Consciousness Is (Probably) Still Only in the Brain, Even Though Cognition Is Not

Luis H. Favela
Department of Philosophy and Cognitive Science Program
University of Central Florida, Orlando, USA

Abstract

There is increasing theoretical justification and empirical support for non-brain-centric approaches to cognition. The body, non-biological tools, and environment are understood as playing causally significant roles in or are constitutive of many instances of cognition. Although not without critics, such non-brain-centric approaches are doing so well that some argue that not only is cognition situated, embodied, extended, and distributed (cognition \textit{SEED}) but so too is consciousness (consciousness \textit{SEED}). Here “cognition” refers to an organism’s abilities to engage with its world, which includes perceiving and acting skillfully, as well as capacities such as decision-making, planning, and reasoning. “Consciousness” refers to states of a system with subjective phenomenal character. Some defend Consciousness \textit{SEED} by appealing to affordances and complex systems theory. I argue that these do not support the claim that cognition \textit{SEED} entails consciousness \textit{SEED}. I then present phenomenological and neurophysiological considerations to think that consciousness is (probably) still only in the brain, even though cognition is not.

1. Introduction

The mind sciences are increasingly recognizing the role of the body (Edelman 2006), non-neural physiology (Caluwaerts \textit{et al.} 2013), non-biological tools (Favela and Chemero 2016), and the environment (Izquierdo and Beer 2016) in their research. This is evident in cognitive science (Chemero 2009), neuroscience (Sporns 2011), and psychology (Glenberg 2010). Even critics of more radical claims – such as the body or environment being constitutive of cognition (Eliasmith 2012, Shapiro 2013) – at least acknowledge the causal role that non-neural processes play. Those more open to seemingly radical conceptions of the mind and related phenomena have urged the philosophical and scientific communities to at least consider cognition as situated, embodied, extended, and
distributed (cognition\textsuperscript{SEED}). That cognition is cognition\textsuperscript{SEED}, and not cognition\textsuperscript{CRUM} (computational-representational understanding of mind; Thagard 2005), is a hard pill for some to swallow. Yet, there is an accumulation of conceptual considerations and empirical evidence.

Although not without critics, such non-brain-centric research programs are doing quite well. So well, that some wonder if other mind-related phenomena that are traditionally treated as brain-centric could actually be non-brain-centric. Following from the idea that cognition is cognition\textsuperscript{SEED}, some have asked: If cognition is situated, embodied, extended, and distributed, could it be that consciousness is too? Could more than the brain or central nervous system constitute consciousness? Could consciousness be more causally intertwined than is usually considered to its environment, body, non-neural tools, or other systems such as other humans, animals, or computers? What follows is a defense of a negative answer to such questions: Although there are good reasons and evidence for thinking that cognition is non-brain-centric (cognition\textsuperscript{SEED}), the reasons for thinking the same is true for consciousness are not nearly as compelling. In short, consciousness is (probably) not consciousness\textsuperscript{SEED}. Consciousness is (probably) only in the brain (consciousness\textsuperscript{Brain}).

Before defending this position, in the next section I define what I mean by “cognition” and “consciousness”. In Section 3 I present an argument for consciousness\textsuperscript{SEED} in the form of Silberstein and Chemero’s (2012, 2016) “extended cognitive-phenomenological systems”. The section concludes with a critique of one of their primary reasons for believing in consciousness\textsuperscript{SEED}: the Gibsonian ecological psychology concept of affordances. Section 4 presents arguments that consciousness is consciousness\textsuperscript{Brain}. These reasons stem from phenomenological and neurophysiological considerations. My goal is to motivate believers in cognition\textsuperscript{SEED} to think that – based on current evidence – it is highly likely that consciousness is a brain-centric phenomenon.

On the other hand, the aim of this paper is not to defend cognition\textsuperscript{SEED}, which has been done many times (e.g., Chemero 2009, Favela 2014, Favela and Martin 2017). For present purposes, I operate on the assumption that there are good conceptual and empirical reasons to support cognition\textsuperscript{SEED} and use this as the basis of my discussion on whether consciousness is consciousness\textsuperscript{Brain} or consciousness\textsuperscript{SEED}.

2. Defining “Cognition” and “Consciousness”

It has been claimed that it is either “confusion or brazenness” for anybody to treat cognition as anything other than a cause of behavior or to abandon descriptions of such causes in terms of representations (e.g., Shapiro 2013, pp. 362f). The history of psychology and cognitive science
demonstrates that it is not true that the account of cognition as a cause of behavior and as representational has been the “thoroughly entrenched” and “dominant” understanding of cognition (Shapiro 2013, p. 362). Examining the history of the term shows that “cognition” has been defined in significantly different ways (cf. Chaney 2013, Greco 2012, Watrin and Darwich 2012).

We do not need to look far back to see this fact. When discussing the purview of psychological science, William James treated “cognition” and “mind” as interchangeable (James 1885). James utilized “cognition” and “mind” to refer to the manner in which an organism interacts with the world. In this way, cognition is akin to an organism’s mindedness (Anderson 2009) or sense-making capacities (Thompson 2007). Cognition, mindedness, and sense making capture the meaningful behavior that begins with perceptual capacities that guide action, and is always involved in an organism’s being-in-the-world. Note that “behavior” is not limited to third-person, objective bodily action, but is a general term referring to the activities of an organism, which include readily observable body-environment perception-action, but also more typically considered first-person activities such as decision-making, planning, and reasoning.

Following this non-computational-representational understanding of mind (Thagard 2005), when I utilize the term “cognition”, I refer to a wider range of phenomena than are implicated in computational-representational terms. “Cognition” refers to an organism’s meaningful activities related to its being-in-a-world (Favela and Martin 2017). What I have in mind is comparable to some enactivist conceptions of sense-making and cognition (cf. Thompson 2007, 2011, Varela et al. 1991). Though what I mean by “cognition” is clearly similar to enactivism, I do not use the term “enactive” because I do not wish to elicit challenges associated with enactive cognitive science, particularly those related to commitments to autopoiesis.

Although attempts have been made to define it “innocently” (Schwitzgebel 2016), the term “consciousness” may be even more controversial than “cognition”, leadings some in the field to claim “it is better to avoid a precise definition” (Crick 1994, p. 20, italics in original). I follow Edelman and Tononi (2000) in my utilization of “consciousness”. First, consciousness is not an object or substance, but a process (p. 9). Second, every conscious state shares three fundamental and general properties: privateness, unity, and informativeness.

Privateness refers to the quality of phenomenal experience whereby “each conscious event is a process that has a single ‘point of view’ ” (p. 23). Unity refers to the quality of every conscious state to be a phenomenal scene that is a single and unified whole. Informativeness refers to the idea that every conscious state is a selection from incalculable numbers of possible alternative states, and that these states have different behavioral
consequences (p. 18). These fundamental and general properties of all conscious states are intended to expand on William James’ conception of consciousness as personal, changing while continuous, and selective in terms of what it focuses upon (cf. Edelman 1992, pp. 6, 111, Edelman and Tononi 2000, p. 18).

These fundamental properties of consciousness may not be totally innocent, but at the very least they begin to gesture at what it is that makes a phenomenon count as consciousness. Nagel (1974) referred to it in terms of the “what it is like”, or the distinct quality of consciousness demonstrated by its singular perspective. Subjectivity is the “most important and characteristic feature of conscious mental phenomena” (p. 436). It is this subjectivity that Chalmers (1995) refers to when he talks about the “hard problem of consciousness”, namely, the question of why there are any subjective phenomenal states. In short, I utilize “consciousness” here to refer to an organism’s subjective, phenomenal experience; where “subjective” refers to my own experiential character of having a point of view, and “phenomenal experience” refers to states of a system characterized by an informative and integrated, yet diverse (Tononi and Koch 2015) condition that goes away in deep sleep, under anesthesia, and after death.

To briefly review: When I say cognition, I mean the abilities an organism has to engage with its world. In this way, cognition is comparable to mindedness or sense-making, and primarily includes the ability to perceive and act skillfully in a world via bodily engagement with an environment, but also consists of capacities more traditionally considered cognitive such as decision-making, planning, and reasoning. When I say consciousness, I mean an organism’s subjective phenomenal states, where subjectivity captures the first-personal aspect, and phenomenal refers to the phenomenological properties of states that, if one were in deep sleep, under anesthesia, or deceased, they would not have.

I do not pretend to have provided a comprehensive definition of cognition or consciousness. I do not think necessary and sufficient conditions for either can be given. Moreover, if we are to be good naturalists and wish to give scientifically respectable accounts of these phenomena, then we need to be open to the possibility of revising our terms. With that said, we need a general grasp of how these terms are currently used. In the next section, I present an argument for consciousnessSEED: Silberstein and Chemero’s (2012, 2016) extended cognitive-phenomenological systems.

3. Extended Cognitive-Phenomenological Systems

Silberstein and Chemero (2012, 2016) attempt to respond to the “hard problem of consciousness” with appeal to extended cognitive-phenomeno-
logical systems. In their earlier work on the topic, Silberstein and Chemero (2012) argue that if cognition and consciousness are understood as a single phenomenon, then phenomenology can be maintained while eliminating the qualia part of the hard problem. They provide four reasons for treating cognition and consciousness as a single process (p. 41): First, it is deflationary in that the metaphysical hard problem concerning why we have phenomenology turns into the empirical problem explaining how “extended phenomenology-cognition works”. Second, searches for neural correlates of consciousness are treated as a “no-go” because consciousness is not just a neural phenomenon. Third, in treating consciousness as fundamentally relational – i.e., phenomenology is not just in the head or the external world – extended phenomenological-cognition is neither reductionist nor requires appealing to qualia. Fourth, such an approach “has substantial support from research in the cognitive sciences” – though it is unclear to me what that research is.

Underlying these four reasons is their understanding of qualia as a concept that facilitates treating cognition in indirect, mental representational terms. They argue that the Gibsonian ecological psychology concept of “affordances” can do the work of eliminating the need to appeal to qualia or mental representations. In later work on the topic, Silberstein and Chemero (2016) admit that even if their earlier attempt (2012) allows them to eliminate qualia by treating cognition and intentionality in terms of extended complex systems – with affordances doing much of the work to preserve the phenomenological aspect of the system – the hard problem remains, namely, accounting for the subjective experiences of the system.

Although I am very friendly to this approach and laud the attempt, in the end I do not think it succeeds in accounting for the subjectivity experienced by some systems. There are two primary reasons why the attempt is unsuccessful. First, while Silberstein and Chemero (2016) acknowledge that appealing to affordances in their earlier work does not address the hard problem in toto (i.e., qualia and subjectivity), I argue that appealing to affordances does not even address the experiential aspect of consciousness. Second, even though the relational nature of cognitive systems underwrites cognition$^{SEED}$ (cf. Silberstein and Chemero 2016), I argue that it does not logically or empirically follow that subjectivity extends beyond the brain.

Silberstein and Chemero (2012) claim that a complex systems approach allows us to understand cognition in terms of “non-linearly coupled brain-body-niche systems” (p. 35). They then present two “conceptual and methodological advances” that can be made from this perspective: The first advance is basically a restatement of arguments and evidence for cognition$^{SEED}$, specifically, that cognition is fundamentally interdependent on brain, body, niche dynamics, and that such an approach negates the need to appeal to computations or representations. I grant that cog-
nition is cognition$^{SEED}$, thus, I find no disagreement with this first point.

The second advance is that, following from this understanding of cognition$^{SEED}$, and motivated by the four reasons stated above, cognition and consciousness are now viewed as a single phenomenon. From such a view, they argue that the qualia aspect of the hard problem is eliminated. It is the second “advance” that I find wanting, primarily due to their reliance on the Gibsonian ecological psychology concept of affordances to do the work of answering the subjectivity aspect of the hard problem.

Affordances are relational features of animal-environment systems (Chemero 2009, Favela and Chemero 2016, Gibson 1979/1986). They are opportunities for action in an animal-environment system, such as whether a slope is stand-on-able (Fitzpatrick et al. 1994), an object is catchable (Oudejans et al. 1996), a chair is sit-on-able (Mark 1987), a staircase is step-on-able (Warren 1984), or a door is pass-through-able (Warren and Whang 1987). Gibsonian ecological psychology can be summed up by three main principles: First, perception is direct and not mediated by mental representations. Second, perception and action are inseparable aspects of perceptual systems. Third, since perception is for action (and vice versa) and does not involve indirect mental representations or a computational inferential step to imbue the representation with meaning, then the contents of perception must be meaningful. Since perception is of affordances, then affordances must be meaningful.

Gibson argued that affordances are meaningful in virtue of their relational nature. For example, a cup is meaningful to me as graspable based on its properties to hold liquid, to fit the dimensions of my hand, near enough to be seen and grabbed, etc. My perception of the cup is due as much to the properties of the cup as to me, and in this way has meaning without needing to be inferred as for “grasping, drinking from, etc.” via an inferential, mentally computed process. Gibson wrote some notoriously esoteric statements about affordances. For example (Gibson 1979/1986, p. 129):

An affordance is neither an objective property nor a subjective property; or it is both if you like. An affordance cuts across the dichotomy of subjective-objective and helps us to understand its inadequacy. It is equally a fact of the environment and a fact of behavior. It is both physical and psychical, yet neither. An affordance points both ways, to the environment and to the observer.

It is this neither being objective nor subjective property (or both) of affordances that Silberstein and Chemero (2012) lean on in their first attempt to address the hard problem. They claim that to agree that qualia must be appealed to in order to explain perception is tantamount to accepting the need to appeal to indirect mental representations (p. 37). But, if cognition and consciousness are a single phenomenon, then the same ar-
Arguments that apply to treating perception as relational and direct must also apply to consciousness qua phenomenology. Silberstein and Chemero say as much when they claim that, since affordances are what animals perceive, and the animal’s niche just is its environment full of affordances, and since affordances are neither and both subjective and objective, then it follows that an animal’s cognition and consciousness are neither and both subjective and objective (p. 43). It is this phenomenological realism, as sustained by affordances, that Silberstein and Chemero claim necessitates in “the entire system, including the environment as experienced, [to be] required to account for phenomenology-cognition” (p. 43). This account is problematic as a basis for responding to the hard problem.

The reason this “phenomenological realism” is problematic as a basis for responding to the hard problem is that affordances are more justifiably understood as primarily cognitive phenomena and not also as phenomena that are conscious. Silberstein and Chemero want to deflate the hard problem, specifically as it involves problematic metaphysical postulates like qualia. Their four reasons stated above can be understood as a sort of eliminative materialist position. They are eliminative in the sense that terms like “qualia” do not refer to anything in the natural world, and they are materialist in that empirical research is the arbitrator of ontology. Accordingly, they claim that the hard problem qua qualia problem does not have the empirical credence to be worth addressing.

However, another major aspect of the hard problem involves accounting for subjectivity, that is, how it is that a system can have a first-person perspective. For deflationary reasons, Silberstein and Chemero (2012, pp. 46f) claim that their account does not address this part of the hard problem. To claim that a system has first-person subjectivity is to hold that the subject has first-person awareness of “objects”, where objects are independent of the system. However, because affordances provide a way to understand perception as cutting across subject-object divisions, the distinction between first-person subject and external object dissolves. Consequently, it is not the subject or the object that is fundamental, but presence that is fundamental, where “presence” refers to the co-dependent relation of perceiver/subject and perceived/object (Silberstein and Chemero 2016, p. 193). I do not think the concept of affordances can do the work to adequately respond to the subjectivity portion of the hard problem. The reason is that the role affordances play in the empirical investigation of perception-action does not necessitate that affordance perception involves any phenomenal character of experience. To put this objection in Silberstein and Chemero’s terms, affordances do not require presence to be successfully perceived and utilized.

---

1 Neither Silberstein nor Chemero (as far as I know) ever refer to their claims as eliminative materialist, though others have (e.g., Ramsey 2013).
It does not follow that when something in the environment is afforded an animal that the animal needs to be conscious of the affordance. This may be a metaphysical possibility, but the empirical evidence does not support it. A number of studies have demonstrated that affordances need not have a phenomenal character of experience in order to be perceived and utilized by a cognitive system (e.g., Anderson et al. 2002, Derbyshire et al. 2006, Makris and Yarrow 2014, McBride et al. 2012). To say that an animal is not conscious of an affordance is to say that it is not having a subjective phenomenal experience of attending to the affordance. Such would be the case, for example, by not having a cup in your field of vision but still reaching for it and successfully grasping it. It is an empirical question whether or not subjective phenomenal states always coincide with utilizing affordances or if all affordances can be successfully utilized whether or not they are accompanied by subjective phenomenal states.

We are now faced with the following questions: Why are animals conscious of some affordances and not others? To complicate matters more, if there is functional equivalence between an affordance both when it is consciously perceived and when it is non-consciously perceived, why is the affordance ever consciously perceived at all? Consciousness is highly taxing on the brain and body in terms of energy consumption (Shulman et al. 2009). Thus, from considerations of both natural selection and

---

2 When I say that an animal does not need to be conscious of an affordance in order to successfully utilize it, I intend to include both the affordance qua directly perceivable opportunity for action (e.g., grasp-able cup or pass-through-able doorway), and the feature of the world that specifies affordances, namely information. In Gibsonian ecological psychology, “information” does not refer to computational information or Shannon information. “Information” in ecological psychology refers to energy distributions (e.g., light and sound) specific to the layout of the world, and those are revealed to an animal via its perceptual capacities. For example, the ambient optical array is information in the sense that the structure of light in an area can specify affordances for a creature able to detect that form of energy. The affordance grasp-able is specified by the way light refracts from the surface of an object (e.g., a coffee cup on my desk illuminated via the fluorescent tubes above my head). Although not central to my arguments, I draw attention to the relationship between affordances and information in order to note that when I say an animal is not conscious of an affordance, I also mean not conscious of the features of the world that play causal roles in perceiving affordances.

3 An adequate discussion of the controversial relationship between attention and consciousness goes far beyond the scope of the current work. With that said, it is worth mentioning that it is important to my critique of Silberstein and Chemero’s “phenomenological realism” that I treat conscious states as involving a higher degree of attention and/or reportability than is required of successful perception or utilization of affordances. Part of what makes “affordances” such a powerful idea is the breadth of application, namely, affordances are perceived and utilized by a wider range of creatures than are typically treated as being conscious, (e.g., earthworms; cf. Darwin 1881/1985). Consequently, contra Block (2007), I do not think that the experiments by Landman et al. (2003) or Sperling (1960) are convincingly interpreted as demonstrating that phenomenology overflows the mechanisms of reportability. But that is an argument for another time.
thermodynamics, minimal and/or efficient energy consumption would be a property of successful cognitive systems (cf. Friston 2010, Moralez and Favela 2016). Accordingly, it would seem that as long as an affordance can be successfully utilized, it would be better – that is, in terms of energy consumption – for the animal to do so without consciously perceiving it. This is starting to sound like the “hard problem of affordances”: Why are any affordances consciously perceived when it seems they can be successfully utilized in more energetically efficient ways non-consciously? The problems with appealing to phenomenological realism do not stop here.

Another issue with Silberstein and Chemero’s appeal to phenomenological realism as a way to address the hard problem of consciousness is that it does not satisfactorily address a version of the dreaded coupling-constitution fallacy (Adams and Aizawa 2008, Kirchhoff 2014). According to the coupling-constitution fallacy, although your iPhone is causally related (i.e., coupled) to your cognitive system as an external memory storage for telephone numbers, it is not constitutive of your cognitive system, that is, your iPhone is not part of a process that constitutes your cognition. For present purposes, I will grant that compelling replies have been made to save cognition from the coupling-constitution fallacy (Favela and Chemero 2016, Wagman and Chemero 2014). Nonetheless, it is not clear that those same replies save consciousness.

Silberstein and Chemero (2012) argue that the entire brain-body-niche system is necessary to explain cognition in the sense that all parts are causally and constitutively necessary. They did already claim that cognition and consciousness are a single phenomenon (p. 35). Notwithstanding an appeal to affordances as crosscutting the subjective/objective divide, their case for viewing cognition in terms of extended phenomenological-cognitive systems does not generalize to consciousness qua subjective phenomenal experience. Affordances arise as relational properties among animal-environment systems, and ecological psychology and dynamical systems theory (i.e., radical embodied cognitive science; Chemero 2009) provide a very strong framework for investigating and understanding cognition as cognition. But it does not follow that consciousness is entailed by the seemingly justifiable claim that non-neural parts of the body and the environment constitute the cognitive system. Consciousness may likely be consciousness in the weaker causal sense, but the body or environment does not constitute it, because for organisms like us, consciousness is primarily in our brain. In the next section I provide reasons why even those friendly to the idea of cognition should think consciousness is consciousness and not consciousness.
4. Defending Consciousness via Phenomenological and Neurophysiological Considerations

If we want to build a theory about a phenomenon, a reasonable place to start is with the typical instances of said phenomenon. Following Edelman and Tononi’s (2000) Jamesian conceptualization of consciousness as a process that is generally private, unified, and informative, an appropriate example is my own subjective phenomenological experience. Besides what I report via language, gestures, or automatic physiological responses, my conscious experience belongs to me alone and is from my perspective. It is unified in that the sound of music, taste of coffee, and images are blended into an unfractured experience (cf. Langland-Hassan 2008).

Under typical and healthy conditions, my conscious experience is more akin to looking at a single painting of a jungle (trees, bushes, a lion, a gazelle, etc.), than as looking at the items depicted in the aforementioned painting disunified and painted on separate canvases that are each placed at different locations in my field of view (some canvases have a tree, another has a bush, another a lion, etc.). Moreover, my phenomenological experience is informative: there are differences that make differences in what I perceive. That is, though the experience is unified, it is a unity of differences. Back to the single, unified painting of a jungle: though the images are on the same canvas and present as-if a single “jungle gestalt”, each tree, bush, lion, and gazelle is identifiable as different aspects of the same scene.

Perhaps I have a limited imagination, but is this not what we would expect consciousness to be like? Natural selection favors organisms that have phenomenological experiences from a single point of view, unified, and informative. From the perspective of a phenomenological analysis, it seems reasonable to think that organisms with fractured phenomenologies would be at an evolutionary disadvantage. Experiencing self from non-self is one of the earliest forms of consciousness (cf. Edelman 1989, Feinberg and Mallatt 2016). According to Edelman and Tononi (2000), of all the values in an organism’s life, “one of the most important global values related to survival [over the course of the history of evolution] was the continuity and coherence of the self” (p. 110). By “global value,” they refer to something that is important to the entire system. Thus, the self is treated as one of the most important categories and values to the organism’s system.

Note that what counts as “self” varies from species to species and culture to culture. The sense of self of an ant, for example, may be more distributed among its colony than a wolf’s self among its pack. Similarly, the sense of self of a human living in twenty-first century United States may be more isolated than that of a human living in sixteenth century Japan. Though the self is identified more or less individually or
collectively, the organism (e.g., ant, wolf, human in the US, human in Japan, etc.) still has a single, subjective point of view.

From a human perspective, my consciousness is constituted by my own privileged perspective: no other organism, human or not, experiences my phenomenology at this spatial and temporal location. Not only is my subjective character of experience isolated from others’, it is also unique within my system. I do not share my system with other first-person perspectives. A healthy, typically developing human has one “I” at a time from which they perceive and act in the world. I am confident there are no other “I’s” nested within me for a number of reasons.

First, I do not have reason to think, nor have I been diagnosed with a condition such as severe schizophrenia that could loosen my confidence in my self as being the source of non-externally caused auditory sounds. Second, my body does not move in ways that surprise me, for example, my left hand does not randomly grab my coffee mug and turn it upside-down on my own head, as if controlled by another “I”. Third, when I go to sleep, another “I” does not wake up and live his life until he goes to sleep and the “I” typing this paper then gets his turn. There are obvious cases where people do not have total control of their body (e.g., multiple sclerosis), and there are people who do things when they are asleep (such as walking, eating, or talking). Nonetheless, these cases are attributable to reasons other than another nested “I”. The primary reason there are no nested “I’s” could be grounded in human neurophysiology.

Although we have a very long way to go in understanding the role of the central nervous system in cognition, and perhaps an even longer way to go until we understand its relation to consciousness, progress is being made. Tononi’s integrated information theory of consciousness (IIT) is a step forward in our understanding of consciousness (Tononi 2012). Based on early work by Edelman and Tononi (2000) and Tononi and Sporns (2003), and more recently fine-tuned by Tononi and Koch (2015), IIT has gone through a number of revisions. Across the revisions, though, the following remains central (Tononi et al. 2016, p. 450):

IIT does not start from the brain and ask how it could give rise to experience; instead, it starts from the essential phenomenal properties of experience, or axioms, and infers postulates about the characteristics that are required of its physical substrate. Moreover, IIT presents a mathematical framework for evaluating the quality and quantity of consciousness.

The five axioms with their corresponding postulates are (Tononi and Koch 2015):

- intrinsic existence (consciousness exists), cause-effect power upon itself;
• composition (consciousness is structured), the parts must have cause-effect power on the system;
• information (consciousness is specific), the particular causal structure;
• integration (consciousness is unified), the experience is irreducible;
• exclusion (consciousness is spatially and temporally definite), the experience is irreducible at its maximum state.

A conscious experience is quantified as the maximally irreducible value measured as integrated information and represented as Φ. A full explication of IIT is far beyond the scope of the present work. What is worth highlighting for current purposes is that IIT attempts to provide a mathematically rigorous and empirically testable way a physical system would have to be in order to sustain a particular state of consciousness. In other words, working from the phenomenology (axioms) back to the mechanisms (postulates), IIT provides an account of the informational structure that would undergird a phenomenal state.

Two features of IIT make it particularly appealing for my purposes: First, IIT starts with the phenomenon. Although the axioms could be criticized for being a priori commitments, they are based on reasonable approximations of what is common across a range of phenomenological experiences. From there, the postulates attempt to facilitate the mathematical formalism necessary to quantify how IIT is instantiated by a physical system. Second, IIT provides an explanation as to why each system has only one conscious perspective at a time, that is, the maximum Φ value.

A major challenge faced by some of the main theories of consciousness – for example, identity theory and functionalism – is to explain why those neurons are conscious at a particular time, and, related, why those same neurons are not conscious when removed from the system? IIT answers these questions by attributing consciousness to the maximum Φ value located within a system at any given moment. Moreover, the richness of consciousness results from the amount of nonlinearly (Tononi and Sporns 2003) and dynamically (Albantakis and Tononi 2015) involved parts of the system. Consequently, IIT would predict that the physical substrate of human consciousness is primarily located in the thalamocortical system (Edelman and Tononi 2000, Ward 2011), and possibly include the claustrum (Crick and Koch 2005, Mathur 2014).

What makes the thalamocortical system so central to consciousness – and high Φ values – is the amount of connectivity and integration. The thalamocortical system has connections spread through various cerebral cortical areas via nonlinearly, dynamic reentrant connections. The amount and speed of connections, along with the self-organized nature
Consciousness Is (Probably) Still Only in the Brain

of this system, serve as compelling reasons to think that it is the seat of consciousness.

There is no doubt that, even if the thalamocortical system is the central location of consciousness, it is causally related to other parts of the brain, as well as the body and environment. But it is the nature of the dynamics of the connections that makes the thalamocortical system the likely place where consciousness comes together. The dynamics of the physical substrate of consciousness are interaction dominant. Systems are interaction dominant when the dynamics of the interactions among the parts supersede those that the parts would have separately (Holden et al. 2009, Richardson and Chemero 2014).

I think understanding IIT in terms of interaction-dominant dynamics can explain why areas of the brain can be densely packed with cells and highly connected, yet do not serve as physical substrates of consciousness. It also explains how the same neurons can contribute to a conscious state in one instance but be unconscious in another. In order to participate in the part of the system that is conscious by contributing to the maximum Φ value, it is not merely a matter of having a large number of neurons, nor of having those neurons be highly connected – though both help. What matters is that the system that underlies consciousness is interaction dominant such that the global state of the system constrains and affects what the local parts do, while simultaneously the local parts constrain and affect the global state of the system in a strong reciprocal manner.

This kind of strong reciprocity exhibited by interaction-dominant systems also justifies describing such systems as non-modular or non-decomposable. When a system is modular, decomposable, and component-dominant, then the properties of the system result from the properties of the component parts interacting in an additive and linear manner. In this way, the capacities of the system can be derived from the capacities of the parts in isolation. On the other hand, the properties of interaction-dominant systems result from multiplicative, nonlinear, and irreducible interactions among components. Accordingly, although my left arm, stomach, and right eye are causally related to what happens in the thalamocortical system, the dynamics of those parts of my body do not affect the structure of the thalamocortical system anywhere near the degree needed to be a part of that interaction-dominant system.

A reasonable objection can be made at this point: If the sort of multiscale spatial and temporal causality typical of cognition

SEED can give rise to interaction-dominant dynamics across brain, body, and world, then why do I claim that the same does not hold true for consciousness? There are two primary reasons why it does not. First, the weaker reason is that there is no compelling empirical evidence for consciousness

SEED among any scientific communities. The concept of interaction dominance and the
methods for assessing it (e.g., fractal analyses) plays significant roles in a variety of contemporary scientific communities, such as cognitive sciences, ecological psychology, and systems neuroscience. These communities produce compelling evidence that many cognitive phenomena are interaction dominant, for example: interpersonal coordination (Davis et al. 2016), memory processes (Szary et al. 2015), neural activations (Ihlen and Vereijken 2010), reaction timing (van Orden et al. 2003), and visual search (Stephen and Mirman 2010). Researchers from within these communities apply these concepts, methods, and empirical data to support the idea that cognition is cognition

There was a time when the idea of extended cognition, for example, did not have compelling empirical evidence within any scientific community. Accordingly, the best that could be done were attempts to provide clever arguments related to “marks of cognition” and “memory notebooks”. In the end, it would take compelling empirical considerations to arbitrate the issue. That consciousness is consciousness

The second, stronger reason for a negative response to the above question involves a misunderstanding of the way the thalamocortical system is an interaction-dominant system. As discussed above, a system is interaction dominant not merely in light of the causal or constitutive relationships of its parts. A system is interaction dominant when the dynamics of the whole supersede what the parts can do either in isolation or in terms of additive and linear relationships. Moreover, the relationship is so tightly reciprocal that changes in one area of the system will alter the system as a whole, as do changes at the global level alter local dynamics and properties.

The above examples of interaction-dominant cognitive phenomena demonstrate that the very natures of the dynamics of the components are altered to give rise to dynamics not exhibited by the parts in isolation or additively and linearly related. A human heart, for example, is an interaction-dominant system precisely because the dynamics of the components (e.g., valves) are altered when coupled with each other so as to

4 Other than in the titles of cited journals, the word “consciousness” does not show up in Davis et al. (2016), Favela and Chemero (2016), Szary et al. (2015), or Stephen and Mirman (2010). “Consciousness” does appear a few times in the text body of Ihlen and Vereijken (2010) and van Orden et al. (2003). However, in those two articles, “consciousness” is utilized in a manner interchangeable with “cognition” (e.g., Ihlen and Vereijken 2010, p. 437, van Orden et al. 2003, p. 344), namely, they do not utilize “consciousness” qua hard-problem-of-consciousness.
give rise to dynamics irreducible to the parts in isolation (Peng et al. 1995). In this way, the thalamocortical system is interaction dominant in that a change to any one part radically alters the whole. This is one way to understand what Tononi means when he says that a conscious state has cause-effect power on itself and that the experience is irreducible (Tononi and Koch 2015). Of course, the parts of the thalamocortical system will be affected by things outside of it, such as inputs from other parts of the brain. However, those causal relations do not play a constitutive role in the conscious state that emerges from thalamocortical activity.

An example may help illustrate the point. Consider that under certain instances, visual perception (i.e., perceiving and acting in regard to visual information and related physiology) is an interaction-dominant cognitive phenomenon that spans brain, body, and environment across multiple spatial and temporal scales. It is interaction dominant because the components are participants in a reciprocal relationship so tightly integrated that changes to one part of the system percolates throughout the whole. For example, as a dog runs, optic flow alters the presentation of the world, which is related to and can cause alterations to leg speed, which is related to and can cause alterations to brain activity, and vice versa.

On the other hand, visual consciousness (i.e., phenomenal states with a visual character of experience) does not extend beyond the brain. Although visual consciousness can be related to visual perception (e.g., optic flow), the system underlying visual consciousness (i.e., thalamocortical system) does not alter the structure or dynamics of the world the way the cognitive system involved in visual perception does. When cognitive activities such as visual perception are interaction dominant, the causal relationship among brain, body, and world is symmetrical in the sense of reciprocal dynamics altering the state of all components. However, when conscious states such as visual consciousness are interaction dominant, the causal relationships underpinning global dynamics (i.e., the conscious state) remain in the thalamocortical system and thus holds an asymmetrical relationship with the rest of the brain, body, and world.

In short, consciousness qua thalamocortical system is integrated such that it alters itself. The thalamocortical system has reciprocal connections throughout the brain and can be altered by activity from outside itself. But the integration of consciousness qua thalamocortical system does not include changing the world in the reciprocal manner required of interaction dominance. Treating IIT in terms of interaction dominance provides another way to understand how Tononi’s example of a digital camera demonstrates that information is not enough to generate rich conscious experiences, but it requires integration of that information (Balduzzi and Tononi 2009).

Tononi begins by explaining that a photodiode has 1 bit of information because it can only be in one of two states: either on (exposed to light)
or off (not exposed to light). An “idealized digital camera” comprised of one million photodiodes could discriminate among $2^{1,000,000}$ states that correspond to one million bits of information. However, such a digital camera is not generating rich experiences. This is because, although there is a lot of information, there is no integration among the information.

To couch this example in terms of component- and interaction-dominant dynamics: The digital camera is component-dominant because its activities are the mere sum of its individual parts, namely, putting together one million individual photodiodes. As a consequence of being component dominant, the photodiodes are not causally coupled to each other. What happens to one photodiode does not affect the others. However, in order to generate a rich experience with a high $\Phi$ value, the photodiodes would have to be integrated so as to comprise an interaction-dominant system.

If the digital camera was an interaction-dominant system, then each photodiode would participate in systems-level dynamics such that the global state of the system affects the local parts (each photodiode), while the properties of each local part affects the global system. Such a nonlinear and dynamical coupling underlies the kind of information integration required for rich conscious experience. This is what happens in the thalamocortical system and why it is likely the seat of consciousness. Though other parts of the brain (or nervous system, body, etc.) may have more cells or connections, their activities may reduce to a sum of their parts and connections. Though the thalamocortical system may have less cells or connections, its activities are more integrated via interaction-dominant dynamics, such that its properties are not reducible to a mere sum of the parts.

5. Conclusion

That the physical substrate of consciousness would give rise to a single, maximum $\Phi$ value accords with the phenomenon we wish to explain: consciousness is a single, subjective, dynamic, integrated, and highly informative state of a system. Despite the fact that consciousness is causally related to the body and environment, it is primarily constituted by a single, interaction dominant, nonlinearly and reentrantly connected system in the brain. Though cognitive activities can result from interaction-dominant dynamics that span brain-body-environment, such activities do not participate in the core physical substrate of consciousness.

The degree to which other parts of the brain, body, and environment participate in the core physical substrate of consciousness is an empirical issue. One hypothesis is that the thalamocortical system exhibits dynamics at spatial and temporal scales that do not allow for other parts of the brain, body, or environment to participate in its interaction-dominant
system. The particular spatial and temporal scales at which cognition occurs could be what facilitates its being able to so readily incorporate various parts of the body, non-biological tools, and the environment.

Consciousness may be a far more limited phenomenon in the sense that it does not readily admit of physical substrates of varying spatial and temporal scale. Such neurophysiological facts, along with the phenomenological considerations discussed above, are why we should think that consciousness is (probably) still only in the brain, even though cognition is not.\textsuperscript{5}

Acknowledgments

I would like to first thank Scott Jordan and Lana Kühle for inviting me to participate in the “Reality of Experience” workshop (Chicago, April 2016), which provided me an opportunity to present and discuss the ideas outlined in this paper with a great group of people. Second, I would like to thank Tony Chemero and Mason Cash for discussions related to various parts of the paper. Third, I would like to thank Mary Jean Amon for comments and edits. Finally, I would like to thank two anonymous reviewers for many insightful critiques that I believe improved this paper a great deal.

References


\textsuperscript{5}It was not until after I first submitted this manuscript for review that I came across Andy Clark’s (2009) similarly titled article “Spreading the Joy? Why the Machinery of Consciousness is (Probably) Still in the Head”. Though our conclusions are very much the same (“I conclude that the case for ECM [extended conscious mind] is at best unproven and that the machinery of conscious experience is (probably) all in the head / CNS”; Clark 2009, p. 987), I would like to think that I arrived at mine via a different but complementary route.


Received: 15 September 2016
Revised: 25 January 2017
Accepted: 28 January 2017
Reviewed by two anonymous reviewers.