

COMPETING MODELS OF CONSCIOUSNESS

David Rosenthal
Philosophy and Cognitive Science
Graduate Center, City University of New York

REVISED PENULTIMATE VERSION

Word count: 2,248

Graziano and colleagues (forthcoming) sketch some suggestive ways in which Graziano's attention-schema theory (AST) of consciousness (Graziano 2013, 2019; Webb and Graziano 2015) connects with three other theories: the global-workspace theory (GWT) of Baars (1997) and Dehaene and Naccache (2001), higher-order-thought (HOT) theory (Rosenthal 2005), and illusionism (Frankish 2016).

According to Graziano and colleagues, these "theories should not be viewed as rivals, but as partial perspectives on a deeper mechanism" (1). The deeper mechanism they have in mind is that of AST, which they urge constitutes "a kind of standard model that covers the basics if not the details" of consciousness (1). I'll raise difficulties for this irenic construal of these four theories, and argue that they are in important ways theoretical competitors.

Graziano and colleagues see the term 'consciousness' as applying to two distinct phenomena. One is "how select information reaches a state in which it is bundled centrally and can impact output systems." They call this i-consciousness, in contrast with what they call m-consciousness, "a more mysterious, extra, experiential essence that people claim accompanies the informational content" that occurs in i-consciousness (2). The 'm' of 'm-consciousness' reflects their view that this notion is inherently mysterious.

With that distinction in place, Graziano and colleagues urge that the central issue for a theoretical treatment of consciousness is "whether m-consciousness, the more ethereal notion of consciousness that people intuitively believe they have, is accurate or instead is an imperfect model of i-consciousness" (2). I'll raise concerns about both these notions.

To begin with, i-consciousness is not something we would naturally classify as any kind of consciousness at all. I-consciousness is exclusively an information-processing notion, and all the relevant processing can occur without being in any way conscious. Indeed, Graziano and colleagues themselves refer to i-consciousness as merely "the content of consciousness" (2), and the content of consciousness is typically something that can occur either consciously or not.

As Graziano and colleagues note, GWT is in effect "an account of i-consciousness" (2). So the foregoing is also a difficulty for that theory. GWT holds that a state is conscious if its content reaches a global workspace, making it available to downstream processing across various modules. But being globally available is not itself a way of being conscious.

There is doubtless considerable overlap between the states we classify as conscious and those that are globally available for downstream processing. But even if that overlap were total, there would be no reason to see being globally available as a way for a state to be conscious.

And the overlap is far from total. Visual states remain conscious well into the periphery, however scant in detail, and conscious peripheral states are rarely available for global processing. Similarly for conscious background auditory sensations. And unconscious thoughts and desires sometimes have global effects. These considerations also apply to Block's (1995) notion of access consciousness. I-consciousness and access consciousness are not types of consciousness at all.

Graziano and colleagues hint at this when they note that GWT "leaves unexplained how people end up believing they have a subjective experience" (13). Their oblique formulation, about "people ... believing they have a subjective experience," is because they see subjective experience as m-consciousness, which they see in turn as mysterious and ethereal. But the difficulty about i-consciousness is worse than that. I-consciousness is independent of anything we naturally classify as consciousness.

Because GWT can't explain why people believe they have subjective experience, Graziano and colleagues invoke something they think of as a higher-order (HO) theory. They see HO theory in its most general form as holding that one's being aware of a stimulus "requires higher-order information about one's own internal processes in addition to lower-order information about the stimulus." And noting that there is debate about just "what that higher-order information is" (13), they construe it along the lines of AST.

The information processing of i-consciousness, they urge, is dominated by attentional factors. So the HO information they posit is "[a] model of one's own attention" (13), of the attentional processes that dominate and guide i-consciousness. That model is the attention schema of AST, construed by "analogy to the body schema that is theorized to contribute to the control of movement" (8). Since this attention schema is a model of attentional processes that figure in i-consciousness and GWT is a theory of i-consciousness, GWT provides a theory of the "lower-order information about the stimulus." And because they posit the attention schema as a HO model of that lower-order information, they present AST as a synthesis of GWT with a type of HO theory.

Graziano and colleagues see this synthesis as supporting their claim that AST, GWT, and HO theory "should not be viewed as rivals, but as partial perspectives on a deeper mechanism." But the HO theory that figures in this synthesis only nominally resembles contemporary HO theories of consciousness. Those theories rely not on an internal model of information processing, but on our awareness of psychological states that we naturally classify as conscious. HO theories rely on what I have called (2005) the transitivity principle, which holds that a psychological state is conscious only if one is in some suitable way aware of that state. Suppose somebody is in some psychological state but is in no

way aware of being in it, and would sincerely deny its presence. That psychological state is plainly not a conscious state. And that implies the transitivity principle.¹

On the transitivity principle, a state's being conscious consists in one's being aware of one's being in that state. But there is no circularity in appealing to that higher-order awareness to explain consciousness. On higher-order-thought theory, one is aware of the state by having a thought, independently of any conscious inference, that one is in that state; others have appealed to a higher-order perception. Since thoughts and perceptions need not occur consciously, we can explain them independently of consciousness. And the relevant HO thoughts or perceptions would themselves almost never be conscious; their being conscious would require a third-order state that resulted in one's being aware of them, and that would be very rare.

Though HO theorists do invoke different ways of being aware of conscious states, Graziano and colleagues are wrong to claim that there is debate about what "higher-order information" such theories posit. That information is always that one is in the target state. By contrast, the attention schema of AST is just an internal model of attentional factors that dominate psychological processing; it involves no higher-order awareness of the sort invoked by HO theories. Contemporary HO theories simply do not figure in the synthesis proposed by Graziano and colleagues.

The transitivity principle, moreover, is well-motivated. Since a state of which one is totally unaware fails to be conscious, some awareness of a state is necessary for it to be conscious. There is no reason, by contrast, why an attention schema would result in any kind of consciousness. Simply being a HO schema is not enough. Graziano and colleagues urge that "[a] model of one's own attention would be useful for predicting one's own behaviour, and would also be useful for regulating attention itself" (13). But such predicting and regulating need not occur consciously. Predicting one's own behavior is sometimes just having unconscious expectations about that behavior, and most regulating of attention occurs unconsciously. There is no reason to expect an attention schema by itself to result in consciousness.

But Graziano and colleagues don't in fact think that the attention schema does result in consciousness. Rather, it results just in m-consciousness, the "imperfect model of i-consciousness" they regard as so mysterious. The attention schema results not in subjective experience itself, but only in "people ... believing they have a subjective experience," in an illusory belief in an ethereal, mysterious m-consciousness.

As Graziano and colleagues note, the idea that we need explain only people's beliefs in subjective experience and not subjective experience itself "resembles the illusionist view of consciousness" (4) (Frankish 2016). And folding in that view completes their proposal to reconcile the four theories of consciousness. According to illusionism, there are no phenomenal properties as conceived by theorists such as Block and Nagel (1974), and the

¹ Indeed, one advocate of GWT insists on the importance of the higher-order awareness invoked by the transitivity principle for a state to be conscious (Naccache 2018, p. 2).

sense that there is something it's like to have conscious experiences is an illusion. Illusionism does not claim there are no experiences; it rejects only the tacit theoretical implications of the notions many theorists invoke of phenomenal consciousness and of there being something it's like to have experiences. Similarly, Graziano and colleagues regard m-consciousness as in effect illusory, as "a simplified, cartoon version of the social world in which agents ... possess an invisible, energy-like or plasma like mental essence" (3).

Graziano and colleagues rely on some experimental work (e.g., Guterstam et al 2019) to show that people do actually think of consciousness in such figurative, quasi-literary terms. But though subjects may sometimes describe their conscious experiences in these ways, it's unlikely that those descriptions literally reflect tacit theories they hold. Rather, people simply find it easiest to describe conscious experience in such evocative ways, relying on others to understand metaphor for what it is. Even when experimental work does elicit such figurative remarks, we should be highly cautious in taking those remarks to reflect literal beliefs.

And consciousness is in any case not an illusion. On our commonsense conception, consciousness is simply the way our mental lives subjectively appear to us. It is mental appearance. The stream of consciousness is a stream of appearances of psychological states that we seem to be in. And those appearances are themselves real; it is not illusory that we seem subjectively to be in various psychological states.

It may well be that we sometimes seem subjectively to be in states that we're not actually in; it may seem subjectively that we see something red when our actual visual state is of something orange. So the way our psychological states subjectively appear to us may in that way sometimes be illusory. But whatever disparity does occur between the psychological states we're actually in and those we subjectively seem to be in does not support seeing consciousness itself as illusory. The subjective appearances are real psychological occurrences.

Illusionism is right that the notions of phenomenal consciousness and phenomenal properties, as they occur in much current literature, do carry problematic theoretical baggage. That problematic baggage consists in holding that the property of a state's being conscious is intrinsic to the state's other mental properties, something not present in folk theory. But that theoretically contentious and optional claim can readily be rejected without the radical and implausible claim that consciousness itself is illusory (e.g., Rosenthal 2010). The problematic implications result only from holding that we know about mental qualitative character in the first instance from the way it is conscious. The apparent mystery about qualitative consciousness is due solely to that assumption, not to the figurative ways people sometimes describe their conscious experiences.

And that assumption is misguided. We know about mental qualitative character mainly from the role it plays in perceiving, and perceiving can occur without being conscious. Indeed, we cast even our first-person descriptions of what conscious experiences are like for us in third-person perceptual terms, by appeal to the types of object that perceptually

result in the relevant experiences. A satisfactory account of conscious qualitative character in terms of perceptual role need involve nothing mysterious, ethereal, or illusory (Rosenthal 2010; 2005, chs. 5-7).

Consciousness is the way our mental lives appear to us subjectively. Those subjective appearances are the higher-order awarenesses that HO theories posit to explain consciousness. There is nothing illusory about the occurrence of these higher-order awarenesses, and no need to posit an attention schema to explain consciousness.

REFERENCES

- Baars, Bernard J. (1997), In the Theater of Consciousness: The Workspace of the Mind, Oxford: Oxford University Press.
- Block, Ned (1995), "On a Confusion about a Function of Consciousness," Behavioral and Brain Sciences 18, 2 (June): 227-247.
- Dehaene, Stanislas and Lionel Naccache (2001), "Towards a Cognitive Neuroscience of Consciousness: Basic Evidence and a Workspace Framework," Cognition 79, 1-2 (April): 1-37.
- Frankish, Keith (2016), "Illusionism as a Theory of Consciousness," Journal of Consciousness Studies 23, 11-12: 11-39.
- Graziano, Michael S. A. (2013), Consciousness and the Social Brain, Oxford: Oxford University Press.
- Graziano, Michael S. A. (2019), Rethinking Consciousness: A Scientific Theory of Subjective Experience, New York: W. W. Norton & Company.
- Graziano, Michael S. A., Arvid Guterstam, Branden J. Bio, and Andrew I. Wilterson (forthcoming), "Toward a Standard Model of Consciousness: Reconciling the Attention Schema, Global Workspace, Higher-Order Thought, and Illusionist Theories," Cognitive Neuropsychology.
- Guterstam, Arvid, Hope H. Kean, Taylor W. Webb, Faith S. Kean, and Michael S. A. Graziano (2019), Proceedings of the National Academy of Sciences, 116, 1 (January 2): 328-333.
- Naccache, Lionel (2018), "Why and How Access Consciousness Can Account for Phenomenal Consciousness," Philosophical Transactions of the Royal Society B: Biological Sciences, 373, 1755 (September 9): 20170357.
- Nagel, Thomas (1974), "What Is It Like to Be a Bat?," The Philosophical Review 83, 4 (October): 435-50.

Rosenthal, David (2005), Consciousness and Mind, Oxford: Clarendon Press.

Rosenthal, David (2010), "How to Think about Mental Qualities," Philosophical Issues: Philosophy of Mind, 20 (October): 368-393.

Webb, Taylor W., and Michael S. A. Graziano (2015), "The Attention Schema Theory: A Mechanistic Account of Subjective Awareness," Frontiers in Psychology 6, Article 500 (April 23): 1-11.