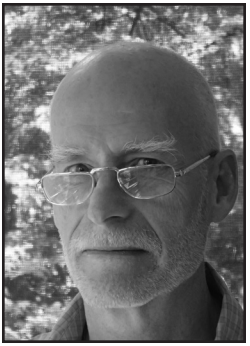
The Problem of Animal Consciousness



Peter Carruthers
UNIVERSITY OF MARYLAND

Romanell lecture delivered at the ninety-second Pacific Division meeting of the American Philosophical Association in San Diego, CA, on March 29, 2018.

I. THE TARGET QUESTION

The kind of consciousness that forms our topic is so-called *phenomenal* consciousness. This is the sort of consciousness that is *like something* to undergo, or that has a distinctive subjective *feel*. Phenomenal consciousness is a species of mental-state consciousness. It is mental states (seeing a sunset, hearing a dog bark, smelling cinnamon) that can be phenomenally conscious. *People* are phenomenally conscious derivatively, by virtue of undergoing phenomenally conscious states. It is phenomenal consciousness that is thought to give rise to the "hard problem" of consciousness.[1] For it seems one can conceive of a zombie—a creature that is like oneself in all physical, functional, and representational respects except that it lacks *this* feeling (the distinctive feeling of the smell of cinnamon). Likewise, there seems to be an unbridgeable explanatory gap between all physical, functional, and representational facts and one's current conscious experience. Hence many have been tempted to conclude that phenomenal consciousness involves properties (often called "qualia") that cannot be reduced to any combination of physical, functional, or representational ones.

It is important to emphasize that the concept of phenomenal consciousness is a first-person one. The various locutions employed ("like something to undergo," "subjective feel," "qualitative character," and so on) are all intended just to draw one's attention to one's own phenomenally conscious experiences. Acquaintance with the latter is a necessary condition for grasping the concept, and no definition or third-personal explanation could confer understanding of the concept. Hence

philosophical zombies don't just lack phenomenal consciousness itself; they must also lack the *concept* of phenomenal consciousness.[2]

Phenomenal consciousness is at least conceptually distinct from *access* consciousness.[3] (Whether there is any real distinction between the properties picked out by these concepts is another matter, to which we return in later sections. On the global-workspace view I defend, there isn't: phenomenal consciousness *is* access-conscious nonconceptual content.) Both are forms of mental-state consciousness: it is mental states that can have phenomenal properties, or that can be accessible to enter into decision-making, reasoning, and verbal report. Whereas the concept of phenomenal consciousness is a first-personal one, *access* consciousness, in contrast, is functionally defined, and the concept could be fully understood by a zombie. A mental state is said to be access conscious if it is accessible to a wide range of other systems for further processing, specifically those involved in issuing verbal reports, in decision-making, in reasoning, and in the formation of long-term memories.

Mental-state consciousness (whether access or phenomenal) should be distinguished from *creature* consciousness, which can be either *transitive* or *intransitive*.[4] Whenever a creature (whether human or animal) is aware of some object or event in its environment or body, it can be said to be (transitively) conscious of that object or event. Put differently, a creature is transitively conscious of an object or event when it *perceives* that object or event. It is debatable whether or not transitive creature consciousness requires mental-state consciousness. For it is debatable whether the perceptual states that enable a creature to be aware of its environment must be conscious ones (in either the access or phenomenal sense). At any rate, it is worth noting that there are many kinds of case where one would pre-theoretically ascribe creature consciousness to an agent—since the agent is displaying flexible perceptual sensitivity to the environment—but where the states in virtue of which it acts as it does are *not* conscious ones. (Examples include sleepwalking and swift visually guided motor control.) This point will be elaborated on in Section II.

Intransitive creature consciousness, on the other hand, is a matter of being awake rather than asleep, or conscious as opposed to comatose. When the creature in question is a human person, then intransitive creature consciousness would normally implicate some or other form of mental-state consciousness. Whenever one is awake one is normally undergoing some conscious mental state or other. But the reverse need

not be true. It seems that dreams are conscious mental states, even though the dreaming subject is asleep, and hence creature *un*conscious.

Note that both forms of creature consciousness admit of degrees. One can be more or less aware of the properties of a stimulus, and one can be more or less awake. Likewise, the concept of access consciousness allows for degrees. A mental state could be available to more, or to fewer, of the systems for reasoning, reporting, remembering, and so on. Phenomenal consciousness, in contrast, is all-or-nothing. It is hard even to conceive of a case of experience that is partly *like something* to undergo, partly not. Indeed, even if one is only partially awake, some of the states one is in are determinately phenomenally conscious—it is *like something* to be barely awake. And even though one's awareness of an object can be more or less detailed, or more or less rich and vivid (contrast looking at something in full light versus looking at it when so dimly lit that one can barely make it out), it is fully—unequivocally—*like something* to be looking at a dimly lit object, even if one is aware of many fewer properties of it.

Consciousness can fade, of course. But in this case it is surely (intransitive) creature consciousness that fades. As one slips into sleep, or is slowly rendered unconscious by anesthesia, one will be subject to *fewer* conscious mental states and/or one will have states with increasingly impoverished contents. But so long as one retains some degree of intransitive creature consciousness one will have *some* phenomenally conscious mental states. And no matter how impoverished their contents, it will be determinately *like something* to be in them. Even if all that remains is an indistinct impression of flickering light, or a vague impression of people talking in the distance, the experience of a faint flickering light, or of indeterminate voices, is unequivocally phenomenally conscious.

Our question about nonhuman animals, then, isn't whether animals can be awake, half-awake, or asleep. (Of course they can.) Nor is it about whether animals can be perceptually sensitive to the properties of their environments. (The obvious answer is that they often are.) Our question is whether the *mental states* of animals are ever conscious; specifically, whether they are ever *phenomenally* conscious. And if they are, which ones, in which species of creature? And how would we know?

II. THE NEED FOR A THEORY

It is natural to think that we need a theory of consciousness, or at least a theory of the neural correlates of consciousness, if we are to address the question of phenomenal consciousness in animals. But Michael Tye has recently argued that this is a mistake.[5] For provided we think conscious states play a causal role in the production of behavior, we can apply Newton's Principle: similar effects have similar causes. So when we see an animal exhibiting behavior similar to that displayed by human beings—and specifically, when we see an animal showing flexible perceptually grounded sensitivity to properties of its environment or its own body—then we can conclude that it enjoys conscious experience. (At least, this is so unless we know of some additional fact that might undermine—or *defeat*—Newton's Principle in any given case. Much of Tye's book is occupied with discussion of possible defeaters, including the question whether absence of a cerebral cortex is a defeater for conscious experience.)

Unfortunately, this approach is unsound. For it neglects a wide range of environmentally sensitive forms of behavior in humans that are at least to some degree flexible, but are caused by *un*conscious perceptual states. Consider, for example, a baseball batter facing a fastball traveling at 96 miles per hour. The distance between the pitching plate and the batting plate is 60 feet, hence there are about 56 feet from the point at which the pitcher releases the ball to the point of contact for the batter. The fastball will travel that distance in 400 milliseconds. In that 400-millisecond interval the batter has to estimate the trajectory of the ball (will it pass through the strike-zone or not?), initiate a swing, and then plan and execute the motor instructions for a specific type of swing, aimed at the expected location of the ball. But our best estimate of the time that elapses between presentation of a stimulus (in this case, light hitting the retina from the ball as it exits the pitcher's hand) and conscious experience is somewhere around 350 milliseconds.[6] It is almost certain, therefore, that the swing is initiated *before* the batter's perception of the trajectory of the ball becomes conscious, and that the execution of the swing is likewise guided by unconscious perceptual states throughout. It seems that while the batter's swing is *accompanied* by conscious experience, it isn't caused by it. So Newton's Principle cannot apply. This means that when we observe an animal responding to its environment in this sort of swift online manner we cannot use Newton's Principle to infer that it enjoys conscious experience.

These points are vindicated and explained by the discovery of two visual systems in humans.[7] As is now familiar to many people, visual processing

proceeds in two distinct cortical streams. (Similar distinctions can be drawn with respect to other sensory systems.[8]) The ventral stream runs forward from visual cortex through the lower portions of the temporal lobes, whereas the dorsal stream runs upward through the parietal lobes. Processing within the ventral stream results in object recognition, with the relative positions of objects represented in allocentric space. Processing in the dorsal stream, in contrast, is coded in limb-centered and body-centered spatial coordinates, and is specialized for online guidance of action. Moreover, the two streams utilize distinct cellular pathways that remain distinct from one another all the way back to the retina. The ventral steam is a continuation of the parvocellular pathway, whereas the dorsal stream is a continuation of the magnocellular pathway. Notably, given the points about speed made in the previous paragraph, transmission of signals through the magnocellular (dorsal) pathway is significantly faster than through the parvocellular one.

Multiple lines of evidence suggest that, while the two visual streams interact with one another in their early stages, their outputs are distinct. For example, damage to the dorsal stream can result in optic ataxia, in which visually guided action becomes halting and inaccurate while conscious experience is fully intact; whereas damage to the ventral stream can result in agnosia (an inability to consciously identify objects), while leaving visually guided motor control intact. Indeed, while the outputs of the ventral stream can be conscious, the outputs of the dorsal stream are inaccessible to consciousness.[9]

Consider, in particular, D.F., a patient with bilateral temporal-lobe damage who has been extensively studied.[10] D.F. suffers from complete visual-form agnosia. While she can still experience colors and textures (and hence might be able to guess at the identity of a banana from its distinctive yellow color and mottled texture), she can no longer experience the shapes or orientations of objects. For example, she cannot recognize a banana from a line drawing of one, she is at chance when judging the orientation of a pencil (whether upright or horizontal), and she is at chance when judging whether a block of wood is square or oblong. But her reaching-and-grasping behavior is indistinguishable from normal. She will orient her hand appropriately when grasping a horizontally held or vertically held pencil, using a normal finger grip (but without being able to report the pencil's orientation in advance). And she is just as accurate as neurotypical people when posting a letter through a letter box arranged at various angles (while remaining at chance when consciously judging those angles).

It is also thought to be the dorsal visual stream that underlies the well-known phenomenon of *blindsight*, in which people with damage to primary visual cortex become blind (that is to say: incapable of conscious experience) in a portion of their visual field, while being well above chance in identifying simple letters and patterns of movement presented in their blind field.[11] For there are projections from the retina to parietal cortex via the superior colliculus (bypassing primary visual cortex), whereas there are no such subcortical projections direct to temporal cortex. Moreover, we know that the monkey Helen, who had the whole of primary visual cortex surgically removed, could nevertheless move around and pick up objects normally, while being unable to identify objects except by touch or taste.[12]

It appears to be the dorsal stream, too, that underlies habitual sensorimotor action. There is good reason to think that habitual behavior can occur without being initiated or guided by conscious experience, and we know that there are projections from the dorsal visual stream to dorsolateral striatum, which is the locus of sensorimotor control in habitual responding.[13] Indeed, as control of behavior shifts during learning from goal-oriented to habitual responding, one can trace a corresponding shift in the networks involved from those centered on the ventral striatum to those involving the dorsolateral striatum.[14] This comports well with one's everyday experience. One can engage in routine activities (like driving a car along a well-known route) without conscious awareness of the stimuli involved in the activity, and while one's conscious mind is wholly engaged with other matters.

It seems likely that sleepwalking implicates a similar dorsal route. While there are a number of different kinds of motor-involving sleep disorder, sleepwalking is the sort that takes place during the two deepest forms of sleep, in the absence of any of the rapid eye-movements (REMs) that would normally indicate the presence of dreaming. (That said, sleepwalkers do sometimes report dreamlike experiences after waking from a sleepwalking episode.[15]) Indeed, sleepwalking (and in some documented instances, even sleep *driving*) takes place with prefrontal cortex and most other regions of cortex fully suppressed (as in non-REM sleep generally). Yet the eyes are open and there is local activation of sensory cortices, the amygdala, and motor regions of cingulate cortex.[16]

Finally, in this catalog of environmentally sensitive actions that can take place in the absence of conscious experience, one should include the behavioral components of emotional states such as fear and anger. Although such emotions generally give rise to distinctive forms of conscious experience (at least in humans), the behaviors in question are

initiated directly and swiftly and in the absence of conscious experience by subcortical circuits that include especially the amygdala.[17] It is these circuits that are responsible for the fight-or-flight response, setting in train a variety of physiological changes (increased heart rate and breathing rate, and so on), activating emotion-expressive facial expressions (the fear-face, the anger-face, and so on), and initiating adaptive forms of behavior guided by the affordances of the environment (fleeing, freezing, fighting, and more). Such emotion-expressive actions will automatically run through to completion unless inhibited by executive commands. Emotionally arousing stimuli will attract one's attention, of course; hence one will soon undergo conscious experiences of the relevant events. No doubt such experiences play a role in sustaining and/or modulating one's ongoing emotions. But they aren't necessary (indeed, they aren't present) at the outset.

There are thus a wide variety of actions in the human case that display perceptual sensitivity to the environment, that may sometimes be accompanied by conscious experience, but are actually caused by perceptual states that are deeply *un*conscious. However, what about actions of the sort that *are* caused by conscious experience in humans? Would it be possible to apply Newton's Principle if we could find the same or similar behavior in animals? In the absence of a good theory of consciousness, however, it is impossible to know which those actions actually are. For example, suppose we had evidence that some animals (like humans) can engage in prospection of the future. Suppose they can solve problems by mentally rehearsing the actions open to them, sustaining the resulting images in working memory, and evaluating likely success on that basis. Still, how are we to know (in the absence of a good theory of consciousness) whether the underlying factors that make these images conscious in the human case are among those that cause the choice? It might be that only a subset of the factors that render human imagery conscious are actually involved in causing the choice; and it might be that only the latter subset is involved in reflective choosing in other animals. Indeed, it might be the case that consciousness is epiphenomenal, not in general, but in respect of the type of behavior in question.

To make the point more concrete, suppose that phenomenal consciousness depends on some threshold being crossed for the extent of informational integration of the states in question with others. (Integrated-information theories of consciousness will be considered in Section III.) It would then be possible that the images deployed by animals when engaging in prospection fall below that threshold, whereas human imagery is generally above it. Or suppose that phenomenal

consciousness depends on the presence of higher-order awareness of the states in question. (Higher-order theories of consciousness will be considered in Section V.) Then it might be just the images themselves, interacting via working memory with evaluative systems, that explains choice, not the accompanying—consciousness-determining—higher-order thoughts. Until we know what the relevant factors are underlying conscious experience, it is impossible to know. And that means we need a theory of consciousness.

It is plain that Tye is mistaken. In order to apply Newton's Principle correctly, we need a theory of the difference between conscious and unconscious forms of perception. Armed with such a theory, we can then raise (and investigate) the question whether both forms of perception are manifested in animals. This is our topic going forward.

We will focus on the set of theories that have the resources to account for the distinction between conscious and unconscious mental states (setting aside theories such as panpsychism,[18] as well as those that draw on subatomic quantum indeterminacies,[19] which seem incapable of doing so). There are four such theories that are taken seriously in the current literature. One is integrated-information theory, proposed initially by Giulio Tononi.[20] Another identifies phenomenally conscious experience with the contents of a form of fragile short-term memory distinct from working memory, proposed by Ned Block.[21] A third is higher-order thought theory, defended (in different forms) by David Rosenthal and the present author.[22] Finally, there is global-workspace theory, which has been proposed and elaborated by Bernard Baars, Michael Tye, Stanislas Dehaene, Jesse Prinz, and others.[23] We will consider and briefly evaluate these four kinds of account in turn, noting their implications for the question of animal consciousness as we go.

III. INTEGRATED INFORMATION

According to integrated-information theory, consciousness can be measured by the extent to which information is integrated in the brain via functional interdependence and complex reverberating feedback loops.[24] My discussion of this view will be brief. One immediate problem with it is that integrated information is a matter of degree, whereas (as we noted above), phenomenal consciousness is either categorically present or categorically absent. This suggests that the theory might better serve as an account of creature consciousness (whether transitive, intransitive, or both), which likewise admits of degrees. Indeed, the theory's proponents themselves tout the benefits of integrated-

information theory in accounting for partially conscious states of light anesthesia, and in accounting for specific experiences (e.g., perceptions of red as opposed to blue), which will correspond to distinct integrated-information packages.[25]

Moreover, since information integration is a graded notion, and is present at all different levels of cognition, the theory implies a limited form of panpsychism, as its proponents also note.[26] For even paradigmatically unconscious blindsight-like states will contain some degree of informational integration, as will states of subsystems within one's cerebellum and spinal cord. Indeed, even a single neuron will be to some degree phenomenally conscious, on this view, since it serves to integrate the information received via its dendrites. But these claims are wholly unmotivated. Why would anyone want to claim that activity in one's spinal cord is a little bit phenomenally conscious, except when dictated by a theory? For there are no first-person facts here that we know of that need to be explained.

Integrated-information theory *could* be developed in such a way as to capture the first-person distinction between conscious and unconscious mental states, however, at least up to a point. The first-personally available difference between conscious and unconscious perception will reduce to the distinction between states that are *highly* informationally integrated (because globally broadcast, involving reverberating information integrated among a number of different regions of prefrontal, temporal, and parietal cortices, as well primary sensory cortical regions) and those that are less so (because they have only local or specialized effects). But then it is the global broadcasting, not the informational integration as such, that explains why conscious states are available to be remembered and reported, as well as entering into planning, whereas unconscious states are not. So to the extent that the theory can account for mental-state consciousness, it only does so by piggybacking off the global-workspace view. Yet at the same time the theory entails a kind of panpsychism that is theoretically unmotivated, and does no explanatory work.

What integrated-information theory *does* do, is imply that phenomenal consciousness will be widespread across the animal kingdom. All creatures will be to some degree phenomenally conscious (as will many of their parts), including not only insects like ants and bees, but also bacteria. This may match some people's pre-theoretical intuitions. But notice that the claim is not just that all creatures have a *few* phenomenally conscious states, or are phenomenally conscious infrequently. Rather, it is that the states that exist in a bee when it detects the scent of nectar,

for example, are to some *small degree* phenomenally conscious, and are *like something* for the bee to undergo to some small extent. Such claims are hard to make sense of. For as we noted earlier, our concept of phenomenal consciousness is all-or-nothing. One wants to insist: granted the bee's discriminatory capacities may be crude, and its representation of the world highly fragmentary and indeterminate, but either the mental states of the bee when detecting nectar have *feel* and are *like something*, or they aren't. They surely can't be *to some small degree* ***like something*** for the bee.

Notice that it makes perfectly good sense to say that a bee's perception of the world—its transitive creature consciousness of the world—is a matter of degree, and is quite limited. It is surely correct that a bee is aware of *much less* of the world around it than we are. Indeed, our perception of the world is orders of magnitude richer. But as we noted in Section I, even the most fragmentary and indeterminate of conscious perceptual states (such as a vague impression of sound or of light) is nevertheless determinately phenomenally conscious. So the plausibility of degrees of transitive creature consciousness across species does nothing to support a parallel claim about degrees of phenomenal consciousness.

IV. FRAGILE SHORT-TERM MEMORY

Ned Block identifies phenomenal consciousness with the contents of a form of fragile visual short-term memory (as well as with the contents of the equivalent early sensory memory buffers in non-visual modalities).[27] These contents are said to be richer than those accessible (via working memory) for planning or reporting. Indeed, the main evidence offered in support of the short-term-memory view is that people claim to see more details in a briefly presented stimulus than they can thereafter report; but when probed on specific items with a cue after stimulus-offset, they display awareness of much more than the limited number of items that can be sustained in working memory, seemingly bearing out their subjective reports.[28] As a result, Block thinks that phenomenal consciousness should be identified with the contents of fragile-short-term memory, whereas access consciousness and reportability comprise the contents of working memory.

Critics have replied that people's intuition that they see the stimulus in rich detail can be explained away, in part by appealing to background scene statistics that are known to be swiftly computed by the visual system.[29] What is consciously experienced, on this view, isn't the

complete detailed display of letters in the stimulus. Rather, it comprises a statistical distillation of the scene, with a content somewhat along the lines of "a bunch of letter-like shapes in a grid." The detailed letters only *become* conscious when people are provided with a local post-stimulus cue, enabling attentional signals to broadcast a subset of the rich contents of fragile short-term memory. In support of such a view, we know that the statistical properties of background scenes in change-blindness experiments require very little in the way of attentional resources to be consciously experienced, in contrast with focal objects.[30]

Moreover, there is evidence that people can only report about four items even when the cue for reporting coincides with stimulus offset, obviating any need for memory.[31] People can be shown an array of colored squares for a full second, whereupon just one of the squares remains illuminated. But that square is now divided into two colored halves, one of which is the original color of the square and the other of which is drawn from elsewhere in the display. The task is just to indicate which is the original color. Participants still show the standard four-item limit (two in each hemisphere), which is also present in their neural EEG signatures that flatten off at that four-item limit during stimulus presentation. These findings seem plainly inconsistent with the short-term-memory view. If one's perception of a colored square were already conscious, then one would think that one would immediately know which side of the square had changed when half of it is replaced by a differently colored rectangle.

In addition, there is recent evidence from a number of different retro-cuing paradigms that seems inconsistent with the short-term-memory account of consciousness.[32] According to the latter, conscious experience should be time-locked to the occurrence of the stimulus, emerging swiftly through a feed-forward sweep of sensory processing, and be independent of attention. Indeed, attention is merely the gateway to access consciousness, on this view. Moreover, the reason why a cue that follows stimulus offset only enables one to access a subset of one's comparatively rich conscious experience is said to derive from strict limits on the scope of attention and working memory. But in the retro-cuing experiments there was just a single stimulus—a faint set of oriented lines (a Gabor patch) set near the limits of visual detection occurring in one of two possible locations—whose location could either be pre-cued or post-cued, either validly or invalidly. The finding was that valid post-cues increased conscious detection, as well as improving people's discrimination of the orientation of the lines. But why, on a short-term-memory account, would a post-cue improve conscious detection? It should already be conscious (contained in fragile short-

term memory), in which case post-cue attention should be irrelevant, unless it was too faint to make it into short-term memory, in which case it won't be available to be targeted by attention. The data suggest quite strongly that, in contrast with the fragile-short-term-memory theory, conscious experience is attention-dependent.

It seems there are no good reasons for thinking that the short-term-memory account of phenomenal consciousness is true, whereas there *are* good reasons for taking it to be false. But what would such a view imply about the distribution of consciousness across the animal kingdom? This question is by no means easy to answer (although Block himself takes for granted that all mammals, at any rate, would qualify).[33] This is because it is left unclear which facts about short-term memory are supposed to carry the explanatory burden. Taken in one way, the implications of the account for animal consciousness become essentially the same as those of the global-workspace theory (see Section VII), since short-term memory is partly individuated through its functional relationship with working memory, and hence with reasoning, decision-making, and verbal report. (There are multiple reverberating memory stores in the human mind that don't qualify for consciousness, on Block's account. What is distinctive of fragile short-term memory is that people *report* awareness of its contents, even if they can't describe them fully.) But taken in other ways, the implications of the short-term-memory account might be highly restrictive (if something about the neural realization of short-term memory in humans and/or apes is what explains the properties of conscious experience) or quite liberal (if the neural realizations that do the explaining are widely shared). At this point the question of the distribution of phenomenal consciousness in the animal kingdom is wide open, on a short-term-memory account.

V. HIGHER-ORDER THOUGHT THEORY

We now consider the class of higher-order theories of consciousness. These come in two basic varieties. The first claims that it is the actual presence of a higher-order thought about a perceptual state that renders that state phenomenally conscious.[34] This view faces a number of severe difficulties. One is that it is hard to know what one should say when the content of perception and the content of the higher-order thought don't align with one another.[35] What happens to one's phenomenally conscious experience if one undergoes a perception of red but *believes* that one is perceiving orange? Another difficulty is to explain how higher-order thoughts (which are fully conceptual states, note) can give rise to the distinctive fine-grained character of phenomenally conscious

experience.[36] And yet another problem is to explain why we should be entertaining multiple higher-order thoughts about our perceptual states at every moment of our waking lives.[37] On a global-workspace account, it is easy to explain why we should always be subject to some or other phenomenally conscious state. This is because the central workspace serves to focus the entire organism on the most relevant stimuli or the most relevant activities whenever one is awake.[38] But what could possibly be the point of continually entertaining higher-order thoughts about (some subset of) our current perceptual states? And notice that the answer to this question cannot be "to provide flexibility in responding." For it is the first-order contents that guide planning and decision-making, not generally the higher-order knowledge that one has such contents.

The other main variety of higher-order approach to consciousness is dual-content theory, once defended by the present author.[39] On this view, every globally broadcast perceptual or percept-like state is both a first-order representation of the world or one's own body *and* a higher-order nonconceptual representation of the first-order percept. Each globally broadcast percept with the content *red*, for example, will at the same time have the content *seeming red* or *experience of red*. This is said to be a byproduct of the fact that globally broadcast contents are available to a higher-order thought system (or "mindreading" faculty) that is *capable* of entertaining thoughts about one's own mental states, combined with the truth of some or other kind of consumer semantics. (According to the latter, the content of a given representation depends not just on its normal causes but also on what the consumer systems for that representation are apt to do with it.) Dual-content theory avoids all the difficulties that attend its actualist cousin. But it requires us to buy into a specific theory of content-determination, and there is little or no direct evidence to favor it over a first-order global-workspace view. Moreover, the arguments that might be thought to support dual-content theory over the latter do not, in reality, do so, as has recently been argued.[40] Hence it is rational to prefer the simpler form of first-order account.

The implications of higher-order thought theory for the question of consciousness in animals are likely to be quite restrictive. Even if one takes the most generous interpretation of the evidence for higher-order self-directed thought in non-human creatures, such thoughts are likely limited to other primates, as well as perhaps a few other highly social mammals and birds.[41] But arguably this interpretation is much too generous, and the data that are often taken to show self-monitoring abilities in animals are better explained in first-order terms, as I have

argued elsewhere.[42] It may be that only humans, or perhaps humans and other species of great ape, are phenomenally conscious, if either of these forms of higher-order thought approach are correct.

VI. THE GLOBAL WORKSPACE

A global-workspace account of consciousness was first developed in detail by Bernard Baars, and was originally formulated in cognitive (rather than neural-network) terms.[43] A close relative of Baars's theory (published subsequently but independently) is Michael Tye's PANIC theory.[44] (PANIC stands for Poised Abstract Nonconceptual Intentional Content.) The basic idea of both approaches is that some perceptual and perception-like contents (including visual and auditory imagery) are globally broadcast and received by a wide range of other systems in the mind, whereas others have more specialized or local uses. Conscious states are those that are broadcast to systems for reporting, planning, reasoning, decision-making, and remembering, whereas unconscious states are those that are *not* so broadcast, although they can have a variety of other roles or effects, such as online guidance of movement. In effect, the theory identifies phenomenal consciousness with access consciousness.

Subsequent work has expanded the evidence-base for global-workspace theory to include especially the neural networks involved. One important finding has been that global broadcasting in humans appears to be an all-or-nothing phenomenon. There is a step-function underlying the global workspace. Either activation levels in the neural populations in question remain below threshold, in which case no global broadcasting occurs (although there can be some additional local or specialized effects); or those activation levels hit threshold, and full global broadcasting results.[45] This means that the workspace account can explain the all-or-nothing character of phenomenal consciousness. (In contrast, fragile-short-term-memory theory can't, except by stipulating that there is a particular level of activity in the neural systems that realize short-term memory that provide a threshold for consciousness.) In this respect, then, there is a good mesh between the explaining theory and the target to be explained.

Much of the work on the neural correlates of consciousness has used minimal contrasts between conscious and unconscious conditions in combination with EEG, MEG, and/or fMRI.[46] Minimal contrasts are ones where the stimuli remain fixed but where consciousness is sometimes present, sometimes absent. One such case is binocular rivalry, where

distinct images are presented to the two eyes, only one of which is generally conscious at any one time. A close relative of binocular rivalry is continuous flash suppression, where an image (of a face, say) can be presented continuously to one eye while the other eye is presented with a flashing Mondrian-type stimulus. The latter dominates initially, while the other stimulus emerges into consciousness from the noise. Briefly presented stimuli can also be backward masked to render them unconscious, while nevertheless being processed quite deeply, extracting high-level content that can prime subsequent behavior. If the intensity and timing parameters are set correctly, it can be arranged so that a backward-masked stimulus of a given intensity and duration is only visible on 50 percent of the trials. (The so-called *inattentional blink* can be arranged similarly.)

The general finding in such experiments is that unconscious stimuli give rise to local reverberations in visual cortex and some higher-level association areas of temporal cortex, whereas stimuli that are consciously experienced give rise to much more widespread coordinated activity, involving both prefrontal and parietal cortices as well.[47] Note that the prefrontal cortex is thought to house the main hubs of the networks responsible for decision-making, reasoning, and verbal report, and that it interacts heavily with the hippocampus and medial temporal cortex in the formation of long-term memories.[48] So we appear to have a neural-network-based account of all of the main components of access consciousness, at least.

Critics have sometimes objected that the minimal-contrast methodology controls for stimulus factors but not for behavioral or downstream cognitive ones.[49] For in order to report that a stimulus is visible on a given trial, participants need to say so, or press a button to indicate so, or respond in some other way that differs from how they respond when the stimulus remains unconscious. Hence there is ongoing debate over the question whether prefrontal cortex is specifically involved in consciousness itself, as opposed to its causes and effects.[50]

All sides in these debates should accept that specific regions of occipital, temporal, and parietal cortices are critically involved in processing and representing particular sensory contents, of course, such as faces, colors, sounds, touch, and so on (that is, everyone should accept a role for fragile visual-short-term memory, and its equivalent in nonvisual sense modalities). But according to global-workspace theory, these contents need to be broadcast more broadly—and to prefrontal cortex, in particular—in order to become conscious. Notice, however, that since these contents are thought to be broadcast quite

widely, their representation in prefrontal cortex is likely to be diffuse and distributed. As a result, one can explain why prefrontal content-related activity sometimes doesn't show up in brain-imaging studies, as has recently been pointed out.[51] Likewise, the same observation can explain why the prefrontal MEG response to a briefly presented stimulus should be less predictive of conscious experience than the MEG signal emanating from visual areas.[52] For visual processing of the stimulus is concentrated in visual cortex, of course, whereas (when conscious) the resulting contents will be broadcast to multiple cortical sites, distributed in a variety of functionally distinct regions.

Moreover, studies using cellular recording techniques together with binocular flash suppression *have* shown that the content of a stimulus is represented in neuronal activity in prefrontal cortex, even under passive viewing conditions (but not when that stimulus is perceptually suppressed).[53] Although this study was conducted with monkeys, and so cannot provide direct evidence of prefrontal involvement in *phenomenal* consciousness, it does imply that the prefrontal activity observed in instances of conscious experience in humans can't just be a matter of action-preparation. For in this study no actions were required.

In addition, a recent set of experiments attempted to control for both behavioral and attentional factors.[54] Participants were first trained to use all eight fingers when indicating the spatial position of a stimulus on a clock-face, with a unique finger used for each position. They were then presented with the same stimuli under conditions of backward masking that rendered the stimulus invisible on about 50 percent of the trials, but they were nevertheless required to guess at a location on *every* trial. Since these "guesses" on invisible trials were accurate at levels well above chance, one can infer that something like blindsight was involved. The investigators then trained pattern classifiers on the resulting EEG and MEG data to determine where in the brain the information about the specific location of the stimulus was represented. The finding was that during consciously seen trials the spatial content reached much further into prefrontal cortex, and was processed much more deeply, than in blindsight cases that shared the same motor response.

In what follows I will be assuming that global-workspace theory provides the most promising account of the neural correlates of phenomenal consciousness, at least, and will be investigating its implications for the question of consciousness in nonhuman animals.

VII. THE IMPLICATIONS OF GLOBAL-WORKSPACE THEORY FOR ANIMALS

Notice, first, that although global broadcasting is all-or-nothing in humans, it will admit of degrees across species. Or more accurately (since the term "global workspace" was introduced in the first instance for the human case), processes in the minds of animals will *more or less closely resemble* human global broadcasting. Recall that globally broadcast contents are made available to a wide range of cognitive systems. They are made available for verbal reporting, to give rise to higher-order awareness, to participate in executive functions of reasoning, planning, and decision-making, as well as to long-term memory-forming systems. Availability for verbal report is unique to human beings, of course. And it seems likely that there can only be higher-order awareness of one's own mental states in a relatively small class of nonhuman creatures (perhaps only great apes, or perhaps primates more generally; or maybe extending to some other social creatures such as dolphins and elephants).[55] Moreover, what we refer to collectively as "executive function" is really a set of different capacities that are at least partly independent of one another.[56] These functions include selecting from among competing action-schemata, mentally rehearsing actions, inhibiting actions, forming intentions for the future, implementing intentions, switching between tasks, directing attention, searching memory, and modulating emotion. It seems quite likely that some of these capacities will be present in some creatures, but absent in others. In addition, all of the receiver systems for globally broadcast information in humans will admit of degrees of internal complexity across species, and will likewise involve differing degrees of conceptual richness and sophistication.

Moreover, given the numbers, complexity, and complex relationships among the set of systems to which phenomenally conscious states are broadcast in humans, it seems almost inevitable that similarities and differences in those systems across species will be complex, multifaceted, and cross-cutting. Indeed, it seems quite unlikely that there is an objective linearly ordered similarity-space waiting to be discovered. On the contrary, there will be a multidimensional similarity *network*, with the minds of some species resembling a global broadcasting architecture in some respects, whereas others resemble it in others.

What implications does all this have for the distribution of phenomenal consciousness across the animal kingdom? Much depends on whether one is a realist or an irrealist about qualia. Suppose, first, that one is

a qualia realist. In that case one thinks that global broadcasting is a theory of the neural *correlates* of consciousness (in humans), not an explanatory theory of phenomenal consciousness itself. One will think that globally broadcast representations have properties (qualia) that cannot be reduced to any combination of functional and representational ones. And in that case (I argue) there will be a fact of the matter about which animal species (if any) are phenomenally conscious—with minds that contain qualia-properties—and which (if any) are not. However, it becomes impossible to know which species these are. In contrast, if one is a qualia *ir*realist and maintains that globally broadcast nonconceptual content is what phenomenal consciousness *is*, it turns out that there is no fact of the matter about which species of animal (if any) have phenomenally conscious mental states.

Before developing these points, let me emphasize that I am now—and henceforward in this discussion—construing qualia realism rather more weakly than many people do. In particular, qualia realism is intended to encompass the views of someone like Ned Block, who denies that phenomenal properties can be reduced to any combination of functional and representational ones, while holding open that they might be reductively explained by some set of physical or neural properties. So qualia realism, as herein understood, is compatible with physicalism. The intended contrast is just with the claim that phenomenal consciousness *reduces* to globally broadcast nonconceptual content.

On a qualia realist account, then, phenomenal consciousness *doesn't* reduce to globally broadcast nonconceptual content. Rather, there is an additional kind of property that attaches to such states, and it is this that is responsible for their phenomenal properties—what they are *like*, their *feel*, and so on. One possibility is that these properties only attach to a state in the presence of the full suite of consumer systems for global broadcasting in humans (including capacities for higher-order thought and verbal reporting). In that case phenomenal consciousness will be unique to humans. Alternatively, states might acquire phenomenal properties whenever broadcast to some specific subset of those systems, thereby fixing the set of animal species that qualify as phenomenally conscious. Since qualia are real, and really attach to some states but not others, there will be a fact of the matter.

How could we learn what those facts are, however? Recall that phenomenal consciousness is first-personal. We can only really know of its presence by acquaintance in ourselves, or by report from others. So we know that other humans have phenomenally conscious states, but only because they can tell us about them, become puzzled, like us,

by the nature of the properties involved, and so on. We can learn, on this basis, that *some* of the receiver systems for global broadcasting in humans aren't necessary for consciousness. Specifically, we can learn that long-term memory is inessential, since even severely amnesic patients can talk to us about their experiences and reflect on them in the required first-personal way. But once capacities for language and higher-order thought are removed from the picture (the former uniquely human, the latter likely quite restricted in distribution across species) we inevitably lack the kind of first-person evidence to warrant belief in the presence or absence of qualia.

What if we could completely knock out capacities for language and higher-order thought on a temporary basis (perhaps using some future and more effective form of transcranial magnetic stimulation, or using some kind of invasive but temporary neural enervation of the systems involved)? If, on recovery, people reported that they continued to enjoy phenomenally conscious experience, then wouldn't that demonstrate that language and reflective thought aren't necessary for consciousness? No, it would not. For these reports would, of course, be retrospective ones, grounded in memory of one's perceptual contents while capacities for reporting were knocked out. But in order to report the contents of a memory, that memory needs to be activated and globally broadcast. It thereby becomes available to linguistic report and higher-order thought. So on the hypothesis that it is these, specifically, that are necessary for a content to acquire the properties of qualia, those memory-contents will thereby *become* phenomenally conscious, even if they hadn't been when originally undergone.

Notice that we can't resolve these issues by appeal to Newton's Principle, either. Identifying behavior that is caused by conscious experience in ourselves, and observing exactly that behavior in another animal, we cannot conclude that the animal, too, has conscious experience. For one can't tell by introspection whether it is specifically the qualia properties that play the relevant causal role in ourselves, or rather the globally broadcast nonconceptual contents to which they attach. Nor can we tell which specific components of the latter are responsible for both the behavior and the presence of qualia. It may be that one subset of the components of human global broadcasting is shared with the animal in question and causes the behavior, whereas it is another subset that is responsible for the state in question acquiring qualia.

I conclude that if one is a qualia realist, while maintaining that global-workspace theory is the best account of the neural correlates of consciousness, then the question of animal consciousness becomes

intractable. For there is no way to know which components of the workspace are sufficient for the presence of qualia.

What, then, if one is a qualia *ir*realist? What if one believes that phenomenally conscious experience just *is* globally broadcast nonconceptual content? The most plausible way to develop such a view is to combine it with the phenomenal concept strategy for dealing with the "hard problem" thought experiments. If one maintains that phenomenal concepts are higher-order acquaintance-based indexicals, grounded in one's access to the nonconceptual contents in question, then one can explain why phenomenal consciousness should seem so puzzling. For in that case one will be able to think, "There could be a being like me in all physical, functional, and representational respects who nevertheless lacks *this sort* of state." Likewise, no amount of physical, functional, or representational information about someone will entail or otherwise warrant a judgment that embeds the concept *this sort of state*, thus giving rise to an apparent explanatory gap. And so on.[57]

Suppose, then, that phenomenal consciousness reduces to globally broadcast nonconceptual content. That leaves intact, of course, the claim that the *concept* of phenomenal consciousness is a first-person one. (Indeed, this is a critical component in the phenomenal concept strategy.) So consider a particular phenomenal concept *this-R*, employed while one thinks about a visual experience of redness. The truth-condition for the claim that a species of animal capable of color vision also undergoes *phenomenal* experience of red, is then of the form: *Species X undergoes perceptual states of the same sort as* ***this-R***. If one is a realist about color qualia, then it is clear what "same sort" here amounts to: it means that the same qualia-property is present in the mind of the animal in question. But if one is a qualia irrealist, then "same sort" here can only mean a perceptual state with the same nonconceptual content entering into the same global broadcasting relation as *this-R*. And then the problem is, what counts as "sameness" of global broadcasting? In order for another animal's attended and broadcast perceptual representations to qualify as *the same*, how similar must the set of consuming systems be to those involved in human global broadcasting? It is unclear that this question admits of a factual answer.

Recall that the similarities and differences between human global broadcasting and global-broadcasting-like activity in other species of animal are likely to be complex, multifaceted, and cross-cutting. How could the question whether any of these qualify as *the same* be settled through anything other than stipulation? Someone might answer by claiming that global broadcasting of a certain sort is a natural kind,

providing a core set of components underlying the manifest differences across species. Indeed, there is some reason to think that top-down attention, and its role in providing a coordinating "center" around which human and animal cognition is organized, constitutes just such a kind. For it seems to be deeply preserved and largely homologous across multiple vertebrate species, at least.[58] But why should one think that the extension of one's first-person phenomenal concepts track such a kind? For if there is one thing that everyone has agreed on, at least since Saul Kripke's famous work on the topic,[59] it is that terms referring to conscious mental states aren't used as natural-kind terms. In contrast, it is generally agreed that our concepts for substances like water *are* natural-kind ones. Even before we knew anything about chemistry, we used the concept *water* to refer to the underlying nature or essence of the recognizable stuff that fills our lakes and rivers (H_2O). But our concepts for the felt qualities of our conscious experiences aren't like that. We don't use them with the intention of referring to whatever natural kind underlies those experiences, whatever that might turn out to be, and however that kind might be presented in other creatures. On the contrary, we mean to refer just to those felt qualities themselves.

Another possibility would be to introduce a categorical, ungraded, concept of global broadcasting, stipulating that an animal's percepts will count as globally broadcast if the processes involved are more similar to human global broadcasting than they are to any form of human *un*conscious perceptual state. But why should we think that phenomenal consciousness tracks the extension of this concept, rather than coinciding with full-blown human global broadcasting (or the extension of yet another related concept, come to that)? What we choose to say here seems to be a matter of stipulation, not discovery. Which is to say there is no antecedent fact of the matter.

Moreover, it isn't *this* categorical property (nor the "centering" natural-kind property sketched two paragraphs back) that figures in the explanation of the conscious/unconscious divide among human percepts, nor in the reductive explanation of our ensnarement by "hard problem" type thought-experiments. For what explains the conscious/unconscious divide in ourselves is that conscious contents are globally broadcast to a specific set of systems (many of which won't be present, or will only be partially present, in other animals). And the reductive explanation of the explanatory gap specifically requires a set of higher-order phenomenal concepts, together with capacities for discursive reflective reasoning, both of which are likely to be uniquely human.

Recall that we are supposing that globally broadcast nonconceptual content is what phenomenal consciousness *is*. Hence there is no special property that gets added when a perceptual state is globally broadcast—there are no qualia. As we transition from species whose cognitive architecture is quite unlike that of human global broadcasting through species whose networks are increasingly similar to our own, nothing lights up, and nothing magical appears. There is just nonconceptual content that is available to a greater range of systems, or to systems with greater internal complexity or conceptual sophistication, that is all. In the human case there is a big introspective difference between states that are conscious and states that aren't, of course, and the difference is akin to "lighting up." For globally broadcast states are ones that we are immediately aware of having, whereas un-broadcast states are ones that we aren't aware of having at all (except through third-person-type interpretative inferences). But introspective availability is just one facet of global broadcasting in humans. It isn't supposed to *constitute* phenomenal consciousness (that would turn the global workspace account into a higher-order theory, rather than a first-order one). And more important, "lighting up" is fully explained (we are supposing) by functional differences between the nonconceptual contents involved in the two types of case, not by the appearance of any new kind of property (qualia).

I conclude that if one combines global-workspace theory with qualia irrealism, then not only is there no fact of the matter about which species of animal are conscious, but the question of animal consciousness is no longer of any significant interest. There are, of course, important questions about cognitive organization across species. And it may be important to chart the similarities and differences between humans and other animals, as well as among other animal species themselves. This is, after all, the subject-matter of comparative psychology. And among the questions that comparative psychologists can (and should) address is the question of global-broadcasting-like arrangements in other species. But once we know all of those details, there is nothing further to be known. Hence the question whether those arrangements are *close enough* to the human case to qualify the animal for phenomenal consciousness isn't something that admits of a substantive (factual) answer.

VIII. CONCLUSION

I have argued that the question of animal consciousness cannot be tackled directly and a-theoretically by deploying a version of Newton's

Principle. Rather, answers to that question have to be theory dependent. Of the four theory-types we have canvassed, integrated-information theory is the least plausible, while seeming to imply that even bacteria are a little bit phenomenally conscious. Short-term-memory theory needs to be taken seriously, but faces significant difficulties; yet its implications for the question of animal consciousness are wholly unclear. Higher-order thought theories, likewise, face significant problems, while implying that the distribution of phenomenal consciousness across the animal kingdom is likely to be quite limited. Finally, I have suggested that global-workspace theory is the most promising approach. But when extended, as is generally intended, into a fully reductive account of phenomenal consciousness, then it means that there is no fact of the matter about animal consciousness. Moreover, it denudes the question of animal consciousness of all interest. If phenomenal consciousness just *is* globally broadcast nonconceptual content (and hence there are no qualia-properties), then no additional property gets added to the world as the mental architectures of nonhuman animals more and more closely resemble the human global broadcasting network, beyond the functional and representational similarities in question.

ACKNOWLEDGMENTS

Thanks to Lizzie Schechter and Bénédicte Veillet for comments on a previous version of this article, as well as to all those who participated in discussion of my presentation at the APA.

NOTES

1. D. Chalmers, *The Conscious Mind* (Oxford: Oxford University Press, 1996).

2. D. Chalmers, "Phenomenal Concepts and the Explanatory Gap," in *Phenomenal Concepts and Phenomenal Knowledge*, ed. T. Alter and S. Walter (Oxford: Oxford University Press, 2006).

3. N. Block, "A Confusion about the Function of Consciousness," *Behavioral and Brain Sciences* 18 (1995): 227–47; and "Consciousness, Accessibility, and the Mesh between Psychology and Neuroscience," *Behavioral and Brain Sciences* 30 (2007): 481–99.

4. D. Rosenthal, *Consciousness and Mind* (Oxford: Oxford University Press, 2005).

5. M. Tye, *Tense Bees and Shell-Shocked Crabs: Are Animals Conscious?* (Oxford: Oxford University Press, 2017).

6. S. Dehaene, *Consciousness and the Brain* (New York: Viking Press, 2014); and S. Marti and S. Dehaene, "Discrete and Continuous Mechanisms of Temporal Selection in Rapid Visual Streams," *Nature Communications* 8:1955 (2017), DOI: 10.1038/s41467-017-02079-x. Note however, that these studies assume that phenomenal consciousness coincides with access consciousness. Given other theories of consciousness, the facts of timing might be different. But the point made here is certainly one that should be of concern to Tye, since his own theory of consciousness involves a form of access-consciousness of the sort investigated by Dehaene and colleagues. Moreover, the fact that the timing of

conscious experience depends on which theory of consciousness is correct just further reinforces the need for theory, in this context.

7. M. Milner and M. Goodale, *The Visual Brain in Action* (Oxford: Oxford University Press, 1995).

8. For the case of audition, see J. Rauschecker and B. Tian, "Mechanisms and Streams for Processing of 'What' and 'Where' in Auditory Cortex," *Proceedings of the National Academy of Sciences* 97 (2000): 11800–806; and G. Hickok and D. Poeppel, "Dorsal and Ventral Streams: A Framework for Understanding Aspects of the Functional Anatomy of Language," *Cognition* 92 (2004): 67–99.

9. For a recent review, see M. Goodale, "How (and Why) the Visual Control of Action Differs from Visual Perception," *Proceedings of the Royal Society B* 281 (2014): 20140337.

10. Milner and Goodale, *The Visual Brain in Action*.

11. L. Weiskrantz, *Blindsight* (Oxford: Oxford University Press, 1986).

12. N. Humphrey, "Vision in a Monkey without Striate Cortex: A Case Study," *Perception* 3 (1974): 241–55. It could well be that blindsighted people don't generally act as the monkey did because they *believe* themselves to be blind. It should be noted, however, that what initially suggested to investigators that blindsight in humans might be a possibility was the observation that a blindsight patient could walk across a cluttered room much more easily than one might have expected; see Weiskrantz, *Blindsight*.

13. J. Lisman and E. Sternberg, "Habit and Nonhabit Systems for Unconscious and Conscious Behavior: Implications for Multitasking," *Journal of Cognitive Neuroscience* 25 (2013): 273–83; J. Saint-Cyr, L. Ungerleider, and R. Desimone, "Organization of Visual Cortical Inputs to the Striatum and Subsequent Outputs to the Pallido-Nigral Complex in the Monkey," *Journal of Comparative Neurology* 298 (1990): 129–56; W. Wood and D. Rünger, "Psychology of Habit," *Annual Review of Psychology*, 67 (2015): 289–314.

14. P. Redgrave, M. Rodriguez, Y. Smith, M. Rodriguez-Oroz, S. Lehericy, H. Bergman, Y. Agid, M. DeLong, and J. Obeso, "Goal-Directed and Habitual Control in the Basal Ganglia: Implications for Parkinson's Disease," *Nature Reviews Neuroscience* 11 (2010): 760–72; A. Burton, K. Nakamura, and M. Roesch, "From Ventral-Medial to Dorsal-Lateral Striatum: Neural Correlates of Reward-Guided Decision-Making," *Neurobiology of Learning and Memory* 117 (2015): 51–59.

15. D. Oudiette, L. Smaranda, M. Pottier, M-A. Buzare, A. Brion, and I. Arnulf, "Dreamlike Mentations During Sleepwalking and Sleep Terrors in Adults," *Sleep* 32 (2009): 1621–27.

16. M. Terzaghi, I. Sartori, L. Tassi, V. Rustoni, P. Proserpio, G. Lorusso, R. Manni, and L. Nobili, "Dissociated Local Arousal States Underlying Essential Clinical Features of Non-rapid Eye Movement Arousal Parasomnia: An Intracerebral Stereo-Electroencephalographic Study," *Journal of Sleep Research* 21 (2012): 502–506; and P. Januszko, S. Niemcewicz, T. Gajda, D. Wolynczyk-Gmaj, P. Justyna, N. Gmaj, T. Piotrowski, and W. Szelenberger, "Sleepwalking Episodes are Preceded by Arousal-Related Activation in the Cingulate Motor Area: EEG Current Density Imaging," *Clinical Neurophysiology* 127 (2016): 530–36. Note here, too, that the inaccessibility of percepts to prefrontal cortex during sleepwalking only shows that those percepts are unconscious if access consciousness and phenomenal consciousness are coextensive.

17. J. LeDoux, *The Emotional Brain* (New York: Simon and Schuster, 1996); and "Semantics, Surplus Meaning, and the Science of Fear," *Trends in Cognitive Sciences* 21 (2017): 303–306; J. Panksepp, *Affective Neuroscience* (Oxford: Oxford University Press, 1998).

18. G. Strawson, "Realistic Monism: Why Physicalism Entails Panpsychism," *Journal of Consciousness Studies* 13, no. 10-11 (2006): 3–31.

19. S. Hameroff and R. Penrose, "Orchestrated Reduction of Quantum Coherence in Brain Microtubules: A Model of Consciousness," *Mathematics and Computers in Simulation* 40 (1996): 453–80.

20. G. Tononi, "Consciousness as Integrated Information: A Provisional Manifesto," *Biological Bulletin* 215 (2008): 216–42; and G. Tononi and C. Koch, "Consciousness: Here, There, and Everywhere?" *Philosophical Transactions of the Royal Society B* 370 (2015): 20140167.

21. N. Block, "Consciousness, Accessibility, and the Mesh between Psychology and Neuroscience"; and "Perceptual Consciousness Overflows Cognitive Access," *Trends in Cognitive Science* 12 (2011): 567–75.

22. Rosenthal, *Consciousness and the Mind*; and P. Carruthers, *Phenomenal Consciousness* (Cambridge: Cambridge University Press, 2000).

23. B. Baars, *A Cognitive Theory of Consciousness* (New York: Cambridge University Press, 1988); *In the Theater of Consciousness* (Oxford: Oxford University Press, 1997); and "The Conscious Access Hypothesis: Origins and Recent Evidence," *Trends in Cognitive Sciences* 6 (2002): 47–52; M. Tye, *Ten Problems of Consciousness* (Cambridge, MA: MIT Press, 1995); and *Consciousness, Color, and Content* (Cambridge, MA: MIT Press, 2000); S. Dehaene, *Consciousness and the Brain*; and J. Prinz, *The Conscious Brain* (Oxford: Oxford University Press, 2012).

24. Tononi, "Consciousness as Integrated Information"; Tononi and Koch, "Consciousness."

25. Tononi and Koch, "Consciousness."

26. Ibid.

27. Block, "Consciousness, Accessibility, and the Mesh between Psychology and Neuroscience"; and "Perceptual Consciousness Overflows Cognitive Access."

28. G. Sperling, "The Information Available in Brief Visual Presentations," *Psychological Monographs: General and Applied* 74 (1960): 1–29; R. Landman, H. Spekreijse, and V. Lamme, "Large Capacity Storage of Integrated Objects before Change Blindness," *Vision Research* 43 (2003): 149–64; and I. Sligte, H-S. Scholte, and V. Lamme, "Are There Multiple Visual Short-Term Memory Stores?" *PLoS ONE* 3 (2008): e1699.

29. M. A. Cohen and D. Dennett, "Consciousness Cannot Be Separated from Function," *Trends in Cognitive Sciences* 15 (2011): 358–64; M. A. Cohen, D. Dennett, and N. Kanwisher, "What Is the Bandwidth of Perceptual Experience?" *Trends in Cognitive Sciences* 20 (2016): 324–35.

30. M. A. Cohen, G. Alvarez, and K. Nakayama, "Natural-Scene Perception Requires Attention," *Psychological Science* 22 (2011): 1165–72.

31. H. Tsubomi, K. Fukuda, K. Watanabe, and E. Vogel, "Neural Limits to Representing Objects Still Within View," *The Journal of Neuroscience* 33 (2013): 8257–63.

32. L. Thibault, R. van den Berg, P. Cavanagh, and C. Sergent, "Retrospective Attention Gates Discrete Conscious Access to Past Sensory Stimuli," *PLoS One* 11 (2016): e0148504; and C. Sergent, V. Wyart, M. Babo-Rebelo, L. Cohen, L. Naccache, and C. Tallon-Baudry, "Cueing Attention After the Stimulus Is Gone Can Retrospectively Trigger Conscious Perception," *Current Biology* 23 (2013): 150–55; Y. Xia, Y. Morimoto, and Y. Noguchi, "Retrospective Triggering of Conscious Perception By an Interstimulus Interaction," *Journal of Vision* 16, no. 3 (2016).

33. N. Block, "The Harder Problem of Consciousness," *Journal of Philosophy* 99 (2002): 391–425.

34. Rosenthal, *Consciousness and the Mind*.

35. P. Mandik, "Beware of the Unicorn: Consciousness as Being Represented and Other Things That Don't Exist," *Journal of Consciousness Studies* 161 (2009): 5–36; J. Weisberg, "Misrepresenting Consciousness," *Philosophical Studies* 154 (2011): 409–33.

36. N. Block, "The Higher-Order Approach to Consciousness Is Defunct," *Analysis* 71 (2011): 419–31.

37. Carruthers, *Phenomenal Consciousness*.

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

39. Carruthers, *Phenomenal Consciousness*; and *Consciousness: Essays from a Higher-Order Perspective* (Oxford: Oxford University Press, 2005).

40. P. Carruthers, "In Defense of First-Order Representationalism," *Journal of Consciousness Studies* 24, no. 5-6 (2017): 74–87.

41. J. D. Smith, "Studies of Uncertainty Monitoring and Metacognition in Animals and Humans," in *The Missing Link in Cognition*, ed. H. Terrace and J. Metcalfe (Oxford: Oxford University Press, 2005); J. D. Smith, M. Beran, J. Couchman, M. Coutinho, and J. Boomer, "The Curious Incident of the Capuchins," *Comparative Cognition & Behavior Reviews* 4 (2009): 61–64; and S. Shettleworth, *Fundamentals of Comparative Cognition* (Oxford: Oxford University Press, 2012).

42. P. Carruthers, "Meta-Cognition in Animals: A Skeptical Look," *Mind & Language* 23 (2008): 58–89; P. Carruthers and J. B. Ritchie, "The Emergence of Metacognition: Affect and Uncertainty in Animals," in *Foundations of Metacognition*, ed. M. Beran, J. Brandl, J. Perner, and J. Proust, 76–93 (Oxford: Oxford University Press, 2012); and P. Carruthers, "Are Epistemic Emotions Metacognitive?" *Philosophical Psychology* 30 (2017): 58–78.

43. Baars, *A Cognitive Theory of Consciousness*; *In the Theater of Consciousness*; and "The Conscious Access Hypothesis."

44. Tye, *Ten Problems of Consciousness*; and *Consciousness, Color, and Content*.

45. S. Dehaene and L. Naccache, "Towards a Cognitive Neuroscience of Consciousness: Basic Evidence and a Workspace Framework," *Cognition* 79 (2001): 1–37; A. Del Cul, S. Baillet, S. Dehaene, "Brain Dynamics Underlying the Nonlinear Threshold for Access to Consciousness," *PLOS Biology* 5 (2007): e260; Marti and Dehaene, "Discrete and Continuous Mechanisms of Temporal Selection in Rapid Visual Streams"; and B. van Vugt, B. Dagnino, D. Vartak, H. Safaai, S. Panzeri, S. Dehaene, and R. Roelfsema, R., "The Threshold for Conscious Report: Signal Loss and Response Bias in Visual and Frontal Cortex," *Science* 360, no. 6388 (2018): 537–42.

46. See Dehaene, *Consciousness and the Brain*, for a review.

47. Ibid.

48. R. Buckner, W. Kelley, and S. Petersen, "Frontal Cortex Contributes to Human Memory Formation," *Nature Neuroscience* 2 (1999): 311–14; A. Preston and H. Eichenbaum, "Interplay of Hippocampus and Prefrontal Cortex in Memory," *Current Biology* 23 (2013): R764-R773.

49. S. Frässle, J. Sommer, A. Jansen, M. Naber, and W. Einhäuser, "Binocular Rivalry: Frontal Activity Relates to Introspection and Action But Not to Perception," *The*

Journal of Neuroscience 34 (2014): 1738–47; M. Pitts, S. Metzler, and S. Hillyard, "Isolating Neural Correlates of Conscious Perception from Neural Correlates of Reporting One's Perception," *Frontiers in Psychology* 5 (2014): 1078; and M. Michel, "Methodological Artefacts in Consciousness Science," *Journal of Consciousness Studies* 24, no. 11-12 (2017): 94–117.

50. M. Boly, M. Massimini, N. Tsuchiya, B. Postle, C. Koch, and G. Tononi, "Are the Neural Correlates of Consciousness in the Front or in the Back of the Cerebral Cortex? Clinical and Neuroimaging Evidence," *Journal of Neuroscience* 37 (2017): 9603-9613; B. Odegaard, R. Knight, and H. Lau, "Should a Few Null Findings Falsify Prefrontal Theories of Conscious Perception?" *The Journal of Neuroscience* 37 (2017): 9593–9602.

51. Odegaard et al., "Should a Few Null Findings Falsify Prefrontal Theories of Conscious Perception?"

52. L. Andersen, M. Pedersen, K. Sandberg, and M. Overgaard, "Occipital MEG Activity in the Early Time Range (<300 ms) Predicts Graded Changes in Perceptual Consciousness," *Cerebral Cortex* 26 (2016): 2677–88.

53. T. Panagiotaropoulos, G. Deco, V. Kapoor, and N. Logohetis, N., "Neuronal Discharges and Gamma Oscillations Explicitly Reflect Visual Consciousness in the Lateral Prefrontal Cortex," *Neuron* 74 (2012): 924–35.

54. M. Salti, S. Monto, L. Charles, J-R. King, L. Parkkonen, and S. Dehaene, S., "Distinct Cortical Codes and Temporal Dynamics for Conscious and Unconscious Percepts," *eLife* 4 (2015): e05652.

55. Shettleworth, *Fundamentals of Comparative Cognition*; Carruthers and Ritchie, "The Emergence of Metacognition."

56. A. Miyake, N. Friedman, M. Emerson, A. Witzki, A. Howerter, and T. Wager, "The Unity and Diversity of Executive Functions and Their Contributions to Complex 'Frontal Lobe' Tasks: A Latent Variable Analysis," *Cognitive Psychology* 41 (2000): 49–100; A. Diamond, "Executive Functions," *Annual Review of Psychology* 64 (2013): 135–68.

57. P. Carruthers and B. Veillet, "The Phenomenal Concept Strategy," *Journal of Consciousness Studies* 14, no. 9-10 (2007): 212–36; and Prinz, *The Conscious Brain*, chapter 9.

58. Carruthers, *The Centered Mind*.

59. S. Kripke, *Naming and Necessity* (Oxford: Blackwell, 1980).