International Journal of Clinical and Health Psychology 23 (2023) 100380

Contents lists available at ScienceDirect

International Journal of Clinical and Health Psychology

journal homepage: www.elsevier.es/ijchp

A biphasic relational approach to the evolution of human consciousness

Steven C. Hayes [a,*], Stefan G. Hofmann [b]

[a] University of Nevada, Reno, USA
[b] Philipps-Universität Marburg, Germany

ARTICLE INFO

Keywords:
Consciousness
Evolutionary science
Processes
Dimension

ABSTRACT

Background/objective: Human consciousness is arguably unique, and its features are hard to explain. Continuous and discrete accounts of consciousness are commonly viewed as incompatible, but both have limitations. Continuous accounts cannot readily account for what appears to be unique about human consciousness; discrete accounts have a hard time explaining how human consciousness could have evolved. The present position paper shows how both continuous and discontinuously elements can be combined.
Method: A biphasic model is constructed by unifying complex systems theory, the evolution of symbolic reasoning as a relational extension of human cooperation, and evolutionary science. The application of this approach to modern views of consciousness is then explored.
Results: Our analysis suggests that human consciousness may be viewed as a discontinuous event, that emerged from continuous foundations. This biphasic account contains processes that can be targeted clinically. For example, developmentally delayed children with problems in consciousness may be helped by targeting the processes the present account suggests are important at different levels of complexity.
Conclusions: This biphasic relational approach fits with the evolutionary record and with data on human cognitive development. It may be useful in guiding clinical intervention.

Introduction

The topic of human consciousness has gathered researchers' and theorists' attention from the beginning of the human sciences (e.g., James, 1890). It is widely agreed to be one of the most difficult scientific problems to solve (e.g., Chalmers, 1995).

Many scientific accounts of consciousness have taken on the burden of explaining how these human abilities evolved (e.g., Graziano, 2021). The evolution of human consciousness has commonly either been approached either as a continuous process (e.g., Ginsburg & Jablonka, 2019) or as a more discrete development (Baars, 1988). Both approaches have weaknesses. Continuous accounts cannot readily account for what appears to be unique about human consciousness. Discrete accounts have a more difficult time explaining how specific contextual and biopsychosocial processes have combined over time to foster human consciousness.

The weaknesses can become practically evident when human development goes awry. For example, suppose some children show notable deficits in expected features of consciousness such as self-knowledge or self-referencing (e.g., Grisdale, Lind, Eacott, & Williams, 2014). If an account is purely continuous, it may not explain how or why a typical developmental trajectory failed to occur. Conversely if it is purely discrete, it may not explain how to begin to remediate these deficits if they are based on more elemental skills.

We propose in this paper to combine modern behavioral and cognitive perspectives and to integrate them within an evolutionarily sensible model that is biphasic – that is, both continuous and discrete. Such a biphasic approach could be more useful in solving practical matters if it helps understand more of the processes that impact human consciousness and thus can help address human problems that may depend on these processes. That is particularly likely if ontogenetic development follows in some ways the phylogenetic and cultural development that led to human consciousness in the first place. While not a given, that possibility itself can be explored in applying the model.

Biphasic accounts can fit with evolutionary theory because the evolutionary record shows that gradual transitions are often punctuated by large steps (Frazzetta, 2011). Gould (1988) famously argued against thinking of evolution as the gradual directional change of a billiard ball as compared to the punctuated equilibrium of Galton's polyhedron, and indeed sometimes evolutionary steps are so large as to be virtually discontinuous or novel.

Viewing evolution through the lens of complexity theory helps integrate the incremental and discontinuous (e.g., Pigliucci, 2008) because phase changes in networks and subnetworks are regularly encountered in the analysis of complex systems (e.g., Paperin, Green, & Sadedin, 2011). For example, changes in the epigenome can produce sudden system failures to suppress transposable elements, resulting in very rapid genetic change (Zeh, Zeh, & Ishida, 2009).

* Corresponding author at: Behavior Analysis Program, Department of Psychology /298, University of Nevada, Reno, NV 89557-0062, USA.

https://doi.org/10.1016/j.ijchp.2023.100380
Received 3 October 2022; Accepted 10 February 2023
Available online 7 March 2023

Human consciousness appears to entail qualitatively distinct features as will be explored below but because all discontinuous changes in evolution must build in some ways on what went before, it is important to consider what continuous processes set the stage for these emergent qualities of human consciousness. The flight of birds provides an analogy. Flying or not flying is discrete but feathers evolved and diversified in non-avian dinosaurs well before the origin of either birds or flight due to continuous functions such as thermal insulation or water repellency, not their aerodynamic effects (Prum & Brush, 2002).

In the area of human consciousness, a continuous account may help avoid explanations that are evolutionarily implausible. For example, theories of human consciousness commonly point to the development of a relatively large and complex human brain as being key. A recent review (Sattin et al., 2021) found that the neural correlates of consciousness were the most frequently studied feature of the 29 theories of consciousness they identified. These neurobiological developments, however, had to come *before* human consciousness itself and the resulting survival advantages that human consciousness provides. Prenatal growth rates (a predictor of brain size) accelerated in hominid evolution soon after the human-chimpanzee divergence about six to eight myo (Monson, Weitz, Brasil, & Hlusko, 2022), becoming more human like than chimpanzee like about 1 myo, well before indicators of human consciousness emerge in the fossil record. Large, complex brains immediately create physiological resource demands and notable difficulties in childbirth, just to name two costs in terms of survival. In the context of an evolutionary time scale, these notable costs would have *prevented* the evolution of the large and complex brains necessary for human consciousness unless they also had more immediate and incremental positive impact on other functions. Those additional functions may be suggested by continuous processes that set the stage for consciousness per se.

Method for developing a biphasic approach

What follows in this position paper is an analysis of how human higher cognition and symbolic reasoning may have evolved, presented initially in the form of a continuous account of the evolution of awareness. We argue that in hominins symbolic reasoning may have emerged as a relational extension of human cooperation based on multi-level selection.

We then consider the challenge of human consciousness to psychology and argue that a continuous account cannot fully address these features. Finally, after a brief review of modern approaches to human consciousness we examine how self-relations enter into human consciousness. Our biphasic relational account addresses modern views of consciousness and provides an empirical avenue for its analysis.

The evolutionary continuity of awareness

Given that the very term "consciousness" pulls for attention to *human* experiences of the qualia of consciousness, we propose in this section to first focus on a central aspect of human consciousness that presents itself as an initial target for a continuous account: "awareness." Awareness varies from lower forms of cognition involving sensation or perception, to somewhat higher forms involving learning or expectancies, even before higher forms of cognition that are reached in human symbolic reasoning or to a self-reflective sense of conscious or sense of self. We can think of this continuity of awareness as simply *the relative ability to respond to oneself and the internal and external environment, and the regularities within and between them* (Hayes, 2018).

Stated that way, the evolution of greater awareness is highly likely to be the product of the evolution of increasingly complex life forms. Sensation provides an example. Eyes appear to have independently evolved as many as 40 times during metazoan development (Schwab, 2018). This makes sense because responding to the electromagnetic spectrum can be evolutionarily adaptive in many contexts for such basic tasks as food gathering, reproduction, or avoiding predation. Learning and memory provides another example. Learning can be functionally defined as "changes in the behavior of an organism that result from regularities in the environment of the organism" (De Houwer, Barmes-Holmes, & Moors, 2013). An animal that can detect and respond to, say, associative stimulus-stimulus regularities is more "aware" than one that cannot, and indeed all organisms that have evolved since the Cambrian period, 545 myo, show these forms of learning and memory (Ginsburg & Jablonka, 2010). These functions are not yet at the level of human consciousness, but the evolution of awareness did not stop half a billion years ago. We argue that the next step forward in awareness by the genus *Homo* laid a needed cognitive foundation for what became an even more qualitatively distinct step into human consciousness itself.

The Emergence of Symbolic Reasoning

Many theories of human consciousness agree that human higher cognition and symbolic reasoning is involved in consciousness (see Sattin et al., 2021). Most are clear, however, that human consciousness is not simply reducible to an information processing system operating on formal symbols alone, such as might occur with artificial intelligence (AI). Searle (1980) classic *Chinese room* thought experiment illustrates the problem. Imagine an English-speaking person who has no knowledge about the Chinese language being alone in a room. In this room is an English instruction manual to process any incoming information. The person then receives some Chinese characters through a slot in the door and is asked to process them according to the English instruction manual to produce other Chinese characters as output. Given enough time, the person should be able to produce the output based on the instruction manual without understanding any of the content of the Chinese writing. Based on the generated output alone, an outside observer who knows Chinese will then mistakenly assume that there is a Chinese speaker in the room.

Just as the person in the Chinese room, computers can use syntactic rules following a set of instructions or machine learning algorithms to manipulate symbols, but they do not "understand" the meaning of these symbols. Syntactical rules alone cannot substitute for the psychological qualities of symbolic reasoning or of communication between conscious entities. This example illustrates the difference the selection of information for global broadcasting (which is accomplished by the English speaker in the Chinese room) and the self-monitoring of those computations and their first-person meaning. Only the latter is associated with the subjective experience (the qualia of consciousness) and that is not part of AI nor any other "brute force" approach to symbol use (Dehaene, Lau, & Kouider, 2017).

In order to know more of what that sense of meaning and self-monitoring entails and how it may have evolved it helps to consider how symbolic reasoning normally emerges in human development and what can be done when it does not.

Consider the relation of a word and its referent in its simplest form. If a word is functioning as a symbol for a referent there must be a two-way street of meaning between them. For example, if a child is taught for the first time to emit a characteristic sound (perhaps the word "dog") in the presence of a known object (perhaps a picture of a familiar dog) the child must then without explicit training to be able to orient toward the picture when hearing "dog" and for the word to carry some of the functions of the dog (e.g., smiling if the picture is that of a pet; fear if the picture if of a dog that bit the child). If that does not occur, we would say that the word is just being "parroted" without being understood. As that term suggests, parrots and many other animals can be taught to emit characteristic sounds or gestures in the presence of an object, but they do not then show a bidirectional relation between that sound or gesture and the object itself, nor do they experience the functions of the object via occurrence of the symbol.

A well-studied phenomenon that maps onto this simplest form of symbolic meaning is "stimulus equivalence" (Sidman & Tailby, 1972). Using the terms drawn from Relational Frame Theory (Hayes, Barnes-Holmes, & Roche, 2001) to describe it, stimulus equivalence requires at

least two integrated instances of the two-way street of meaning or "mutual entailment" (ME), an instance of the combination of these relations in what is called "combinatorial entailment" (CE), and a resulting transfer or transformation of stimulus function (ToF).

An example of ME is orienting toward a dog when hearing "dog" without specific training to do so, following only training in saying "dog" when seeing a dog. As an example of CE, if the child above is also taught that a "dog" makes the sound "bow wow," equivalence relations will emerge between all three events in all directions (e.g., hearing "bow wow" the child could pick out a picture of a dog without training). In terms of ToF if the child was bitten by a dog even the words "bow wow" may now elicit robust fear. If only direct conditioning were the issue, that would be unlikely since it would require robust two-stage backward conditioning.

There are no well accepted and replicated instances of nonhuman animals showing stimulus equivalence, nor its ME, CE, or ToF components, even with extensive training. The so-called language trained chimpanzees respond to such tests at chance levels (Dugdale & Lowe, 2000). Similarly, children who fail to show stimulus equivalence are universally severely language delayed (Devany, Hayes, & Nelson, 1986).

In normal language development, stimulus equivalence is followed rapidly by other types of bidirectional and combinatorial relations (Hayes et al., 2001). For example, a child hearing a new word that is different from known words will look for novel objects and will derive a ME relation between the two events (Lipkens, Hayes, & Hayes, 1993). As children are taught additional relational cues these cognitive networks expand in complexity. A child may originally learn concepts such as "bigger than" with physical comparisons (e.g., a house is bigger than a mouse), but eventually these become applicable to any event without regard to their appearance (e.g., children can learn that a dime is "bigger than" a nickel; see Berens & Hayes, 2007 for an experimental demonstration). This is why ToF involves stimulus *transformation* because in symbolic thought the impact of events are actively *changed* by the underlying cognitive network. For example, let's imagine a person who knew American coinage was repeatedly given mild electric shocks when a nickel appeared but was never shocked in the presence of other coins. This individual would likely show much less arousal to a penny than a nickel because a penny is putatively "smaller" (in value) but may show even more arousal to a dime than a nickel (despite never having been shocked in its presence) because the dime is putatively 'larger" (in value) regardless of its smaller size (for an experimental demonstration of this effect see Dougher, Hamilton, Fink, & Harrington, 2007). Such effects cannot be simply associative (a penny and dime are each equally associated with a nickel), nor are they based on relative physical size. Instead, they are based on higher cognitive processes (value of the coin) that actively change the impact of events based on their symbolic meaning.

In human evolution a process of symbolic meaning led to a more complex form of awareness that sits at the edge of human consciousness itself: the symbolic construal of oneself, the environment, and regularities between and within them. Importantly, this is not just an account of the distant evolutionary past – it has concrete applied implications in the present. We now know that when children with notable developmental disabilities fail to show ME and CE and are then trained in ME and CE, they can often acquire these abilities to a notable degree and their social and intellectual awareness increases accordingly (Belisle, Dixon, & Stanley, 2018; Dixon et al., 2021). Typically developing children can also considerably increase their intellectual and reasoning skills with such training when it is applied to more complex relations (Colbert, Tyndall, Roche, & Cassidy, 2018).

This evolutionarily new form of cognitive awareness is not yet fully what we mean by human consciousness but it is foundational and it produces a kind of state change in the cognitive control of behavior. When children develop ME, standard functional analytic assessments are 3 ½ times less likely to be able to identify reinforcers for maladaptive behavior (Belisle et al., 2017). In other words, when the processes of symbolic meaning emerge or are trained in special needs children, direct contingency accounts do a poor job of explaining their behavior.

How might symbolic meaning have evolved

To date only humans have been shown to have ME, CE, and ToF cognitive functions. If these higher cognitive skills are foundational to human consciousness, it may explain part of why only humans have it. It is not uncommon for non-human animals to emit characteristic sounds or calls in the presence of an animal, object, event, or state of motivation (Manser, 2010) but these tend to be relatively fixed both in the speaker and in the response of the listener (Tomasello, 2008). Non-human speakers and listeners communicate in interlocking systems but "listeners acquire information from signalers who do not, in the human sense, intend to provide it" (Seyfarth & Cheney, 2003, p. 168). Human symbolic meaning, conversely, is relational. Human beings are extraordinarily adept at relational reasoning (Penn, Holyoak, & Povinelli, 2008) which forms the core of human intelligence (Chuderski, 2022). In human symbolic systems these relations are *shared* between speakers and listeners, which will be important to our account.

In accounting for this ability and documenting why it matters to human consciousness it is worth considering the known human advantages in areas of culture, cognition, and cooperation. These so-called "three Cs" are characteristic of all aspects of human evolution (Wilson, 2007). It has been previously argued (Hayes & Sanford, 2014) that the evolution of symbolic meaning can be usefully viewed as an extension of human cooperation, leading then to cultural changes and to further progress in symbolic thinking. The core of the argument is as follows: Human beings are social primates with notable strengths in cooperation that far outdistance other primates. Some of the basic tools of cooperation are present at birth or develop soon after. Young infants show social referencing and joint attention even before symbolic language emerges and use this information to infer intentionality in others (Tomasello, Carpenter, Call, Behne, & Moll, 2005). For example, if an infant has been playing with toys with an adult, when "clean up time" comes if that adult points to a slightly out of reach toy the child can grasp, the baby will try to grasp the toy and put it in the clean-up box. If an adult who has not been playing comes in and points to the same toy the baby will try to grasp the toy and give it to the adult (Liebal, Behne, Carpenter, & Tomasello, 2009). Infants recognize cooperation in others and will reward it in others without training to do so – for instance an infant with a limited supply of special treats will give some of them away to a nice puppet but not to a mean puppet (Hamlin, Wynn, Previde & Mahajan, 2011). Human infants also have basic theory of mind skills, that can be modified with learning and that improves through the pre-school years and beyond (Wellman, 2018).

Even before symbolic thinking emerged these advanced cognitive skills could foster greater cooperation among hominins based on multilevel selection (MLS). MLS suggests that under most conditions selfish advantages to individuals drive evolutionary adaptation but in a small number of conditions when the selfishness of individuals are suppressed and the benefits of cooperation are notable, selection can shift to the group level (Wilson & Wilson, 2008). These shifts are often associated with major evolutionary transitions such as the shift from single-celled organisms to multicellular organisms, or the development of the role of mitochondria in eukaryotic cells. The understanding of others' intentions, for example, could foster greater cooperation based on MLS, *before* symbolic thinking emerged as such. Virtually all hominins showed notably strong signs of cooperation such as in hunting practice, range of foraging, task specialization, and group or social network size as compared to other primates (Fuentes, 2018; Layton, O'Hara, & Bilsborough, 2012).

If we assume that the skills that we see in preverbal infants today were present as psychological tools of pre-verbal human hominins we can return to the moment with a mutually entailed relation between a symbol and referent first occurred in a highly cooperative hominin band. Imagine an adult member of such a troop emitting a characteristic sound in the presence of an object – much as happens in many nonhuman animals. If that adult later emitted that same sound and pointed toward the object that is slightly out of reach, why *wouldn't* listeners provide the object?

Thus, a purely social form of ME could have emerged based on little more than the component skills of cooperation and social connection that pre-verbal human infants display today. This represents a kind of *relational extension* of awareness as we are considering it so far, in two senses of the term "relational." First, it is a social extension based on the *mutual cooperative relation of a speaker and listener*: a request expressed with a name for an object can lead to the social satisfaction of that request with an object that accords with that name. Second, it would begin the human journey of relational reasoning in which a name is related to a referent. From the perspective or point of view of the speaker the object *is* [name] but based on the perspective or point of view of the listener the name *is* [object].

A troop that created a set of such relations would have a kind of proto-language that would extend its cooperative abilities substantially as objects and actions could increasingly be structured socially and from afar. Calling out "apple" across a ravine might lead troop members on the other side to gather apples and bring them back; calling out the name of an approaching predator visible to another but not to a potential victim might lead to successful avoidance of predation; and so on.

At the point at which the perspective of a speaker and listener is simultaneously present within the same skin of an individual, an internalized skill would exist of a bidirectional "___ is ___" relational frame that could apply to any name of anything and vice versa. Several additional relations (hierarchical or part-whole relations, temporal relations, spatial relations, comparative relations, and so on) could then gradually emerge within a hominin community that culturally supported these developments, and a true verbal community could begin to form. For example, if a named speaker is larger than a named listener, the arbitrary comparative relation (e.g., Joe is larger than Sam; Sam is smaller than Joe) might emerge based on the ToF of the non-arbitrary comparative relation between the two within the troop. At this point, human cognition would have a cognitive problem-solving impact within the troop as members began to imagine and compare alternative outcomes for different courses of action using a network of cognitive relations, freed somewhat from the physical form of the related events.

This relational account can address why hominin brain development deviated from other primates. Even before symbolic reasoning emerged, the survival advantages of the expansion of social cooperation, social network complexity, and understanding of intentionality could have strongly and incrementally advantaged the evolution of larger and more complex brain structures. There is substantial evidence linking these two areas in all primates (Adolphs, 2009) but that is especially true in the genus *Homo* (Fuentes, 2018). There are a number of distinctive features of the human brain such as a pronounced frontal cortex that are known to be involved in the capacity for multimodal convergence and integration of cognitive processes. As compared to nonhuman primates, humans have a distinctive ventrolateral frontal pole difference in interregional interactions within the ventral lateral frontal cortex that are related to higher cognition (Neubert et al., 2014). There is also evidence to suggest that the complexity in the pyramidal cell phenotype in the prefrontal cortices are uniquely linked to human cognitive processing (Elston, 2003). Importantly, several of these neurobiological features are linked to theory of mind skills and social network size as well (e.g., Lewis, Rezaie, Brown, Roberts, & Dunbar, 2011), suggesting a possible source of selection pressure *before* symbolic reasoning per se evolved, especially if there is a social and cooperative basis of the evolution of human cognition itself. As cooperation transitioned to cognitive problem-solving, with its known cultural and survival advantages, further neurobiological specialization could evolve within the *Homo* genus and *sapiens* species.

Human consciousness: implications for psychology

As we prepare to move to the discrete part of our process-oriented biphasic account, in this section we propose to consider what modern philosophers and theorists have said about human consciousness and the challenge it presents for psychology.

Explanatory gap of the mind-body problem

The heart of the problem of consciousness is the issue of subjective experience. Although difficult to define, conscious human beings know *what it feels like* to be conscious. Humans are not merely conscious, they are conscious of their consciousness. Most philosophers agree that the subjective experience of *what it is like* to be in a particular consciousness state (the qualia of consciousness) cannot simply be reduced to its physical attributes, such as brain circuitries and nerve cell activities. This seemingly intractable *mind-body problem* has been famously illustrated in Nagel (1974) essay entitled, *what is it like to be a bat*. Chalmers (1995) later described this as the *hard problem* to describe the explanatory gap between the physical world and the subjective experience.

States of consciousness

There is general agreement that consciousness is not a monolithic construct. Rather, there are different states of consciousness that may be grouped into at least two classes: *global states* and *local states*. Global states (or levels) concern an organism's overall subjective profile and are associated with changes in arousal and behavioral responsiveness. Examples of global states are wakefulness, dreaming, and sedation. In contrast, local states relate to a particular conscious perception, emotion or thought (also known as the conscious contents). These local states can be described at different levels of granularity, ranging from low-level perceptual features (e.g., the color red), to objects (e.g., a rose), or to complete multimodal perceptual features (e.g., the beauty of the red rose).

Selfhood

Perhaps the most remarkable feature of the human species is the construct of the self and its ability to psychologically project and conceptualize oneself in space and time. By being both, object and subject (because the self perceives itself), it also introduces a circularity in its definition. The self is a subset of a local state of consciousness, which encompasses the experience of selfhood, including one's own emotions, intentions, volition and free will, body ownership, explicit autobiographical memory, etc. *Self-consciousness* is the conscious awareness of the self, which requires a self-reflective cognitive ability of not merely awareness but also awareness of awareness. The ability to attribute mental states to others (mentalization and theory of mind skills) requires the capability to distinguish between the self and non-self and to infer states of consciousness in others.

The self is a highly complex construct. In his *principles of psychology*, William James (1890/1983) posited that awareness of the self (self-awareness) implies that the self is both object and subject, because there is an aspect of the self that knows (i.e., the knower) and an aspect that wants or needs to be known. The latter has been referred to as the *Me-self*, and the former as the *I-self*. Aspects of the *I-self* include self-awareness, self-agency (i.e., the ownership of ones' thoughts and behaviors), self-coherence (i.e., perceiving oneself as a single, stable entity), and self-continuity (i.e., perceiving oneself as the same person over time). According to James, aspects of the Me-self includes the material Me, the spiritual Me, and the social Me. Because there are many different social settings, there are also many different social Me's. Building of these foundations at least two aspects of the self can be inferred: The *core self*, which is the relatively stable perception of ones' own persona, and the *social self*, which is more context-dependent and influenced by social role expectations (Hofmann & Doan, 2018). This distinction is more than mere theoretical because it opens new ways of understanding and treating emotional disorders that are differentially associated with adaptations of the core self or the social self. For example, infants who cry

when food is taken away express general distress as the result of removal of a pleasant stimulus. Older children, in contrast, may display anger in addition to general distress in similar situations because the toddler now has developed a sense of self as an agent. Once the individual has developed a *theory of mind* and can understand that other individuals have different sensations, emotions, and thoughts from one's own, even more complex emotions, such as shame, can develop. Therefore, the experience of emotion is hinged upon conceptual contents of self and other. Early in development, a stimulus reflexively activates the core self, which may be conceived as the unconscious representations of the self, distinct from the physical and social environment. As the self develops, the individual's unspecific affect is molded by contextual factors, such as social and environmental experiences, into specific emotional concepts. As a result, for treatment to be most effective, age-appropriate and developmentally relevant contextual factors need to be considered.

Phenomenal and functional properties of consciousness

The self is closely tied to consciousness. The phenomenal properties of consciousness must be distinguished from its functional properties. The former term refers to the experiential character of consciousness (the qualia of what it's like to be conscious). In contrast, the functional aspect of consciousness is about the role that mental states play in the cognitive economy of an organism by virtue of being conscious. Here, the term *function* encompasses both teleological functions (as shaped by evolution) and dispositional functions (i.e., the role a process plays in the operation of a larger system of which it is a part of). For example, consciously perceiving a red rose is associated with a range of possible functions, such as the ability to behave flexibly with respect to it (e.g., touching it, smelling it, picking it up, painting it, stepping on it, etc.). Depending on the specific scenario, this may then form an episodic memory of the event and provide the conditions to generate a verbal report about the experience.

Prominent contemporary theories of consciousness

There are many different theoretical accounts on consciousness (e.g., Francken et al., 2022; Graziano, 2022; Michel et al., 2019; Seth, 2018). Out of those, we decided to focus on three theory clusters that are most relevant to our approach, namely the global workspace theories, higher-order theories, and integrated information theory (e.g., Seth & Bayne, 2022). We will briefly summarize these approaches and highlight those aspects we believe align with our view.

Global workspace theories (GWTs)

GWTs are based on the blackboard architecture of AI in which the blackboard is the central resource through which specialized processors share and receive information. For example, "I am conscious of the red rose" means that there is a relationship between my cognitive system and a specific object (the rose). This object may then be selected for further verbal or nonverbal processing. For example, I may smell or say "look at that beautiful rose." The percept of the rose has thus become conscious and globally available because the information is available and can trigger a repertoire of actions, thoughts, etc.

This basic idea of GWTs are that sensory information gains access to consciousness when it is 'broadcast' within an anatomically widespread neuronal workspace that is implemented across higher-order cortical association areas, with a particular emphasis on the prefrontal cortex (Mashour et al., 2020). Access to the global workspace is achieved through nonlinear network 'ignition' in which recurrent processing amplifies and sustains neuronal representations (Dehaene, Sergent, & Changeux, 2003). Once ignited, signals are then amplified, allowing them to enter the workspace (and thus become conscious).

Conscious mental states are those that are globally available to a wide range of cognitive processes, including attention, evaluation, memory, and verbal report (Baars, 1988). That way, conscious states guide behaviors and cognitions in a flexible, context-dependent way. This closely aligns with our view as it relates consciousness to specific cognitive processes that have been influenced by the continuous evolution of awareness and relational learning (De Houwer, Finn, Raemaekers, Cummins, & Boddez, 2022), such as attention, evaluation, and working memory. Consistent with GWTs, we also believe that human consciousness is a discontinuous event that emerges from continuous processes. Consciousness and working memory are intimately related because attended working memory items similarly use the global workspace (Mashour, Roelfsema, Changeux, & Dehaene, 2020).

Higher order theories (HOTs)

It has been noted that global availability constitutes one, but not the only, aspect of consciousness (Dehaene et al. 2017). A second aspect of consciousness is the self-referential relationship in which the cognitive system is able to monitor its own processing. This is often referred to as meta-cognitions - the ability to form internal representations of one's own knowledge and abilities. Because this implies a high-order level of processing (and to distinguish it from first-order theories, such as the GWT), these theories have been referred to as higher order theories (HOTs).

HOTs assume that a mental state is conscious due to these meta-representations. Meta-representations have as their targets other representations. For example, a representation with the content "I am conscious of the red rose" is a meta- representation of another representation (the rose), and this meta-representation is conscious awareness. HOT has been applied to visual experiences (Odegaard, Chang, Lau, & Cheung, 2018), emotional states, such as anxiety (LeDoux & Brown, 2017), and perceptual decision (Morrison, 2016). HOTs emphasize anterior cortical regions, especially the prefrontal cortex (Lau & Rosenthal, 2011). Based on HOTs, some contents may be non- or un-conscious if they are not the targets of appropriate meta-representational states, whereas other contents are necessarily conscious if they are accompanied by appropriate meta-representational states. It has been noted that HOTs rarely focus on global states of consciousness and they do not address the functions of consciousness (Seth & Bayne, 2022). As we will describe below, our views on meta-cognitions closely align to HOTs with an emphasis on the processes associated with consciousness including the relational nature of perspective-taking and sense of self.

Integrated information theory (IIT)

IIT proposes that consciousness should be understood in terms of the *cause–effect power* of a complex system, resulting in *integrated information*. Consciousness and integrated information can be quantified as phi (Φ), which measures how much information is generated by a system as a whole, compared with its parts independently. IIT assumes that consciousness is widely distributed throughout nature, including in many non-biological systems, and might even occur in systems that are as simple as cell phones or computers (Tononi, 2008). We are mentioning IIT because while it is widely cited it is among the approaches that are farthest from our account. IIT contradicts such philosophical arguments as Searle's Chinese room thought experiment and fails to address the qualia of human consciousness. More specifically, Searle (2013) argued that IIT cannot explain consciousness because the information in question is observer-relative. Therefore, a photodiode, for example, cannot be conscious because the information in the photodiode is only relative to a conscious observer who knows what it does. In essence, IIT is a radically continuous account that defines away the hard problem of consciousness itself rather than addressing it. We believe that complexity of a system may be a necessary, but it is not a sufficient condition for consciousness.

A biphasic relational approach to human consciousness

At this point we can return to an approach to the evolution of human consciousness based in particular on a foundation of relational processes. A number of key features of human consciousness need to be addressed including meta-cognitive awareness; awareness of awareness; self-monitoring; meta-representation; global availability; construction of sense of self, both a core self and social self; or the experiential aspects of consciousness. We have argued that even before full human consciousness emerges human cognition is relational, and that cognitive relations are mutual, combinatorial, not bound by formal properties of related events, and can profoundly transform the biopsychosocial functions of any event it touches (see our earlier discussion of ME, CE, and ToF). Its relational nature is arguably two-fold, based on the perspective-taking relation of speakers and listeners, and the ability to relate any event to any other event in any useful way.

The latter quality means that there is nothing to prevent symbolic processes from applying to perspective taking or theory of mind skills themselves. If the two-way street of symbolic meaning emerged from the cooperative relation between a speaker and listener, once those roles are fully internalized, different perspectives can be generated internally. Take, for example, the perspective-taking relations of person, place, and time (McHugh & Stewart, 2012). With sufficient Theory of Mind skills, we will not only know that we experience events now from here – we can cognitively alter that perspective of "I / Here / Now" and imagine how others see us, how we will view ourselves when old, or how it would be to observe "here" from over "there."

The transformation of stimulus functions that results from cognitive changes in perspective means we can imagine the future, reconstruct the past, imagine worlds that have never been, and place ourselves as conscious beings into these imagined universes, altering our own experience on that basis. We can sense what a just world might feel like, or what the world we leave our grandchildren will be like if we do not alter global climate change. We can put ourselves in the bodies of children in a war zone, or imagine a world without hate.

The combination of symbolic reasoning and its extension to cognitive perspective taking requires a change in how we have been discussing awareness. Once human perspective taking is itself a flexible relational cognitive process, it affords a core sense of self and the meta-cognitive awareness of awareness or consciousness of consciousness that fosters the qualia of human consciousness. A person can metaphorically look at their own internal processes of awareness in flight, as if the core self can stand aside to look back on oneself. The sense of self that has been termed the "observing self" (Deikman, 1983), is based on perspective-taking (McHugh & Stewart, 2012) and has long been argued to be central to mindful awareness and human spirituality (Hayes, 1984). This view is relational in yet a third sense: the relation of oneself to one's own experience.

This change requires a different and discrete approach to human consciousness itself. Human consciousness can be seen as the ability to self-reflectively cognitively construct and respond to oneself, the internal, external, and social environment, and the regularities within and between them. In other words, built on the foundation of evolved higher cognitive processes, the key features of human consciousness reviewed earlier can begin to be accounted for by the meta-cognitive awareness of awareness that occurs when these processes turn back upon themselves.

Viewed this way, a biphasic relational approach to human consciousness has profound implications. The ongoing stream of human action can now be readily partitioned into such dimensions as emotion, cognition, attention, motivation, sense of self, or overt action because consciousness itself creates these partitions. For example, contingent operant relations (antecedents, behaviors, consequences) can now be cognitively partitioned into what is expected, hoped for, unfair, successful, skillful, shameful, and so on. Contingencies of reinforcement become contingencies of meaning based on how consequences are interpreted. Personhood is more easily judged and categorized in oneself and others ("I am like this, while you are like that") while at the same time pure awareness as an aspect of consciousness is experienced as featureless ("I am") which moves that aspect of self beyond the realm of judgment and closer to domains that are usually called *spiritual* or *transcendental*.

These features of human consciousness blend on biopsychosocial evolution at all scales and ancient biophysiological processes will be harnessed to new purposes. For example, the brain will set about filtering basic sensory and sensorimotor information to fit a self-narrative – or will open these gates under the influence of mindfulness or attentional training, or with the influence of psychedelic assisted therapy producing changes in sense of self (Hayes, Law, Malady, Zhu, & Bai, 2020).

There is an ineffable quality to human consciousness. Once consciousness means in part the perspective from which experience is experienced, consciousness has no edges or limits that one can become conscious of. It is not possible to know what it is like fully to be unconscious, at least not consciously. That means that human consciousness is not thing-like and it cannot fully be the object of reflection because it is the very context of reflection – which helps explain spirituality, the noetic quality of consciousness, and altered state of consciousness (Hayes, 1984).

Human cognitive abilities can give rise to a more content-based sense of self, which is almost the exact opposite of its noetic qualities. This is one reason our model may be important in an applied sense. It is easy for awareness of one's own experience to establish a kind of "ego tunnel" in which first-person experience creates an illusion of self in a content-based sense (Metzinger, 2010). A full-blown narrative emerges that then psychologically and neurobiologically narrows the full range of conscious experience and ancient brain systems can be mobilized to protect this narrative-based self, creating greater psychological rigidity and psychopathology.

Altered states of consciousness, conversely, often have expansive qualities that tap into boundlessness across time, place, or person. This makes sense once it is understood that all cognitive relations imply a bidirectional continuity. Relationally speaking "here" implies "there" and "now" implies "then" just as much as "big" implies "little" or "hot" implies "cold." What is a polarity at the level of content, however, is a unity at the level of process. The spatial content of "here" and "there", for example, is a single spatial perspective that when viewed as a process can collapse into "everywhere." Anything that is literally everywhere cannot be distinguished from nowhere. In the same way "now" and "then" can collapse into "always / never" at a process level; "this thing" or "that thing" into "everything / nothing"; or "me" and "you" into "everyone / no one". That can help explain why, under extraordinary psychological conditions, such as those being explored in psychedelic assisted psychotherapy, spiritual experiences, or a sense of oceanic awareness, fosters a sense of absolute conscious unity across time, place, and person. That very sense has in turn been shown to be important as a process of change in psychedelic assisted therapy for persons struggling with mental distress (Ko, Knight, Rucker, & Cleare, 2022).

Mindfulness meditation, meta-cognitive approaches, acceptance and commitment therapy, mindfulness-based cognitive therapy, self-compassion training, and virtually all other forms of modern evidence-based therapy directly touch upon the core features of human consciousness as we have analyzed it here. Indeed, a recent review showed that the mediators of change in psychotherapy more generally center on such processes (Hayes et al., 2022).

We are arguing here that a meta-cognitive awareness of awareness is central to human consciousness. When augmented in therapy, that aspect of human consciousness may be able to facilitate such processes of known importance as self-compassion, self-acceptance, perspective-taking, values-choices, and the conscious interconnection between people including between the therapist and client. A recent post treatment interview with a successful case of treatment with a person with severe body dysmorphic disorder (Pickard, Lumby, & Deane, 2021) puts words to how this approach to consciousness might land in therapy. The case involved a person who was feeling as thought she was "monstrously

deformed and ugly" (p. 236). After a successful course of therapy she explained to her therapist what change felt like:

"There was sort of like a compassionate attention to the suffering, and then there was a shift to identification of my values. So, sitting in the space of me struggling with that sense of identity and not knowing . . . you were modelling that way of being in the world and I found that more helpful than anything else in terms of recovery, to see someone modelling, you know, creating a bounded, consistent space that I could internalise." (p. 236-237).

While consciousness itself may be an ally in psychotherapy, the biphasic model also suggests that such therapy is unlikely to be of benefit with person who are not sufficiently developed in the continuous aspects of our relational model such as in children with notable developmental delays. In that case they need training in these more elemental cognitive processes – there is some recent evidence supporting that idea (Gilsenan, Yi, Hinman, Barron, & Dixon, 2021). Thus, a process-based relational approach to human consciousness comports with applied interventions currently being explored across the range of processes noted in our model.

Conclusions

In conclusion, we argue that processes of change emerge in human psychology that amount to processes of managing human consciousness itself (Hayes, Ciarrochi, Hofmann, Chin, & Sahdra, 2022). By understanding human consciousness in this way, a wider variety of new psychological interventions can emerge that target these processes (Hofmann, Hayes, & Lorscheid, 2021).

We are arguing that a discrete event has occurred for humanity. Human consciousness is no longer on a continuum, even if it evolved from features that begin in the awareness of the simplest living creatures. When symbolic relations combined with perspective-taking skills to become self-reflective, an additional evolutionary step was taken. That additional step carried humanity into empty space on the cliff's edge of the evolution of awareness, and everything that went before is now part of something fundamentally different.

Funding

Dr. Hofmann receives financial support by the Alexander von Humboldt Foundation (as part of the Alexander von Humboldt Professur), the Hessische Ministerium für Wissenschaft und Kunst (as part of the LOEWE Spitzenprofessur), and NIH/NIMH R01MH128377, NIH/NIMHU01MH108168, Broderick Foundation/MIT, and the James S. McDonnell Foundation 21st Century Science Initiative in Understanding Human Cognition – Special Initiative.

References

Adolphs, R. (2009). The social brain: Neural basis of social knowledge. *Annual Review of Psychology*, *60*, 693–716. doi:10.1146/annurev.psych.60.110707.163514.

Baars, B. J. (1988). *A cognitive theory of consciousness*. Cambridge, UK: Cambridge University Press.

Belisle, J., Dixon, M. R., & Stanley, C. R. (2018). The mediating effects of derived relational responding on the relationship between verbal operant development and IQ. *Behavior Analysis in Practice*, *11*, 411–416. doi:10.1007/s40617-018-0215-2.

Berens, N. M., & Hayes, S. C. (2007). Arbitrarily applicable comparative relations: Experimental evidence for a relational operant. *Journal of Applied Behavior Analysis*, *40*, 45–71. doi:10.1901/jaba.2007.7-06.

Chalmers, D. J. (1995). Facing up to the problem of consciousness. *Journal of Consciousness Studies*, *23*, 200–219.

Chuderski, A. (2022). Fluid intelligence emerges from representing relations. *Journal of Intelligence*, *10*, 51. doi:10.3390/jintelligence10030051.

Colbert, D., Tyndall, I., Roche, B., & Cassidy, S. (2018). Can SMART training really increase intelligence? A replication study. *Journal of Behavioral Education*, *27*, 509–531. doi:10.1007/s10864-018-9302-2.

De Houwer, J., Barnes-Holmes, D., & Moors, A. (2013). What is learning? On the nature and merits of a functional definition of learning. *Psychonomic Bulletin and Review*, *20*, 631–642. doi:10.3758/s13423-013-0386-3.

De Houwer, J., Finn, J. M., Raemaekers, M., Cummins, J., & Boddez, Y. (2022). Thinking of learning phenomena as instances of relational behavior. *Learning and Behavior.*

Deikman, A. J. (1983). *The observing self*. Boston: Beacon Press.

Dehaene, S., Lau, H., & Kouider, S. (2017). What is consciousness and could machines have it? *Science*, *358*, 486–492. doi:10.1126/science.aan8871.

Dehaene, S., Sergent, C., & Changeux, J. P. (2003). A neuronal network model linking subjective reports and objective physiological data during conscious perception. *Proceedings of the National Academy of Science*, *100*, 8520–8525. doi:10.1073/pnas.1332574100.

Devany, J. M., Hayes, S. C., & Nelson, R. O. (1986). Equivalence class formation in languageable and languagedisabled children. *Journal of the Experimental Analysis of Behavior*, *46*, 243257. doi:10.1901/jeab.1986.46-243.

Dixon, M. R., Belisle, J., Hayes, S. C., Stanley, C. R., Blevins, A., Gutknecht, K. F., Partlo, A., Ryan, L., & Lucas, C. (2021). Evidence from children with autism that derived relational responding is a generalized operant. *Behavior Analysis in Practice*, *14*, 295–323. doi:10.1007/s40617-020-00425-y.

Dougher, M. J., Hamilton, D. A., Fink, B. C., & Harrington, J. (2007). Transformation of the discriminative and eliciting functions of generalized relational stimuli. *Journal of the Experimental Analysis of Behavior*, *88*, 179–197. doi:10.1901/jeab.2007.45-05.

Elston, G. N. (2003). Cortex, cognition and the cell: New insights into the pyramidal neuron and prefrontal function. *Cerebral Cortex*, *13*, 1124–1138. doi:10.1093/cercor/bhg093.

Francken, J. C., Beerendonk, L., Molenaar, D., Fahrenfort, JJ,, Kiverstein, J. D., Seth, A. K., & Gaal, v. (2022). An academic survey on theoretical foundations, common assumptions and the current state of consciousness science. *Neuroscience of Consciousness*. doi:10.1093/nc/niac011 niac011.

Frazzetta, T. H. (2011). Flatfishes, turtles, and bolyerine snakes: Evolution by small steps or large, or both? *Evolutionary Biology*, *39*, 30–60. doi:10.1007/s11692-011-9142-7.

Fuentes, A. (2018). How humans and apes are different, and why it matters. *Journal of Anthropological Research*, *74*, 151–167. doi:10.1086/697150.

Ginsburg, S., & Jablonka, E. (2010). The evolution of associative learning: A factor in the Cambrian explosion. *Journal of Theoretical Biology*, *266*, 11–20. doi:10.1016/j.jtbi.2010.06.017.

Ginsburg, S., & Jablonka, E. (2019). *The evolution of the sensitive soul: Learning and the origins of consciousness*. Cambridge, MA: MIT Press.

Graziano, M. S. (2022). A conceptual framework for consciousness. *Proceedings of the National Academy of Sciences*, *119*, e2116933119. doi:10.1073/pnas.211693311.

Grisdale, E., Lind, S. E., Eacott, M. J., & Williams, D. M. (2014). Self-referential memory in autism spectrum disorder and typical development: Exploring the ownership effect. *Consciousness and Cognition*, *30*, 133–141. doi:10.1016/j.concog.2014.08.023.

Hayes, S. C. (1984). Making sense of spirituality. *Behaviorism*, *12*, 99–110.

Hayes, S. C., Ciarrochi, J., Hofmann, S. G., Chin, F., & Sahdra, B. (2022). Evolving an idionomic approach to processes of change: Towards a unified personalized science of human improvement. *Behaviour Research and Therapy*, *156*, 104155. doi:10.1016/j.brat.2022.104155.

Hayes, S. C., Law, S., Malady, M., Zhu, Z., & Bai, X. (2020). The centrality of sense of self in psychological flexibility processes: What the neurobiological and psychological correlates of psychedelics suggest. *Journal of Contextual Behavioral Science*, *15*, 30–38. doi:10.1016/j.jcbs.2019.11.005.

Hofmann, S. G., & Doan, S. N. (2018). *The social foundations of emotion: Developmental, cultural, and clinical dimensions*. Washington, DC: American Psychological Association.

Hofmann, S. G., Hayes, S. C., & Lorscheid, D. (2021). *Learning process-based therapy: A skills training manual for targeting the core processes of psychological change in clinical practice*. Oakland, CA: Context Press /New Harbinger Publications.

James, W. (1890). *Principles of psychology*. /1983, Cambridge, MA: Harvard University Press.

Ko, K., Knight, G., Rucker, J. J., & Cleare, A. J. (2022). Psychedelics, mystical experience, and therapeutic efficacy: A systematic review. *Frontiers in Psychiatry*, *13*, 917199. doi:10.3389/fpsyt.2022.917199.

Lau, H., & Rosenthal, D. (2011). Empirical support for higher-order theories of conscious awareness. *Trends in Cognitive Science*, *15*, 365–373. doi:10.1016/j.tics.2011.05.009.

Layton, R., O'Hara, S., & Bilsborough, A. (2012). Antiquity and social functions of multilevel social organization among human hunter-gatherers. *International Journal of Primatology*, *33*, 1215–1245. doi:10.1007/s10764-012-9634-z.

LeDoux, J. E., & Brown, R. (2017). A higher- order theory of emotional consciousness. *Proceedings of the National Academy of Sciences*, *114*, E2016–E2025. doi:10.1073/pnas.1619316114.

Lipkens, G., Hayes, S. C., & Hayes, L. J. (1993). Longitudinal study of derived stimulus relations in an infant. *Journal of Experimental Child Psychology*, *56*, 201–239. doi:10.1006/jecp.1993.1032.

Manser, M. B. (2010). The generation of functionally referential and motivational vocal signals in mammals. In: S. M. Brudzynski (Ed.), *Handbook of Behavioral Neuroscience: Vol.* 19, 477-486. doi: 10.1016/B978-0-12-374593-4.00043-7.

Mashour, G. A., Roelfsema, P., Changeux, J. P., & Dehaene, S. (2020). Conscious processing and the global neuronal workspace hypothesis. *Neuron*, *105*, 776–798. doi:10.1016/j.neuron.2020.01.026.

McHugh, L &, & Stewart, I. (2012). *The self and perspective taking: Contributions and applications from modern behavioral science*. Oakland, CA: New Harbinger Publications.

Michel, M., Beck, D., Block, N., et al. (2019). Opportunities and challenges for a maturing science of consciousness. *Nature Human Behavior*, *3*, 104–107. doi:10.1038/s41562-019-0531-8.

Monson, T. A., Weitz, A. P., Brasil, M. F., & Hlusko, L. J. (2022). Teeth, prenatal growth rates, and the evolution of human-like pregnancy in later Homo. *Proceedings of the National Academy of Sciences*, *119*,(41) e2200689119. doi:10.1073/pnas.2200689119.

Morrison, J. (2016). Perceptual confidence. *Analytic Philosophy*, *57*, 15–48. doi:10.1111/phib.12077.

Nagel, T (1974). What is it like to be a bat? *The Philosophical Review*, *83*, 435–450. doi:10.2307/2183914\.

Neubert, F. X., Mars, R. B., Thomas, A. G., Sallet, J., & Rushworth, M. F. (2014). Comparison of human ventral frontal cortex areas for cognitive control and language with areas in monkey frontal cortex. *Neuron*, *81*, 700–713. doi:10.1016/j.neuron.2013.11.012.

Odegaard, B., Chang, M. Y., Lau, H., & Cheung, S. H. (2018). Inflation versus filling- in: why we feel we see more than we actually do in peripheral vision. *Philosophical Transactions of the Royal Society of London, B Biological Science: 373* 20170345. doi:10.1098/rstb.2017.0345.

Paperin, G., Green, D. G., & Sadedin, S. (2011). Dual-phase evolution in complex adaptive systems. *Journal of the Royal Society Interface*, *8*, 609–629. doi:10.1098/rsif.2010.0719.

Penn, D. C., Holyoak, K. J., & Povinelli, D. J. (2008). Darwin's mistake: Explaining the discontinuity between human and nonhuman minds. *Behavioral and Brain Sciences*, *31*, 109–178. doi:10.1017/S0140525×08003543.

Pickard, J. A., Lumby, C., & Deane, F. P. (2021). True beauty lies within: Therapist interview of a client who received acceptance and commitment therapy for body dysmorphic disorder. *Clinical Psychologist*, *25*, 234–239. doi:10.1080/13284207.2021.1948304.

Prum, R. O., & Brush, A. H. (2002). The evolutionary origin and diversification of feathers. *The Quarterly Review of Biology*, *77*, 261–295. doi:10.1086/341993.

Sattin, D., Magnani, F. G., Bartesaghi, L., Caputo, M., Fittipaldo, A. V., Cacciatore, M., Picozzi, M., et al. (2021). Theoretical models of consciousness: A scoping review. *Brain Sciences*, *11*, 535. doi:10.3390/brainsci11050535.

Schwab, I. (2018). The evolution of eyes: major steps. The Keeler lecture 2017: Centenary of Keeler Ltd. *Eye*, *32*, 302–313. doi:10.1038/eye.2017.226.

Searle, J. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences*, *3*, 417–457. doi:10.1017/S0140525×00005756.

Searle, J. (2013). Can a photodiode be conscious? Reply to Christof Koch and Giulio Tononi. New York Review of Books, March 7, 2013, https://www.nybooks.com/articles/2013/03/07/can-photodiode-be-conscious/

Seth, A. K. (2018). Consciousness: The last 50 years (and the next). *Brain and Neuroscience Advances*, *2*. doi:10.1177/2398212818816019.

Seth, A. K., & Bayne, T. (2022). Theories of consciousness. *Nature Review Neuroscience*, *23*, 439–452. doi:10.1038/s41583-022-00587-4.

Seyfarth, R. M., & Cheney, D. L. (2003). Signalers and receivers in animal communication. *Annual Review of Psychology*, *54*, 145–173. doi:10.1146/annurev.psych.54.101601.145121.

Tomasello, M. (2008). *Origins of human communication*. Cambridge, MA: The MIT Press.

Tomasello, M., Carpenter, J., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences*, *28*, 675–735. doi:10.1017/s0140525×05000129.

Tononi, G. (2008). Consciousness as integrated information: A provisional manifesto. *The Biological Bulletin*, *215*, 216–242. doi:10.2307/25470707.

Wellman, H. W. (2018). Theory of mind: The state of the art. *European Journal of Developmental Psychology*, *15*, 728–755. doi:10.1080/17405629.2018.1435413.

Wilson, D. S. (2007). *Evolution for everyone: How Darwin's theory can change the way we think about our lives*. New York: Delta.

Wilson, D. S., & Wilson, E. O. (2008). Evolution "for the good of the group". *American Scientist*, *96*, 380–389. doi:10.1007/s10818-008-9049-2.

Zeh, D. W., Zeh, J. A., & Ishida, Y. (2009). Transposable elements and an epigenetic basis for punctuated equilibria. *Bioessays*, *31*, 715–726. doi:10.1002/bies.200900026.