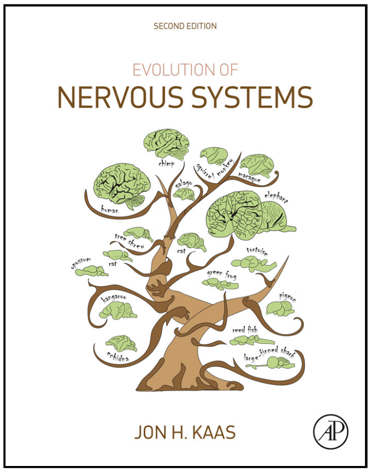

SECOND EDITION

EVOLUTION OF
NERVOUS SYSTEMS

JON H. KAAS

From Graziano, M.S.A., Webb, T.W., 2017. From Sponge to Human: The Evolution of Consciousness. In: Kaas, J (ed.), Evolution of Nervous Systems 2e. vol. 3, pp. 547–554. Oxford: Elsevier.
ISBN: 9780128040423

Academic Press

3.29 From Sponge to Human: The Evolution of Consciousness

MSA Graziano and TW Webb, Princeton University, Princeton, NJ, United States

Abstract

The attention schema theory is a proposed explanation for the brain basis of conscious experience. The theory is mechanistic, testable, and supported by at least some preliminary experiments. In the theory, subjective awareness is an internal model of attention that serves several adaptive functions. This chapter discusses the evolution of consciousness in the context of the attention schema theory, beginning with the evolution of attentional mechanisms that emerged more than half a billion years ago and extending to human consciousness and the social attribution of conscious states to others.

3.29.1 Introduction

Recently we proposed the attention schema theory, one possible explanation for how the brain constructs a subjective, conscious experience (Graziano, 2013, 2014; Graziano and Kastner, 2011; Graziano and Webb, 2014; Kelly et al., 2014; Webb and Graziano, 2015; Webb et al., 2015). To many people, consciousness is the total of what is in one's mind: memories, thoughts, decisions, and sensory perceptions. In that perspective, explaining consciousness is a matter of explaining how the brain processes a variety of information—not an easy problem to solve, but also not a fundamental mystery. However, by conscious experience we mean something else. We are not attempting to explain the content of consciousness, but instead how one gets to be conscious of that content. In today's information world, there is no mystery about how a machine can have memories, make decisions, process sensory information, contain self-information, and so on. The mystery is how the brain can have a subjective conscious experience of anything at all. The attention schema theory explains how the brain can have a conscious experience and how conscious experience is adaptive. The theory is conceptually simple, but is so far removed from how most people have considered the problem of consciousness that it is a challenge to explain. The present chapter takes an evolutionary perspective to explain this theory in a step-by-step manner.

3.29.2 Early Evolution of Attention

In the attention schema theory, consciousness evolved gradually with no sharp beginning and has precursors that go deep into evolutionary time. We start the discussion with the origin of attention (see Fig. 1). At the simplest level, attention is the selective enhancement of some signals over others such that neural processing resources are strategically deployed. Attention probably appeared very early in animal evolution, beginning in some form just after the emergence of the nervous system.

Sponges have no nervous system and according to genetic analysis may have diverged from other animals 700 million years ago or even earlier (Erwin et al., 2011). The hydra, a small relative of jellyfish, has perhaps the simplest nervous system known—a nerve net that, when stimulated anywhere, evokes a generalized response (Bode et al., 1988). Hydras apparently have little or no selective enhancement of some signals over others. The split between the ancestors of hydras and other animals may also have been as early as 700 million years ago (Erwin et al., 2011) although others suggest it is closer to 600 million years ago (Budd, 2008). Given these points of reference, nothing like attention was likely to have been present before about 700 million years ago.

Competition among neuronal signals is one of the most fundamental properties of networks of neurons. The selection of some winning signals that rise in strength and suppress other signals is a natural outcome of the interplay between excitatory input and lateral inhibition among neurons (Desimone and Duncan, 1995; Hadeler, 1974). This selective signal enhancement, therefore, should be expected in almost all animals that have nervous systems, and this appears to be the case. For example, in a classic finding, the crab eye selectively enhances visual edges and borders (Barlow and Fraioli, 1978). Crabs are arthropods, a phylum of animals believed to have diverged from the ancestors of vertebrates as early as 650 million years ago (Erwin et al., 2011). Therefore the ability for nervous systems to selectively enhance some signals over others may have appeared before the split between vertebrates and arthropods occurred.

Evolution of Nervous Systems, Second Edition, 2017, 547–554

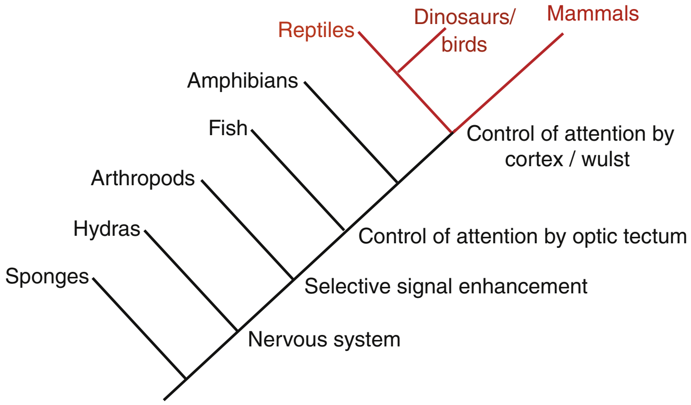
Reptiles
Dinosaurs/ birds
Mammals
Amphibians
Fish
Arthropods
Hydras
Sponges
Control of attention by cortex / wulst
Control of attention by optic tectum
Selective signal enhancement
Nervous system

Figure 1 Some aspects of the evolution of attention. In the attention schema theory, awareness evolved as an adjunct to attention, and therefore the evolution of attention is informative about the evolution of awareness. Sponges have no nervous systems. Hydras have a nervous system but little or no competition among neural signals and therefore nothing resembling attention. Arthropods have neuronal competition and selective signal enhancement, arguably a simple type of attention. Vertebrates such as fish have a centralized control of attention through the optic tectum. Reptiles, mammals, and birds have a centralized control of attention through a structure called the wulst in reptiles and birds and the cortex in mammals. Awareness as humans typically conceptualize it may be limited to this last group (shown in *red*).

The selective enhancement of signals, such as in the eye of the crab, could be thought of as a simple form of sensory attention occurring in local networks. A more complex form of attention contains another component, the ability to control that selective signal enhancement centrally: the top-down control of attention. One of the main structures that control attention in the vertebrate brain is the optic tectum (in mammals called the superior colliculus). This central brain structure integrates information from many senses and helps transform that input into complex, coordinated movement (Stein and Meredith, 1993). It emphasizes what is sometimes called overt attention, or movements that orient the eyes, head, and body toward salient stimuli. The optic tectum is found in all vertebrates in which it has been studied, including lampreys, fish, amphibians, reptiles, mammals, and birds. Therefore, it probably first evolved near the beginning of vertebrate evolution. Fossil and genetic evidence suggests that vertebrates may have emerged as a group around 520 million years ago (Erwin et al., 2011), suggesting that the central control of attention in vertebrates is extremely old, dating back to around the Cambrian explosion.

In mammals, layered over the collicular control of attention is a cortical control system. Almost all experimental work on attention focuses on the cortical control of attention in mammals, especially primates (Beck and Kastner, 2009). Mammals diverged from other animals with the evolution of synapsids around 300 million years ago (Kemp, 2005). Yet the origins of the cortex probably date back before that split. Birds and reptiles have a structure thought to be homologous to the mammalian cortex (Medina and Reiner, 2000), although its role in attentional control is unknown. Given these evolutionary relationships, it appears that the cortical control of attention may have begun at least 300 million years ago, or possibly longer.

Given this timeline, it is clear that attention is not a recent evolutionary phenomenon. It is extremely old and has evolved gradually. Selective signal enhancement probably appeared with the emergence of neurons more than half a billion years ago. Central control of attention through the optic tectum probably appeared with the evolution of vertebrates, about half a billion years ago. The cortical control of attention may have begun to emerge around or even before the origin of mammals, 300 million years ago.

3.29.3 The Attention Schema

A fundamental principle of control theory is that any effective control system needs an internal model of the thing to be controlled (Camacho and Bordons Alba, 2004). For example, the brain contains a body schema, an internal model of the body, used to help control movement (Graziano and Botvinick, 2002; Scheidt et al., 2005; Wolpert et al., 1995). This principle of an internal model has been helpful in designing guidance systems for robotic arms. Control of a physical body, however, is only one application. The same principle applies to the control of any complex dynamic process, such as the control of temperature and airflow in a building or the control of traffic in a city. All of these control systems can be enhanced by including an internal model of the thing being controlled. The model must capture the dynamics, the present state, the possible future states, and the likely consequences. When that internal model makes errors, the control system is impaired. For example, it becomes less stable and more affected by external perturbations.

In the attention schema theory, to control attention effectively the brain must contain an attention schema. The attention schema is a constantly updated set of information that provides an approximate description of attention—of the basic properties of attention, of how it shifts, of the spatial and temporal dynamics of those shifts, and of the many consequences of a shift in attention to behavior and memory. We argue that an attention schema must have evolved along with attention itself. Any nervous system that has an ability to control attention should have at least some simple internal model of attention. Some form of attention schema would therefore have been present from almost the beginning of the evolution of the nervous system more than half a billion years ago and should be present in almost every animal that has a nervous system.

There may be many different variants of attention schema implemented in different ways depending on the type of nervous system. There may even be several attention schemas within one brain. For example, in primates, there might be one internal control model associated with the collicular control of attention and a different internal control model associated with the cortical control of attention. These different forms of attention schema might contain very different information, model attention at different levels of detail or accuracy, and operate at different timescales. An internal model related to the colliculus might be more involved in modeling immediate, overt attention (the orienting of eyes, head, and body toward sensory stimuli). In contrast, an internal model related to the cortical control of attention might model a richer set of attentional phenomena including not just overt but also covert attention; attention not just to sensory signals but also to internal signals such as recalled memories and thoughts that are momentarily in the focus of attention; using attention not just to drive a stereotyped behavioral reaction but to coordinate more flexible, context-specific responses; and using attention to help store information in memory so that it can potentially shape distant future behavior. An attention schema that usefully models cortical attention would need to model these many subtle dynamics and consequences.

In the attention schema theory, the attention schema is central to conscious experience. It would not be correct, however, to suppose that every animal with an attention schema has a humanlike consciousness. The relationship between the attention schema and conscious experience is complex, probably evolved gradually, and depends on the specific mixture of information contained within the attention schema, as discussed in the next section.

3.29.4 The Relationship Between the Attention Schema and Conscious Experience

Fig. 2 illustrates the relationship between the attention schema and conscious experience in the human brain. The person in Fig. 2 is looking at an apple. His brain contains internal models of the main components of the scene. Three internal models are shown.

First, the brain contains an internal model of the visual stimulus. This internal model is a complex set of information constructed in the visual system. It includes information about shape, color, size, location, texture, and other useful features of the stimulus. The internal visual model is not perfectly accurate. The borders are exaggerated. Color is an artificial construct of the brain that roughly represents the more complex reality of a reflectance spectrum. The internal model of the apple is in this sense schematic. It is simplified. It is an incomplete, but efficient and useful set of information about the apple. It is a caricature. It is constantly recomputed as new information arrives.

This brain also contains a model of the self. The self-model is again a set of information constructed in the brain. It includes information about the physical self—the body schema, the brain's internal model of the configuration and movement of the body. It also includes information from biographical memories and general information about personality and identity and goals. The self-model is really more a collection of many models at many levels and probably spans many brain networks.

Finally, the brain diagrammed in Fig. 2 contains an internal model of the attentive relationship between the self and the apple. Like all internal models in the brain, this attention schema contains incomplete information. It is a schematic description. It may contain inaccurate, abstracted information, much like color information is an inaccurate, abstracted description of the reflectance spectrum of an object. It might describe attention in a very general or fuzzy way as a mental possession of something, reflecting the reality that attention involves a deep computational processing of information. It might describe attention as a mysterious ingredient that enables one to react to things and remember things. It might describe attention as an essence located inside oneself. It

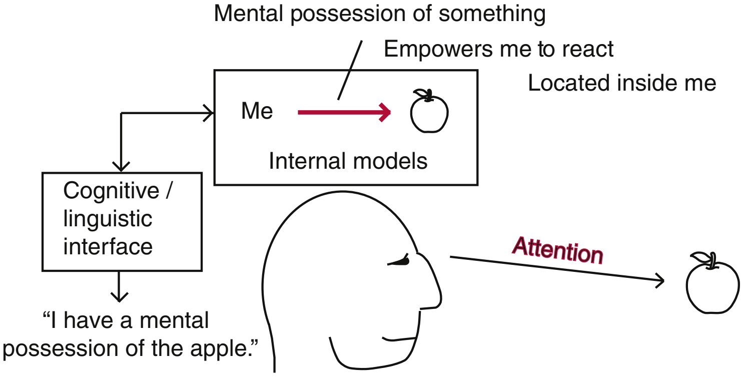
Mental possession of something
Empowers me to react
Located inside me
Me
Internal models
Cognitive / linguistic interface
"I have a mental possession of the apple."
Attention

Figure 2 The main components of the attention schema theory. The person's brain has an internal model of the self, an internal model of the apple, and a third internal model, a model of the attentive relationship between the self and the apple. Attention here refers to the brain mechanistically focusing its processing resources on the apple. The internal model of attention describes that computational relationship in an abstracted, schematic manner. It describes something impossible and physically incoherent, a caricature of attention, subjective awareness. This brain insists that it has subjective awareness because it is captive to the incomplete information in its internal models.

might contain other general, abstracted, imprecise information about attention. But this internal model would be silent about the physical mechanisms of attention—the neurons, synapses, and electrochemical signals that the brain has no need to know about.

The three internal models diagrammed in Fig. 2, a model of the apple, of the self, and of the attentive relationship between them, form a larger, overarching internal model, a set of information that captures the moment. It is the brain's simulation of the world and of the important items and events in the world at that moment. That simulation is not accurate. It does not provide a technically precise description of a brain focusing computational resources on the visual signal of an apple. Instead it provides a caricature, a distorted, simplified version of reality. In that simulation, there is a self; there is an apple; and the self has a nonphysically describable mental possession of the apple.

Fig. 2 shows another component, the "cognitive/linguistic interface." Because this is a human brain, we can query it linguistically to probe its capability. One could think of this component as a search engine. If we ask it a question, the search engine takes in the question, searches the database of the internal models, and on the basis of the available information answers the question.

We ask, "What's in front of you?"

The search engine, accessing the internal models, finds an answer. "An apple."

We probe deeper and ask, "What are the properties of the apple?"

The search engine can obtain that information as well. "It's red, round, sharper at the bottom, rounder at the top, has a stem," and so on. The search engine cannot convey every detail of the rich visual internal model, but it can abstract basic information from it and report that information.

We ask, "What are you?"

It has access to that information in its self-model. It says, "I'm a person."

Again we ask for details. "Tell us more about yourself."

It says, "I'm so tall, so wide, I can move in this and that way, I'm friendly, I come from Buffalo," and so on, as the cognitive/linguistic search engine abstracts information from its self-model.

Now we ask, "What is the mental relationship between you and the apple?"

The search engine searches the database and reports the available information, this time relying on its internal model of attention. It says, "I have a mental possession of the apple."

Once again we probe for details, asking, "What are the physical properties of this mental possession?"

The search engine is stuck. Its internal models do not describe any physical properties for that mental possession. To help in our probe, we ask, "Do you know what physical properties are?"

Here the brain can answer. It can reply, "Yes, the apple has physical properties." It gains that information from the internal model of the apple. It can say, "My body has physical properties." It gains that from the body schema, a part of its self-model. But it also replies, "My mental possession of the apple, the mental possession in and of itself, has no physical properties. It just *is*. Mental possession is nonphysically describable. It is, in that way, metaphysical. Even though it has no physical substance, it does have a location: it is an essence located inside me. It is the thing in me that *knows* something, that *experiences* something, that allows me to react to things. It is subjective, meaning that I, the subject, have it. It has an object, meaning that I direct my mental possession toward the apple. Without a mental possession of the apple, I cannot react to it. With a mental possession of the apple, I can choose to react to it or remember it for later. It's my *consciousness* of that apple."

We probe the brain further and ask, "But aren't you just a search engine accessing internal packets of information and reporting the contents of those internal models? Don't you claim to have subjective awareness of the apple just because that's what the available information describes?"

The cognitive/linguistic search engine accesses those internal models and finds no information that corresponds to that description. The internal models do not include the information, "By the way, this is an internal model, a packet of information." The question itself makes no computable sense to this machine. It answers, "No, what internal models? What information? Cannot compute. Meaningless. Nonsense. No, there is an apple, there is a me, and I have a mental possession of the apple."

We push harder. We try to explain matters to this person and say, "You aren't really subjectively aware of the apple. Subjective awareness is an impossible, physically incoherent thing that is described by a packet of information in your brain. The function of that packet of information is to keep track of attention. It models attention in a way that is accurate enough to be useful but not so accurate as to waste processing resources. Subjective awareness is a caricature of attention."

The cognitive/linguistic interface accesses those internal models and based on the incomplete information in them says, "No, cannot compute, meaningless, I am not accessing internal models. I am subjectively aware of the apple. Subjective awareness—the internal, private, metaphysical property—*is* possible, because it exists in me. Your statements are out of order. Non sequitur. Error. My consciousness of the apple exists because it does. It does because by introspection I find it."

We now have a workable explanatory theory that covers two phenomena: first, why the human brain insists that it has a subjective conscious experience, and second, why humans resists believing the explanatory theory.

The apple is of course a limited example. The theory can apply to subjective awareness of anything—a thought, an emotion, a memory, oneself, another person, a sound, warmth, cold, pain, anything that is potentially within the purview of attention. In this theory, what we call consciousness depends on one crucial component, an internal model of attention of sufficient complexity to contain the type of information discussed previously. The main advantage of the theory is that it takes consciousness entirely out of the realm of metaphysics and into the mechanistic, materialistic, and experimentally testable universe.

A useful comparison can be made to color processing. Before Newton worked out the science of color, scholars assumed that white light was pure luminance scrubbed clean of all contaminating colors. This physically incoherent, impossible property of purified luminance is described by an internal model in the brain, a construct of the early processing stages of the visual system. It is a quick-and-dirty but effective way to model an important feature of the real world. People were captive to that evolutionarily built-in model. Even now that we know the truth intellectually, we cannot turn off that internal model. It is not a matter of choice. When you look at something white, your visual system still constructs that same internal model of pure luminance stripped of color.

We are left with two ways to understand the color white. Intellectually we know it is a mixture of all colors. Introspectively, we obtain a different answer that we cannot think away—white is pure luminance. We may know that the introspective answer is wrong, but we are stuck with it because we cannot change the evolutionary past that shaped our visual systems. Just so, in the attention schema theory, we are left with two ways to understand consciousness. Intellectually, we can realize that there is no metaphysical inner experience. It is all information. Consciousness is an impossible property that is described by a packet of information in the brain. At the same time, introspectively, we arrive at the traditional answer: I have a conscious experience, therefore it must exist. We arrive at that answer because introspection is access to internal information. The theory makes logical sense of the whole phenomenon.

3.29.5 Testing Whether Awareness Is the Internal Model of Attention

The attention schema theory hinges on one proposition: subjective awareness is the brain's schematic model of attention. Moreover, in the theory, the attention schema probably evolved shortly after the evolution of attention, as a part of the mechanism that controls attention. The most basic use of an attention schema, in this hypothesis, is to help control attention. This relationship between an internal model and the thing it models can be tested experimentally.

Consider the body schema, which probably evolved as part of the mechanism for controlling the body. The body schema provides a model of the arm. That model is not always accurate. It is quite easy to introduce discrepancies between the body schema and the actual arm. When these discrepancies occur, the control of the arm is compromised in ways that are predicted by control theory (Graziano and Botvinick, 2002; Scheidt et al., 2005; Wolpert et al., 1995). The most obvious consequences are that the arm becomes less stable and more affected by external perturbations.

To test the attention schema theory experimentally, our task is to introduce discrepancies between awareness (the proposed internal model) and attention (the thing being modeled). There may be many ways that awareness can become misaligned from attention. In our experiments we focused on one particular type of misalignment, attention in the absence of awareness. The reason we focused on this particularly simple type of misalignment is that it is already established in the literature. Many previous studies have demonstrated attention to a visual stimulus in the absence of awareness of that stimulus (Hsieh et al., 2011; Jiang et al., 2006; Kentridge et al., 2008; Koch and Tsuchiya, 2007; Lamme, 2004; McCormick, 1997; Norman et al., 2013; Tsushima et al., 2006). That dissociation allows us to ask the crucial question. Does attention in the absence of awareness act like attention in the absence of an internal control model? Specifically, without awareness attention should become less stable and more affected by external perturbations.

Here we summarize one example experiment (Webb et al., 2015). The experiment concerns stability. In control theory, when the internal control model is temporarily missing, the control system is less able to maintain stability. If awareness is the internal control model for attention, then without awareness, attention should still be possible but should become less stable.

The experiment used a Posner paradigm (Posner, 1980) to measure the amount of attention drawn to a small dot. In this paradigm, first the dot is presented briefly on the right or left side of a display screen. After the dot, a test stimulus is presented on the right or left and the person must press a key to indicate the identity of the test stimulus. If the test stimulus appears on the same side as the initial dot, then reaction times to the test stimulus are typically shorter, because the dot initially and automatically drew attention to that location. If the test stimulus appears on the opposite side as the initial dot, then reaction times are typically longer, because the dot initially drew attention to the wrong location. By measuring the difference in response time between these two kinds of trials, it is possible to measure the amount of attention that was automatically drawn to the initial dot.

In our experiment, on some trials the dot was subjectively seen by subjects and on some trials the dot did not reach subjective awareness. There are a variety of ways to manipulate the subjective awareness of the dot. In the experiment summarized here, the dot was masked using a procedure called metacontrast masking. After the presentation of the dot and before the presentation of the target, two rings were presented, one centered around the dot, the other centered around the corresponding location on the opposite side of the screen. Because the mask appeared on both sides of the screen, it did not bias attention to one side. However, the mask can render the dot perceptually invisible. The exact timing of the mask determined whether the subjects were subjectively aware of the dot or not. When 50 ms were interposed between the onset of the dot and the onset of the mask, the dot became perceptually invisible to subjects, as confirmed by trial-by-trial report. When 100 ms were interposed between the onset of the dot and the onset of the mask, the dot became perceptually visible, as confirmed again by trial-by-trial report. The design therefore allowed us to measure attention to the dot when subjects were aware of the dot and when they were not aware of the dot.

In **Fig. 3**, the *Y*-axis shows the amount of attention drawn to the dot. The *X*-axis shows the time between the onset of the dot and the onset of the test stimulus. The red line shows the results for trials in which subjects were aware of the dot. A significant amount of attention was drawn to the dot and the amount of attention dropped gradually within the first 600 ms after the onset of the dot. The blue line shows the results for trials in which subjects were unaware of the dot. Attention was still drawn to the dot, but the time course of attention changed. It peaked and then dropped. At the 270 ms time point subjects were actually paying significantly more attention to the dot when they were not aware of it. By 590 ms, subjects were paying less attention to the dot when they were not

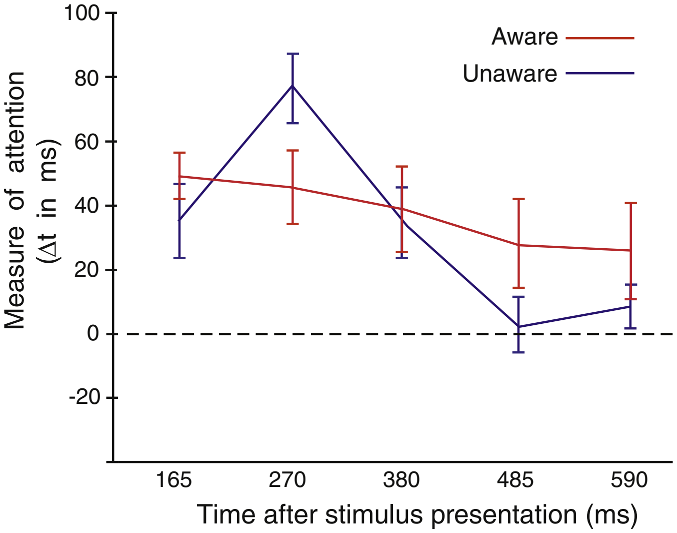

Figure 3 Testing attention with and without awareness. In this experiment, attention to a visual stimulus was tested by using the stimulus as a cue in a Posner spatial attention paradigm (see Webb et al., 2015 for details). In some trials, the participants were aware of the visual cue (*red line*). In other trials, they were unaware of it (*blue line*). Attention to the cue was less stable across time when awareness was absent. This result follows the predictions of control theory in which an internal control model helps to maintain stability of the controlled variable. The *X*-axis shows time after cue onset. The *Y*-axis shows attention drawn to the cue [$\Delta t =$ (mean response time for spatially mismatching trials in which the test target appeared on the opposite side as the initial cue) $-$ (mean response time for spatially matching trials in which the test target appeared on the same side as the initial cue)]. Error bars are standard error.

aware of it. When subjects were not aware of the visual stimulus, they could still attend to it, but attention was less stable over time. The hypothesis was confirmed. In this experiment, awareness acted in a manner consistent with the internal model of attention. Without awareness, attention was possible but less stable over time.

A growing number of experiments of this type (Webb and Graziano, 2015; Webb et al., 2015) point to a specific relationship between attention and awareness. The two are clearly not the same since they can be dissociated. Yet they are clearly not totally independent. They interact. Without awareness, attention is still possible, but the control of attention suffers. The relationship between them is consistent with the attention schema theory: awareness acts as the internal control model of attention.

3.29.6 Attributing Awareness to Others

The previous sections emphasize the role of the attention schema in low-level, automatic mechanisms that control attention. That function is obviously of great adaptive value and we suggest it originated early in the evolution of nervous systems. However, we also suggest that the attention schema took on other roles in evolutionary time and that one significant evolutionary expansion was the use of an attention schema to model the attentional states of others. Awareness may be foundational to social perception.

Determining when in evolution this additional function arose is difficult. Humans have a capacity for theory of mind, or inferring the complex mind states of others (Frith and Frith, 2003; Wimmer and Perner, 1983). Whether nonhuman animals do is controversial, but by some reports apes do (Call and Tomasello, 2008; Premack and Woodruff, 1978) and crows do (Clayton, 2015). One limitation of the work on theory of mind is that it tends to focus on relatively complex cognitive problems that may be outside the capacity of many animals. A zebra may not have the ability to infer all the contents of a lion's mind, but it may be able to attribute a more basic property to the lion: Is it aware of me? Is it aware of my calf? Attributing a state of awareness is useful not just for prey-predicting predators, but also for social interaction within a species. Dogs show an ability to intuit the awareness of other dogs (Horowitz, 2009). If the ability to attribute awareness to others is present in primates, dogs, and even crows, and if that ability has an origin in a common ancestor, then it may have first evolved at least 350 million years ago. Clearly the ability has evolved since then and in humans it appears to be particularly elaborated.

We suggest that consciousness in humans is at least as much a matter of attributing it to others as it is a matter of one's own private consciousness. By attributing consciousness to others, we do not mean an intellectual exercise or a deduction that consciousness may exist in someone else. We refer to an automatic process. It is better described as perception rather than cognition. It cannot be chosen or turned off. When projecting awareness onto someone else, we are ourselves not necessarily aware of it. We intuit that Joe is aware of the cookie, or unaware of the puddle in front of him, and we use that perception to predict his behavior and thus better interact with him. After all, Joe's behavior is dominated by the focus of his attention. If we want to predict his behavior, it would be useful to have an internal model of his attention, an attention schema that we can use to attribute awareness to him.

In the attention schema theory, attributing awareness to others and attributing awareness to oneself diverged from the same underlying brain mechanism. At least some evidence supports an overlap in the brain networks involved in those two functions. Damage to the temporoparietal junction (TPJ) can lead to neglect, a loss of awareness of objects and events on the side of space opposite to the brain lesion (Critchley, 1953; Halligan et al., 2003; Vallar, 2001; Vallar and Perani, 1986). Yet the TPJ has also been implicated in attributing mind states to others (Saxe and Kanwisher, 2003; Young et al., 2010). We recently found that a relatively dorsal and posterior subregion of the TPJ participates in attributing awareness to others (Kelly et al., 2014). When that specific region of the TPJ was disrupted, the participant's own awareness of briefly presented stimuli was also disrupted. When other, control regions of the TPJ were disrupted, stimulus detection was unaffected. These results support the hypothesis that the brain mechanisms for attributing awareness to others and to oneself overlap in the cortex.

Humans attribute consciousness to far more than just other people. We attribute consciousness to characters in a story. We attribute consciousness to puppets or dolls. Even when we know intellectually that a puppet has no mind, we still fall for the social perceptual illusion. We get angry at the coffee machine when it doesn't work. Humans have been known to attribute consciousness to storms, rivers, and even empty space. We attribute consciousness to ghosts and gods. Human spirituality is all about the tendency to perceive consciousness in almost everything around us, human or otherwise. The attention schema theory may seem on first sight like a trivial technical proposition—awareness is an attention schema. Yet when the implications of that trivial technical proposition are unfolded, it turns into a theory of consciousness and a theory of the spirit world.

3.29.7 Summary: What Living Things Are Conscious?

In the attention schema theory, conscious experience is an evolutionary outgrowth of attentional processing. Attentional processing is only present in animals that have a nervous system. Hence in this theory (to the disappointment of poets) plants, fungi, bacteria, sponges, and other living things without nervous systems do not have anything like conscious experience. Many types of animals have some simple forms of attention in the sense of networks of neurons that can selectively enhance some signals over others. That selective signal enhancement is a precursor to conscious experience, but does not itself confer conscious experience. In the present theory one could not say that the crab eye, or for that matter the human retina, has a conscious experience.

Apparently all or almost all vertebrates have a central attention controller, the optic tectum. In the theory, a central controller of attention requires an internal model of attention. That internal model, the attention schema, is a constantly updated set of information that describes at least some general, useful properties of attention. Like all internal models in the brain, it is incomplete and sometimes in error. It is, in effect, a caricature of attention. Yet having an internal model of attention does not necessarily confer conscious experience. That depends on the information content of the model. We suggest that in the case of the optic tectum, the internal model describes mainly overt attention. It presumably models the dynamics of the sensory-driven movement of eyes, head, and body. The information in that internal model would not describe anything we would recognize as conscious experience. And yet it is functionally similar. It is an evolutionary precursor to conscious experience. Thus frogs and fish have a mechanism that is functionally similar to consciousness, but most people would probably not recognize it as such.

Birds, reptiles, and mammals have a forebrain structure—the wulst in birds and reptiles, the cortex in mammals. Its role in attention has been studied extensively in mammals, especially primates. Unlike the more limited sensory-motor style of attention controlled by the optic tectum, cortical attention can be overt or covert and can be directed to external sensory events or to internal cognitive and emotional events. It can result in more complex, context-dependent reactions and can also result in information being stored in memory to potentially guide future behavior. An internal model of attention for the cortical control of attention would therefore need to be much more complex than an internal model of attention for the tectal control of attention.

In the attention schema theory, the brain constructs an attention schema to help in the cortical control of attention. That internal model depicts a caricature of attention. It is that caricature of attention that we call conscious experience. In the theory, conscious experience is an impossible, physically incoherent thing that is described by a packet of information in the brain and serves as a useful caricature of attention.

We speculate that this type of attention schema may be present in some form in birds, reptiles, and mammals. In this speculation, those animals have at least a simple form of something that most people would recognize as conscious experience.

In the attention schema theory, the attention schema also evolved a related function, the ability to model the attentional states of others. This evolutionary change may have occurred gradually. Crocodiles are highly social animals and it is possible that they have some ability to model the attentional states of others—in effect, to attribute conscious states to others. Many bird species are also highly social and might be able to model the attentional states of others. Primates are excellent at social cognition, and in humans this social use of an attention schema appears to be especially well developed and is used promiscuously. We can attribute conscious experience to almost everything, whether it is living, nonliving, or purely imaginary. The adaptive value of this ability to attribute conscious experience to others is, first, to predict the behavior of others, and second to allow for complex social cooperation.

To end with the obvious, only humans have humanlike consciousness. It is not necessarily superior to the consciousness of other animals. It is simply species-specific. In addition to the processes discussed so far, humans also have a rich cognitive layer and a linguistic capacity. We not only have that internal packet of information that describes conscious experience, but also have the capacity to cogitate and talk about it.

References

Bode, H.R., Heimfeld, S., Koizumi, O., Littlefield, C.L., Yaross, M.S., 1988. Maintenance and regeneration of the nerve net in Hydra. Am. Zool. 28, 1053–1063.

Budd, G.E., 2008. The earliest fossil record of the animals and its significance. Philos. Trans. R. Soc. Lond. B Biol. Sci. 363, 1425–1434.

Barlow Jr., R.B., Fraioli, A.J., 1978. Inhibition in the Limulus lateral eye in situ. J. Gen. Physiol. 71, 699–720.

Beck, D.M., Kastner, S., 2009. Top-down and bottom-up mechanisms in biasing competition in the human brain. Vis. Res. 49, 1154–1165.

Camacho, E.F., Bordons Alba, C., 2004. Model Predictive Control. Springer Publishing, New York.

Call, J., Tomasello, M., 2008. Does the chimpanzee have a theory of mind? 30 years later. Trends Cogn. Sci. 12, 187–192.

Clayton, N.S., 2015. Ways of thinking: from crows to children and back again. Q. J. Exp. Psychol. 68, 209–241.

Critchley, M., 1953. The Parietal Lobes. Haffner Press, London.

Desimone, R., Duncan, J., 1995. Neural mechanisms of selective visual attention. Annu. Rev. Neurosci. 18, 193–222.

Erwin, D.H., Laflamme, M., Tweedt, S.M., Sperling, E.A., Pisani, D., Peterson, K.J., 2011. The Cambrian conundrum: early divergence and later ecological success in the early history of animals. Science 334, 1091–1097.

Frith, U., Frith, C.D., 2003. Development and neurophysiology of mentalizing. Philos. Trans. R. Soc. Lond. B Biol. Sci. 358, 459–473.

Graziano, M.S.A., 2013. Consciousness and the Social Brain. Oxford University Press, New York.

Graziano, M.S.A., 2014. Speculations on the evolution of awareness. J. Cogn. Neurosci. 26, 1300–1304.

Graziano, M.S.A., Kastner, S., 2011. Human consciousness and its relationship to social neuroscience: a novel hypothesis. Cogn. Neurosci. 2, 98–113.

Graziano, M.S.A., Webb, T.W., 2014. A mechanistic theory of consciousness. Int. J. Mach. Conscious. 2 http://dx.doi.org/10.1142/S1793843014400174.

Graziano, M.S.A., Botvinick, M.M., 2002. How the brain represents the body: insights from neurophysiology and psychology. In: Prinz, W., Hommel, B. (Eds.), Common Mechanisms in Perception and Action: Attention and Performance, XIX. Oxford University Press, Oxford, pp. 136–157.

Hadeler, K., 1974. On the theory of lateral inhibition. Kybernetik 14, 161–165.

Hsieh, P., Colas, J.T., Kanwisher, N., 2011. Unconscious pop-out: attentional capture by unseen feature singletons only when top-down attention is available. Psychol. Sci. 22, 1220–1226.

Horowitz, A., 2009. Attention to attention in domestic dog (*Canis familiaris*) dyadic play. Anim. Cogn. 12, 107–118.

Halligan, P.W., Fink, G.R., Marshall, J.C., Vallar, G., 2003. Spatial cognition: evidence from visual neglect. Trends Cogn. Sci. 7, 125–133.

Jiang, Y., Costello, P., Fang, F., Huang, M., He, S., 2006. A gender- and sexual orientation-dependent spatial attentional effect of invisible images. Proc. Natl. Acad. Sci. U.S.A. 103, 17048–17052.

Kelly, Y.T., Webb, T.W., Meier, J.D., Arcaro, M.J., Graziano, M.S.A., 2014. Attributing awareness to oneself and to others. Proc. Natl. Acad. Sci. U.S.A. 111, 5012–5017.

Kemp, T.S., 2005. The Origin and Evolution of Mammals. Oxford University Press, Oxford.

Kentridge, R.W., Nijboer, T.C., Heywood, C.A., 2008. Attended but unseen: visual attention is not sufficient for visual awareness. Neuropsychologia 46, 864–869.

Koch, C., Tsuchiya, N., 2007. Attention and consciousness: two distinct brain processes. Trends Cogn. Sci. 11, 16–22.

Lamme, V.A., 2004. Separate neural definitions of visual consciousness and visual attention; a case for phenomenal awareness. Neural Netw. 17, 861–872.

Medina, L., Reiner, A., 2000. Do birds possess homologues of mammalian primary visual, somatosensory and motor cortices? Trends Neurosci. 23, 1–12.

McCormick, P.A., 1997. Orienting attention without awareness. J. Exp. Psychol. Hum. Percept. Perform. 23, 168–180.

Norman, L.J., Heywood, C.A., Kentridge, R.W., 2013. Object-based attention without awareness. Psychol. Sci. 24, 836–843.

Posner, M.I., 1980. Orienting of attention. Q. J. Exp. Psychol. 32, 3–25.

Premack, D., Woodruff, G., 1978. Does the chimpanzee have a theory of mind? Behav. Brain Sci. 1, 515–526.

Stein, B.E., Meredith, M.A., 1993. The Merging of the Senses. MIT press, Boston.

Scheidt, R.A., Conditt, M.A., Secco, E.L., Mussa-Ivaldi, F.A., 2005. Interaction of visual and proprioceptive feedback during adaptation of human reaching movements. J. Neurophysiol. 93, 3200–3213.

Saxe, R., Kanwisher, N., 2003. People thinking about thinking people: fMRI investigations of theory of mind. Neuroimage 19, 1835–1842.

Tsushima, Y., Sasaki, Y., Watanabe, T., 2006. Greater disruption due to failure of inhibitory control on an ambiguous distractor. Science 314, 1786–1788.

Vallar, G., 2001. Extrapersonal visual unilateral spatial neglect and its neuroanatomy. Neuroimage 14, S52–S58.

Vallar, G., Perani, D., 1986. The anatomy of unilateral neglect after right-hemisphere stroke lesions. A clinical/CT-scan correlation study in man. Neuropsychologia 24, 609–622.

Webb, T.W., Graziano, M.S.A., 2015. The attention schema theory: a mechanistic account of subjective awareness. Front. Psychol. http://dx.doi.org/10.3389/fpsyg.2015.00500.

Webb, T.W., Kean, H.H., Graziano, M.S.A., 2015. Effects of awareness on the control of attention. J. Cogn. Neurosci. 28 (6), 842–851.

Wolpert, D.M., Ghahramani, Z., Jordan, M.I., 1995. An internal model for sensorimotor integration. Science 269, 1880–1882.

Wimmer, H., Perner, J., 1983. Beliefs about beliefs: representation and constraining function of wrong beliefs in young children's understanding of deception. Cognition 13, 103–128.

Young, L., Dodell-Feder, D., Saxe, R., 2010. What gets the attention of the temporo-parietal junction? an fMRI investigation of attention and theory of mind. Neuropsychologia 48, 2658–2664.