
This article has appeared in the journal Animal Sentience, a peer-reviewed journal on animal cognition and feeling. It has been made open access, free for all, by WellBeing International and deposited in the WBI Studies Repository. For more information, please contact wbisr-info@wellbeingintl.org.
Caterpillars, consciousness and the origins of mind

Arthur S. Reber

Department of Psychology
University of British Columbia

Abstract: A novel framework for the origins of consciousness and mind, the Cellular Basis of Consciousness (CBC), is presented. The model is based on a simple, perhaps radical axiom: subjectivity is an inherent feature of particular kinds of organic form. Experiential states, including those denoted as "mind" and "consciousness," are present in the most primitive species. The model has several conceptual and empirical virtues, among them: (a) it (re)solves the problem of how minds are created by brains — also known as the "Hard Problem" (Chalmers 1995) — by showing that the apparent difficulty results from a category error; (b) it redirects the search for the origins of mind from complex neural structures to foundational biomechanical ones; and (c) it reformulates the long-term research focus from looking for "miracle moments" where a brain is suddenly capable of making a mind to discovering how complex and sophisticated cognitive, emotional and behavioral functions evolve from more primitive ones.

Keywords: mind, consciousness, hard problem, evolutionary biology, locomotion

Arthur S. Reber is Broeklundian Professor of Psychology, Emeritus, Brooklyn College and Graduate Center of the City University of New York, He is Fellow of Association for Psychological Science, American Association for the Advancement of Science, Fulbright Foundation, and currently Visiting Professor, Psychology, University of British Columbia. His research is on implicit learning: the process through which knowledge about the world about us is picked up largely independently of awareness of both the process and products of that learning. http://academic.brooklyn.cuny.edu/userhome/psych/areber/
Introduction

This essay revives and updates a model of the origins of consciousness that I first put forward some twenty years ago (Reber 1997). That paper had two aims. The first was to critique the notion that consciousness could be explained by a version of philosophical functionalism that maintained that mind could be viewed as "hardware independent." This view, now called "computationalism" (Scheutz 2002), is based on the assumption that if you could capture the functional, computational features that make up human cognition, you would have a genuine artificial intelligence capable of instantiating human mental states and exhibiting genuine consciousness and subjectivity.

The second aim was to put forward an alternative, bio-centric vision of consciousness by embedding it in a larger model of evolutionary biology. The key assumption was that consciousness and mind were inherently dependent on biological forms and functions, that mental states are intrinsically hardware dependent, and hence that a conscious computer could not exist.

Here, I jettison the first aim. The points I argued for are now largely accepted. Few take the hardware-independent assumption of computationalism seriously today and the arguments of those who do (Kurzweil 2013) tend to be more "futuristic" speculations than serious proposals. James Reggia's recent (2014) overview identified several current computational models but concluded that

"... no existing approach to artificial consciousness has presented a compelling demonstration of phenomenal machine consciousness, or even clear evidence that artificial phenomenal consciousness will eventually be possible."

The fact that artificial entities are capable of behaving in ways that appear to be analogs of things that organisms with minds do in no way implies that they have minds. Poker-playing bots can beat many humans at the game but it is not because they "know" anything about poker (Bowling, et al. 2015; Sandholm 2010). The analogy is empty.

In the rest of this essay I will resuscitate and update the original claims about the origins of consciousness and put them into the larger context of animal sentience. Because many of the arguments in this essay are derived from Reber (1997), there will be a certain amount of "self-plagiarizing."

For stylistic reasons I use a number of quasi-synonyms for the umbrella term "animal sentience." References will be made to "mind," "consciousness," "phenomenal state(s)," "subjective experience," "qualia" and "mental state(s)." They are all meant to draw attention to the same underlying element, that all organisms have shared internal experiences and feelings, that all animals share a broad spectrum of sentience. "Conscious state" is a synonym of "phenomenal state" and both carry the same denotative meaning as "sentient
state." I am well aware of the many lexicographic niceties here (see Reber 1993, Reber, Allen & Reber 2009) but the arguments made will stand (or fall) on evolutionary biological principles, not linguistic ones.

**The Origins of the CBC Model**

Intellectual voyages have starting points. This one began when I engaged with a small, green caterpillar munching on one of my basil plants. Initially, I looked at it through objectivist eyes. Its actions seemed routine, repetitive, and limited in scope. There were no prayers to the god of basil, no joining with conspecifics to explore new ways to harvest the crop, no sense of mental states or cognitions. Yet its behavior didn't seem all that mechanical. It wasn't much of a stretch to view it as engaging in something akin to a choice process — determining where to chew next, how large a piece to tear out, assessing whether it is safe to eat now or better to check for predators. I looked at the other leaves, each clipped in much the same fashion as one might expect, but each also noticeably different, as though novel decisions had been made as the caterpillar worked its way from leaf to leaf. I couldn't shake the feeling that not only did my caterpillar have a mind, but it had consciousness and that both of these phenomenal states were essential elements of its being. And I began to worry about the anthropomorphic fallacy.

**Appreciating and Adapting Anthropomorphism**

We behavioral and cognitive scientists have been trained to view anthropomorphism as a seductive demon at which to shake objectivistic garlic.1 We were cautioned against endowing species with limited affective or cognitive capacity functions that they likely did not have. I think we've gone too far. Two dogs romping, rolling around in the grass, and bounding over and around each other look happy. I have no doubt they are and that their internal experiences are on an evolutionary continuum with mine. Avian species learn patterns implicitly (Chen, van Rossum & ten Cate 2015; ten Cate & Okanoya 2012), and their data are remarkably similar to those from experiments on implicit pattern learning with adult (Reber 1993) and infant humans (Gomez & Gerken 2000; Saffran, Aslin & Newport 1996). Crows fashion tools (Weir, Chappell & Kacelnik 2002) and certainly look like they have intentionality. It is far more cumbersome to explain their behavior without assuming they possess a primitive version of what many vigorously maintain is a purely human mental function (see Heyes & Dickinson 1990 for an overview). When we encounter a wounded creature, even one with whom our common ancestry is in the distant past, we feel an empathic connection — and that affective reaction is a hint that we apprehend the

---

1 There are more than a few instances where anthropomorphism has led us astray. Perhaps the most famous is the horse "Clever Hans" whose trainer, having became convinced that Hans had many human cognitive abilities, carried out methodologically flawed tests that initially appeared to prove it (see Pfungst & Rosenthal 1965).
shared emotive valence, the common sensory core. That starfish curling its arms when cut experiences pain and it is on a continuum with my pain when my arm is cut.2

The move toward appreciating the hidden strengths in anthropomorphism has long been championed by Frans de Waal (see de Waal 2016) who has argued that acknowledging cross-species empathy is not only acceptable, it's an important tool in the researcher's armamentarium. Where caution needs to be exercised is when the analysis is of a process or mechanism that does not satisfy this species-continuity, where the functions are those with distinct evolutionary histories or when any apparent continuity is missing. For example, the cognitive mechanisms and their underlying neural substrates that allow for processes like self-referencing, apprehending nested dependencies, inferential reasoning, metacognitive analysis3 and the like should still be approached with the standard caution against anthropomorphic over-reach.

The CBC’s Evolutionary Framework – A Simple Axiom

Axiom: Mind and consciousness are not unique features of human brains. They are grounded in inherent features present in simpler forms in virtually every species. Any organism with flexible cell walls, a sensitivity to its surrounds and the capacity for locomotion will possess the biological foundations of mind and consciousness.4

Complexity has its roots in simplicity. Evolution has a pyramidal schema. Older forms and functions lie at the base, the more recently evolved ones toward the zenith (Reber 1989, 1992, 1993). In virtue of the nature of pyramidal systems, the older structures and the behaviors and processes that utilize them will be relatively stable, showing less individual-to-individual and species-to-species variation. They will also, in virtue of their foundational status, be robust and less likely to be lost. Adaptive forms and functions are not jettisoned; they are modified and, if the selection processes are effective, they will become more complex and capable of greater behavioral and mental flexibility and power.

2 In a paper provocatively titled "Why fish do not feel pain," Key (2016) maintains that our experience of pain is bound up with specific neurobiological structures and cortical pathways not found in the genus piscine. However, from the CBC perspective, because fish and humans have common evolutionary roots, the underlying subjective experience shares a common core and hence a common affective state. The particular physiological mechanism in use is not an issue. There are multiple ways to experience pain. This argument is developed in more detail below.

3 Metacognition may need to be removed from this list. Rosati and Santos (2016) reported circumstances under which rhesus monkeys know when they do not have information they know they need and shift food-searching strategies to gain the additional knowledge.

4 While recent work (Dener, Kacelnik & Shemesh 2016) shows that plants make risk-sensitive root-growth "decisions" based on temporal variation in nutrients, I am excluding plants and fungi on the grounds that they have rigid cell walls composed of cellulose, hemicellulose and pectin (plants) or chitin (fungi), and lack the capacity for endogenous locomotion. Sheets-Johnstone (1999) made a similar case in The Primacy of Movement. The actions of a bacterium as it moves, feeds, stops, starts and changes direction led her to conclude that "...a form of corporeal consciousness is present" (p. 75). Cognitivists and philosophers of mind, she argued, have been so "mesmerized by brains ... (that) the larger creaturely world of which humans are a part is forgotten, egregiously slighted or arrogantly distorted" (p. 77).
In short, behaviors, functions and forms that evolved early will become established, secure and robust. They maintain their roles as core features of species that evolved later on the same biological platform. If we have minds and consciousness and subjective experience, then we will see rudimentary versions of them in our pre-hominid ancestors. If we have perceptual and memorial functions, we will find their roots in earlier, less-complex species. If we have subjective, phenomenal experiences, qualia, pains and pleasures, allures and revulsions, precursors of these endogenous states will be found in simpler organisms.

The Solution to the Hard Problem

This stance is radically different from the one usually taken by cognitive scientists and philosophers of mind. More typical is the one that, quoting Harnad (2016), asks

"... if organisms do have minds, the hard problem is to explain how and why such survival/reproduction machines would evolve minds: What is the added causal role and adaptive value of having a mind, over and above the causal role and adaptive value of just having the behavioral capacities themselves?"

(Emphasis in original.)

Casting the question in this way is in line with the view of Chalmers (1995), who initiated the current fascination with the "hard problem." So long as it is posed in this manner, it will still look "hard" because it assumes that there is some "added" element that comes from having a mind. However, from the CBC perspective the answer is easily expressed. Organisms have minds, or the precursors of what we from our philosophy of mind perspective think of as minds, because they are an inherent component of organic form. What gets "added" isn't ontologically novel; it's a gradual accretion of functions that are layered over and interlock with pre-existing ones.

The problem we've been struggling with — without much success — is the result of a category error. In contemporary philosophy and the cognitive sciences, what is traditionally denoted by the term "mind" is a recently evolved, multi-functional cluster of operations that occurs in humans and perhaps in some other species. All other forms of subjectivity are excluded. This stance has spawned a very difficult — perhaps unsolvable — problem: to ascertain where along the species continuum this mind-thing appears and what are the neural organizational properties that allow it. The CBC framework cuts the conceptual Gordian knot. All experience is mental. All organisms that experience have minds, all have consciousness.

The first benefit of this re-positioning of the problem is that the "hard problem" suddenly isn't hard anymore. It is replaced by other problems, ones that will almost certainly turn out to have their own difficult features; but it will be a very different kind of enterprise. Instead of trying to grasp the neuro-complexities in brains that give rise to minds, we can redirect
the focus toward understanding how particular kinds of basic, primitive organic forms came to have the bio-sensitivity that is the foundation of subjectivity.

It’s clear that this argument requires a commitment to a biological reductionism\(^5\) along the lines of that adopted by physicists. In modern particle physics, matter is treated as having features that are an inherent aspect of its physical form. When you get nuclear particles, you get the weak and strong nuclear forces; when you get mass, you get the warp in space-time we call gravity; when you get electrons, you get the electromagnetic force. That’s it; and if there turn out to be more fundamental properties, if there is a deeper level, the argument would hold there instead. You start with a big bang with the properties of ours some 13.8 billion years ago and you end up with four forces: weak and strong nuclear, electromagnetic, and gravitational.

Physicists don’t have any trouble with this kind of fundamental reductionism (see Carroll, 2016) but social scientists do and that can create problems. We have traditionally found ourselves pulled toward complexity and then we have to work to unpack its many components and reveal the various interactions among them — dutifully assessing effect sizes and partialing out variance. Often this approach works. It’s how we’ve done some of our best science. But when it comes to issues like consciousness and the origins of mind, it’s been an impediment to progress because, often without realizing it, we were confronted with the problem of Emergentism.

**On the Emergentist’s Dilemma**

Harnad (2016), once again, phrased the issue succinctly and, perhaps unwittingly, revealed the problem.

"... evidently organisms do have minds, for some reason or other — or at least some species do: Which ones? Clearly our own species does. Which others? And what does it mean to 'have a mind'?"

Buried in this seemingly innocent statement is a deep problem, one that virtually no one working on the general issue of mind and consciousness seems to have come to grips with. It is the *Emergentist’s Dilemma* and it is a nearly unsolvable one unless we break out of the approach we’ve historically taken.

Is it reasonable to say, as I did above, that gravity is an inherent property of matter or that the strong nuclear force is a fundamental property of an atomic nucleus? I think so, if by this we mean the same kind of thing that I do when I argue that consciousness is an inherent

---

\(^5\) Reductionism is, properly, an umbrella term for a cluster of scientific principles that are best viewed from a pragmatic perspective. That is, reductionist exercises are not mad, unthinking plunges toward hypothetical basic principles. They are coherent voyages carried out with an eye as to which level of analysis carries the possibility of increased explanatory power for the questions being asked. Similar arguments for a psycho-neural reductionism have been put forward by John Bickle (2008).
property of particular kinds of organic form. The parallel is inviting. Particular kinds of
physical matter carry with them particular properties. Particular kinds of biological stuff
carry with them particular properties. In one case the properties are forces; in the other
they are subjective experiences.

The standard approach described by Harnad leads to asking what properties of brains
permit minds to emerge. Emergentist arguments have ranged from the simple-minded to
the sophisticated. My favorite example of the former is from an early science fiction story
(Clarke 1961; Levi 1968) where the world’s telephone system became so complex that it
reached "critical mass" for consciousness. This new, very artificial mind announced its
arrival with a birth cry: every telephone in the world rang simultaneously.

The more sophisticated versions are exemplified by the work of neuroscientists like
Edelman and Damasio, and neuro-computationalists like Penrose and Hameroff. Edelman
(1989, 1992), for example, maintained that consciousness could emerge from brain
functions based on the principles of neural Darwinism. He argued that a critical feature in
the development of consciousness was the evolution of two segregated neural systems: one
geread toward internal, hedonic processes of regulation and the other focused on encoding
external information about the state of the world. Put simply, he hypothesized separate
neurological systems for distinguishing self from non-self, me from thee.

Damasio’s (2010) model focused on identifying the neural structures in the brain and the
specific particular pathways which, he maintained, have causal roles in creating subjectivity
and the concept of self. For Damasio subjectivity itself emerges as a property of
neurologically-based structures (brains) and, presumably, is absent in brains that lack these
structures.

The Penrose and Hameroff model (Hameroff & Penrose 1995; Penrose & Hameroff 1995) is
built on the two-pronged argument that consciousness must involve non-computational
elements and that such internal states may be captured by a proposed quantum coherence
in the microtubules found in all cells. This move allowed Penrose (1994) to entertain the
notion that consciousness can be contained in single-celled organisms and does not require
neuronal systems – an argument with which I am comfortable. The problem, however, is
that there is little evidence for the hypothesized quantum effects within cells, in particular
neural cells in human brains, and good reasons for suspecting that they are not possible
(Litt et al. 2006).

But these and other models that speculate about how brains could create consciousness all
suffer a fatal flaw: they require a miracle. They all need to specify the properties of complex
neural states that give rise to conscious states; they have to identify the neural or
computational properties that permit a brain to make a mind. And worse, they need to
explicate why, when a neurological system crosses some threshold of complexity or hits
upon the right set of interconnecting neural centers and pathways, it suddenly "makes
consciousness." It is reminiscent of the classic Sidney Harris cartoon of two scientists
standing at a blackboard covered with mathematical arcana except for one beckoning blank spot. One scientist points to the spot and says, "And then a miracle happens." Admittedly, the CBC also calls for a miracle, but it's a very small one and far more tractable than the ones that current neuroscientists and philosophers of mind have put on the table for us.

**Common Threads with Other Areas of Research**

While these arguments may appear radical, they really aren't. The CBC's evolutionary framework is a familiar one. It is precisely the approach we've always taken to domains of study like emotion, memory and perception. Kandel's seminal work on the biology of memory (e.g., Castellucci, Pinsker, Kupfermann & Kandel 1970) was carried out on a rather simple species, *Aplysia californica*. Research on the foundations of vision focused on the simple ommatidia of crabs and starfish (Garm & Nilsson 2014). The approach to human emotion has traditionally been to view our experiences as extensions of similar endogenous states seen across the species spectrum. A sea-slug may only have a rather limited form of memorial representation but it shares a common biological core with ours. A horseshoe crab or a starfish may carve up the visual space differently from us but we can find a common bond in the adaptation of light-absorbing pigments to provide a species with information needed to navigate a complex environment.

We've been unwilling to take this step with topics like consciousness and mind. It's time to do so. It's time to change the questions asked — and, as has oft-been noted, such a shift in epistemic inquiry can be the opening to finding better answers. We need a reductionist theory of biological systems and an understanding of the circumstances under which a subjective state accompanies just the right kind of biological state. The real issue is not how particular kinds of complex organizations of neural structures give rise to mental life, but how organic systems of varying complexity permit the evolution of cognitive systems of varying richness: how more sophisticated forms of consciousness and mental states develop as biological/neurological structures become more richly evolved and complex.

**On Key and Pain in Fish**

Recently in this journal, Brian Key (2016) put forward the argument that we cannot know (for certain) whether fish feel pain or whether what they experience is in any demonstrable way comparable to what we feel. Key's thesis certainly hit many a nerve. Dozens of scholars wrote commentaries, both supportive and critical, and Key penned several responses.

The issues raised by Key concern the cortical and sub-cortical structures involved in processing pain, the knowability of the subjective states of other organisms, the problems associated with interpretations of overt behavior, and what I regard as old-fashioned worries over anthropomorphism. No resolution was reached.

The CBC framework renders Key's arguments moot. Every organism feels pain. Pain is what an organism experiences when making contact with an object or event in the environment.
that is causing or can cause tissue damage. It is marked as a negative experience by the organism acting to remove itself from that stimulus — a key adaptive response with clear selection advantages.

Key might wish to argue that there is no way to know what the organism is experiencing or that it can never be known to be akin to my pain when I touch a hot stove. This may be true, but it is of little consequence and skirts the core issue. It would be like arguing that there is no way to know whether one of Kandel’s sea-slugs really had a memory of its experiences, or whether a horseshoe crab really had a visual sensory experience of the food source it quickly moved toward. In all such cases, we’re merely confronted with an evolutionarily early manifestation of a function that lies on a continuum and is experienced by all organisms with common genetic roots.

Key’s position, moreover, runs smack into the Emergentist’s Dilemma — identifying the point in species development where this non-pain becomes pain. It is far more likely that all the varying experiences with negative valence lie along a continuum where properties are added, nuances are developed and novel representations are acquired with increasing neural complexity.

**Coda — Deflecting Mysterianism**

The CBC framework solves or perhaps better, resolves the "hard problem." Mind and consciousness become end-points on continua and not distinct ontological entities. The category error is corrected.

Physicists are fond of noting there they have a fairly deep understanding of quantum mechanics and equally solid theoretical foundations of macro-matter but they still don’t know how to get from the quantum level to a coffee cup. Admittedly, the CBC framework presents us with an analog of this conundrum but it’s one that will be a lot easier to get a handle on than the one Chalmers put before us; and it rescues us from the tendency to drift into the dualism that he advocated.

In his largely positive review of physicist Sean Carroll’s *The Big Picture*, philosopher Barry Loewer (2016) noted that Carroll’s reductionist argument\(^6\) is problematic owing to its failure to address the issue of how "thought, consciousness and free will fit into physical theory." Statements of this kind reflect that dualist tug, the one that keeps getting us into trouble just because it makes mysterianism look attractive. But Carroll is right — well, almost. The place to focus is not on quantum effects but on the way fundamental physical laws operate to create living organic forms and how these primitive forms instantiate subjectivity.

---

\(^6\) Carroll’s thesis is that the only coherent approach to understanding our universe is to recognize that everything, including social phenomena, economics, poetry and morality, must be viewed as reducible to fundamental laws of quantum mechanics and relativity.
Reactivity, consciousness, intentionality, memorial representation, creativity and all the rest are the manifestations of particular kinds of organic stuff. When you get simple collections of this stuff, you get primitive, subjective reactivity; when you get complexly organized and structured stuff, you get complex mentation and representation, self-reflection and modulation, insight and ecstasy, despair and depression. No new ingredients are needed: simply an understanding of how the old ones function and permit the evolution of the newer.

Finally, is there a downside to the CBC approach? In a sense there is. Should it become the industry standard, philosophers of mind and cognitive neuroscientists will have to abandon that small guilty pleasure that comes from contemplating the seemingly vast gap between mind and matter and renounce the little mysterian tweaks that accompany asking questions like "how does the brain make the mind."

Call for Commentary: Animal Sentience publishes Open Peer Commentary on all accepted target articles. Target articles are peer-reviewed. Commentaries are editorially reviewed. There are submitted commentaries as well as invited commentaries. Commentaries appear as soon as they have been revised and accepted. Target article authors may respond to their commentaries individually or in a joint response to multiple commentaries.

Instructions: http://animalstudiesrepository.org/animsent/guidelines.html
References


De Waal, F. (2016) *Are We Smart Enough to Know How Smart Animals Are?* New York: Norton.


