International Journal of Machine Consciousness
Vol. 6, No. 2 (2014) 1−14

DOI: 10.1142/S1793843014001316

World Scientific
www.worldscientific.com

A Mechanistic Theory of Consciousness

Michael S. A. Graziano* and Taylor W. Webb†

Department of Psychology, Princeton University,
Princeton, NJ 08544, USA
**graziano@Princeton.edu*
†*taylor.w.webb@gmail.com*

Recently we proposed a theory of consciousness, the attention schema theory, based on findings in cognitive psychology and systems neuroscience. In that theory, consciousness is an internal model of attention or an "attention schema". Consciousness relates to attention in the same way that the internal model of the body, the "body schema", relates to the physical body. The body schema is used to model and help control the body. The attention schema is used to model and help regulate attention, a data-handling process in the brain in which some signals are enhanced at the expense of other signals. We proposed that attention and the attention schema co-evolved over the past half-billion years. Over that time span, the attention schema may have taken on additional functions such as promoting the integration of information across diverse domains and promoting social cognition. This paper summarizes some of the main points of the attention schema theory, suggests how a brain with an attention schema might conclude that it has a subjective awareness, and speculates that the same basic properties can be engineered into machines.

Keywords: Attention; internal model; awareness.

1. Introduction

Recently we proposed a theory of consciousness, the attention schema theory, based on findings in cognitive psychology and systems neuroscience. We argued that the theory helps to make sense of a large body of experimental work [Graziano, 2013, 2014; Graziano and Kastner, 2011; Kelly *et al.*, 2014]. If the theory is correct, it suggests that subjective experience is scientifically understandable, mechanistic, and can be artificially reconstructed. In the theory, consciousness is more than a philosophical flourish. It is one of the tools that brains use to process information. Neuroscientists will never fully understand how the brain works without understanding consciousness, and engineers will never build fully capable computers without designing them into some version of the same tool. Although the attention schema theory was formulated from the perspective of psychology and neuroscience, it might be of interest in other areas of expertise. The purpose of the present paper

is to communicate some of the main points of the theory to an audience outside neuroscience.

Consciousness is difficult to study because of its diverse connotations. To some people, consciousness is the sum total of a person's memories. To others, it is an awareness of oneself at any moment in time. Many researchers focus on the qualia of sensory events such as color or touch. Some study altered states of consciousness including dreaming or meditation. Others study pathologies of consciousness caused by brain damage. All of these aspects of consciousness are legitimate topics of study.

The approach to consciousness taken here, however, is focused on the central piece of the puzzle. How does the brain become aware of anything at all, whether it is memory, self, or sensory event? What is subjective experience? Not all information in the brain reaches awareness. Most of it does not. What makes the difference between merely *processing* information and being *aware* of it? The following sections outline the attention schema theory, introducing it through an evolutionary perspective. The paper provides only a cursory summary of some aspects the theory. A more complete exposition is provided in the book *Consciousness and the Social Brain* [Graziano, 2013].

2. The Evolution of Attention

To explain the attention schema theory, it is necessary to begin with the process of attention and the mechanisms by which it is controlled in the brain. Although these mechanistic issues may seem far removed from the more ethereal issues of awareness and subjective experience, the link between attention and awareness will hopefully become clear in the following sections.

The word "attention" is used in many ways both colloquially and scientifically. Here we use a specific, neuroscientific definition [Beck and Kastner, 2009; Desimone and Duncan, 1995]. Attention is a selection process by which some signals in the brain are enhanced in strength at the expense of other, competing signals. The boosted signals have a bigger impact on downstream systems. Those signals are more deeply processed, more likely to be stored in memory for later use, and more likely to alter behavioral output. In this definition, attention is a data-handling method.

The earliest neural networks may have lacked anything like attention. For example, hydras appear to have an undifferentiated nerve net incapable of selective signal enhancement. Hydras may have branched from other animals about 600 million years ago (MYA), though that number is not certain [Budd, 2008]. Nervous systems that use some form of selective signal enhancement can be found in almost all other animals that have been studied including crabs, flies, birds, and people [e.g., Barlow and Fraioli, 1978; Beck and Kastner, 2009; Mysore and Knudsen, 2013; van Swinderen, 2012]. These phyla and classes of animals are thought to have diverged from each other in the late Cambrian during the so-called Cambrian explosion, around 550–520 MYA. It is therefore a reasonable guess that the earliest forms of

AQ: Pls check should it be 520-550.

August 18, 2014 11:56:22am WSPC/258-IJMC 00131 ISSN: 1793-8430 Page Proof

attention evolved roughly between 600 and 520 MYA. Attention then presumably increased in complexity and sophistication in the past half billion years of evolution.

Visual attention in humans and monkeys is the most heavily studied example of attention [Beck and Kastner, 2009; Desimone and Duncan, 1995]. In the primate visual system, attention is many-layered. Competition among signals occurs within and between multiple layers of processing, including subcortical nuclei and many interconnected cortical visual areas. The competition is also biased or influenced by signals that impinge on the visual system. For example, if you are looking at a pile of change on the table, the visual representation of a dime might rise in signal strength and temporarily win the competition. One way the dime's visual representation might be boosted is if light sparkles from the dime, providing what is termed a bottom-up bias. A second way the dime's representation might be boosted is if you are engaging a cognitive process to find dimes, providing what is termed a top-down bias. This complicated interaction of competing signals and biasing signals results in a constantly shifting attentional state in which one or another visual representation wins the competition of the moment and is more fully processed.

3. The Body Schema and the Attention Schema

In the roughly half-billion-year timespan during which attention evolved, the brain presumably evolved ever more sophisticated mechanisms to control attention. The attention schema theory focuses on one aspect of this regulation of attention. To control something, it is useful to have a model or simulation of the thing to be controlled. The usefulness of an internal model is now a generally recognized principle of control engineering [Franklin *et al.*, 1989; Jacobs, 1993].

For example, the general wants to control his army. To help, he has a model army of plastic men and tanks on a map. The model is not very accurate but helps in keeping track and making predictions. Indeed a crucial aspect of a control model is that it does not need to be perfectly accurate. It can be a cartoonish, approximate depiction and still provide benefit to the control system.

A good example of a control model constructed by the brain is the body schema, or internal model of the body. It is worth outlining some of the key features of the body schema in detail because of its close relationship to the sense of self and consciousness.

Regions of the brain that span the somatosensory system, the visual system, and the motor system, integrate many sources of information to construct an internal model or simulation of the body [Graziano and Botvinick, 2002; Hwang and Shadmehr, 2005; Kawato, 1999; Wolpert *et al.*, 1995]. That model is constantly updated. It keeps track of body segments, their sizes, shapes, joint angles, speed, force, the tension on muscles, and other properties. The model can also help to make predictions a few seconds into the future.

The body schema is notoriously inaccurate in two ways.

First, the body schema lacks physical detail. It lacks information on the specific bone structure inside the body, on muscle attachment points and wrapping geometry,

on how the proteins myosin and actin bind and pull against each other to produce muscle force, and so on. The body schema contains no detailed physical or mechanistic information. It is a surface model. It depicts the surface shape of the body and a few need-to-know items such as the overall hinged structure of the limbs and the movement of joints.

Imagine that an outer space alien discovers humans but lacks access to a body for dissection. The alien, however, has a brain-reading device that can read the information contained in the body schema. The alien scientist foolishly thinks he can use the body schema to inform himself about the actual human body. Alas the alien arrives at some peculiar conclusions. He concludes that the human body is magical. It is magical in this sense — it can move in elaborate ways, but contains no internal mechanism or structure to support that movement. That is how the body schema describes the body. The body schema is intrinsically inaccurate, like a cartoon sketch.

But more than that, the body schema sometimes makes outright mistakes. A person's arm can be in one position and the body schema can register it in a different position. Dissociations between the body and the body schema are quite easy to produce and form the basis of many standard somatosensory illusions [e.g., Botvinick and Cohen, 1998; Lackner, 1988].

Why does the brain have such a sloppy model of the body? The answer is presumably a balance between cost and benefit. It takes processing time and energy, as well as neuronal space in the brain, to compute a body schema. To optimize survival, the brain needs something that can be computed fast and on the fly. It is adaptive to have a quick and dirty model as long as it is good enough to get by most of the time.

Many of the same principles evident in the body schema are theoretically transferrable to an attention schema. Because a brain has a need to control its own attention, theoretically it aught to construct a model of attention, or an attention schema. That model should be a constantly updated description of what attention is, what it means for a brain to attend to something, what the possible consequences of attention are, and what signals in particular are the focus of attention at the moment. That model is likely to be quick and dirty, lacking any detail about the neuronal mechanism of attention, and sometimes flat out wrong, but nonetheless useful as a rough model of the brain's state of attention. In the next section, we explore the psychological implications of an internal model of attention and how it may relate to subjective awareness.

4. Properties of the Attention Schema

We suggest that the attention schema gradually co-evolved with attention over the last half-billion years. Presumably the attention schema began as something quite simple and then grew in sophistication. Perhaps simple forms of an attention schema are present in flies or sea slugs. But to understand the attention schema from a psychological perspective, it is useful to consider a type of animal with a more complex brain that evolved more recently.

August 18, 2014 11:56:22am WSPC/258-IJMC 00131 ISSN: 1793-8430 Page Proof

Suppose a monkey looks at and attends to a banana. Again, by "attention", what is meant here is something mechanistic. Visual signals related to the banana win a competition in the brain and rise in strength. The stronger signals then drive downstream processes such as memory, response choice, and the sensory guidance of behavior.

If the attention schema theory is right, however, the monkey's brain does more than pay attention to the banana. It also constructs a schematic model of that state of attention. The model would require the following three chunks of information.

First, the brain must construct a model of the banana including information on its color, three-dimensional shape, location in space, and other object-defining properties. This model is probably mostly constructed in the visual system.

Second, the brain must construct a model of the monkey. Perhaps that self-model is partly the body schema.

Third, the brain must construct a model of the specific relationship between subject and object, a model of attention itself.

In this theory, the monkey's brain constructs a large, multi-part, internal model that says in effect, "There is a me, there is a banana in front of me, and in specific I am paying attention to that banana." The internal model of attention must link together something like that information.

A monkey has some capacity for higher cognition. When his higher cognition receives information from that internal model, what does it learn? Cognition is only as well-informed as the internal models that feed into it. It can do no better than that. In a sense, cognition is captive to the brain's internal models. Higher cognition is like the space-alien scientist noted in the last section, the one that gains information about the physical body only by accessing the incomplete information in the body schema, and therefore mistakenly concludes that the body is magical. The monkey's higher cognition gains information about the state of attention only by accessing the incomplete information in the attention schema.

The attention schema would certainly not describe attention in a physically accurate way. The model would lack any of the mechanistic details of neurons and signal competition. The monkey has no need to know that it has neurons and signals, synapses or neurotransmitters. Instead the model would contain sketchy and superficial information about attention. It would describe attention as a magical state of knowing. Here, we mean "magical" in the sense used in the previous section: A process that lacks any physical or mechanistic basis. The model would depict a state of knowing without any physical basis for that knowing.

The model would depict that magical state of knowing as hovering inside the body. It is a part of the monkey's own self, wedded to his body schema. The model would also depict some of the basic implications of that magical state of knowing: It implies an ability to choose to act on the banana, and an ability to remember the banana for future reference.

An attention schema would depict a mental possession or subjective *experience* of the banana. It is useful to keep in mind the meaning of the word "subjective". There

is a subject, a me. There is an object, the banana. And there is a relationship between the two: The subject has mental possession of the object and thus is enabled to act in certain ways with respect to the object.

When that monkey's higher cognition introspects, or accesses the data in that internal model, the data informs it that there is a self and the self has a subjective awareness, or experience, of the banana in front of it. The monkey's cognition has no means to doubt this information. Nothing tells it that this information comes from an inner construct. Nothing tells it that the construct is a cartoonish depiction of something else. Nothing tells it that it is being fed any information at all. Higher cognition learns only that subjective experience exists, is here, is inside, and has possessed that banana.

The monkey is aware of the banana.

The theory is of course not specific to bananas. It works as well for a sound or a touch, a memory or a thought. The monkey attends to item X. The monkey also constructs an internal model of that state of attention. If higher cognition accesses that internal model, it is informed that there is a self and the self has a subjective awareness of X.

This account of awareness arguably has a certain inevitability to it. Brains engage in attention. To control attention, in control theory, there aught to be an internal model of it, or an attention schema. That attention schema would necessarily leave out the physical details. It would depict a state of *knowing* that is non-physical, without mechanism. And higher cognition would be captive to that internal model. The creature would be certain that it has subjective awareness and would have no basis for understanding the true source of that certainty. The theory explains how a brain can arrive at the conclusion that it is aware of something without even knowing that it has arrived at a conclusion or that the conclusion derives from computation. This account is in many ways similar to the account of Gazzaniga [1970] in which awareness is a self-explanatory narrative. It is also similar to the account of Dennett [1992] in which ineffable experience is replaced by computation.

In the attention schema theory, awareness is not an illusion. It is better described as a caricature. A caricature is a distorted depiction of something real. The process of attention does physically exist. The brain's model of it, however, is not entirely accurate, and therefore introspection gives us a distorted understanding of attention that we report as an ethereal awareness.

5. The Relationship Between Awareness and Attention

If the theory is correct, then awareness and attention should relate to each other in the following three ways.

First, awareness and attention should usually covary. If you are attending to something, then in most circumstances you should also be aware of it. This match between awareness and attention is indeed usually present [Posner, 1994; Merikle and

August 18, 2014 11:56:22am WSPC/258-IJMC 00131 ISSN: 1793-8430 Page Proof

Joordens, 1997; Mack and Rock, 1998; Mole, 2008; De Brigard and Prinz, 2010; Prinz, 2011].

Second, awareness should differ from attention in certain key ways. Just as the body schema can sometimes become misaligned from the body due to inaccuracies inherent in any internal model, awareness should sometimes become misaligned from attention. It should be possible to pay attention to something by all physiological measures and yet fail to be aware of it. Many studies have now confirmed that indeed it is possible to pay attention to an item and yet have no reportable awareness of it [Baars, 1997; McCormick, 1997; Kentridge *et al.*, 1999; Lambert *et al.*, 1999; Ivanoff and Klein, 2003; Lamme, 2003; Woodman and Luck, 2003; Kentridge *et al.*, 2004; Ansorge and Heumann, 2006; Jiang *et al.*, 2006; Koch and Tsuchiya, 2007; Mele *et al.*, 2008; Mulckhuyse and Theeuwes, 2010; van Boxtel *et al.*, 2010]. It may seem counter-intuitive to pay attention to something and yet be unaware of it. But attention is a mechanistic process in the brain, like the regulation of blood flow or the growth of new synapses. It is a process of signal enhancement. Awareness, in contrast, is in the form of knowledge that is represented in the brain and can at least sometimes be reported. Awareness acts, in effect, like the brain's sometimes-wrong knowledge of its state of attention.

Third, when the brain attends to an item and yet is not aware of it, according to the theory, the brain has a temporarily faulty internal model of its attentional state. Therefore, the control of attention should suffer. By analogy, when the brain lacks a clear internal model of the arm, the control of the arm is compromised. The arm may be difficult to move to a new position or difficult to maintain in one position against external perturbations [Scheidt *et al.*, 2005]. Just so, if you are attending to a visual stimulus but unaware of it, your attention may be harder to disengage from the stimulus, or may be unduly influenced by inconsequential features of the stimulus. This third hypothesis about the relationship between awareness and attention — that in the absence of awareness, the control of attention should act as though it has lost its internal model — is one of the most crucial predictions of the theory. We are currently testing it in human psychophysical studies.

6. Integration of Information

Many scholars believe that a defining feature of consciousness is its integration of information across different domains [e.g., Baars, 1983; Crick and Koch, 1990; Damasio, 1999; Engel and Singer, 2001; Newman and Baars, 1993; Schiff, 2008; Tononi, 2008]. Although this integration of information is not the central contention of the attention-schema theory, the theory is nonetheless compatible with the integration hypothesis. Indeed, the theory may provide a simple explanation for why consciousness tends to be integrative.

The brain constructs models, or simulation, or updatable descriptions, of things in the real world. Those models themselves are made of smaller components. For example, for the visual system to construct a model of a red apple, it must link

together its model of the color red with its model of other stimulus features such as the shape or movement of the apple. This is integration of information. Color, however, is domain specific. It is not typically bound to information in other domains. Unless you have a condition called synesthesia, you do not literally see sounds as colored, see emotions as colored, or see mathematical thoughts as colored. Color can be linked to other visual information, but not typically to information outside the visual domain. It does not serve as a useful domain-general hub — a model to which models of many other kinds can be linked.

But in the attention schema theory, the brain does construct a model that is domain general. Attention is relevant to almost all domains of information processed in the brain — to vision, sound, a touch on the skin, emotion, thought, memory, or whatever the signals may be to which you are attending. In the attention schema theory, the brain constructs a model of attention and links it to a model of the attended item. That model of attention, the attention schema, is therefore an integrative hub. It is domain-general — a model that is linkable to almost any category of information.

Evolution is opportunistic. Sometimes a trait that evolves for one function takes on other functions. Perhaps the attention schema evolved first as a way of helping to control one's attention. We propose that a second obvious adaptive advantage of an attention schema is its ability to link information across domains. In this theory, awareness evolved initially as part of the control mechanism for attention and then allowed for an increase in intelligence by promoting domain-general integration of information.

7. Social Cognition

Over the half-billion years of its evolution, the attention schema may have taken on many adaptive functions. We proposed that it was gradually modified to model, monitor, and predict the attentional states of other animals [Graziano, 2013, 2014; Graziano and Kastner, 2011; Kelly *et al.*, 2014]. In this suggestion, we attribute awareness to other people as a means of modeling their attentional states, just as we attribute it to ourselves to model our own attentional states.

For example, Bill pays attention to a hamburger in front of him. That mechanistic process of attention leads to certain external signs on Bill such as his gaze direction, facial expression, body language, and verbal cues. If you are observing Bill, then based on a synthesis of those many cues you attribute awareness to him. You have an internal model informing you that Bill is aware of the hamburger.

Arguably, your ability to attribute awareness to someone else is foundational to all other social thinking. Maybe you think Bill is angry. It is difficult to attribute anger to him unless you first understand that he is aware of the unpleasant thing that aught to make him angry. You cannot predict that he will shout at you unless you first understand that he is aware of you. Maybe you think Bill intends to reach out and grasp something. You cannot make that attribution of intention unless you

understand that he is aware of the object to be grasped. Maybe you think that someone else thinks that you think that he is lying to you. That complicated back and forth of social cognition depends on understanding that the other person has such a thing as awareness and is aware of you, of your likely thoughts, and of his own thoughts. Social cognition makes no sense and has no foundation without the underlying attribution of awareness.

It is not yet clear when animals evolved the ability to attribute awareness to each other. Since many species of birds are highly social, perhaps birds can attribute awareness to other birds [Thom and Clayton, 2013]. Certainly many mammals can, including humans. The last common ancestor of birds and mammals lived approximately 350 MYA, and therefore a reasonable guess is that the social attribution of awareness first appeared sometime before that — though of course it could have evolved independently in both groups.

In this extension of the attention schema theory, awareness first evolved to help control one's own attention, and then gradually expanded into another use that has ended up defining us humans socially and culturally. It gave us our concept of mind and allowed us to live immersed in a society of the minds of other people.

In the human brain, there is some evidence of overlap between the areas responsible for attributing awareness to others and the areas necessary for one's own awareness. This overlap in function is particularly evident in an area of the cerebral cortex called the temporo-parietal junction (TPJ), more or less just above the ears and about an inch in. The TPJ has been a scientific puzzle because of an apparent conflict between two competing lines of research. In one line of research, it is involved in constructing models of other people's minds [e.g., Brunet *et al.*, 2000; Ciaramidaro *et al.*, 2007; Fletcher *et al.*, 1995; Gallagher *et al.*, 2000; Goel *et al.*, 1995; Saxe and Kanwisher, 2003; Saxe and Wexler, 2005; Vogeley *et al.*, 2001]. In another line of research, the TPJ is involved in attention and awareness [e.g., Astafiev *et al.*, 2006; Corbetta *et al.*, 2000; Mitchell, 2008; Shulman *et al.*, 2010]. Damage to the TPJ can even cause a severe and long-lasting deficit in awareness called hemispatial neglect [Karnath *et al.*, 2001; Vallar and Perani, 1986]. In neglect, typically damage to the right side of the brain causes a loss of awareness of anything to the left side of the body.

Why should a region of the cortex be involved in social cognition in some experiments and in attention and awareness in other experiments? One possible reason might be that this brain region participates in computations about awareness, whether you are attributing awareness to yourself or to someone else. It would not be correct to claim that the TPJ is the source of all computations related to awareness. However, it may play a role.

We recently conducted an experiment to test this hypothesis more directly [Kelly *et al.*, 2014]. The experiment involved two stages. First, people were scanned in an MRI machine to measure brain activity. The subjects looked at a picture of a cartoon face that was next to an object and rated how aware the cartoon person seemed to be

August 18, 2014 11:56:22am WSPC/258-IJMC 00131 ISSN: 1793-8430 Page Proof

of the object. In this task, certain areas of the brain became active above control levels. One area of activation was consistently within the TPJ.

In the second part of the experiment, the same people were taken out of the scanner environment and placed in a different testing room. The hotspot in the TPJ that was identified in the first part of the experiment was then targeted with a technique called transcranial magnetic stimulation (TMS). In that technique, a magnetic pulse is directed through the skull to temporarily disrupt brain function in a small patch of tissue, approximately 1 cm wide. In this experiment, disrupting the TPJ on one side of the brain disrupted the subject's ability to report dots flashed on a screen on the other side of space. The effect was not general to the entire TPJ. Instead, disruption of the specific hotspot obtained in the first part of the experiment was necessary. When the disruption was targeted to another site, 2 cm away but still within the larger area of the TPJ, the effect was no longer obtained.

One way to summarize this experiment is that specific areas of the brain became active when a person looked at someone else and answered the question, "Is he aware of the item next to him?" When the same brain regions were disrupted, the person was less able to answer the question, "Am I aware of the item in front of me?" This finding helps to support the hypothesis that awareness has taken on a social role at least in humans. A system in the human brain participates in computations about awareness whether you are attributing it to yourself or to someone else.

8. Some Thoughts on Machine Consciousness

In the attention schema theory, consciousness is more than a philosophical flourish. It is a fundamental part of the data processing machinery of the brain. If the theory is correct, then awareness is an internal model of attention and is crucial for the proper regulation of attention. In addition, awareness has taken on ever-expanding roles through evolutionary time including promoting the integration of information across different domains and promoting social cognition.

All of these functions are as useful to artificial intelligence as they are to human intelligence. They are also amenable to engineering. Every process described in this paper could be built, though probably at first only at a simple level.

There is no fundamental or theoretical limit to stop computer scientists from building a device that employs a human-like attention. In that process, signals compete at a local and global level. Winning signals rise in strength and have a disproportionate effect on memory and response choice.

There is also no fundamental or theoretical limit to stop engineers from adding an attention schema to help that artificial device predict and therefore regulate its own attention. That attention schema could contain simplifying information, modeling attention as though it were an ectoplasmic and magical substance that can reach out and "know" or "experience" things while being physically seated inside the machine itself.

There is no theoretical limit to stop engineers from adding the equivalent of higher cognition, a general purpose processor that is informed by the internal models computed within deeper levels of the device.

Given these pieces, we would have a machine that is convinced it has subjective awareness. If that higher cognition has access to language production, then the machine could tell us that it has awareness. It would report that when it introspects it finds awareness inside itself. It just knows it. Awareness is supplied to it *a priori*, like a Kantian prior. It would behave, in these respects, like any person.

The device could be designed to attribute awareness not just to itself, but to others as well. In that way, the machine would have a better basis for predicting the behavior of others and also a more human-like social capability as it attributes spirit to the beings around it.

The naïve approach of waiting to see if computers become conscious as they become more complicated has not yet yielded a satisfactory result. It may be more effective to design a machine in such a way that it concludes it has consciousness and can report that conclusion. The machine could use that self-model to regulate its own data flow and to understand the behavior of others.

If Deep Blue can beat Gary Kasparov, and Watson can win at Jeopardy, then a computer that contains the essential components of consciousness is easily within present technology. A concerted effort with sufficient resources could build such a device, perhaps within a decade.

Acknowledgment

Supported by the Princeton Neuroscience Institute.

References

Ansorge, U. and Heumann, M. [2006] "Shifts of visuospatial attention to invisible (metacontrast-masked) singletons: Clues from reaction times and event-related potentials," *Adv. Cogn. Psych.* **2**, 61–76.

Astafiev, S. V., Shulman, G. L. and Corbetta, M. [2006] "Visuospatial reorienting signals in the human temporo-parietal junction are independent of response selection," *European J. Neurosci.* **23**, 591–596.

Baars, B. J. [1983] "Conscious contents provide the nervous system with coherent, global information," in *Consciousness and Self-Regulation*, eds. Davidson, R. J., Schwartz, G. E. and Shapiro, D. (Plenum Press, NY), p. 41.

Baars, B. J. [1997] "Some essential differences between consciousness and attention, perception, and working memory," *Conscious. Cogn.* **6**, 363–371.

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

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

Botvinick, M. and Cohen, J. [1998] "Rubber hands 'feel' touch that eyes see," *Nature* **391**, 756.

Brunet, E., Sarfati, Y., Hardy-Baylé, M. C. and Decety, J. [2000] "A PET investigation of the attribution of intentions with a nonverbal task," *Neuroimage* **11**, 157–166.

August 18, 2014 11:56:22am WSPC/258-IJMC 00131 ISSN: 1793-8430 Page Proof

Budd, G. E. [2008] "The earliest fossil record of the animals and its significance," *Philos. Trans. Roy. Soc. London B, Biol. Sci.* **363**, 1425−1434.

Ciaramidaro, A., Adenzato, M., Enrici, I., Erk, S., Pia, L., Bara, B. G. and Walter, H. [2007] "The intentional network: How the brain reads varieties of intentions," *Neuropsychologia* **45**, 3105−3113.

Corbetta, M., Kincade, J. M., Ollinger, J. M., McAvoy, M. P. and Shulman, G. L. [2000] "Voluntary orienting is dissociated from target detection in human posterior parietal cortex," *Nature Neurosci.* **3**, 292−297.

Crick, F. and Koch, C. [1990] "Toward a neurobiological theory of consciousness," *Seminars Neurosci.* **2**, 263−275.

Damasio, A. [1999] *The Feeling of What Happens: Body and Emotion in the Making of Consciousness* (Harcourt, New York).

De Brigard, F. and Prinz, J. [2010] "Attention and consciousness," *Wiley Interdiscip. Rev. Cogn. Sci.* **1**, 51−59.

Dennett, D. C. [1992] *Consciousness Explained* (Back Bay Books, New York).

Desimone, R. and Duncan, J. [1995] "Neural mechanisms of selective visual attention," *Ann. Rev. Neurosci.* **18**, 193−222.

Engel, A. K. and Singer, W. [2001] "Temporal binding and the neural correlates of sensory awareness," *Trends Cogn. Sci.* **5**, 16−25.

Fletcher, P. C., Happé, F., Frith, U., Baker, S. C., Dolan, R. J., Frackowiak, R. S. and Frith, C. D. [1995] "Other minds in the brain: A functional imaging study of 'theory of mind' in story comprehension," *Cognition* **57**, 109−128.

Franklin, G. F., Powell, J. D. and Workman, M. L. [1989] *Digital Control of Dynamic Systems* (Addison-Wesley, Boston).

Gallagher, H. L., Happé, F., Brunswick, N., Fletcher, P. C., Frith, U. and Frith, C. D. [2000] "Reading the mind in cartoons and stories: An fMRI study of 'theory of mind' in verbal and nonverbal tasks," *Neuropsychologia* **38**, 11−21.

Gazzaniga, M. S. [1970] *The Bisected Brain* (Appleton Century Crofts, New York).

Goel, V., Grafman, J., Sadato, N. and Hallett, M. [1995] "Modeling other minds," *Neuroreport* **6**, 1741−1746.

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. and Botvinick, M. M. [2002] "How the brain represents the body: Insights from neurophysiology and psychology," in *Common Mechanisms in Perception and Action: Attention and Performance XIX*, eds. Prinz, W. and Hommel, B. (Oxford University Press, Oxford), pp. 136−157.

Graziano, M. S. A. and Kastner, S. [2011] "Human consciousness and its relationship to social neuroscience: A novel hypothesis," *Cogn. Neurosci.* **2**, 98−113.

Hwang, E. J. and Shadmehr, R. [2005] "Internal models of limb dynamics and the encoding of limb state," *J. Neural Engrg.* **2**, S266−S278.

Ivanoff, J. and Klein, R. M. [2003] "Orienting of attention without awareness is affected by measurement-induced attentional control settings," *J. Vis.* **3**, 32−40.

Jacobs, O. L. R. [1993] *An Introduction to Control Theory* (Oxford University Press, Oxford).

Jiang, Y., Costello, P., Fang, F., Huang, M. and He, S. [2006] "A gender- and sexual orientation-dependent spatial attention effect of invisible images," *Proc. Natl. Acad. Sci. USA* **103**, 17,048−17,052.

Karnath, H. O., Ferber, S. and Himmelbach, M. [2001] "Spatial awareness is a function of the temporal not the posterior parietal lobe," *Nature* **411**, 950−953.

Kawato, M. [1999] "Internal models for motor control and trajectory planning," *Curr. Opin. Neurobiol.* **9**, 718−727.

Kelly, Y. T., Webb, T. W., Meier, J. D., Arcaro, M. J. and Graziano, M. S. A. [2014] "Attributing awareness to oneself and to others," *Proc. Natl. Acad. Sci. USA* **111**, 5012−5017.

Kentridge, R. W., Heywood, C. A. and Weiskrantz, L. [1999] "Attention without awareness in blindsight," *Proc. Biol. Sci.* **266**, 1805−1811.

Kentridge, R. W., Heywood, C. A. and Weiskrantz, L. [2004] "Spatial attention speeds discrimination without awareness in blindsight," *Neuropsychologia* **42**, 831−835.

Koch, C. and Tsuchiya, N. [2007] "Attention and consciousness: Two distinct brain processes," *Trends Cogn. Sci.* **11**, 16−22.

Lackner, J. R. [1988] "Some proprioceptive influences on the perceptual representation of body shape and orientation," *Brain* **111**, 281−297.

Lambert, A., Naikar, N., McLachlan, K. and Aitken, V. [1999] "A new component of visual orienting: Implicit effects of visual information and subthreshold cues on covert attention," *J. Exp. Psychol. — Human Percep. Perform.* **25**, 321−340.

Lamme, V. A. [2003] "Why visual attention and awareness are different," *Trends Cogn. Sci.* **7**, 12−18.

Mack, A. and Rock, I. [1998] *Inattentional Blindness* (MIT Press, Cambridge).

McCormick, P. A. [1997] "Orienting attention without awareness," *J. Exp. Psychol. — Human Percep. Perform.* **23**, 168−180.

Mele, S., Savazzi, S., Marzi, C. A. and Berlucchi, G. [2008] "Reaction time inhibition from subliminal cues: Is it related to inhibition of return?" *Neuropsychologia* **46**, 810−819.

Merikle, P. M. and Joordens, S. [1997] "Parallels between perception without attention and perception without awareness," *Conscious. Cogn.* **6**, 219−236.

Mitchell, L. P. [2008] "Activity in the right temporo-parietal junction is not selective for theory-of-mind," *Cerebral Cortex* **18**, 262−271.

Mole, C. [2008] "Attention in the absence of consciousness?" *Trends Cogn. Sci.* **12**, 44−45.

Mulckhuyse, M. and Theeuwes, J. [2010] "Unconscious attentional orienting to exogenous cues: A review of the literature," *Acta Psychol.* **134**, 299−309.

Mysore, S. P. and Knudsen, E. I. [2013] "A shared inhibitory circuit for both exogenous and endogenous control of stimulus selection," *Nature Neurosci.* **16**, 473−478.

Newman, J. and Baars, B. J. [1993] "A neural attentional model for access to consciousness: A global workspace perspective," *Concepts Neurosci.* **4**, 255−290.

Posner, M. I. [1994] "Attention: The mechanisms of consciousness," *Proc. Natl. Acad. Sci. USA* **91**, 7398−7403.

Prinz, J. [2011] "Is attention necessary and sufficient for consciousness?" in *Attention: Philosophical and Psychological Essays*, eds. Mole, C., Smithies, D. and Wu, W. (Oxford University Press, Oxford), pp. 174−204.

Saxe, R. and Kanwisher, N. [2003] "People thinking about thinking people: fMRI investigations of theory of mind," *Neuroimage* **19**, 1835−1842.

Saxe, R. and Wexler, A. [2005] "Making sense of another mind: The role of the right temporo-parietal junction," *Neuropsychologia* **43**, 1391−1399.

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

Schiff, N. D. [2008] "Central thalamic contributions to arousal regulation and neurological disorders of consciousness," *Annals NY Acad. Sci.* **1129**, 105−118.

Shulman, G. L., Pope, D. L., Astafiev, S. V., McAvoy, M. P., Snyder, A. Z. and Corbetta, M. [2010] "Right hemisphere dominance during spatial selective attention and target detection occurs outside the dorsal frontoparietal network," *J. Neurosci.* **30**, 3640−3651.

Thom, J. M. and Clayton, N. S. [2013] "Re-caching by Western scrub-jays (*Aphelocoma californica*) cannot be attributed to stress," *PLoS One* **8**(1), e52936.

Tononi, G. [2008] "Consciousness as integrated information: A provisional manifesto," *Biologic. Bullet.* **215**, 216−242.

Vallar, G. and 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.

van Boxtel, J. J., Tsuchiya, N. and Koch, C. [2010] "Consciousness and attention: On sufficiency and necessity," *Front. Psychol.* **1**, 217.

van Swinderen, B. [2012] "Competing visual flicker reveals attention-like rivalry in the fly brain," *Front. Integ. Neurosci.* **6**, 96.

Vogeley, K., Bussfeld, P., Newen, A., Herrmann, S., Happé, F., Falkai, P., Maier, W., Shah, N. J., Fink, G. R. and Zilles, K. [2001] "Mind reading: Neural mechanisms of theory of mind and self-perspective," *Neuroimage* **14**, 170−181.

Wolpert, D. M., Ghahramani, Z. and Jordan, M. I. [1995] "An internal model for sensorimotor integration," *Science* **269**, 1880−1882.

Woodman, G. F. and Luck, S. J. [2003] "Dissociations among attention, perception, and awareness during object-substitution masking," *Psychol. Sci.* **14**, 605−611.