comm.* 12(1): 2930. doi:10.1038/s41467-021-23266-x

Libet B, Alberts WW, Wright EW, et al. 1964. Production of threshold levels of conscious sensation by electrical stimulation of human somatosensory cortex. *J Neurophysiol*. 27:546-578. doi:10.1152/jn.1964.27.4.546

Libet B, Wright EW, Feinstein B, et al. 1979. Subjective referral of the timing for a conscious sensory experience: a functional role for the somatosensory specific projection system in man. *Brain*. 102:193-224. doi:10.1093/brain/102.1.193

Libet B. 1985. Unconscious cerebral initiative and the role of conscious will in voluntary action. *Behav Brain Sci.* 8:529-539. doi:10.1017/S0140525X00044903

Lipowski ZJ. 1989. Delirium in the elderly patient. *N Engl J Med*. 320(9):578-582. doi:10.1056/NEJM198903023200907

Lou HC, Changeux JP, Rosenstand A. 2017. Towards a cognitive neuroscience of self-awareness. *Neurosci Biobehav Rev*. 83:765-773. doi:10.1016/j.neubiorev.2016.04.004

Mashour GA, Roelfsema P, Changeux JP, et al. 2020. Conscious processing and the global neuronal workspace hypothesis. *Neuron*. 105:776–798. doi:10.1016/j.neuron.2020.01.026

Melcher D, Kumar D, Srinivasan N. 2020. The role of action intentionality and effector in the subjective expansion of temporal duration after saccadic eye movements. *Sci. Rep.* 10:16922. doi:10.1038/s41598-020-73830-6

Mesulam MM. 1999. Spatial attention and neglect: parietal, frontal and cingulate contributions to the mental representation and attentional targeting of salient extrapersonal events. *Philos. Trans R. Soc. Lond., B, Biol Sci*. 354:1325–1346. doi:10.1098/rstb.1999.0482

Mesulam MM, Thompson CK, Weintraub S, et al. 2015. The Wernicke conundrum and the anatomy of language comprehension in primary progressive aphasia. *Brain*. 138:2423–2437. doi:10.1093/brain/awv154

Metcalfe J. 1986. Premonitions of insight predict impending error. *J. Exp. Psychol.: Learn. Mem. Cogn*. 12:623–634. doi:10.1037/0278-7393.12.4.623

Michel M, Doerig A. 2020. A new empirical challenge for local theories of consciousness. *Mind Lang*. 2020:1– 16. doi:10.1111/mila.12319

Michel M, Morales J. 2020. Minority reports: Consciousness and the prefrontal cortex. *Mind Lang*. 35:493– 513. doi:10.1111/mila.12264

Moulton ST, Kosslyn SM. 2009. Imagining predictions: Mental imagery as mental emulation. *Philos. Trans R. Soc. Lond., B, Biol Sci*. 364:1273–1280. doi:10.1098/rstb.2008.0314

Moscovitch M. 1995. Models of consciousness and memory. In: M. Gazzaniga M. eds. *The cognitive neurosciences*. Cambridge, Massachusetts: MIT Press. Cambridge, 1341.

Mundy ME, Downing PE, Dwyer DM, et al. 2013. A critical role for the hippocampus and perirhinal cortex in perceptual learning of scenes and faces: Complementary findings from amnesia and FMRI. *J. Neurosci.* 33:10490–10502. doi:10.1523/JNEUROSCI.2958-12.2013

Murray EA, Wise SP, Baldwin MKL, et al. 2020. *The evolutionary road to human memory*. Oxford, UK: Oxford University Press.

Nichols RM, Loftus EF. 2019. Who is susceptible in three false memory tasks?. *Memory*. 27:962–984. doi:10.1080/09658211.2019.1611862

Nisbett RE, Wilson TD. 1977. Telling more than we can know: Verbal reports on mental processes. *Psychol. Rev.* 84:231–259. doi:10.1037/0033-295X.84.3.231

Ochsner KN, Knierim K, Ludlow DH, et al. 2004. Reflecting upon feelings: An fMRI study of neural systems supporting the attribution of emotion to self and other. *J. Cogn. Neurosci.* 16:1746–1772. doi:10.1162/0898929042947829

Odegaard B, Knight RT, Lau H. 2017. Should a few null findings falsify prefrontal theories of conscious perception? *J. Neurosci.* 37:9593–9602. doi:10.1523/JNEUROSCI.3217-16.2017

Oliver C, Richards C. 2015. Practitioner Review: Self-injurious behaviour in children with developmental delay. *J Child Psychol Psychiatry*. 56(10):1042-1054. doi:10.1111/jcpp.12425

Otto TU, Ögmen H, Herzog MH. 2006. The flight path of the phoenix – the visible trace of invisible elements in human vision. *J. Vis*. 6:1079–1086. doi:10.1167/6.10.7

Paller KA, Creery JD, Schechtman E. 2021. Memory and Sleep: How Sleep Cognition Can Change the Waking Mind for the Better. *Annu. Rev. Psychol.* 72:123–150. doi:10.1146/annurev-psych-010419-050815

Persaud N, McLeod P, Cowey A. 2007. Post-decision wagering objectively measures awareness. *Nat. Neurosci.* 10:257–261. doi:10.1038/nn1840

Pilz KS, Zimmermann C, Scholz J, et al. 2013. Long-lasting visual integration of form, motion, and color as revealed by visual masking. *J Vis*. 13:12. doi:10.1167/13.10.12

Pinker S. 2007. *The Language Instinct: How the Mind Creates Language.* New York, New York: Harper Perennial Modern Classics.

Poppel E, Held R, Frost D. 1973. Residual visual function after brain wounds involving the central visual pathways in man. *Nature*, 243:295–296. doi:10.1038/243295a0

Popper KR, Eccles JC. 1977. *The Self and Its Brain: An Argument for Interactionism.* 1st ed. Berlin, Germany: Springer-Verlag.

Posner JB, Saper CB, Schiff ND, et al. 2019. *Plum and Posner's Diagnosis and Treatment of Stupor and Coma.* 5th ed. Oxford, UK: Oxford University Press.

Prior H, Schwarz A, Güntürkün O .2008. Mirror-induced behavior in the magpie (Pica pica): Evidence of self-recognition. *PLoS Biol.* 6:e202. doi:10.1371/journal.pbio.0060202

Raby CR, Alexis DM, Dickinson A, et al. 2007. Planning for the future by western scrub-jays. *Nature*. 445:919-921. doi:10.1038/nature05575

Ramsøy TZ, Overgaard M. 2004. Introspection and subliminal perception. *Phenomenol. Cogn. Sci.* 3:1–23. doi:10.1023/B:PHEN.0000041900.30172.e8

Reyna VF, Brainerd CJ. 1995. Fuzzy-trace theory: An interim synthesis. *Learn. Individ. Differ.* 7:1–75. doi:10.1016/1041-6080(95)90031-4

Renoult L, Irish M, Moscovitch M, et al. 2019. From knowing to remembering: The semantic-episodic distinction. *Trends Cogn. Sci*. 23:1041–1057. doi:10.1016/j.tics.2019.09.008

Repovs G, Baddeley A. 2006. The multi-component model of working memory: Explorations in experimental cognitive psychology. *Neuroscience,* 139:5–21. doi:10.1016/j.neuroscience.2005.12.061

Rodriguez N, Kintzer E, List J, Lypson M, et al. 2021. Implicit bias recognition and management: Tailored instruction for faculty. *J. Natl Med Assoc*. 113:566-575. doi:10.1016/j.jnma.2021.05.003

Science Non-Fiction. 2016. *Hitting a fastball requires more than just quick reactions* [online report]. Available at https://sciencenonfiction.org/2016/05/23/hitting-a-fastball-requires-more-than-just-quick-reactions/comment-page-1/. Accessed May 30, 2021

Sanders K, Osburn S, Paller KA, et al. 2019. Targeted memory reactivation during sleep improves next-day problem solving. *Psychol. Sci.* 30:1616–1624. doi:10.1177/0956797619873344

Schacter DL, Addis DR, Buckner RL. 2007. Remembering the past to imagine the future: the prospective brain. *Nat Rev Neurosci*. 8:657-661. doi: 10.1038/nrn2213. PMID: 17700624

Schacter DL, Dodson CS. 2001. Misattribution, false recognition and the sins of memory. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* 356:1385–1393. doi:10.1098/rstb.2001.0938

Schacter DL, Gilbert DT, Knock MK, et al. 2019. *Psychology.* 5th ed. New York, New York: Worth Publishers.

Scharnowski F, Rüter J, Jolij J, et al. 2009. Long-lasting modulation of feature integration by transcranial magnetic stimulation. *J. Vis.* 9:1–1, 110. doi:10.1167/9.6.1

Schneider E, Züst MA, Wuethrich S, et al. 2021. Larger capacity for unconscious versus conscious episodic memory. *Curr Biol*. 31:3551-3563.e9. doi:10.1016/j.cub.2021.06.012

Schurger A, Sitt JD, Dehaene S. An accumulator model for spontaneous neural activity prior to self-initiated movement. *Proc. Natl. Acad. Sci. U.S.A*. 109:E2904-E2913. doi:10.1073/pnas.1210467109

Sergent C. 2018. The offline stream of conscious representations. *Phil. Trans. R. Soc. B.* 373:20170349. doi:10.1098/rstb.2017.0349

Sergent C, Dehaene S. 2004. Is consciousness a gradual phenomenon?. *Psychol. Sci.* 15:720–728. doi:10.1111/j.0956-7976.2004.00748.x

Sergent C, Wyart V, Babo-Rebelo M, et al. 2013. Cueing attention after the stimulus is gone can retrospectively trigger conscious perception. *Curr. Biol*. 23:150-155. doi:10.1016/j.cub.2012.11.047

Sikora-Wachowicz B, Lewandowska K, Keresztes A, et al. 2019. False recognition in short-term memory - Age-differences in confidence. *Front. Psychol.* 10:2785. doi:10.3389/fpsyg.2019.02785

Simons DJ. 2010. Monkeying around with the gorillas in our midst: Familiarity with an inattentional-blindness task does not improve the detection of unexpected events. *i-Perception*. 1:3–6. doi:10.1068/i0386

Simons JS, Peers PV, Hwang DY, et al. 2008. Is the parietal lobe necessary for recollection in humans? *Neuropsychologia.* 46:1185–1191. doi:10.1016/j.neuropsychologia.2007.07.024

Smith R, Lane RD. 2015. The neural basis of one's own conscious and unconscious emotional states. *Neurosci Biobehav Rev*. 57:1–29. doi:10.1016/j.neubiorev.2015.08.003

Sperling G. 1960. The information available in brief visual presentations. *Psychol. Monogr*. 74:1–29. doi:10.1037/h0093759

Stiles NRB, Li M, Levitan CA, et al. 2018. What you saw is what you will hear: two new illusions with audiovisual postdictive effects. *PLoS One* 13:e0204217. doi:10.1371/journal.pone.0204217

Suddendorf T, Corballis MC. 2007. The evolution of foresight: What is mental time travel, and is it unique to humans?. *Behav Brain Sci*. 30:299-351. doi:10.1017/S0140525X07001975

Sy JL, Miao HY, Marois R, et al. 2021. Conscious perception can be both graded and discrete. *J Exp Psychol Gen*. 150:1461-1475. doi:10.1037/xge0001009

Teasdale G, Jennett B. 1974. Assessment of coma and impaired consciousness. A practical scale. *Lancet*. 2:81–84. doi:10.1016/s0140-6736(74)91639-0

Thilo KV, Walsh V. 2002. Vision: When the clock appears to stop. *Curr. Biol*. 12:R135–R137. doi:10.1016/S0960-9822(02)00707-8

Thomas V, Davidson M, Zakavi P, et al. 2017. Simulated forward and backward self motion, based on realistic parameters, causes motion induced blindness. *Sci Rep*. 7:9767. doi:10.1038/s41598-017-09424-6

Timmermans B, Cleeremans A. 2015. How can we measure awareness? An overview of current methods. In: Overgaard M, ed. *Behavioral Methods in Consciousness Research*. Oxford, UK: Oxford University Press.

Tononi G, Boly M, Massimini M, et al. 2016. Integrated information theory: from consciousness to its physical substrate. *Nat. Rev. Neurosci.* 17:450–461. doi:10.1038/nrn.2016.44

Tordjman S, Celume MP, Denis L, et al. 2019. Reframing schizophrenia and autism as bodily self-consciousness disorders leading to a deficit of theory of mind and empathy with social communication impairments. *Neurosci Biobehav Rev.* 103:401–413. doi:10.1016/j.neubiorev.2019.04.007

Trübutschek D, Marti S, Ojeda A, et al. 2017. A theory of working memory without consciousness or sustained activity. *Elife*. 6:e23871. doi:10.7554/eLife.23871

Tulving, E., 1985. Memory and consciousness. *Can. Psychol.* 26 (1), 1–12. doi.org/10.1037/h0080017

Tulving E. 1999. *Memory, Consciousness, and the Brain: The Tallinn Conference*. Abington-on-Thames, UK: Psychology Press, 66-68.

Vinding M, Jensen M, Overgaard M. 2014. Distinct electrophysiological potentials for intention in action and prior intention for action. *Cortex*, 50:86–99. doi:10.1016/j.cortex.2013.09.001

Vonk, J. 2020. A fish eye view of the mirror test. *Learn Behav* 48, 193–194. doi:10.3758/s13420-019-00385-6

Weiskrantz L, Warrington EK, Sanders MD, et al. 1974. Visual capacity in the hemianopic field following a restricted occipital ablation. *Brain*. 97:709–728. doi:10.1093/brain/97.1.709

Wierzchoń M, Asanowicz D, Paulewicz B, et al. 2012. Subjective measures of consciousness in artificial grammar learning task. *Conscious. Cogn*. 21(3), 1141–1153. doi:10.1016/j.concog.2012.05.012

Wilson JE, Mart MF, Cunningham C, et al. 2020. Delirium [published correction appears in Nat Rev Dis Primers. 2020 Dec 1;6(1):94]. *Nat Rev Dis Primers*. 6(1):90. Published 2020 Nov 12. doi:10.1038/s41572-020-00223-4

Wolk DA, Dickerson BC, Alzheimer's Disease Neuroimaging Initiative. 2011. Fractionating verbal episodic memory in Alzheimer's disease. *NeuroImage*. 54:1530–1539. doi:10.1016/j.neuroimage.2010.09.005

Yaron I, Melloni L, Pitts M, et al. 2022. The Consciousness Theories Studies (ConTraSt) database: analyzing and comparing empirical studies of consciousness theories. *[published online ahead of print, 2022 Feb 21]. Nat Hum Behav. 10.1038/s41562-021-01284-5. doi:10.1038/s41562-021-01284-5*

Zhang X, Zhaoping L, Zhou T, et al. 2021. Neural activities in v1 create a bottom-up saliency map. *Neuron*. 73:183-192. doi:10.1016/j.neuron.2011.10.035

Zheng A, Montez DF, Marek S, et al. 2021. Parallel hippocampal-parietal circuits for self- and goal-oriented processing. *Proc Natl Acad Sci U S A*. 118(34):e2101743118. doi:10.1073/pnas.2101743118