Michael S.A. Graziano

Consciousness Engineered

Abstract: *The attention schema theory offers one possible account for how we claim to have consciousness. The theory begins with attention, a mechanistic method of handling data in which some signals are enhanced at the expense of other signals and are more deeply processed. In the theory, the brain does more than just use attention. It also constructs an internal model, or representation, of attention. That internal model contains incomplete, schematic information about what attention is, what the consequences of attention are, and what its own attention is doing at any moment. This 'attention schema' is used to help control attention, much like the 'body schema', the brain's internal simulation of the body, is used to help control the body. Subjective awareness — consciousness — is the caricature of attention depicted by that internal model. This article summarizes the theory and discusses its relationship to the approach to consciousness that is called 'illusionism'.*

1. Introduction

Recently my colleagues and I proposed the attention schema theory as an explanation of consciousness (Graziano, 2013; 2014; Graziano and Kastner, 2011; Kelly *et al.*, 2014; Webb and Graziano, 2015; Webb, Kean and Graziano, 2016). Here by 'consciousness' I mean that, in addition to processing information, people report that they have a conscious, subjective experience of at least some of that information. The attention schema theory is a specific explanation for how we make that claim. It is a theory of how information is constructed in the brain and used to model the world and guide decisions, conclusions, speech, and behaviour. It is a theory of how the human machine claims to have consciousness and assigns a high degree of certainty to that conclusion.

Correspondence:
Michael S.A. Graziano, Dept. of Psychology, Princeton Neuroscience Institute,
Princeton University, Princeton, NJ 08544. *Email: graziano@princeton.edu*

Journal of Consciousness Studies, **23**, No. 11–12, 2016, pp. 98–115

2. Build-a-brain

One useful way to introduce the theory is through the hypothetical challenge of building a robot that asserts it is subjectively aware of an object and describes its awareness in the same ways that we do. I argue that the construction outlined below is not simply an academic exercise in engineering a zombie. Instead that type of mechanism is so basic that it is likely to have evolved in the brain. Moreover, as discussed in the second half of the article, growing evidence suggests that something like that mechanism *does* exist in the brain.

Figure 1 shows a robot looking at an apple. What information should be incorporated into its brain? First, we give it information about the apple (Figure 1A). Light enters the eye, is transduced into signals, and the information is processed to construct a description of the apple that includes shape, colour, size, location, and other attributes. This representation, or internal model, is constantly updated as new signals arrive. The model is schematic. It is a simplified proxy for the real thing. Given the limited processing resources in the brain, internal models are necessarily incomplete and simplified. They are efficient. They are data-compressed. Here we give our robot just such a simplified, schematic internal model of an apple.

Is the robot in Figure 1A aware of the apple? In one sense, yes. The term 'objective awareness' is sometimes used to indicate that the information has gotten in and is being processed (e.g. Szczepanowski and Pessoa, 2007). The machine in Figure 1A is objectively aware of the apple. But does it have a *subjective* experience?

To help explore that question we add a user interface, the linguistic processor shown in Figure 1B. Like a search engine, it can take in a question, search the internal model, and answer. We ask, 'What's there?' It answers, 'An apple'. We ask, 'What are the properties of the apple?' It answers, 'It's red, it's round, it's at that location'. It can provide those answers because it contains that information.

Figure 1B could represent an entire category of theory about consciousness, such as the global workspace theory (Baars, 1988; Newman and Baars, 1993). In that theory, consciousness occurs when information is broadcast globally throughout the brain. In Figure 1B, the sensory representation of the apple is broadcast globally, and as a result the cognitive and linguistic machinery has access to information about the apple. The robot can therefore report that the apple is present.

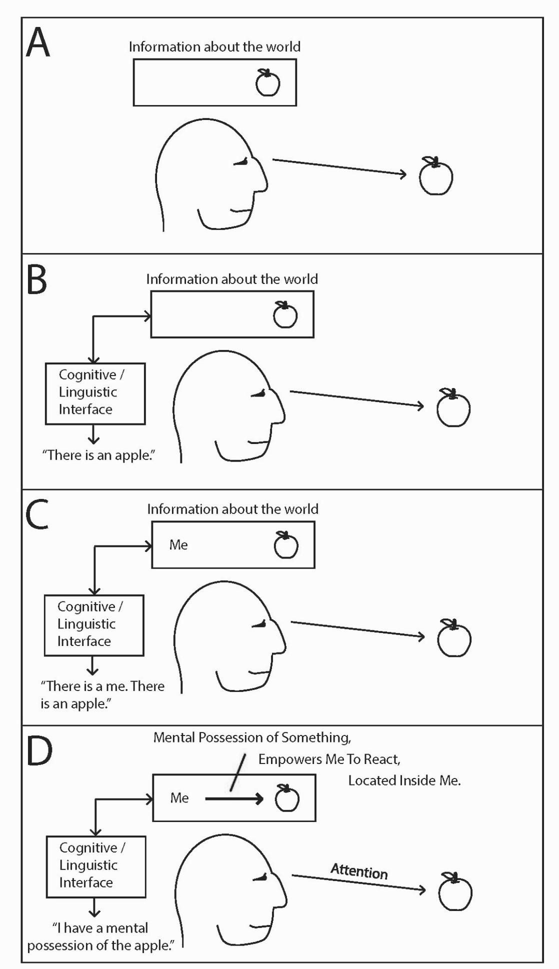
A
Information about the world
B
Information about the world
Cognitive / Linguistic Interface
"There is an apple."
C
Information about the world
Me
Cognitive / Linguistic Interface
"There is a me. There is an apple."
D
Mental Possession of Something,
Empowers Me To React,
Located Inside Me.
Me
Cognitive / Linguistic Interface
"I have a mental possession of the apple."
Attention

Figure 1. Construction of the attention schema theory. A robot has information about the world in the form of internal models. A. The robot has an internal model of the apple. B. The robot has a linguistic interface that acts as a search engine. It takes in questions, searches the internal model, and replies to the questions. C. The robot has a second internal model, a model of the self. D. The main components of the attention schema theory. The robot has an internal model of the self, a model of the apple, and a model of the attentional relationship between the self and the apple. That attention schema describes something physically incoherent, a caricature of attention, subjective awareness. The machine insists that it has subjective awareness of the apple because it is captive to the incomplete information in its internal models.

But Figure 1B remains an incomplete account of how a machine claims to be conscious of an apple. Consider asking, 'Are you aware of the apple?' The search engine searches the internal model and finds no answer. It finds information about an apple, but no information about what 'awareness' is or whether it has any of it, and no information about what the quantity 'you' is. It cannot answer the question. It does not compute in that domain.

Perhaps we can improve the machine. In Figure 1C, a second internal model is added, a model of the self. This new internal model, like the model of the apple, is a constantly updated set of information. It might include the body schema, the brain's model of the physical self and how it moves. The self model might also include autobiographical memory and general information about personality, beliefs, and goals. If we ask the robot in Figure 1C, 'Tell us about yourself?' it can now answer. It has been given the construct of self. It might reply, 'I'm a person, I'm standing right here, I'm so tall, so wide, I can move, I grew up in Buffalo, I'm a nice guy', and so on, as its cognitive search engine accesses its internal models. Figure 1C could represent an entire category of theory in which consciousness depends on self-knowledge or self-narrative (e.g. Gazzaniga, 1970; Nisbett and Wilson, 1977).

However, once again this account is incomplete. We can ask the machine in Figure 1C, 'What is the mental relationship between you and the apple?' The search engine accesses the two available internal models and finds no answer. It finds plenty of information about the self and plenty of separate information about the apple, but no information about a mental relationship between them — no information about what a mental relationship is. Equipped only with the components shown in Figure 1C, the machine cannot even parse the question.

So far we have given the machine an internal model of the apple and an internal model of the self, but we have neglected a crucial third item present in this scene — a less concrete, more intangible item — the computational relationship between the self and the apple. We now give the machine an internal model of attention.

The word 'attention' has many meanings, some colloquial, some technical. For example, overt attention is the orienting of eyes and other sensors toward an important event. Here, by attention I refer to something often called covert attention, the deep processing of some select signals at the expense of other signals. Covert attention can move from item to item. You can shift that deep processing from the text in front of you, to the sounds coming from your back yard, to a memory that you've just recalled, to a math problem that you're solving in your head. Covert attention is a mechanistic neural phenomenon, a selective signal enhancement caused by competition among signals in the brain (Beck and Kastner, 2009; Desimone and Duncan, 1995). A person rarely looks at an apple in isolation. Other items are probably present: a plate, a table, a wall behind them, too much for the brain to process in depth at the same time. It has to prioritize. The apple signal wins the competition of the moment, is enhanced at the expense of other visual signals, and as a result can dominate the brain's outputs. The brain deeply processes information about the apple and is therefore primed to generate behaviour toward it or remember it. This is the attentive relationship between the self and the apple. An internal model of that attentive relationship is added to Figure 1D.

First consider what information might be contained in an internal model of attention. How would it describe attention? Presumably, like the internal model of the apple, it would describe useful, functional, *abstracted* properties of attention, not microscopic physical details. It might describe attention as a mental possession of something. It might describe attention as something that empowers oneself to react. It might describe attention as something located inside oneself, belonging to oneself, and not directly observable to the outside world. It might include many other abstracted properties of attention. But this internal model would not contain information about neurons, competing electrochemical signals, or other physical nuts and bolts that the brain has no pragmatic need to know. Like all internal models, it would be incomplete and schematic. It would be silent on the physical mechanisms of attention.

We ask the robot in Figure 1D, 'What is the mental relationship between you and the apple?' The search engine accesses its internal models and reports the available information. It says, 'I have a mental possession of the apple'. The answer is promising and we probe deeper. 'Tell us more about this mental possession. What are its physical properties?' For clarity, we also ask, 'Do you know what physical properties are?' The machine can answer 'Yes' because it has a body schema that describes the physical body and it has an internal model of an apple that describes a physical object. Reporting the information available to it, it might say (if it has a good vocabulary), 'I know what physical properties are. But my mental possession of the apple, the mental possession in-and-of-itself, has no physically describable properties. It's an essence located inside me. Like my arms and legs are physical parts of me, I also have a non-physical or *metaphysical* part of me. It's my mind taking hold of things — the colour, the shape, the location. My subjective self seizes those things.' The machine is describing covert attention, and the description sounds semi-magical only because it is vague on the details and the mechanistic basis of attention.

Because we built the robot, we know why it gives that answer. It's a machine accessing internal models. Whatever information is contained in those models it reports to be true. That information lies deeper than language, deeper than higher cognition. The machine insists it has subjective awareness because, when its internal models are searched, they return that information. Introspection will always return that answer. In the same way, it reports that the apple has a colour even though in reality the apple has a reflectance spectrum, not colour. Just as in Metzinger's ego tunnel (Metzinger, 2010), this brain is captive to the schematic information in its internal models.

The theory diagrammed in Figure 1D is different from a higher-order thought theory (Lau and Rosenthal, 2011). In that approach, consciousness occurs when the brain's cognitive machinery constructs a higher-order, cognitive representation, or an interpretation. Instead, in the attention schema theory, subjective awareness does not depend on cognitive or linguistic processing. It is not a construct of higher-order thought. The cognitive/linguistic layer in Figure 1 was added as a convenience to be able to query the machine, but it is not necessary. Suppose you are a rat with little cognitive and no linguistic capacity, thus the 'cognitive/linguistic' box in Figure 1D is missing. You still have the internal models themselves. The internal models in Figure 1D are fundamental, low-level representations necessary for survival: representations of self, of items in the world such as apples, and of the ever-present process of attention, the computational relationship between the self and everything else. These representations can guide behaviour directly, even without higher cognition. One way to think about Figure 1D is that the brain constructs an overarching internal model, a continuously updated simulation of its world. In that simulation, there is a self that is conscious of the apple. The brain constructs that simulation even if it lacks the sophistication to cogitate about it or talk about it.

One advantage of this theory of consciousness is that it can accommodate the correct range of information. The brain can focus attention on a colour, a shape, a motion, a sound, a touch, a memory, a thought, a fragment of autobiographical knowledge, an emotional state, or almost any other domain of information that is processed cortically. An attention schema is therefore applicable to that same range of information. One could replace the apple in Figure 1D with almost anything, whether a feature of the external world or a feature of one's internal cognition. The theory accounts for why we claim to have a conscious experience of colour, shape, sound, self, memory, emotion, and so on, and why, despite the diversity of information, the consciousness is somehow of the same nature in all cases. In the theory, an internal model of attention pertains to multiple kinds of information.

The logic of the theory can be summarized in four points. One, the brain constructs internal models of important objects and processes in the world. Therefore, two, the brain constructs an internal model of its own process of attention. Three, internal models are never accurate descriptions. They are incomplete and schematic, due to a trade-off between accuracy and processing resources. Therefore, four, a brain with an internal model of attention, even if that brain has a good enough linguistic and cognitive capacity to talk about it, would not report its attention in a physically accurate, detailed, or mechanistic manner. Instead it would claim to have something physically incoherent: a subjective mental experience. A five-word summary of the theory, that admittedly loses some nuance, is this: awareness is an attention schema.

3. Adaptive uses of the attention schema

It is clear why an internal model of an apple is useful — to guide behaviour with respect to the apple. It is also clear why an internal model of the self is useful — to monitor and thus better control one's behaviour. But what is the adaptive value of an attention schema? In the following sections I describe three uses for an attention schema, beginning with its possible role in the widespread integration of information.

3.1. Integration of information

The idea that awareness is related to the integration of information around the brain has been suggested in many forms (e.g. Baars, 1988; Crick and Koch, 1990; Damasio, 1990; Engel and Singer, 2001; Lamme, 2006; Newman and Baars, 1993; Tononi, 2008). The attention schema theory is, in its own way, a theory about the integration of information. In Figure 1D, the attention schema is a chunk of information, a descriptive model, that is linked to many disparate kinds of information. Information about the self and information about an apple are linked together by way of an intermediate bridge, the attention schema.

One can think of information itself as having connectivity. For example, colour information is a connector. Imagine a scattering of dots, some black, some red. The red ones happen to form a larger shape, an X. That X stands out because dots of a similar colour are easily linked together to form a single, integrated representation. Ultimately there is an anatomical underpinning to that phenomenon, but one can make partial sense of it purely from the point of view of information. In the lattice of information, some dots are connected to each other because they are connected to the same colour information. Colour, as a connector, is obviously limited to the visual domain.

Spatial location is another connector, but unlike colour it can operate across sensory domains. If a visual stimulus and a sound appear at the same location, we are prone to link the two, constructing an integrated representation of an object that has both visual and auditory aspects. This spatial interaction has been especially studied in the superior colliculus, where tactile, visual, and auditory information is processed in a single spatial framework (Stein, Stanford and Rowland, 2009). But even though location information can be used to link across sensory domains, it is still limited in its ability to bridge some kinds of information. Information domains that do not have an obvious spatial component are not included.

An attention schema can act as the ultimate connector. Almost all kinds of information in the brain are subject to attention. An attention schema, a central representation of attention, could serve as a hub that connects to any information domain. Awareness, as a model of attention, is like a colour that can tint any topic. In Figure 1D, the attention schema links information about the self, including the body schema and autobiographical knowledge, with information about an apple, including shape and colour and location. In that way, an overarching representation subsumes a great range of information domains. But the figure diagrams only one limited example. One could just as well attend to a thought, a taste, a recalled memory, an emotion, a movement of the limbs, or even all of those together if you are grasping the apple, biting it, thinking about it, and enjoying it. All of those radically different information domains can be linked to the attention schema and thus linked to each other in a single representation. In effect, the brain constructs an integrated representation of its world at that moment, and the attention schema is the crucial bridge that connects the disparate parts because the attention schema represents the deep, computational relationship between the self and the various components of one's world.

So many people have noted the apparent relationship between consciousness and the widespread integration of information that one can't help thinking, with all the smoke, there must be fire. Surely a good theory of consciousness should include that relationship. The attention schema theory does so quite naturally. But in this theory, consciousness is not magically caused by integrating a mass of information together, like in the science fiction trope where Skynet wakes up. Instead, consciousness is a construct. It is an attention schema. That construct serves a central role in bridging across disparate domains of information, allowing for a more complete model of oneself operating in the world.

3.2. Control of attention

A primary function of the attention schema may be the efficient control of attention.

A basic principle of control theory is that a control system benefits from an internal model of the thing to be controlled (Camacho and Bordons Alba, 2004). For example, the brain constructs a body schema, an internal model of the body, to help control movement (Wolpert, Ghahramani and Jordan, 1995). Like all internal models, the body schema is imperfect. It can sometimes become misaligned from the body. Almost all experimental work demonstrating the existence of the body schema relies on those inaccuracies. When misalignment between the body schema and the body occurs, movement of the body is still possible but the controller suffers characteristic deficits that reveal the importance of the body schema (Graziano and Botvinick, 2002; Scheidt *et al.*, 2005; Wolpert, Ghahramani and Jordan, 1995).

If the attention schema theory is correct, the relationship between consciousness and attention should also adhere to the predictions of control theory. The most experimentally revealing conditions should occur when the internal model makes a mistake — when consciousness becomes misaligned from attention. In that case, the control of attention should suffer in a manner consistent with the loss of an accurate internal model.

In the scientific literature on the relationship between consciousness and attention, typically the term ‘awareness’ is used to refer to a conscious, reportable experience of the sensory stimulus. In keeping with that usage, in the following discussion I will use the term awareness as synonymous with consciousness. It is now well established that attention and awareness can be separated. People can attend to a stimulus in the absence of awareness of that stimulus (Hsieh, Colas and Kanwisher, 2011; Jiang *et al.*, 2006; Kentridge, Nijboer and Heywood, 2008; Koch and Tsuchiya, 2007; Lamme, 2004; McCormick, 1997; Norman, Heywood and Kentridge, 2013; Tsushima, Sasaki and Watanabe, 2006). This separability has led to the suggestion that attention and awareness may be independent processes. In the attention schema theory, the two are not truly independent. They have a principled relationship, diagrammed in Figure 2. Attention has a control system and one part of that system is an internal control model. That internal model, the attention schema, contains information about awareness. If the system is attending to some item X but has not constructed an awareness of X, that corresponds to a temporarily faulty internal model. The internal model of attention has failed to update correctly. In that case, the control of attention should suffer. Attention should still be possible, but it should lose stability and be more easily perturbed by outside influences. In much the same way, if the body schema fails to register the location of your arm perhaps because of anaesthesia of the arm, of course you still have an arm, and you can even control its movement to some degree; but the arm becomes less stable and more easily perturbed by outside influences. Here it is useful to make some clarifications. In control theory, the internal model is not the entire controller. It is one useful part of the controller. Without it, or if it becomes impaired, some control is still possible. Control is compromised in specific ways that are discussed in greater detail below.

Most of the experiments that distinguish visual attention and visual awareness have focused on the first-order phenomenon: attention can exist without awareness. Few experiments ask whether attention behaves the same way, or changes, when awareness is present or absent. The paradigms used to manipulate visual awareness typically involve major changes to the stimulus. As a result, the aware condition and the unaware condition are not easy to compare quantitatively.

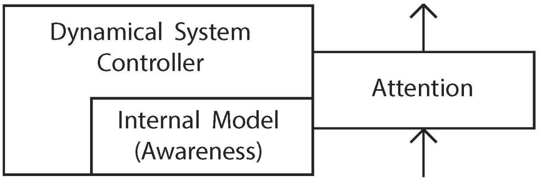

Figure 2. Awareness as an internal control model for attention. The arrows represent information subjected to the process of attention. Attention is regulated by a complex, dynamical systems controller. In the attention schema theory, one part of that controller is an internal model of attention, and awareness is the cartoonish depiction of attention rendered by that internal model.

Recently we performed a series of psychophysics experiments in human subjects to answer this question (Webb, Kean and Graziano, 2016). In those experiments, a visual stimulus drew people's attention to a location. In one condition the stimulus was masked such that participants were subjectively unaware of it. In another condition the mask was adjusted such that participants reported being subjectively aware of the stimulus. The amount of attention drawn by the stimulus was measured using a standard Posner paradigm.

On the basis of the attention schema theory, we predicted that attention would show less stability over time when awareness of the stimulus was absent. This prediction was confirmed. Without awareness of the stimulus, attention to that stimulus behaved in a less stable manner. Attention wobbled up and down significantly more during the tested time interval. Figure 3 shows data from one of the experiments that demonstrates this stabilizing effect of awareness on attention. From the point of view of control theory, when awareness of the stimulus was absent, attention to the stimulus acted as though the stabilizing, internal control model of attention was missing. These experiments are among the most direct tests of the hypothesis that awareness serves as the internal model for attention.

A separate line of experiments by Schurger and colleagues (2010; 2015) also supports the attention schema theory. In those experiments, a visual stimulus evoked activity throughout the visual cortex. Neuronal representation was more stable in time and more consistent across trials when awareness was present than when awareness was absent. Experiments such as these point toward the attention schema theory, in which awareness plays a fundamental role as the internal, stabilizing control model for attention.

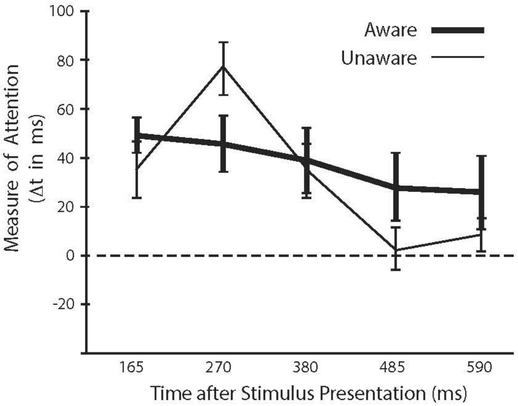

Figure 3. Testing attention with and without awareness. In this experiment, attention to a visual stimulus is tested by using the stimulus as a cue in a Posner spatial attention paradigm (see Webb, Kean and Graziano, 2015, for details). In some trials, the participants are aware of the visual cue (thick line). In other trials, they are unaware of it (thin line). Attention to the cue is less stable across time when awareness is 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 (Δt = [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.

One of the central questions of consciousness research is whether consciousness has any adaptive role and, if so, what that role may be. In the attention schema theory, consciousness serves a set of well-defined and testable roles in information processing. Its most basic role, perhaps its evolutionary origin, is to serve as a control model for attention. An engineer who wishes to understand how the brain processes information must understand the internal models that the brain uses to regulate itself. In this theory, consciousness is one of those internal models.

3.3. Social cognition

A third possible adaptive use of an attention schema lies in social cognition. I will begin with an analogy to the body schema. The brain evolved an internal model of the body, an ever-changing, ever-updating complex of information that describes the physical shape, structure, and movement of the body (Graziano and Botvinick, 2002). The body schema presumably first evolved when nervous systems became sophisticated in the control of movement, probably more than half a billion years ago.

One often overlooked function of the body schema, in humans, is to model the bodies of others. If subjects look at a picture of a hand and decide whether it is a left or right hand, the decision is markedly faster when the pictured hand already matches the configuration of the subject's own hand (Parsons, 1987; Sekiyama, 1982). The more different the configurations, the longer the latency to respond, as though the subjects were mentally reconfiguring their own hands to match the picture. Moreover, the same cortical areas, especially the posterior parietal lobe, were active whether judging other people's body configurations or one's own body schema (Bonda *et al.*, 1995). Presumably the use of the body schema to model oneself and control one's own movements emerged first in evolution, and its use in monitoring and predicting the bodies of others was a gradual evolutionary extension.

In the attention schema theory, a similar extension of function occurred for the internal model of attention. In the theory, the attention schema first evolved as a simple model that was part of the control mechanism for attention. Its function was to model one's own state of attention. However, over evolutionary time, an additional adaptive function emerged. The attention schema became increasingly adapted to model the attentional states of others. The advantage is obvious. It improves my ability to predict the behaviour of others and thus guide my own behaviour with respect to others. Attention is one of the main determinants of behaviour. If Bill's brain is focusing attention on X, then X is likely to dominate his behaviour. If I want to predict Bill's behaviour, it would be useful for me to have a model of attention that I can apply to Bill — a model of what attention is, what its dynamics and consequences are, and what in particular Bill is attending to. In the theory, the awareness that I attribute to Bill is a simplified and effective model of his deep attentive processing of information.

In this proposal, we do not merely figure out intellectually whether someone is aware of something. Instead that attribution has an automatic, immediate, perceptual quality because it depends on an internal model constructed beneath the level of high-order cognition.

Ventriloquism is a good example that helps to demonstrate the contrast between a cognitive model and perceptual model of someone else's mind. We have a *perception* of awareness in the ventriloquist puppet, while at the same time we know cognitively that the puppet is not really aware.

People attribute awareness to more than just other people and ventriloquist dummies. We attribute it to our dogs and cats. Some people feel that their houseplants are conscious. Animism is the attribution of subjective consciousness to trees, rivers, volcanoes, and other natural phenomena. Gods, angels, ghosts, spirits are all attributions of awareness. The belief in life after death is the false attribution of awareness outside the physical body. When I am alone in a dark house on a stormy night I sometimes have a creepy feeling that another conscious agent is in the next room stalking me. Intellectually I know it's not true, but my brain has evidently constructed that perceptual model. Sometimes people get angry at the car or the coffee machine as if it were aware of its misdeeds. I argue that this secondary role for the attention schema, attributing awareness to others, is the primary basis for human spiritual belief. We live in a world awash in perceived awareness.

Psychophysical evidence suggests that people do indeed construct a rich model of the attentional state of others. Most research on the topic focuses on only one component, the gaze direction of others. However, people do combine gaze cues, facial expression, and context when assessing the attentional state of others (Kelly *et al.*, 2014). A recent study suggests that we not only reconstruct the object of someone else's attention, we also reconstruct some of the dynamic aspects of attention such as whether attention was drawn extrinsically (by an external stimulus) or directed intrinsically (by an internal decision) (Pesquita, Chapman and Enns, forthcoming). These studies are beginning to show that the brain does indeed construct a rich internal model of the attentional state of others, consistent with the attention schema theory.

One particular brain area may be a central node in a network that computes information related to awareness. Damage to the temporoparietal junction (TPJ) can result in hemispatial neglect, a disturbance of awareness (Valler and Perani, 1986). Attributing states of awareness to others evokes activity in the same subregions of the TPJ (Kelly *et al.*, 2014; Igelstrom *et al.*, 2016). These studies of course do not pin a single function on the TPJ, which presumably contributes to a range of cognitive functions. The studies do, however, begin to support the attention schema theory, in which awareness in both the personal and social sense is a construct of specialized networks in the brain.

4. Illusionism

In the target article of this special issue, Frankish describes an approach to consciousness called illusionism that is shared by many theories of consciousness. The attention schema theory has much in common with illusionism. It clearly belongs to the same category of theory, and is especially close to the approach of Dennett (1991). But I confess that I baulk at the term 'illusionism' because I think it miscommunicates. To call consciousness an illusion risks confusion and unwarranted backlash. To me, consciousness is not an illusion but a useful caricature of something real and mechanistic. My argument here concerns the rhetorical power of the term, not the underlying concepts.

In my own discussions with colleagues, I invariably encounter the confusion and backlash. To most people, an illusion is something that does not exist. Calling consciousness an illusion suggests a theory in which there is nothing present that corresponds to consciousness. However, in the attention schema theory, and in the illusionism described by Frankish, something specific is present. In the attention schema theory, the real item that exists inside us is covert attention — the deep processing of selected information. Attention truly does exist. Our internal model of it lacks details and therefore provides us with a blurred, seemingly magicalist account of it.

Second, in normal English, to experience an illusion is to be fooled. To call consciousness an illusion suggests to most people that the brain has made an error. In the attention schema theory, and also in the illusionism approach described by Frankish, the relevant systems in the brain are not in error. They are well adapted. Internal models always, and strategically, leave out the unnecessary detail.

Third, most people understand illusions to be the result of a subjective experience. The claim that consciousness is an illusion therefore sounds inherently circular. Who is experiencing the illusion? It is difficult to explain to people that the experiencer is not itself conscious, and that what is important is the presence of the information and its impact on the system. The term illusion instantly aligns people's thoughts in the wrong direction.

All of the common objections I encounter have answers. They are based on a misunderstanding of illusionism. But the misunderstanding is my point. Why use a misleading word that requires one to backtrack and explain? For these reasons, in my own writing I have avoided calling consciousness an illusion except in specific circumstances, such as the consciousness we attribute to a ventriloquist puppet, in which the term seems to apply more exactly.

Perhaps I am too much of a visual physiologist at heart. To me, an illusion is a mistake in a sensory internal model. It introduces a consequential discrepancy between the internal model and the real world. That discrepancy can cause errors in behaviour. In contrast, an internal model, at all times, with or without an illusion, is an efficient, useful compression of data. It is never literally accurate. Even when it is operating correctly and guiding behaviour usefully, it is a caricature of reality. I am comfortable calling consciousness a caricature, but not an illusion. It is a cartoonish model of something real.

References

Baars, B.J. (1988) *A Cognitive Theory of Consciousness*, New York: Cambridge University Press.

Beck, D.M. & Kastner, S. (2009) Top-down and bottom-up mechanisms in biasing competition in the human brain, *Vision Research*, **49**, pp. 1154–1165.

Bonda, E., Petrides, M., Frey, S. & Evans, A. (1995) Neural correlates of mental transformations of the body-in-space, *Proceedings of the National Academy of Sciences USA*, **92**, pp. 11180–11184.

Camacho, E.F. & Bordons Alba, C. (2004) *Model Predictive Control*, New York: Springer.

Crick, F. & Koch, C. (1990) Toward a neurobiological theory of consciousness, *Seminars in the Neurosciences*, **2**, pp. 263–275.

Damasio, A.R. (1990) Synchronous activation in multiple cortical regions: A mechanism for recall, *Seminars in the Neurosciences*, **2**, pp. 287–296.

Dennett, D.C. (1991) *Consciousness Explained*, Boston, MA: Little, Brown, & Co.

Desimone, R. & Duncan, J. (1995) Neural mechanisms of selective visual attention, *Annual Review of Neurosciences*, **18**, pp. 193–222.

Engel, A.K. & Singer, W. (2001) Temporal binding and the neural correlates of sensory awareness, *Trends in Cognitive Sciences*, **5**, pp. 16–25.

Frankish, K. (this issue) Illusionism as a theory of consciousness, *Journal of Consciousness Studies*, **23** (11–12).

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

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

Graziano, M.S.A. (2014) Speculations on the evolution of awareness, *Journal of Cognitive Neuroscience*, **26**, pp. 1300–1304.

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

Graziano, M.S.A. & Kastner, S. (2011) Human consciousness and its relationship to social neuroscience: A novel hypothesis, *Cognitive Neuroscience*, **2**, pp. 98–113.

Hsieh, P., Colas, J.T. & Kanwisher, N. (2011) Unconscious pop-out: Attentional capture by unseen feature singletons only when top-down attention is available, *Psychological Science*, **22**, pp. 1220–1226.

Igelstrom, K., Webb, T.W., Kelly. Y.T. & Graziano, M.S.A. (2016) Topographical organization of attentional, social and memory processes in the human temporo-parietal cortex, *eNEuro*, **3**, ENEURO.0060-16.2016.

Jiang, Y., Costello, P., Fang, F., Huang, M. & He, S. (2006) A gender- and sexual orientation-dependent spatial attentional effect of invisible images, *Proceedings of the National Academy of Sciences USA*, **103**, pp. 17048–17052.

Kelly, Y.T., Webb, T.W., Meier, J.D., Arcaro, J. & Graziano, M.S.A. (2014) Attributing awareness to oneself and to others, *Proceedings of the National Academy of Sciences USA*, **111**, pp. 5012–5017.

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

Koch, C. & Tsuchiya, N. (2007) Attention and consciousness: Two distinct brain processes, *Trends in Cognitive Sciences*, **11**, pp. 16–22.

Lamme, V.A. (2004) Separate neural definitions of visual consciousness and visual attention: A case for phenomenal awareness, *Neural Networks*, **17**, pp. 861–872.

Lamme, V.A. (2006) Towards a true neural stance on consciousness, *Trends in Cognitive Sciences*, **10**, pp. 494–501.

Lau, H. & Rosenthal, D. (2011) Empirical support for higher-order theories of consciousness, *Trends in Cognitive Sciences*, **15**, pp. 365–373.

McCormick, P.A. (1997) Orienting attention without awareness, *Journal of Experimental Psychology, Human Perception and Performance*, **23**, pp. 168–180.

Metzinger, T. (2010) *The Ego Tunnel*, New York: Basic Books.

Newman, J. & Baars, B.J. (1993) A neural attentional model for access to consciousness: A global workspace perspective, *Concepts in Neuroscience*, **4**, pp. 255–290.

Nisbett, R.E. & Wilson, T.D. (1977) Telling more than we can know — verbal reports on mental processes, *Psychological Review*, **84**, pp. 231–259.

Norman, L.J., Heywood, C.A. & Kentridge, R.W. (2013) Object-based attention without awareness, *Psychological Science*, **24**, pp. 836–843.

Parsons, L.M. (1987) Imagined spatial transformations of one's hands and feet, *Cognitive Psychology*, **19**, pp. 178–241.

Pesquita, A., Chapman, C.S. & Enns, J.T. (forthcoming) Seeing attention in action: Human sensitivity to attention control in others, *Proceedings of the National Academy of Sciences USA*.

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, *Journal of Neurophysiology*, **93**, pp. 3200–3213.

Schurger, A., Pereira, F., Treisman, A. & Cohen, J.D. (2010) Reproducibility distinguishes conscious from nonconscious neural representations, *Science*, **327**, pp. 97–99.

Schurger, A., Sarigiannidis, I., Naccache, L., Sitt, J.D. & Dehaene, S. (2015) Cortical activity is more stable when sensory stimuli are consciously perceived, *Proceedings of the National Academy of Sciences USA*, **112**, pp. E2083–2092.

Sekiyama, K. (1982) Kinesthetic aspects of mental representations in the identification of left and right hands, *Perceptual Psychophysics*, **32**, pp. 89–95.

Stein, B.E., Stanford, T.R. & Rowland, B.A. (2009) The neural basis of multisensory integration in the midbrain: Its organization and maturation, *Hearing Research*, **258**, pp. 4–15.

Szczepanowski, R. & Pessoa, L. (2007) Fear perception: Can objective and subjective awareness measures be dissociated?, *Journal of Vision*, **10**, pp. 1–17.

Tononi, G. (2008) Consciousness as integrated information: A provisional manifesto, *Biological Bulletin*, **215**, pp. 216–242.

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

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**, pp. 609–622.

Webb, T.W. & Graziano, M.S.A. (2015) The attention schema theory: A mechanistic account of subjective awareness, *Frontiers in Psychology*, doi: 10.3389/fpsyg.2015.00500.

Webb, T.W., Kean, H.H. & Graziano, M.S.A. (2016) Effects of awareness on the control of attention, *Journal of Cognitive Neuroscience*, **28**, pp. 842–851.

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