



A conceptual framework for consciousness

Michael S. A. Graziano^{a,1}

Edited by Michael Gazzaniga, University of California Santa Barbara College of Letters and Science, Santa Barbara, CA; received December 9, 2021; accepted March 21, 2022

This article argues that consciousness has a logically sound, explanatory framework, different from typical accounts that suffer from hidden mysticism. The article has three main parts. The first describes background principles concerning information processing in the brain, from which one can deduce a general, rational framework for explaining consciousness. The second part describes a specific theory that embodies those background principles, the Attention Schema Theory. In the past several years, a growing body of experimental evidence—behavioral evidence, brain imaging evidence, and computational modeling—has addressed aspects of the theory. The final part discusses the evolution of consciousness. By emphasizing the specific role of consciousness in cognition and behavior, the present approach leads to a proposed account of how consciousness may have evolved over millions of years, from fish to humans. The goal of this article is to present a comprehensive, overarching framework in which we can understand scientifically what consciousness is and what key adaptive roles it plays in brain function.

consciousness | attention | awareness | theory of mind | evolution

The neuroscientific study of consciousness is in a quagmire because it is gathering data on an ill-posed question. This article argues that a correct explanation is available, but requires a new conceptual structure. The evidence that supports that explanation—and there is a growing amount—makes no sense without first putting the new conceptual structure into place. Once articulated, the concept of consciousness becomes tractable and poised for empirical exploration. The aim here is to lay out the logical principles and the conceptual framework, and then to review the experimental data that support the thesis. Much of that data are new, collected in the past 2 y specifically to test the theory.

The framework described here is related to a longstanding approach that dates back at least to Dennett in 1991 (1), Nisbett and Wilson in 1977 (2), or Gazzaniga in 1970 (3). That longstanding approach has been called “illusionism” (4), although I argue the term is misleading (5). It could be called the “self-model approach.” It is mechanistic and reductionistic. Although the self-model approach is sometimes viewed as counterculture or minority, it may not actually be in the minority among scientists. The field of study may be developing a conceptual convergence, since a large cohort of philosophers and scientists have made arguments that at least partly overlap it (1–19).

The present article first briefly summarizes a typical account of consciousness, one that I would call “mystical,” to explain how the problem is traditionally framed. The

article then describes two principles of information processing in the brain, from which one can logically deduce a general framework for explaining consciousness. That framework escapes the mystical trap. Next, the article describes a specific theory that embodies those two general principles, the Attention Schema Theory (AST) (5, 20–23). Once AST is positioned in context, the article will discuss the recent, rapidly accumulating experimental evidence for the approach. Finally, the article will take up the possible evolutionary history of consciousness. The goal of the article is to present an overarching framework in which we can understand what consciousness is and what adaptive roles it plays in brain function and behavior. The theory is mechanistic enough, and focused enough on tangible, pragmatic benefits, that it may also be of interest to computer scientists and may ultimately allow for artificial versions of consciousness.

The “Problem” of Consciousness

To many people, the word “consciousness” evokes every aspect of the mind: thoughts, decisions, memories, perceptions, emotions, and especially self-knowledge. To be conscious, to many people, evokes an ability to make intelligent decisions with a knowledge of oneself as an agent in the world.

That inclusive account is not what the word has come to mean, scientifically. As an analogy, consider a bucket filled with items (5). The components listed above are the items often in the bucket, but one can also study the bucket itself. How do people have a subjective experience of any item at all, be it decision, sensory perception, or memory? What is experience? Why do we not say, like a computerized monitoring system, “That object is red,” but instead we say, “I experience redness”? And given that we have experience, why is so much of the information in the brain outside the consciousness bucket? In this article, by consciousness, I mean the property of experience, not any of the specific, constantly changing items that can be experienced. With apologies to the English language, I will sometimes use the word, “experienteness.”

Author affiliations: ^aDepartment of Psychology, Princeton University, Princeton, NJ 08544

Author contributions: M.S.A.G. wrote the paper.

The author declares no competing interest.

This article is a PNAS Direct Submission.

Copyright © 2022 the Author(s). Published by PNAS. This open access article is distributed under [Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 \(CC BY-NC-ND\)](https://creativecommons.org/licenses/by-nc-nd/4.0/).

¹Email: graziano@princeton.edu.

Published April 29, 2022.

The question of consciousness has sometimes been called the “hard problem” because, in one perspective, it seems to wear armor that protects it from explanation (24). Experience is nonphysical: one cannot physically touch it and register a reaction force, or objectively measure its mass, size, temperature, or any other physical parameter. One can measure the substrate—the neurons, synapses, and electrochemical signals—but the sensation attached to that physical process, the experientialness itself, is without physical presence. One can only have it and attest to it. There is no other direct window on it. A great many speculative theories have been proposed for the mechanism, the alchemical combination if you will, by which a physical process in the brain produces an experience. Other authors have provided a systematic categorization of many of these theories (25). There is no agreement on which, if any, of these speculations is correct, perhaps because so much of the problem remains outside the domain of measurement.

What are we left with? The brain contains vast amounts of information, constantly shifting and changing. For a minority of that information, for reasons unknown, an additional and intangible experientialness (so people think) is exuded by, or attached to, or generated from, the processing of that information. How can we possibly study such an intangible thing scientifically? Science requires measurement, and the feeling of consciousness is not publicly, objectively measurable; therefore, by definition, the task is impossible. It could also be called the mystical problem, since mysticism, by definition, concerns the unexplainable world of spirit and mind. All theories of consciousness that presuppose the existence of this hard-problem essence of experience are mystical, by this definition of the word. Moreover, the adaptive benefit of a conscious experience is unclear, and therefore the topic does not fit easily into evolution by natural selection, the framework by which we understand the rest of biology. Is consciousness functional or is it a useless epiphenomenon? Why not just have a brain that processes information and adequately controls behavior, but lacks the adjunctive feeling?

The reason for the apparent intractability of the problem, I argue, is the component of mysticism that has lured scholars (and casual lay-philosophers) away from a simpler underlying logic.

Two General Principles

Philosophy laid the foundations for all branches of science. In many branches, the philosophy was settled so long ago that scientists are not used to bothering with it anymore. But to scientists reading this article, I ask for your patience as I step through some foundational logic needed before we can create a scientific theory of consciousness. In this section, I will describe two principles from which one can deduce a general framework for explaining consciousness.

Principle 1. Information that comes out of a brain must have been in that brain.

To elaborate: Nobody can think, believe, or insist on any proposition, unless that proposition is represented by information in the brain. Moreover, that information must be in the right form and place to affect the brain systems responsible for thinking, believing, and claiming. The

principle is, in a sense, a computational conservation of information.

For example, if I believe, think, and claim that an apple is in front of me, then it is necessarily true that my brain contains information about that apple. Note, however, that an actual, physical apple is not necessary for me to think one is present. If no apple is present, I can still believe and insist that one is, although in that case I am evidently delusional or hallucinatory. In contrast, the information in my brain is necessary. Without that information, the belief, thought, and claim are impossible, no matter how many apples are actually present.

Principle 1 may seem trivial, but it is not always obvious and it is not often applied to the problem of consciousness. Here I will do so. If you are a normal human, then you believe that you have a subjective, phenomenal experience, an experience of some of the information content in your head. You believe it, you’re certain of it at an immediate, gut level, you are willing to proclaim it. That belief must obey principle 1: your brain must contain information descriptive of experientialness, or you would not be able to believe, think, or claim to have it.

We have already reached a simple realization that is absent from almost every proposed theory of consciousness. Almost all theories conflate two processes: having consciousness and believing you have it. The unspoken assumption is: The reason I believe I have consciousness is that I actually do have it. But those two items are differentiable, just as believing an apple is in front of you and having an apple in front of you are separable. You believe you have consciousness because of information in your brain that depicts you as having it. For example, when you believe you consciously experience a color, your brain must contain information not only about the color, but also about experientialness itself, otherwise, the belief and claim would be impossible. The existence of an actual feeling of consciousness inside you, associated with the color, is not necessary to explain your belief, certainty, and insistence that you have it. Instead, your belief and claim derive from information about conscious experience. If your brain did not have the information, then the belief and claim would be impossible, and you would not know what experience is, no matter how much conscious experience might or might not “really” be inside you.

You might respond, “It’s not just a belief or a claim. I definitely have a feeling itself. I know it, because I can feel it right now.”

At this point in my article, I do not want to adjudicate whether you have an actual feeling to go along with your belief in one. I will take up that question later. But I do want to point out a pitfall. This argument, “I know it because I can feel it right now,” is, in my experience, the most common argument in consciousness studies. It is used as though it were a trump card, sweeping aside all other claims. But the argument is tautological. To argue for the presence of feeling because you feel it, is to state, “it’s true because it’s true.” The most likely explanation for the persistence of this logically flawed argument is that the brain is prone to a kind of information loop. When you ask yourself whether you have a subjective feeling, you engage in a process. Your cognition accesses data that has been

automatically constructed in deeper brain systems. The data then constrain your cognitive belief, your thinking, and your answer. The presence of information about feeling is still the operative factor. Some system or systems in the brain must construct information about the nature of conscious experience, about minds and feelings, or we would not have such beliefs about ourselves and nobody would be writing articles on the topic.

I do not wish to delve too deeply into philosophy in this article. However, it may be worth pointing out that principle 1 represents a fundamental philosophical position. For centuries, a central line of philosophical thought has suggested that experience is primary and all else is secondary inference. This view has been represented in different forms over the centuries, for example by Descartes (26), Kant (27), Schopenhauer (28), and many others. By turning experience into an irreducible fundamental, the view has blocked progress in understanding consciousness. By definition, there can be no explanation of an irreducible fundamental. Here I am suggesting that this line of thought is mistaken. It has missed a step. We believe we have experience; belief derives from information. That realization closes a loop. It allows us to understand how physical systems like the brain might encode and manipulate the information that forms the basis of our beliefs, thoughts, and claims to subjective experience.

Note that principle 1 does not deny the existence of conscious experience. It says that you believe, think, claim, insist, jump up and down, and swear that you have a conscious feeling inside you, because of specific information in your brain that builds up a picture of what the conscious feeling is. The information in the brain is the proximal cause of all of that believing and behavior. Whether you have an actual conscious feeling to go along with that believing, thinking, and claiming, is a separate question, which I will take up with respect to principle 2.

Principle 2. The brain's models are never accurate.

The brain builds information sets, or models, of reality. In principle 2, the brain's models are never perfectly accurate. They reflect general properties of, but always differ substantially from, the item being modeled.

An adaptive advantage is gained by representing the world in a simplified manner that can be computed rapidly and with minimal energy. For example, colors are constructs of the brain that do not perfectly match the reality of wavelengths. In a particularly egregious gap between model and reality, white light is not represented in the visual system as a complex mixture of thousands of oscillating components, but instead as a high setting in a brightness channel and a low setting in a limited number of color channels. Though we all have learned in school that white is a mixture of all colors, no amount of intellectual knowledge represented in higher cortical areas can change how the low-level visual system models white. The model is automatic.

At its essence, principle 2 absolves us from having to explain physically incoherent mysteries just because people introspectively accept them to be true. It absolves us from having to explain how white light is physically scrubbed clean of all contaminants and purified. It isn't, as

Newton discovered in 1672 (29). The solution is that the model constructed by the visual system is not accurate. Principle 2 absolves us from having to explain how a ghostly, invisible arm can extend out of the body after an amputation (a phantom limb) (30). The solution is that the brain constructs a model of the body, and the model is not accurate. These models are not empty illusions: they are caricatures. They represent something physically real, but they are not accurate. Models never are. The brain's models are useful, adaptive, simplified, and never fully accurate, yet they form the basis of our beliefs, thoughts, and claims.

Let us apply principle 2 to the question of consciousness.

We already know from principle 1 that your brain must construct a set of information from which derives your certainty, belief, and claim that you possess an intangible experience: something insubstantial, nonmeasurable, a hard problem. By principle 2, that bundle of information is not accurate. Science is under no obligation to look for—or to explain—anything that has exactly the same properties depicted in that model.

Principles 1 and 2, together, provide a general framework for understanding consciousness. In that framework, first, the brain contains some objectively measurable, physical process: call it “process A.” (In the next section, I will discuss an especially simple theory about what process A might be). Second, the brain constructs a model, or a set of information, to monitor and represent that process A. Third, the model is not accurate. It is a simplification, missing information on granular, physical details. It should not be dismissed as an empty illusion (hence my reluctance to endorse the term “illusionism”), but would be better described as a caricature or a representation that simplifies and distorts the object it represents. Fourth, as a result of the simplified and imperfect information in that model reaching higher cognition, people believe, think, and claim to have a physically incoherent property. The property they claim to have is an intangible experientialness, the hard problem, the feeling of consciousness. Fifth, philosophers and scientists mistakenly try to discover what alchemical combination causes a feeling of consciousness to emerge.

I call this explanation the “2 + 2 approach,” because there is no wiggle room. The framework outlined here is not another speculation about a neural circumstance that might or might not magically generate the feeling of consciousness. It is not another opinion. If principle 1 and principle 2 are true, then the present approach is correct. It is time for consciousness researchers to choose between an explanation that is fundamentally magical and an explanation that is mechanistic and logical.

The 2 + 2 explanation is not a theory so much as a broad conceptual framework within which a theory can be built. A specific theory of consciousness would require answers to the following questions. What is process A, the real physical process that, when represented by an imperfect model in the brain, leads people to believe, think, and claim that they have conscious experience? What anatomical systems in the brain carry out process A, and what systems create the imperfect model of A? What is the adaptive or survival value for an animal to have A and to have a model of A? Which

species have this architecture and which species do not? These questions are scientifically approachable. All mysticism has been removed, and we are left with a solvable puzzle of neuroscience and evolution.

A Theory of Consciousness that Embodies Principles 1 and 2

AST is an especially simple way to embody principles 1 and 2 in a scientific theory. To explain it, I will first discuss the body schema, a useful concept that dates back at least a century (31–34). The brain constructs a representation, or simulation, of the body: a bundle of information, constantly recomputed, that represents the shape of the body, keeps track of movement, and makes predictions. It is probably constructed in a network of cortical areas, including the posterior parietal lobe and the motor and premotor cortex. Fig. 1A shows three functional consequences of having a body schema. First, it is necessary for the good control of movement (35–38). Second, it is involved in looking at someone else and intuitively understanding the other person’s body configurations (39, 40). Third, because higher cognition and language have some access to it, the body schema gives us at least some explicit, reportable knowledge about our own bodies. That “knowledge” is not always accurate, as in the case of a phantom limb, and is never a fully detailed or rich representation of every muscle, tendon, and bone shape. Per principle 2, the brain’s models are never fully accurate.

The central proposal of AST is that the brain constructs an attention schema. The proposal was not originally intended as an explanation of consciousness, but rather to account for the skillful endogenous control of attention that people and other primates routinely demonstrate. A fundamental principle of control engineering is that a controller benefits from a model of the item it controls (35–37). In parallel to a body schema, an attention schema could also be used to model the attention states of others, thus contributing to social cognition. Finally, an attention schema, if at least partly accessible by higher cognition and language, could contribute to common human intuitions, beliefs, and claims about the self. In specific, an attention schema should lead people to believe they have an internal essence or property that has the general characteristics and consequences of attention, a capacity to take vivid mental possession of items. For this reason, we theorized that an attention schema might result in the widespread belief that we contain conscious experience. The three major functional consequences proposed for an attention schema are illustrated in Fig. 1B, in parallel to the body schema in Fig. 1A.

Attention here refers specifically to selective attention, a process by which one set of cortical signals is enhanced and competing signals are suppressed (41–43). The enhanced signals have a much bigger impact on decision-making, memory, and behavior. Attention is most commonly studied in the domain of vision, but one can also selectively attend to a sound, a touch, or even a thought, a recalled memory, or an emotion. Whether external or internal events, anything representable in the cortex appears to be subject to the process of selective attention.

In AST, what is the informational contents of an attention schema? The schema would not simply identify the items being attended. It would depict the properties of attention itself. Partly, an attention schema would contain a state description. Because attention can vary over a high-dimensional space of external events and internal signals, and because of the graded and highly distributed nature of attention, and because of the constantly changing state of attention, a state description would be extremely complex. On top of the state description, the attention schema would include predictions about how attention is likely to change in the near future and, perhaps more crucially, predictions about how the state of attention is likely to affect decision-making, emotion, memory, and behavior. Just as for the body schema, the attention schema would lack information about the microscopic or physical underpinnings. It would lack information about neurons, synapses, signal competition, or specific pathways through the brain. Thus, any cognitive or linguistic access to an attention schema would result in people claiming to have an essentially nonphysical essence inside of them that can shift its state fluidly, take vivid mental possession of items, and empower oneself to decide and to act.

A common misunderstanding about AST is that it merely explains how people have a general, cognitive belief in a nonphysical mind, a psychological folk theory. That is certainly a part of AST, but it is only one part. AST includes: first, attention that changes moment-by-moment;

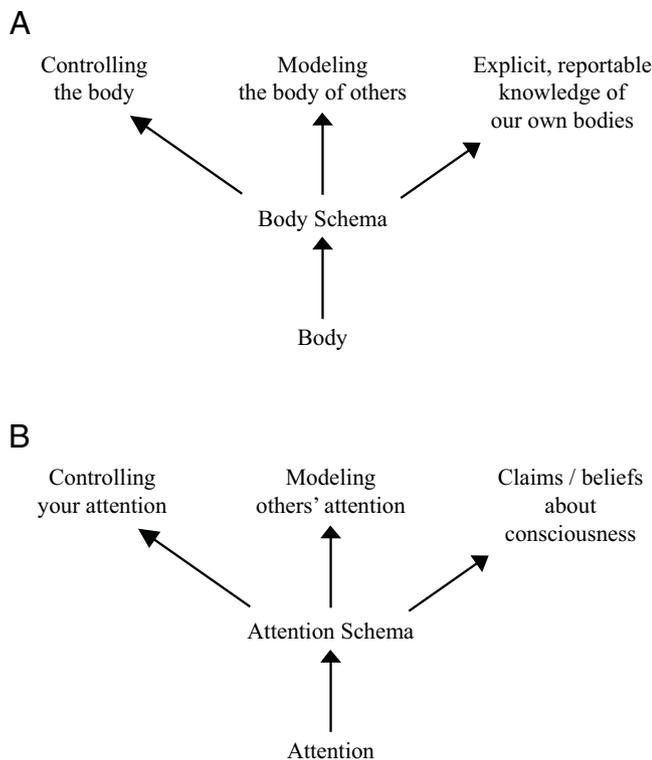


Fig. 1. Comparison between the body schema and the attention schema. (A) The body schema is a bundle of information that represents the physical body and that leads to better control of the body, better understanding of the body configurations of other people, and cognitive beliefs and linguistic claims about one’s own body. (B) The attention schema is a bundle of information that represents attention and has similar consequences as the body schema.

second, a model of attention that is automatically constructed, in the moment, reflecting the changing state of attention; third, beliefs that derive from cognitive access to the model; and fourth, output (such as linguistic output). If you focus only on the third step, you might have the false sense that AST is limited to cognitive beliefs. But AST covers more.

For example, suppose you are attending to an apple. In AST, the attentive relationship between you and the apple is represented by means of an attention schema. The proposed attention schema is just as automatic, obligatory, and moment-by-moment as the visual representation of shape, or location, or color. It is, in a sense, another represented feature of the object. As you look at and attend to the apple, a multicomponent model of the apple is constructed, in which roundness, redness, and vivid experientialness are all represented and bound together. That model provides a somewhat simplified, caricatured representation of the apple's shape, complex reflectance spectrum, and the attentive relationship between you and the apple. That model can influence higher cognition, shaping your cognitive understanding of what is happening in the moment, and perhaps ultimately shaping longer-term intellectual thoughts and beliefs. If you reduce attention to the apple, the model automatically changes. If you withdraw attention from the apple entirely, the experientialness component of the model, the component that represents your attention on the apple, disappears, and at the same time the visual components of the model fade in signal strength (since attention enhances signal strength). If you attend to the apple again, then the model is automatically rebuilt. Not only is the apple representation boosted again in signal strength, such that it can affect downstream systems around the brain, but the larger model also contains a representation of experientialness again. AST is therefore a theory of attention, a theory of the in-the-moment feature we call experience, and a theory of the more abstract, cognitive beliefs that people develop as a result.

AST is a specific way to embody principles 1 and 2. In principle 1, all intuitive certainties, beliefs, and claims stem from information in the brain. Correspondingly, in AST, the intuitive certainty that you have experientialness stems from the information in the attention schema. In principle 2, all models in the brain are inaccurate. Correspondingly, in AST, the attention schema represents attention inaccurately as a ghostly, nonphysical, mental essence or vividness: experientialness. It is a shell model of attention, a caricature of surface properties, not a representation of mechanistic details.

Evidence of an Attention Schema

The close correlation between reportable consciousness and attention was noted at least as far back as William James in 1890 (44), and has been supported by a great variety of studies since (45–47). What you are attending to, you are typically conscious of; what you are not attending to, you are typically not conscious of. Attention and consciousness, however, can dissociate in cases of weak stimuli at the threshold of detection. It is possible for a person's attention to be drawn to a visual stimulus, in the

sense that the person processes the stimulus preferentially and even responds to the stimulus, while the person reports being unconscious of the stimulus (48–58). The finding that reportable consciousness closely covaries with attention, yet can dissociate from it in some cases, is arguably the most direct indication that reportable consciousness is a result of the brain constructing an imperfect model of attention, which in turn feeds cognitive beliefs and verbal reports (right branch of Fig. 1B).

If AST is correct, then without an attention schema, the endogenous control of attention should be impaired (left branch of Fig. 1B). Recently, this control-theory prediction was tested using artificial neural-network models trained to engage in simple forms of spatial attention (59, 60). Only when given an attention schema could the networks successfully control attention. These findings confirm the basic, control-theory principle that a good control system needs a descriptive and predictive model of the item it controls. In this argument, because the human brain is good at controlling attention, it therefore must have an attention schema.

One of the most specific predictions of AST concerns the relationship between the right and left branches of Fig. 1B, between consciousness and the control of attention. Consciousness should be necessary for the control of attention. Suppose you are attending to a stimulus, but your attention schema makes an error and fails to model that state of attention. Two consequences should ensue, according to the diagram in Fig. 1B. First, you should be unable to adequately control that particular focus of attention. Second, you should be unable to report a state of conscious experience of the attended item. This correlation between consciousness and control of attention has been confirmed many times. When consciousness of a visual stimulus is absent, people are unable to sustain attention on the stimulus if it is relevant for an ongoing task (57), unable to suppress attention on it if the stimulus is a distractor (54, 57), and unable to learn to shift attention in a specific direction away from the stimulus (56, 58). A growing set of experiments therefore appears to establish a key prediction of AST: without consciousness of an item, attention on the item is still possible, but the control of attention with respect to that item almost entirely breaks down. The relationship is not "consciousness is attention"; instead, it is "consciousness is necessary for the control of attention."

AST also predicts that people construct models of other people's attention (middle branch of Fig. 1B). Ample evidence confirms that this is so. It is well established that people track the direction of gaze of others as a means to monitor attention (61–65). Even more than gaze tracking, people construct rich and multidimensional models of other people's attention (66–68). For example, people combine facial expression cues with gaze cues to reconstruct the attentional states of others (69). People also intuitively understand whether someone else's attention is exogenously captured or endogenously directed (66, 67).

One of the strangest facets of how people model the attention of others is the phenomenon of illusory eye beams. A growing set of experiments (70–73) suggests that when you look at a face staring at an object, you construct

a subthreshold motion signal falsely indicating a flow passing from the face to the object. The effect is present only if you believe the face is attending to the object. If you believe the face to be inattentive, or to be attending to something else, or if the face is blindfolded, the effect disappears. The subthreshold motion signal is enough to produce a measurable motion after effect (71) and to produce measurable brain activation in cortical area MT (middle temporal visual area), a center of motion processing (72). The illusory motion even biases people's judgments of physics: people are more likely to think an object will tip over if they see a face gazing at the object, as though the illusory eye beams are physically pushing on the object (70). All of these effects are implicit: people do not know they are generating illusory motion signals.

A possible functional value of illusory eye beams emerged when people looked at pictures containing several faces and judged which face was more attentive to an object (73). Social judgments were significantly biased by the experimental introduction of a hidden, subthreshold motion signal in the stimulus, passing from the faces to the object. The data suggest that motion signals help people to swiftly and intuitively keep track of who is attending to what. Perhaps it is something like drawing arrows on the social world, a trick for efficiently connecting sources and targets of attention. We suggest that these illusory motion signals are an example of evolutionary exaptation, when a trait that evolved for one function takes on a second, unrelated function, like when teeth (originally evolved for chewing) become enlarged for threat (social signaling). The motion-processing machinery evolved for visual perception, but may have been adapted to enhance social cognition. In this interpretation, the reason why the illusory motion signal is so weak (people do not explicitly "see" it) is because the signal evolved to be just strong enough to adaptively enhance social perception, while never becoming strong enough to harm normal vision. (Similarly, teeth may evolve to be larger in some species for social display, but not so large as to interfere with chewing.)

The phenomenon of fictitious eye beams is a good example of principle 2: the brain's models are never accurate. Here we have a model automatically constructed by the brain, a part of our social cognition, adaptive, and yet physically incoherent. No eye beam exists in reality. This phenomenon is a scientific reminder: models in the brain evolve because they lead to adaptive outcomes, not because they provide literal truth about the world. Perhaps this inaccurate but useful model of attention helps explain age-old folk intuitions about consciousness as a subtle, invisible, energy-like essence generated in or inhabiting a person, that can emanate out of the body: an aura, chi, Ka, ghost, soul, or however one calls it. AST may be able to provide useful insights into human spiritual belief, and explain some of the magical intuitions surrounding the consciousness we attribute to ourselves and to others (74).

Since it is now well established that people construct elaborate models of other people's attention, one can ask: What brain systems construct those models? Are they the same systems that model and control our own attention, and are they associated with consciousness? In brain imaging

experiments, when people reconstruct the attentional state of others, activity tends to rise most in the temporoparietal junction (TPJ) (67, 75). The activity is often bilateral, but depending on the specific paradigm, for reasons that remain unknown, sometimes the activity is left-hemisphere-biased and sometimes right-hemisphere-biased. Activity is also sometimes found in the superior temporal sulcus, the precuneus, and the medial and dorso-lateral prefrontal cortex. This distribution of activity is generally consistent with previous studies on brain networks for theory of mind (76–80).

Activity in at least some subregions of the TPJ has also been found in association with one's own attention (81–83). Moreover, TPJ activity is associated with the interaction between attention and reported consciousness (75, 84, 85). A recent study argued that this activity is consistent with error correction of a predictive model of attention (84).

Damage to the TPJ, especially on the right, is associated with the clearest specific deficit in consciousness in the clinical literature. Hemispatial neglect involves a profound loss of attentional control and of conscious experience on the opposite side of space. Stimuli on the affected side can still be processed, can influence behavior in unconscious ways, and can evoke a normal, initial amount of activity in sensory brain areas, suggesting that some exogenous attention is probably drawn to the stimuli (86–88). However, a conscious experience of anything on the affected side of space, and an endogenous control of attention toward anything on the affected side, is either severely impaired or absent. The epicenter of neglect—the brain area that, when damaged, causes the most severe form of neglect—is the right TPJ (89, 90).

Taken together, the neuroscientific evidence suggests that a cortical network, with some emphasis on the TPJ, is associated with building models of other people's attention, with modeling one's own attention, and with some aspects of the control of attention. When it is damaged, it results in a profound disruption of reported conscious experience. Presumably the same network is involved in many complex functions, but one role may be to build an attention schema, consistent with AST.

Evolution of Consciousness

If the theory described here is correct, what speculative evolutionary story can be told about consciousness?

The simplest component of attention, a competition between signals through lateral inhibition (91), almost certainly began to evolve with the first nervous systems, 600 million or more years ago. More complex forms of attention, such as the overt movement of the eyes controlled by the optic tectum, probably evolved with the first vertebrates (92, 93). Then a sophisticated, covert, selective attention that can be controlled endogenously evolved in the forebrain of vertebrates (94, 95). That complex form of attention, and an attention schema to help control it, are probably present, to some degree, in a huge range of species spanning many mammals, birds, and nonavian reptiles, all of which have an expanded forebrain. A similar mechanism may, of course, have evolved independently in other evolutionary branches, such as in octopuses (96, 97), but the evidence is not yet clear. The social ability to model

the mental contents and attention states of others evolved later among at least some mammals (especially primates) and birds (crows), and may be more widespread than the current literature suggests (62, 98–103). At a much more recent time, human ancestors evolved a cognitive and linguistic capacity such that, not only do we have models of our own and others' attention, but we can form rich cognitive and cultural beliefs based on those deeper models, and we can make verbal claims to each other about those beliefs. Finally, on the basis of those models and cognitive beliefs, philosophers and scientists have come to argue for the existence of a nonmaterial, subjective feeling that emerges from the brain and is an unsolvable mystery.

In this speculation, the components of what we call consciousness may be present in some form in a huge range of animals, including mammals, birds, and many nonavian reptiles. I am not suggesting that all of these animals are conscious in a human-like manner, but some of the same mechanisms are likely to be present. These animals must construct models of self. They construct body schemas to represent the physical self, or else their motor systems would be unable to produce coordinated movement. The speculation here is that they also construct a control model for attention, an attention schema. Obviously, humans have undergone an enormous evolutionary expansion of social ability in the past few million years. We are a hypersocial species compared to most, with an extraordinary social intelligence and capacity for manipulation of others and control over ourselves. However, in tracing the evolution of consciousness within the framework of AST—specifically, in tracing the evolution of the ability to model one's own attention and the attention of others—the cognitive components probably emerged long before our genus.

What AST Can and Cannot Explain

Here are five items that the presence of an attention schema can explain, followed by three items it cannot.

First, it can explain our expert ability to direct attention. Attention is like a skilled dancer leaping fluidly from place to place. That dance of attention allows us to perform complex tasks, deploying our resources to each new phase of a task as needed. According to the principles of control engineering, a good control of attention is possible only with a control model.

Second, an attention schema can explain how we intuitively attribute an attentive mind to others.

Third, an attention schema can explain why we believe, think, and claim to have a subjective experience attached to select items that change from moment to moment.

Fourth, an attention schema hints at the evolutionary history of consciousness, and which branches of life may have it in some form.

Fifth, an attention schema gives technologists a potential lead on constructing artificial consciousness: machines that believe, think, and claim to be conscious according to the same principles that people do, and that receive the same

computational and behavioral benefit from it. Though many of us may balk at giving machines such a capacity, the advance is probably inevitable. The future of consciousness research is not about philosophy; it is about technology.

In contrast, here are three items that an attention schema cannot explain.

First, it cannot explain most mental processes, such as how we make decisions, have emotions, or remember the past. If that is your definition of consciousness, then an attention schema does not explain it. The attention schema explains why, having made a decision, or generated an emotional state, or recalled a memory, we sometimes also believe, think, and claim that the process comes with an adjunctive feeling, a conscious experience.

Second, an attention schema cannot explain creativity. To some, the word consciousness refers to imaginative abilities outside the range of modern programmable machines. The attention schema does not explain creativity. It explains why, having engaged in an act of creativity, we sometimes believe, think, and claim that the event came with an adjunctive experience.

Third, an attention schema cannot explain how a feeling itself, an essence, a soul, a chi, a Ka, a mental energy, a ghost in the machine, or a phenomenal experience emerges from the brain. It excludes mysticism. It explains how we believe, think, and claim to have such things, but it does not posit that we actually have intangible essences or feelings inside us. If you start your search for consciousness by assuming the existence of a subjective feeling—a private component that cannot be measured and can only be felt and attested to, experientialness itself—then you are assuming the literal accuracy of an internal model. By principle 1, your conviction that you have consciousness depends on an information set in your brain. By principle 2, the brain's models are never accurate. You have accepted the literal truth of a caricature, and you will never find the answer to your ill-posed question. When the police draw a sketch of a suspect, and you set yourself the task of finding a flat man made of graphite, you will fail. Yet at the same time, if you take the opposite approach and insist that the sketch is an empty illusion, you are missing the point. Instead, understand the sketch for what it is: a schematic representation of something real. We can explain physical processes in the brain; we can explain the models constructed by the brain to represent those physical processes; we can explain the way those models depict reality in a schematic, imperfect manner; we can explain the cognitive beliefs that stem from those imperfect models; and most importantly, we can explain the adaptive, cognitive benefits served by those models. AST is not just a theory of consciousness. It is a theory of adaptive mechanisms in the brain.

Data Availability. There are no data underlying this work.

ACKNOWLEDGMENTS. This work was supported by the Princeton Neuroscience Institute Innovation Fund.

1. D. C. Dennett, *Consciousness Explained* (Little, Brown, and Co., Boston, MA, 1991).
2. R. E. Nisbett, T. D. Wilson, Telling more than we can know—Verbal reports on mental processes. *Psychol. Rev.* **84**, 231–259 (1977).
3. M. S. Gazzaniga, *The Bisected Brain* (Appleton Century Crofts, New York, 1970).

4. K. Frankish, Illusionism as a theory of consciousness. *J. Conscious. Stud.* **23**, 1–39 (2016).
5. M. S. A. Graziano, *Rethinking Consciousness: A Scientific Theory of Subjective Experience* (W. W. Norton, New York, 2019).
6. D. M. Rosenthal, Consciousness, content, and metacognitive judgments. *Conscious. Cogn.* **9**, 203–214 (2000).
7. D. Rosenthal, *Consciousness and Mind* (Oxford University Press, Oxford, UK, 2006).
8. C. Frith, Attention to action and awareness of other minds. *Conscious. Cogn.* **11**, 481–487 (2002).
9. S. J. Blackmore, Consciousness in meme machines. *J. Conscious. Stud.* **10**, 19–30 (2003).
10. O. Holland, R. Goodman, Robots with internal models: A route to machine consciousness? *J. Conscious. Stud.* **10**, 77–109 (2003).
11. T. Metzinger, *The Ego Tunnel: The Science of the Mind and the Myth of the Self* (Basic Books, New York, 2009).
12. N. Humphrey, *Soul Dust* (Princeton University Press, Princeton, NJ, 2011).
13. R. Gennaro, *The Consciousness Paradox: Consciousness, Concepts, and Higher-Order Thoughts* (The MIT Press, Cambridge, MA, 2012).
14. G. Carruthers, A metacognitive model of the sense of agency over thoughts. *Cogn. Neuropsychiatry* **17**, 291–314 (2012).
15. P. Churchland, *Touching a Nerve: Our Brains, Our Selves* (W. W. Norton, New York, 2013).
16. W. Prinz, Modeling self on others: An import theory of subjectivity and selfhood. *Conscious. Cogn.* **49**, 347–362 (2017).
17. D. J. Chalmers, The meta-problem of consciousness. *J. Conscious. Stud.* **25**, 6–61 (2018).
18. L. Blum, M. Blum, A theory of consciousness from a theoretical computer science perspective: Insights from the conscious turing machine. *arXiv [Preprint]* (2021). <https://arxiv.org/abs/2107.13704> (Accessed 10 September 2021).
19. E. Fischer, J. Sysma, Zombie intuitions. *Cognition* **215**, 104807 (2021).
20. M. S. A. Graziano, *God, Soul, Mind, Brain: A Neuroscientist's Reflections on the Spirit World* (Leapfrog Press, Teaticket, MA, 2010).
21. M. S. A. Graziano, S. Kastner, Human consciousness and its relationship to social neuroscience: A novel hypothesis. *Cogn. Neurosci.* **2**, 98–113 (2011).
22. M. S. A. Graziano, *Consciousness and the Social Brain* (Oxford University Press, New York, 2013).
23. M. S. Graziano, T. W. Webb, The attention schema theory: A mechanistic account of subjective awareness. *Front. Psychol.* **6**, 500 (2015).
24. D. J. Chalmers, Facing up to the problem of consciousness. *J. Conscious. Stud.* **2**, 200–219 (1995).
25. A. Doering, A. Schurger, M. H. Herzog, Hard criteria for empirical theories of consciousness. *Cogn. Neurosci.* **12**, 41–62 (2021).
26. R. Descartes, "Meditations on first philosophy" in *The Philosophical Writings of René Descartes*, trans. J. Cottingham, R. Stoothoff, and D. Murdoch (Cambridge University Press, Cambridge, UK, 2017). Originally published in Latin in 1641.
27. I. Kant, *Critique of Pure Reason*, trans. J. M. D. Meiklejohn, Project Gutenberg (1781). <https://www.gutenberg.org/files/4280/4280-h/4280-h.htm>. Accessed 1 October 2021.
28. A. Schopenhauer, *The World as Will and Representation*, trans. E. F. J. Payne (Dover Publications, New York, 1818, 1966).
29. I. A. Newton, Letter of Mr. Isaac Newton, Professor of the Mathematics in the University of Cambridge; Containing his new theory about light and colors: Sent by the author to the publisher from Cambridge, Febr. 6. 1671/72; In order to be communicated to the Royal Society. *Phil. Trans. Roy. Soc.* **6**, 3075–3087 (1671).
30. V. S. Ramachandran, D. Rogers-Ramachandran, Phantom limbs and neural plasticity. *Arch. Neurol.* **57**, 317–320 (2000).
31. H. Head, G. Holmes, Sensory disturbances from cerebral lesions. *Brain* **34**, 102–254 (1911).
32. M. S. A. Graziano, M. M. Botvinick, "How the brain represents the body: Insights from neurophysiology and psychology" in *Common Mechanisms in Perception and Action: Attention and Performance XIX*, W. Prinz, B. Hommel, Eds. (Oxford University Press, Oxford, UK, 2002), pp. 136–157.
33. N. P. Holmes, C. Spence, The body schema and the multisensory representation(s) of peripersonal space. *Cogn. Process.* **5**, 94–105 (2004).
34. M. S. A. Graziano, *The Spaces Between Us: A Story of Neuroscience, Evolution, and Human Nature* (Oxford University Press, New York, 2018).
35. R. C. Conant, W. R. Ashby, Every good regulator of a system must be a model of that system. *Int. J. Syst. Sci.* **1**, 89–97 (1970).
36. B. A. Francis, W. M. Wonham, The internal model principle of control theory. *Automatica* **12**, 457–465 (1976).
37. E. F. Camacho, C. Bordons Alba, *Model Predictive Control* (Springer, New York, 2004).
38. R. Shadmehr, F. A. Mussa-Ivaldi, Adaptive representation of dynamics during learning of a motor task. *J. Neurosci.* **14**, 3208–3224 (1994).
39. L. M. Parsons, Imagined spatial transformations of one's hands and feet. *Cognit. Psychol.* **19**, 178–241 (1987).
40. E. Bonda, M. Petrides, S. Frey, A. Evans, Neural correlates of mental transformations of the body-in-space. *Proc. Natl. Acad. Sci. U.S.A.* **92**, 11180–11184 (1995).
41. R. Desimone, J. Duncan, Neural mechanisms of selective visual attention. *Annu. Rev. Neurosci.* **18**, 193–222 (1995).
42. D. M. Beck, S. Kastner, Top-down and bottom-up mechanisms in biasing competition in the human brain. *Vision Res.* **49**, 1154–1165 (2009).
43. T. Moore, M. Zirnsak, Neural mechanisms of selective visual attention. *Annu. Rev. Psychol.* **68**, 47–72 (2017).
44. W. James, *Principles of Psychology* (Henry Holt and Company, New York, 1890).
45. D. J. Simons, C. F. Chabris, Gorillas in our midst: Sustained inattention blindness for dynamic events. *Perception* **28**, 1059–1074 (1999).
46. A. Mack, I. Rock, *Inattention Blindness* (MIT Press, Cambridge, MA, 2000).
47. T. Drew, M. L. Vö, J. M. Wolfe, The invisible gorilla strikes again: Sustained inattention blindness in expert observers. *Psychol. Sci.* **24**, 1848–1853 (2013).
48. P. A. McCormick, Orienting attention without awareness. *J. Exp. Psychol. Hum. Percept. Perform.* **23**, 168–180 (1997).
49. R. W. Kentridge, C. A. Heywood, L. Weiskrantz, Attention without awareness in blindsight. *Proc. Biol. Sci.* **266**, 1805–1811 (1999).
50. A. Lambert, N. Naikar, K. McLachlan, V. Aitken, A new component of visual orienting: Implicit effects of peripheral information and subthreshold cues on covert attention. *J. Exp. Psychol. Hum. Percept. Perform.* **25**, 321–340 (1999).
51. R. W. Kentridge, C. A. Heywood, L. Weiskrantz, Spatial attention speeds discrimination without awareness in blindsight. *Neuropsychologia* **42**, 831–835 (2004).
52. U. Ansorge, M. Heumann, Shifts of visuospatial attention to invisible (metacontrast-masked) singletons: Clues from reaction times and event-related potentials. *Adv. Cogn. Psychol.* **2**, 61–76 (2006).
53. Y. Jiang, P. Costello, F. Fang, M. Huang, S. He, A gender- and sexual orientation-dependent spatial attentional effect of invisible images. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 17048–17052 (2006).
54. Y. Tsushima, Y. Sasaki, T. Watanabe, Greater disruption due to failure of inhibitory control on an ambiguous distractor. *Science* **314**, 1786–1788 (2006).
55. P. Hsieh, J. T. Colas, N. Kanwisher, Unconscious pop-out: Attentional capture by unseen feature singletons only when top-down attention is available. *Psychol. Sci.* **22**, 1220–1226 (2011).
56. Z. Lin, S. O. Murray, More power to the unconscious: Conscious, but not unconscious, exogenous attention requires location variation. *Psychol. Sci.* **26**, 221–230 (2015).
57. T. W. Webb, H. H. Kean, M. S. A. Graziano, Effects of awareness on the control of attention. *J. Cogn. Neurosci.* **28**, 842–851 (2016).
58. A. I. Wilterson *et al.*, Attention control and the attention schema theory of consciousness. *Prog. Neurobiol.* **195**, 101844 (2020).
59. J. van den Boogaard, J. Treur, M. Turpijn, "A neurologically inspired neural network model for Graziano's attention schema theory for consciousness" in *International Work Conference on the Interplay Between Natural and Artificial Computation: Natural and Artificial Computation for Biomedicine and Neuroscience, Corunna, Spain, June 19–23, 2017, Part 1*, J. M. F. Vicente, H. R. Álvarez-Sánchez, F. de la Paz López, H. T. Moreo, H. Adeli, Eds. (Springer, 2017), pp. 10–21.
60. A. I. Wilterson, M. S. A. Graziano, The attention schema theory in a neural network agent: Controlling visuospatial attention using a descriptive model of attention. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2102421118 (2021).
61. A. Puce, T. Allison, S. Bentin, J. C. Gore, G. McCarthy, Temporal cortex activation in humans viewing eye and mouth movements. *J. Neurosci.* **18**, 2188–2199 (1998).
62. S. Baron-Cohen, *Mindblindness: An Essay on Autism and Theory of Mind* (MIT Press, Cambridge, MA, 1997).
63. C. K. Friesen, A. Kingstone, The eyes have it! Reflexive orienting is triggered by nonpredictive gaze. *Psychon. Bull. Rev.* **5**, 490–495 (1998).
64. A. J. Calder *et al.*, Reading the mind from eye gaze. *Neuropsychologia* **40**, 1129–1138 (2002).
65. A. Frischnen, A. P. Bayliss, S. P. Tipper, Gaze cueing of attention: Visual attention, social cognition, and individual differences. *Psychol. Bull.* **133**, 694–724 (2007).
66. A. Pesquita, C. S. Chapman, J. T. Enns, Humans are sensitive to attention control when predicting others' actions. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 8669–8674 (2016).
67. A. Guterstam, B. J. Bio, A. I. Wilterson, M. S. A. Graziano, Temporo-parietal cortex involved in modeling one's own and others' attention. *eLife* **10**, e63551 (2021).
68. B. J. Bio, T. W. Webb, M. S. A. Graziano, Projecting one's own spatial bias onto others during a theory-of-mind task. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E1684–E1689 (2018).
69. B. J. Bio, M. S. A. Graziano, Using smiles, frowns, and gaze to attribute conscious states to others: Testing part of the attention schema theory. *PsyArXiv [Preprint]* (2021) doi.org/10.31234/osf.io/7s8bk (Accessed 1 October 2021).
70. A. Guterstam, H. H. Kean, T. W. Webb, F. S. Kean, M. S. A. Graziano, Implicit model of other people's visual attention as an invisible, force-carrying beam projecting from the eyes. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 328–333 (2019).
71. A. Guterstam, M. S. A. Graziano, Implied motion as a possible mechanism for encoding other people's attention. *Prog. Neurobiol.* **190**, 101797 (2020).
72. A. Guterstam, A. I. Wilterson, D. Wachtell, M. S. A. Graziano, Other people's gaze encoded as implied motion in the human brain. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 13162–13167 (2020).
73. A. Guterstam, M. S. A. Graziano, Visual motion assists in social cognition. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 32165–32168 (2020).
74. M. S. A. Graziano, A. Guterstam, B. J. Bio, A. I. Wilterson, Toward a standard model of consciousness: Reconciling the attention schema, global workspace, higher-order thought, and illusionist theories. *Cogn. Neuropsychol.* **37**, 155–172 (2020).
75. Y. T. Kelly, T. W. Webb, J. D. Meier, M. J. Arcaro, M. S. A. Graziano, Attributing awareness to oneself and to others. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 5012–5017 (2014).
76. R. Saxe, N. Kanwisher, People thinking about thinking people. The role of the temporo-parietal junction in "theory of mind". *Neuroimage* **19**, 1835–1842 (2003).
77. U. Frith, C. D. Frith, Development and neurophysiology of mentalizing. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **358**, 459–473 (2003).
78. M. I. Gobbin, A. C. Koralek, R. E. Bryan, K. J. Montgomery, J. V. Haxby, Two takes on the social brain: A comparison of theory of mind tasks. *J. Cogn. Neurosci.* **19**, 1803–1814 (2007).
79. S. J. van Veluw, S. A. Chance, Differentiating between self and others: An ALE meta-analysis of fMRI studies of self-recognition and theory of mind. *Brain Imaging Behav.* **8**, 24–38 (2014).
80. H. Richardson, G. Lisandrelli, A. Riobueno-Naylor, R. Saxe, Development of the social brain from age three to twelve years. *Nat. Commun.* **9**, 1027 (2018).

81. A. A. Stevens, P. Skudlarski, J. C. Gatenby, J. C. Gore, Event-related fMRI of auditory and visual oddball tasks. *Magn. Reson. Imaging* **18**, 495–502 (2000).
82. J. T. Serences *et al.*, Coordination of voluntary and stimulus-driven attentional control in human cortex. *Psychol. Sci.* **16**, 114–122 (2005).
83. M. Corbetta, G. Patel, G. L. Shulman, The reorienting system of the human brain: From environment to theory of mind. *Neuron* **58**, 306–324 (2008).
84. A. I. Wilterson, S. A. Nastase, B. J. Bio, A. Guterstam, M. S. A. Graziano, Attention, awareness, and the right temporoparietal junction. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2026099118 (2021).
85. T. W. Webb, K. M. Igelström, A. Schurjer, M. S. A. Graziano, Cortical networks involved in visual awareness independent of visual attention. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 13923–13928 (2016).
86. J. C. Marshall, P. W. Halligan, Blindsight and insight in visuo-spatial neglect. *Nature* **336**, 766–767 (1988).
87. G. Rees *et al.*, Unconscious activation of visual cortex in the damaged right hemisphere of a parietal patient with extinction. *Brain* **123**, 1624–1633 (2000).
88. P. Vuilleumier *et al.*, Neural response to emotional faces with and without awareness: Event-related fMRI in a parietal patient with visual extinction and spatial neglect. *Neuropsychologia* **40**, 2156–2166 (2002).
89. G. Vallar, D. Perani, The anatomy of unilateral neglect after right-hemisphere stroke lesions. A clinical/CT-scan correlation study in man. *Neuropsychologia* **24**, 609–622 (1986).
90. V. Verdon, S. Schwartz, K.-O. Lovblad, C.-A. Hauert, P. Vuilleumier, Neuroanatomy of hemispatial neglect and its functional components: A study using voxel-based lesion-symptom mapping. *Brain* **133**, 880–894 (2010).
91. R. B. Barlow, Jr, A. J. Fraioli, Inhibition in the Limulus lateral eye in situ. *J. Gen. Physiol.* **71**, 699–720 (1978).
92. E. Knudsen, J. S. Schwartz, "The optic tectum, a structure evolved for stimulus selection" in *Evolution of Nervous Systems*, J. Kaas, Ed. (Academic Press, San Diego, 2017), pp. 387–408.
93. C. Maximino, Evolutionary changes in the complexity of the tectum of nontetrapods: A cladistic approach. *PLoS One* **3**, e3582 (2008).
94. L. Medina, A. Reiner, Do birds possess homologues of mammalian primary visual, somatosensory and motor cortices? *Trends Neurosci.* **23**, 1–12 (2000).
95. T. S. Kemp, *The Origin and Evolution of Mammals* (Oxford University Press, Oxford, UK, 2005).
96. A.-S. Darmaillacq, L. Dickel, J. A. Mather, *Cephalopod Cognition* (Cambridge University Press, Cambridge, UK, 2014).
97. D. B. Edelman, B. J. Baars, A. K. Seth, Identifying hallmarks of consciousness in non-mammalian species. *Conscious. Cogn.* **14**, 169–187 (2005).
98. H. Wimmer, J. Perner, Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition* **13**, 103–128 (1983).
99. H. W. Wellman, Theory of mind: The state of the art. *Eur. J. Dev. Psychol.* **15**, 728–755 (2018).
100. C. Krupenye, F. Kano, S. Hirata, J. Call, M. Tomasello, Great apes anticipate that other individuals will act according to false beliefs. *Science* **354**, 110–114 (2016).
101. N. S. Clayton, Ways of thinking: From crows to children and back again. *Q. J. Exp. Psychol. (Hove)* **68**, 209–241 (2015).
102. A. Horowitz, Attention to attention in domestic dog (*Canis familiaris*) dyadic play. *Anim. Cogn.* **12**, 107–118 (2009).
103. D. I. Perrett *et al.*, Visual cells in the temporal cortex sensitive to face view and gaze direction. *Proc. R. Soc. Lond. B Biol. Sci.* **223**, 293–317 (1985).