Cognitive Neuropsychology

ISSN: 0264-3294 (Print) 1464-0627 (Online) Journal homepage: https://www.tandfonline.com/loi/pcgn20

Consciousness and the attention schema: Why it has to be right

Michael S. A. Graziano

To cite this article: Michael S. A. Graziano (2020): Consciousness and the attention schema: Why it has to be right, Cognitive Neuropsychology, DOI: 10.1080/02643294.2020.1761782

To link to this article: https://doi.org/10.1080/02643294.2020.1761782

Published online: 20 May 2020.

COGNITIVE NEUROPSYCHOLOGY
https://doi.org/10.1080/02643294.2020.1761782

Routledge
Taylor & Francis Group

ARTICLE COMMENTARY

Consciousness and the attention schema: Why it has to be right

Michael S. A. Graziano

Department of Psychology, Princeton University, Princeton, NJ, USA

ABSTRACT

This article describes some aspects of the underlying logic of the attention schema theory (AST) of subjective consciousness. It is a theory that distinguishes between what the brain actually, physically has, what is represented by information models constructed in the brain, what higher cognition thinks based on access to those models and what speech machinery claims based on the information within higher cognition. It is a theory of how we claim to have an essentially magical, subjective mind, based on the impoverishment and reduction of information along that pathway. While the article can stand on its own as a brief account of some critical aspects of AST, it specifically addresses questions and concerns raised by a set of commentaries on a target article.

ARTICLE HISTORY
Received 27 January 2020
Revised 19 March 2020
Accepted 17 April 2020

KEYWORDS
Consciousness; awareness; attention; theory of mind; social cognition

Introduction

In the target article for this issue (Graziano et al., 2020), my colleagues and I suggested that several current theories of consciousness are compatible with each other, and that the connectivity between them becomes especially clear in the context of the attention schema theory (AST), the mechanistic theory of subjective experience that we proposed. I warmly thank everyone who contributed commentaries responding to that article. Every response presented a useful, well-reasoned point of view, some agreeing with our primary arguments, some directly opposed. In every case, I value the comments and the pointers to a larger literature, and I hope the overarching discussion has been helpful to everyone.

Many of the commentaries supported our arguments or amplified them by adding new ideas to the larger story (e.g., Blackmore, 2020; Dennett, 2020; Frankish, 2020; Prinz, 2020; Romo & Rossi-Pool, 2020; Vernet et al., 2020; Yankulova & Morsella, 2020). Some of the commentaries presented counter-arguments mainly centred on AST itself. If AST is incorrect or seriously incomplete, then it cannot contribute significantly to a standard theory of consciousness. The best way I can respond to these commentaries, therefore, is to explain why AST makes sense. Rather than address each commentary separately, repeating the arguments that the authors expressed better in their own words, I've collapsed the arguments into three main categories. These three concerns about AST were especially well represented and I hear them often.

First, some ask how an attention schema can possibly explain a subjective feeling. How could having a bundle of information in the brain – information that describes attention – cause anyone to be subjectively aware of anything? For example, one argument is that consciousness – subjective experience, the what-it-feels-like component – is something we really do have, not something we merely think we have or say we have. Consciousness is not an illusion but an actuality. Yet AST appears to be an explanation for how a machine "thinks" and "says" it has consciousness, not an explanation for how a brain actually *is* conscious. (Comments that make this point or a similar point include Brown & LeDoux, 2020; Gennaro, 2020; Lane, 2020; Masciari & Carruthers, 2020; Rosenthal, 2020).

Second, and in contrast to the first point, some argue that consciousness is, indeed, an illusion (Blackmore, 2020; Dennett, 2020; Frankish, 2020). Yet the target article seems to describe AST coyly, refusing to call it an illusionist theory when it obviously is one. Why not admit that consciousness is an illusion?

The third, and most common concern about AST is: why focus on attention? Surely consciousness is much larger than attention, encompassing many more

CONTACT Michael S. A. Graziano Graziano@princeton.edu Department of Psychology, Princeton University, Princeton, NJ 08544, USA

processes hosted by the brain. Why not propose a more general "mind schema" or other related schemas instead of only an attention schema? (This question is represented in many of the comments, including: Frankish, 2020; Lane, 2020; Metzinger, 2020; Panagiotaropoulos et al., 2020; Prinz, 2020).

All of these concerns about AST can be addressed by carefully considering the building blocks of the theory. Here I will explain the theory from a new angle, laying out certain aspects of the underlying logic that address these three specific concerns. In the final two sections, I will also address two questions about AST raised in the commentaries that are not concerns so much as requests for clarification.

I am aware that I've titled this reply in a provocative way, but I will try to argue that an attention schema, when properly understood, has an intrinsic logic that is hard to escape, and it is likely to be a crucial part of the larger system we call consciousness.

Real objects and models of them

To start, I will put consciousness aside and discuss an analogy that should be uncontroversial. The analogy will allow me to discuss the difference between having something and thinking you have it. I'll focus on the body schema, and in particular the arm schema – a topic that I studied for many years.

You have an arm, a part of physical reality (see the left side of Figure 1A). The brain constructs a model of the arm – the arm schema (also shown in Figure 1A). The model is information – a simulation of an arm. It is based partly on sensory information coming from the arm, but much of it is internally generated. One demonstration of its internally generated nature is that if your arm is amputated, the model can linger on – a phantom limb.

The arm schema is rich in detail, but not complete. It includes information about the joint degrees of freedom, the overall shape and structure of the arm and hand, inertia, viscosity, how the arm might interact with other body parts (e.g., if you move in a certain way your hand will hit your stomach), and predictive information about how internally generated commands will make the arm move here or there. The model is lacking many details of the arm – it doesn't contain information about individual muscle attachments, or bone shape, or the proteins that cause muscle contraction – information the brain doesn't need to know about. The purpose of an internal model is not to be physically accurate or complete, but to be useful. Sometimes the arm schema makes mistakes, especially if we trick it with laboratory manipulations. For example, the arm may be in one position while the model registers it as somewhere else. These mistakes are typically called illusions. Usually, however, the arm model tracks the arm closely.

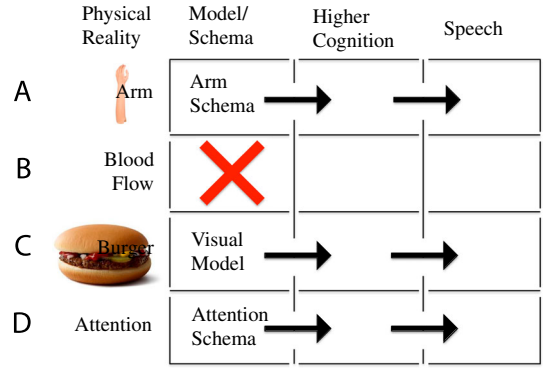

Figure 1. How the brain might model, gain cognitive access to, and speak about, objects or processes in the real world. (A) The arm schema represents a real arm and allows us to think about and talk about our limbs. (B) The complex pattern of blood flow in the brain lacks a model in the brain and thus we cannot directly introspect and report about it, though we can learn about it intellectually. (C) Visual objects such as hamburgers are modelled in the visual system. Information from the visual model can reach cognition and speech. (D) In the attention schema theory, the physical process of attention is represented in the brain by a model, allowing us to directly introspect about, think about, and talk about it. When we do so, we describe it as subjective awareness. The model can make mistakes, hence awareness can sometimes dissociate from attention. [To view this figure in colour, please see the online version of this journal.]

Higher cognitive systems such as working memory can receive at least some of the information from the arm model (see Figure 1A). From higher cognition, at least some of the information can be used by language machinery and turned into speech (again, see Figure 1A).

Because of the pathway diagrammed in Figure 1(A), you can close your eyes, thereby blocking any visual confirmation, and still introspectively know about your arm, think about your arm, and talk about it. You can accurately say, "My arm is at my side. Now it's up in the air".

This account of the arm schema and its relationship to cognition and speech is no doubt oversimplified. The boundaries between processes in the brain are more continuous and less modular than indicated in Figure 1(A), and there are a lot more processes than

are represented here, but the overarching description is essentially correct. Note that I have said nothing about consciousness so far. I have described, in effect, a machine that reconstructs and reports information about its physical body.

I want to point out three obvious properties about the arm and its representation. First, the information becomes impoverished at each step. The real arm is richer in detail and nuance than the model; the model is richer than the information passed to higher cognitive systems; and even less detail and richness can be expressed in speech.

Second, the arm schema is automatic. You can't choose to construct it or choose not to; it's not culturally learned; it's not an intellectual construct; instead, built-in machinery continuously constructs it. Sometimes it reaches your higher cognition and sometimes not, but it is always there.

Third: let's talk about illusions. You can have an illusion if there is a temporary mismatch between the arm schema and reality. A phantom limb is an illusion. So is a laboratory-contrived distortion of the arm schema. A correctly functioning internal model, however, is not normally considered to be an illusion.

Some people may suggest that, philosophically, you could call the arm schema an illusion in all circumstances, because it's never a perfect, or perfectly detailed, representation of the arm. There is always a gap between reality and the model constructed in the brain. Using the term "illusion" in that more inclusive sense makes a philosophical point, but I prefer not to use the word in that way. That philosophical emphasis paints a picture in which the arm schema is a separate entity unmoored to reality, a construct of the brain, a way the brain tricks itself. But the arm schema is anchored to reality and has a specific functional purpose: to represent the arm. When it is doing its job effectively, it is not well described as an illusion. It's a model. It's a representation. It's the brain's useful caricature of the arm. Given the constraints on the brain's processing ability, it is necessarily always true that what we think we have and say we have is a distortion or simplification of what we actually have.

For the sake of clarity through the rest of this piece, I will explain what I mean by four specific terms, corresponding to the four columns, from left to right, in Figure 1. First, when I say something is "really" present, I mean that an actual item in a physically real universe exists; in the case of Figure 1(A), an arm. When I say that you "intuit" something, I use the word in an extremely narrow sense, and mean that the brain has constructed an automatic model of it (the arm schema in Figure 1A). The model is present outside of higher cognition, which can sometimes access it. When I say that you "think" something, I mean that higher cognition has accessed the information and is holding or manipulating it. When I say that you "say" or "claim" something, I am referring to the speech machinery that allows you to make a verbal report (the final box on the right side of Figure 1A).

Now, armed with these terms, let us consider a different physically real object: the pattern of blood flow in the brain (see Figure 1B). The brain has a complex, constantly changing pattern of flow through arteries, veins and capillaries. Evidently, however, the brain does not construct a model to monitor its own blood flow. Or, if such a model exists, it has not been discovered yet, and we have no cognitive access to it. People cannot intuit their own brain's blood flow. We can study it scientifically, read about it in a book, and supply cognition with academic information. We can form intellectual constructs. Then we can think and talk and write about it, as I'm doing now. But people cannot gain cognitive access to an automatic, ongoing model of their own brain's blood flow. The brain has many processes, like blood flow, that objectively exist but have no internal model or direct pathway to cognition and speech. The lesson here is: just because the brain has something, does not mean that we can intuit that we have it, think we have it or say we have it.

Figure 1(C) shows an example of an object that the brain does sometimes model. In this case, the object is a visual stimulus, a hamburger. The visual system in your brain constructs a model filled with visual details, texture and colour – a rich bundle of information. Presuming you attend to the hamburger, boosting the strength on that visual model, your higher cognition may also have access to that information, allowing you to think about and talk about the hamburger.

Now let's consider a final example of a physically real item that the brain can potentially model: attention. Diagrammed in Figure 1(D) is the real item (attention), the brain's model of it (the attention schema), cognitive access to the model and the possibility of verbal report. The same pathway, from a real item

through intermediate steps to a report, is diagrammed here for attention as for concrete objects such as a hamburger or an arm.

I do not like the word attention – it has too many colloquial connotations – but I don't know of a better label. I mean something mechanistic: selective signal enhancement and the consequent deep processing and impact on behaviour. We hypothesized that not only does the brain *have* attention but also builds a model of it: an attention schema. A growing body of data suggests that some type of an attention schema does indeed exist. Some of the evidence is briefly described and cited in the target article (Graziano et al., 2020; see also Guterstam et al., 2018; Pesquita et al., 2016; Tsushima et al., 2006; Vernet et al., 2019; Webb et al., 2016).

A basic principle of control engineering is that, if you want a system to control something, you should give the system a model of the thing to be controlled. To control the arm, the brain needs an arm schema. To control attention, the brain should, in theory, use an attention schema. Attention is, in some ways, like a hand. It moves from item to item, location to location. Your mind grasps the hamburger, or a sound, or a memory. Attention, however, has many more degrees of freedom than an arm. Attention can be spread or focused, it can be directed in space or to abstract features like color and shape, and attention can move through domains that have nothing to do with the external world – you can direct attention to internal thoughts and memories.

Suppose you are attending to the hamburger. To dispel a common misconception, the attention schema is *not* information about the hamburger – that is the job of the hamburger model. The attention schema is information about the dynamics of attention itself. Was attention exogenously attracted, or endogenously directed? How fast does attention move through space? How wide or focused is it? How intense is it? Is it being siphoned to a distractor, away from the desired target? What are the likely consequences of attention on memory? On decision-making? On movement control? What can attention *do*? The attention schema, proposed by AST, is a model of attention itself; it treats attention as an active, dynamic object.

How can we recognize an attention schema? Has it been discovered previously, without being recognized as an attention schema? To recognize it, we can search for a specific trait: an attention schema should correlate closely with attention. If you can find evidence of a process that typically co-varies with attention, but that isn't attention – that can sometimes become dissociated from attention – then that thing is a candidate for an attention schema. In the same way, the arm schema closely tracks the arm, and yet can sometimes slip and dissociate from the arm.

One process that fits this bill is subjective awareness. Usually, attention and awareness covary. Nothing showcases that tight relationship more than inattentional blindness – the many demonstrations that when you are not attending to something, you are not aware of it (Drew et al., 2013; Mack & Rock, 2000; Simons & Chabris, 1999). Attention and awareness track each other most of the time. In my experience, having studied both, it is easier to separate the arm from the arm schema than it is to separate awareness from attention. A common misconception is that attention and awareness are frequently separated, because we are aware of objects in peripheral vision (Gennaro, 2020; Rosenthal, 2020). You can stare directly at object A and still be aware of object B to the side. However, this claim is incorrect and is based on a misconception of what attention is. Attention is not the same as foveation. Covert attention can spread and move around the visual field, from the fovea to the periphery. You are most likely aware of object B because your covert attention has at least partly moved to it. Without some covert attention directed to your peripheral vision, object B is likely to disappear from your awareness – unless it gives a jump or a flicker, or has a high contrast or salience, pulling exogenous attention to itself. Inattentional blindness shows how astonishingly unaware we are of objects all around us, in the field of view, that do not receive at least some attention. Because attention is constantly shifting and moving, spreading, sending out tendrils and refocusing in a protean way across the visual world, we are typically aware of many items both at the fovea and away from it, sometimes at the same time and sometimes in series. Some exceptions do exist, but for the most part, attention and awareness covary; just like, with some exceptions, the arm and the arm schema covary.

We might also recognize an attention schema based on a second characteristic: higher cognition should be able to gain access to an attention schema, and linguistic machinery should be able to

verbally report on it (as diagrammed in Figure 1D). Moreover, when we do think about it and report on it, because we're drawing on the contents of a model, we should find ourselves describing something that superficially resembles attention but that is not exactly the same. What we describe should be a schematized version, lacking the microscopic physical details of real attention.

Again, awareness fits the bill. If you start with attention and ignore its physical details such as neurons, competition between signals, and specific brain areas, and instead give a kind of shell description of it, you'd be left with something suspiciously similar to awareness. That shell description would depict something ethereal in character because it has been stripped of the materialistic details; it would depict a magical mental possession of objects and ideas that gives us the ability to understand and react. This similarity between awareness and a detail-stripped depiction of attention is more completely described in a series of points in the target article (Graziano et al., 2020).

To me, there is very little wiggle room. If it walks like a duck and quacks like a duck, it's a duck. Awareness walks and quacks like an attention schema. Or rather, we *intuit* that we have awareness, we *think* we have awareness, and we *say* we have awareness, because the brain has an attention schema, higher cognitive access to that model, and a linguistic output. AST explains why we intuit and think and say we have awareness, why awareness superficially resembles attention, and why awareness tracks attention at least as closely as the arm schema tracks the arm.

The attention schema does not contain all of consciousness – and yet it contains a crucial piece. To understand what I mean, imagine being aware of the sight of a hamburger; or aware of the sound of a bird; aware of your arm; aware of thinking that 2 + 2 = 4; aware of your own rich memories of yesterday; aware of happiness. Now consider what is common across all those instances of awareness. None of the details are the same, none of the content. Some of those instances involve self-awareness and some involve sensory awareness of the outside world. What is common, in each case, is a subjective experience of something. The sheen of experience, the essence of awareness, the seeming inner eye, the inner feel, is the same. In AST, the attention schema is the information set that tells the system that a property of subjective experience is present. One might say it depicts "experienceness" if such a word is allowed. In the statement, "I am aware of X", the attention schema supplies the information behind the "am aware of" part. Other models supply the vast sets of information behind the "I" and the "X" part.

Consider the case of the hamburger (Figure 1C). The brain constructs a model of the burger, in rich detail. The burger is attended, such that the visual model is enhanced in signal strength and can affect cognition and language. As a result, the machine can report the presence of a burger and describe its details. In AST, a model of attention (Figure 1D) supplies the extra information on the basis of which the machine can claim to have a subjective, what-it-feels-like component superadded to the visual details. One might say the visual qualia of the burger's colour and shape lie at the union of the burger model and the attention schema. Or, more precisely, the system thinks and claims to have a specific burger quale because higher cognition and speech are drawing on the information contained in that combination of the visual model and the attention schema.

Concern 1: But isn't consciousness real?

Of the three concerns that I outlined at the start of this piece, the first was that AST denies conscious experience rather than explaining it. That concern, which has been expressed in many ways by many people, could be put like this: "I have a subjective, conscious experience. It's real; it's the feeling that goes along with my brain's processing of at least some things. I *say* I have it and I *think* I have it because, simply, I *do* have it. Let us accept its existence and stop quibbling about illusions. Our primary question as consciousness researchers is: how is that inner feeling generated?"

That approach, as ubiquitous and as tempting as it sounds, is logically incorrect. It naïvely mishandles the multi-step relationship between having something and thinking you have it. To explain why it is incorrect, let us start with a premise and see where it leads us. Suppose the brain has a real consciousness. Logically, the reason why we *intuit* and *think* and *say* we have consciousness is not because we actually have it, but must be because of something else; it is because the brain contains information that describes us having it. Moreover, given the limitations on the brain's

ability to model anything in perfect detail, one must accept that the consciousness we intuit and think and say we have is going to be different from the consciousness that we actually have. Similar, perhaps, but different, just as the arm schema differs from the arm that we actually have. I will make the strong claim here that this statement – the consciousness we think we have is different from, simpler than, and more schematic than, the consciousness we actually have – is necessarily correct. Any rational, scientific approach must accept that conclusion. The bane of consciousness theorizing is the naïve, mistaken conflation of what we actually have with what we think we have.

The attention schema theory systematically unpacks the difference between what we actually have and what we think we have. In AST, we really do have a base reality to consciousness: we have attention – the ability to focus on external stimuli and on internal constructs, and by focusing, process information in depth and enable a coordinated reaction. We have an ability to grasp something with the power of our biological processor. Attention is physically real. It's a real process in the brain, made out of the interactions of billions of neurons. The brain not only uses attention, but also constructs information about attention – a model of attention. The central hypothesis of AST is that, by the time that information about attention reaches the output end of the pathway (the right side of Figure 1), we're claiming to have a semi-magical essence inside of us – conscious awareness. The brain describes attention as a semi-magical essence because the mechanistic details of attention have been stripped out of the description. AST, therefore, offers a specific hypothesis about the relationship between the consciousness we think we have and the consciousness we actually have.

In our target article, we used the terminology of i-consciousness (the consciousness we objectively and physically have) and m-consciousness (the model of i-consciousness, the mental essence that we intuit and think and say we have). The point of that terminology is to explicitly make the link between the real object and the information about it in our intuition, cognition and speech.

Many of the commentaries on the target article suggested that m-consciousness can be mapped onto phenomenal consciousness and i-consciousness onto access consciousness (Brown & LeDoux, 2020; Frankish, 2020; Gennaro, 2020; Masciari & Carruthers, 2020; Vernet et al., 2020). The two terminologies are similar, and I admit that, for clarity, we should have more explicitly compared the two, although we did explicitly acknowledge that other researchers used terminology different from ours. We chose to use our own terminology to avoid confusion or hidden assumptions that might attach to previous terminology. Access consciousness is the higher cognitive access to and manipulation of information in the brain, although it was originally not necessarily precisely defined and thus there is room for debate about the exact meaning (Block, 1996). In the present account, i-consciousness is attention. It is a specific type of information processing inside the brain characterized by selective enhancement of signals and enhanced broadcasting of information around the brain (the global workspace). Our i-consciousness is therefore a more limited or specifically defined process than the range of cognitive processes sometimes assigned to access consciousness. Phenomenal consciousness is supposed to be an essentially non-physical, personal experience (Block, 1996). Our m-consciousness is closely related. It is a construct, a kind of convenient if imperfect picture the brain builds, to usefully represent i-consciousness. In order to present our own specific argument as clearly as possible, we used our own terminology.

Some of the commentaries suggested that i-consciousness is not consciousness, has nothing to do with consciousness, and shouldn't be labelled as such (e.g., Blackmore, 2020; Rosenthal, 2020). Other commentaries were in agreement with us that i-consciousness is the "real" or "objective" consciousness inside us (e.g., Panagiotaropoulos et al., 2020; Romo & Rossi-Pool, 2020). The central point of AST is the close relationship between i-consciousness and m-consciousness, and that is why we used the same word to refer to both. Imagine you spend a lifetime living in a house, where a picture of your house hangs on the living room wall. Somehow, through all the years, you haven't noticed that the picture is a representation of the house. Now, to your confusion, I'm telling you that they correspond. You have a real house (r-house) and a picture house (p-house). You have a real consciousness (i-consciousness) and a picture of it that the brain constructs for itself (m-consciousness). That is the heart of AST. Those who argue that i-consciousness is not really consciousness, and

wonder why we labelled it that way, miss the point of AST, which is precisely that i-consciousness *is* the real entity from which reports of m-consciousness derive.

Concern 2: But isn't consciousness an illusion?

The second concern that I outlined at the start of the article is about AST as an illusionist theory (e.g., Blackmore, 2020; Dennett, 2020; Frankish, 2020). Illusionist theories emphasize how subjective awareness does not really exist – the brain tricks itself into thinking it has it. Obviously, AST aligns with that perspective. So why not more forcefully admit that AST is an illusionist theory? The reason is that the word "illusion" turns the focus away from the most important concept in AST.

If I say, "I have an arm", and I am not an amputee, then I am not suffering an illusion. I really do have an arm. If I look at a real hamburger and say, "There's a burger", I'm not suffering an illusion; it represents a real burger. Nobody uses the word "illusion" to refer to those instances, even though the brain's models of the arm and the burger are mere caricatures. According to AST, when I say, "I have awareness", I really do have the base reality; I have attention. Calling consciousness an illusion obscures the central point of AST, the importance of attention as the base reality. Awareness is not merely the brain tricking itself. In AST, awareness is a functional account of attention, modified and simplified to fit data constraints.

I do not mean to attack the illusionist perspective. Many illusionist philosophers would look at my account of AST and say, "that's exactly what we mean by consciousness as an illusion. Some aspects of consciousness are a distorted internal account of what we actually have". For that reason, I do not mind if people call AST an illusionist theory. To me, however, the illusionist language sounds unnecessarily dismissive of m-consciousness. The point of AST is to do the opposite – to emphasize the usefulness of m-consciousness as a quick-and-dirty model of i-consciousness. Rather than say that consciousness is an illusion, I would say m-consciousness is a caricature. One defining property of a caricature is that it implies a real object that is being caricatured. A second defining property is that it is a simplification and distortion of the object being caricatured. A third property is that a caricature is made for a reason – it is typically put to some kind of use. To say, "m-consciousness is a caricature of attention", as a slogan, may have less of a rhetorical ring than, "consciousness is an illusion", but it much more closely captures AST.

Concern 3: Why the focus on attention?

The third and most common concern is: why does AST link consciousness specifically to an attention schema (Frankish, 2020; Gennaro, 2020; Lane, 2020; Metzinger, 2020; Panagiotaropoulos et al., 2020; Prinz, 2020)? Why not a memory schema? Or a decision-making schema? Or a mental imagery schema? Or a response schema? Why not just say: consciousness is a mind schema?

I agree with the general idea. The brain must model many aspects of itself. If by "consciousness" you mean the broader content in the mind, then, of course, consciousness contains models of many things far beyond attention. AST, however, deals in one specific component that plays a special role. It is as if, in explaining how a car works, we decided today to focus on the central role of the spark plugs, without dismissing the importance of the rest of engine.

To return to a point made earlier in this piece, all instances of consciousness share a feature: experience. Whether self-awareness or awareness of external stimuli, whether memory or sight or pain, why do we attach "experienceness" to all of these instances? AST does not address what makes experiences richly different from each other. It addresses the commonality, the overarching claim to subjective experience.

The reason why the theory links awareness specifically to attention is straightforward. As noted in the previous sections, subjective awareness resembles attention in almost all its superficial, general properties, its dynamics and consequences. It also correlates moment-by-moment with attention. Almost always, what you attend to, you're aware of, and what you're not attending to, you're not aware of. Awareness does not correspond in the same way to any other function that I know of – not to memory, or emotion, or decision-making or mental imagery. We can sometimes be aware *of* those things. But those items do not covary with awareness. The idea of a response schema (Frankish, 2020) is excellent. The idea of schemas for other internal processes (Panagiotaropoulos et al., 2020; Prinz, 2020) is exactly right. When people make the claim, "I am aware of X", they surely rely on a model of the "I" component

(Lane, 2020). The brain must construct many other self-schemas and we should study them as part of the fabric of consciousness. But awareness has a special relationship to attention. As the arm schema tracks the arm, so awareness tracks attention.

One of the common ways to try to knock down AST is to point out situations in which awareness does not correspond to focused attention on an external stimulus, and then to argue that, therefore, awareness cannot be a model of attention. That counterargument is incorrect. First, attention can be directed to internal states as well as to external stimuli. Whatever it is you are aware of, chances are good that you are also directing some attention to it. Second, the theory does not require that awareness and attention *always* match. That would be like arguing the arm schema does not represent the arm because one can find cases in which the two do not match, such as the case of the phantom limb. Of course the arm schema and the arm can sometimes be dissociated – that is how we know about the existence of the arm schema. Just so, of course awareness and attention should dissociate, perhaps especially in fringe situations like the edge of sleep, meditation or drug-induced states, when the normal mechanism slips. The crucial experimental evidence is not that awareness and attention are *always* in lock-step, and not that attention is proven to be necessary or sufficient for awareness, but that awareness closely tracks attention with relatively little slippage, much as any model in the brain tracks the thing it models. One of the few facts about awareness that has overwhelming experimental support, as noted above, is that attention and awareness closely – but not perfectly – covary.

I noted above that AST was something like the spark-plug theory of how an engine works. It addresses an important component of the machine without dismissing the importance of the rest of engine. AST says that consciousness depends on a particular piece, an attention schema, plugged into the larger system. That piece does not contain the contents of consciousness. The brain must also construct models of color, pain, emotion, self, memory, response and many other items. The attention schema is the piece that allows us to intuit and think and say, "And subjective experience is also present". Without that piece added to the larger system, the very idea of subjective experience would become irrelevant to us. We would not even know what it is, and would not be able to attribute it to ourselves.

Is an attention schema evolutionarily old or unique to humans?

In the final sections, I will address two specific questions that arose in the commentaries. First, do non-human animals have an attention schema (Dennett, 2020)? Here I will take a definite stand: yes. Many must, or they would be unable to control their attention in basic and necessary ways. Any creature that can endogenously direct attention must have some kind of attention schema, and good control of attention has been demonstrated in a range of animals including mammals and birds (e.g., Desimone & Duncan, 1995; Knudsen, 2018; Moore & Zirnsak, 2017). My guess is that most mammals and birds have some version of an attention schema that serves an essentially similar function, and contains some of the same information, as ours does. Just as other animals must have a body schema or be condemned to a flailing uncontrolled body, they must have an attention schema or be condemned to an attention system that is purely at the mercy of every new sparkling, bottom-up pull on attention. To control attention endogenously implies an effective controller, which implies a control model.

Therefore, in AST, just as animals "know" about their own bodies in some deep intuitive sense via their body schemas, they also "know" about a subjective experience inside of them (a detail-poor depiction of their attentional state) via an attention schema. They may, however, lack higher cognitive levels of reflection on those deeper models.

Dennett (2020) suggests that only humans need an attention schema and that dogs do not. I think perhaps the difference in opinion here relates to higher level and lower level models. Humans undoubtedly have layers of higher cognitive models, myths and beliefs and cultural baggage. Much of the ghost mythology that we discussed in our target article (Graziano et al., 2020) is presumably unique to humans, exactly as Dennett suggests. But in AST, many of these human beliefs stem from, or are cultural elaborations of, a deeper model that is built into us and many other animals – an intrinsic model of attention.

Is the attention schema really a higher-order thought?

Some of the commentaries asked whether an attention schema is really an example of a higher-order thought (Brown & LeDoux, 2020; Frankish, 2020; Gennaro, 2020; Rosenthal, 2020). I agree with most of these commentaries that, yes, an attention schema is a kind of higher-order thought, although AST is not typical of most higher-order thought theories.

The ambiguity might stem partly from two different ways to think about higher order. Consider again the burger in Figure 1(C). The visual system constructs a lower-order, perception-type representation. Some of the information in that representation might then reach higher cognition. Here, lower and higher are being used in a specific way: lower is automatic, obligatory, fixed, perceptual, probably represented in sensory brain areas. Higher is flexible and cognitive, intentional, probably represented partly in prefrontal cortex. But this type of "lower" and "higher" is not the same as in the higher-order thought theory. There, a higher-order thought is meta information – information about how information is handled in the brain.

In AST, the attention schema is higher-order in the sense that it is a representation of attention. It is a representation of how information in handled in the brain. But it is also lower-order in a sense. Note that in Figure 1, it is diagrammed at the same level as the visual model of the burger and the arm schema. A person can't intellectually choose to construct an attention schema, and can't choose not to; it is automatic and obligatory, and higher cognition has partial access to it. This ambiguity, in which the attention schema is higher-order in one sense and lower-order in another, is one way in which AST differs from at least some formulations of the higher-order thought theory.

Another difference between AST and many versions of the higher-order thought theory lies in the core explanation of consciousness. In some higher-order thought formulations, awareness or qualia are enabled or lit up when lower-order representations become the target of higher-order thoughts. In those formulations, the higher-order thought theory does not really explain consciousness so much as hypothesize that a mystery of consciousness is switched on under certain circumstances. In AST, conscious experience does not light up or emerge in that manner. The process is fundamentally different. As diagrammed in Figure 1, the physical phenomenon of attention is modelled imperfectly by an attention schema. Following from that model, we intuit, think, and claim to have a subjective experience.

Disclosure statement

No potential conflict of interest was reported by the author(s).

References

Blackmore, S. (2020). But AST really is illusionism. *Cognitive Neuropsychology*. [Epub ahead of print].

Block, N. (1996). How can we find the neural correlates of consciousness? *Trends in Neuroscience*, *19*(11), 456–459. https://doi.org/10.1016/S0166-2236(96)20049-9

Brown, R., & LeDoux, J. (2020). Higher order memory schemas and conscious experience. *Cognitive Neuropsychology*. [Epub ahead of print].

Dennett, D. (2020). On track to a standard model. *Cognitive Neuropsychology*. [Epub ahead of print].

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, *18*(1), 193-222. https://doi.org/10.1146/annurev.ne.18.030195.001205

Drew, T., Võ, M. L., & Wolfe, J. M. (2013). The invisible gorilla strikes again: Sustained inattentional blindness in expert observers. *Psychological Science*, *24*(9), 1848-1853. https://doi.org/10.1177/0956797613479386

Frankish, K. (2020). Consciousness, attention, and response. *Cognitive Neuropsychology*. [Epub ahead of print].

Gennaro, R. (2020). *Cognitive Neuropsychology*. [Epub ahead of print].

Graziano, M. S. A., Guterstam, A., Bio, B. J., & Wilterson, A. I. (2020). Toward a standard model of consciousness: Reconciling the attention schema, global workspace, higher-order thought, and illusionist theories. *Cognitive Neuropsychology*. [Epub ahead of print].

Guterstam, A., Kean, H. H., Webb, T. W., Kean, F. S., & Graziano, M. S. A. (2018). Implicit model of other people's visual attention as an invisible, force-carrying beam projecting from the eyes. *Proceedings of the National Academy of Sciences, U. S. A.*, *116*(1), 328–333. https://doi.org/10.1073/pnas.1816581115

Knudsen, E. I. (2018). Neural circuits that mediate selective attention: A comparative perspective. *Trends in Neurosciences*, *41*(11), 789–805. https://doi.org/10.1016/j.tins.2018.06.006

Lane, T. (2020). Somebody is home. *Cognitive Neuropsychology*. [Epub ahead of print].

Mack, A., & Rock, I. (2000). *Inattentional blindness*. MIT press.

Masciari, C., & Carruthers, P. (2020). What explains the hard problem of consciousness? *Cognitive Neuropsychology*. [Epub ahead of print].

Metzinger, T. (2020). Self-modeling epistemic spaces and the contraction principle. *Cognitive Neuropsychology*. [Epub ahead of print].

Moore, T., & Zirnsak, M. (2017). Neural mechanisms of selective visual attention. *Annual Review of Psychology*, *68*(1), 47–72. https://doi.org/10.1146/annurev-psych-122414-033400

Panagiotaropoulos, F., Wang, L., & Dehaene, S. (2020). *Cognitive Neuropsychology*. [Epub ahead of print].

Pesquita, A., Chapman, C. S., & Enns, J. T. (2016). Humans are sensitive to attention control when predicting others' actions. *Proceedings of the National Academy of Sciences, U. S. A*, *113*(31), 8669–8674. https://doi.org/10.1073/pnas.1601872113

Prinz, W. (2020). The social roots of consciousness. *Cognitive Neuropsychology*. [Epub ahead of print].

Romo, R., & Rossi-Pool, R. (2020). Toward a conscious model of consciousness. *Cognitive Neuropsychology*. [Epub ahead of print].

Rosenthal, D. (2020). Competing models of consciousness. *Cognitive Neuropsychology*. [Epub ahead of print].

Simons, D. J., & Chabris, C. F. (1999). Gorillas in our midst: Sustained inattentional blindness for dynamic events. *Perception*, *28*(9), 1059–1074. https://doi.org/10.1068/p281059

Tsushima, Y., Sasaki, Y., & Watanabe, T. (2006). Greater disruption due to failure of inhibitory control on an ambiguous distractor. *Science*, *314*(5806), 1786–1788. https://doi.org/10.1126/science.1133197

Vernet, M., Japee, S., Lokey, S., Ahmed, S., Zachariou, V., & Ungerleider, L. G. (2019). Endogenous visuospatial attention increases visual awareness independent of visual discrimination sensitivity. *Neuropsychologia*, *128*, 297–304. https://doi.org/10.1016/j.neuropsychologia.2017.08.015

Vernet, M., Quentin, R., Japee, S., & Ungerleider, L. (2020). From visual awareness to consciousness without sensory input: The role of spontaneous brain activity. *Cognitive Neuropsychology*. [Epub ahead of print].

Webb, T. W., Kean, H. H., & Graziano, M. S. A. (2016). Effects of awareness on the control of attention. *Journal of Cognitive Neuroscience*, *28*(6), 842–851. https://doi.org/10.1162/jocn_a_00931

Yankulova, J., & Morsella, E. (2020). Conscious contents: Their unanalysable, arbitrary, and unarbitrary properties. *Cognitive Neuropsychology*. [Epub ahead of print].