

Going Out of My Head: An Evolutionary Proposal Concerning the “Why” of Sentience

Stan B. Klein, Bill N. Nguyen, & Blossom M. Zhang

University of California, Santa Barbara

Author Notes: I want to thank Alba Papa-Grimaldi for her persistent guidance and unwavering support. Ditto Rob Kundendorf. Please address all correspondence to: Stan Klein, Department of Psychological and Brain Sciences, 551 Ucen Road, UCSB, Santa Barbara, CA 93106. Email: [klein@psych.ucsb.edu](mailto:klein@psych.ucsb.edu).

### Abstract

The explanatory challenge of sentience is known as the “hard problem of consciousness”: How does subjective experience arise from physical objects and their relations? Despite some optimistic claims, the perennial struggle with this question shows little evidence of imminent resolution. In this article I focus on the “why” rather than on the “how” of sentience. Specifically, why did sentience evolve in organic lifeforms? From an evolutionary perspective this question can be framed: “What adaptive problem(s) did organisms face in their evolutionary past and how were those challenges met? I argue that sentience was a critical component of the adaptive solution (i.e., adopting an agentic stance) to increasingly complex and unpredictable demands placed on vertebrates approximately 500 million years ago (the so-called Cambrian explosion). One consequence of taking an agentic stance is that it freed the organism from its neural moorings, positioning it within phenomenal space outside its brain.

## Going Out of My Head: An Evolutionary Proposal Concerning the “Why” of Sentience

“Ignoring the origin of things is always a risky matter. It is even more risky in an effort that purports to explain mental events. But that is what has happened in much of the history of psychology and the philosophy of mind.” (Edelman, 1992, p. 33)

Sentience is the capacity to experience life. More formally, it is the subjective or qualitative feeling had when undergoing an experience. An organism is sentient if and only if there is “something it is like” for “organism X to be in a mental state Y” (e.g., Chalmers, 1996; Nagel, 1974; Hacker, 2002). Sentience is what most philosophers have in mind when discussing phenomenal consciousness (e.g., Chalmers, 1996; Klein, 2014a; Strawson, 2009).<sup>1</sup>

The theoretical challenge of sentience is called the hard problem: How does subjective experience arise from physical objects and their relations (e.g., Banick, 2019; Chalmers, 1996; Clement & Malerstein, 2003; Georgalis, 2006; Jackson, 1982). How is experiential reality (that aspect of reality of which we can be most certain; e.g., Gallagher & Zahavi, 2008; Klein, 2015; Midgley, 2014; Strawson, 2009; Wittgenstein, 1958) possible, given that the dictates of contemporary science stipulate that everything from molecules to minds is wholly physical (for reviews, see Crane & Mellor, 1990; Klein, 2016; Strawson, 2009)?

Some argue that the hard problem is, and will remain, intractable in consequence of its incommensurability with the requirements of scientific method and explanation (e.g., Levine,

2001; Wright, 2007). Others attribute its recalcitrance to conceptual limitations of the human mind (e.g., Chomsky, 2016; McGinn, 1991; Plonitsky, 2010). Still others deny the hard problem exists, arguing either (a) a solution already is at hand (e.g., Graziano, 2019, 2022; Kastrup, 2019; Tsuchiya, 2017) or (b) the problem is chimeric -- a quixotic attempt to imbue substance to an ill-formed question (e.g., Carruthers, 2000; Dennett, 1991; Jackson, 2003).

Theorists who accept that there is a hard problem often attempt to circumvent the explanatory impasse by treating consciousness as consisting of sentience along with a host of mental functions (e.g., thoughts, perceptions, concerns, problem solving, language, judgments, evaluations, remembering, social skills – all of which could, in principle, be effected by a philosophical zombie; e.g., Kirk, 1974). Definitional inclusivity provides room to maneuver around sentience while still having something to say about consciousness (e.g., Budson, Richman & Kensinger, 2022; Dennett, 1991; Farthing, 1992; Graziano, 2019; Kotchoubey, 2018).

Unfortunately, theoretic expositions that conflate the objects of consciousness<sup>2</sup> with that which enables those objects to be subjectively realized fail to shed light on the target of inquiry – how is there something it is like for organism X to be in a mental state Y? Sentience makes possible the subjective experience of its intentional objects. As such it cannot be thematized by those objects. Put differently, consciousness must, of logical necessity, be directed toward that which it is not -- some “other” that can serve as its target. To understand consciousness, one cannot transform it into an object of consciousness -- that is, into a non-sentient entity suitable for conscious apprehension (e.g., Earle, 1955, 1972; Husserl, 1964; Kant, 1998; Klein, 2012, 2014a; MacMurray, 1957/1969; Neuhausser, 1990; Rossman, 1991; Strawson, 2005).<sup>3</sup>

Viewing mental objects and processes as aspects of consciousness -- rather than as non-conscious entities that can be subjected to reflective acts of consciousness -- has resulted in a proliferation of types and subtypes often having little in common beyond the designation “consciousness” (e.g., access consciousness, noetic consciousness, auto-noetic consciousness, temporal consciousness, core consciousness, reflective consciousness, primary consciousness, phenomenal consciousness, proto consciousness; e.g., Klein, 2014a, 2020). What is being explained by this fractious cohort is unclear.

A cause of the intermixing of sentience and its objects is that most psychological, philosophical and neuroscientific examinations of consciousness take as their explananda introspective reports and behaviors of humans (typically adult) in varying states of cognitive well-being (for reviews, see Hurlburt, & Schwitzgebel, 2007; Schwitzgebel, 2019). While an anthropomorphic focus has pragmatic utility (relying on the verbal reports of non-humans is a fool’s errand), it has disadvantages. Most important, it targets a product of natural selection that has built-in complexity courtesy of millions of years of evolutionary tinkering. Consciousness in modern humans is overlain with mental processes and faculties that, while capable of being taken as its objects, are not conscious in and of themselves. Little wonder contemporary treatments of consciousness consist in a chaotic, often disjunctive, compilation of theories and micro theories.

### 1. Goals and Investigative Approach

In this article my focus rests squarely on sentience. Since I have no idea “how” subjective experience is possible (i.e., the hard problem of consciousness), I restrict analysis to the question

of “why” evolution favored sentience. What is the survival/reproductive advantage of experiencing life from a first-person perspective?

This question -- “what, if anything, does sentience add to the functioning of an organism?” – presents (in my opinion) a far less formidable (and hence more tractable) challenge than do attempts to explicate the mechanisms that enable non-sentient matter to have subjective experience. For instance, it sidesteps highly contentious, perennial controversies posed by the “materialism vs dualism” debate.

Investigative humility – i.e., training one’s initial efforts on aspects of a problem most compatible with available methods of inquiry, rather than attempting to untangle its stubborn intricacies in one fell swoop – is (again, in my opinion) a prerequisite for progress. This especially is true when investigating consciousness – a phenomenon whose properties are sufficiently enigmatic (e.g., Chalmers, 1996; Humphrey, 2006; McGinn, 1991; Searle, 1997) that even the most basic pre-conditions for reasoned analysis (e.g., knowing what it is one is investigating) prove elusive. Accordingly, to understand what it means for an organism to be sentient, a “prioritize, divide and conquer” strategy may be the best approach.

## 2. The “Why” Question: An Evolutionary Approach to Understanding the Functional Significance of Sentience

There are a number of proposals attempting to explicate the function(s) of consciousness (e.g., Dretske, 1997; Graziano, 2022; Kotchoubey, 2018; Velmans, 2009). Most deal with functionality evidenced by contemporary Homo sapiens. Unfortunately, as discussed above, this is the wrong place to begin an inquiry. Training efforts on the culmination, rather than the

commencement, of consciousness's evolutionary journey presents one with an unruly montage consisting of sentience interwoven with processes and content that stand in relation to consciousness, but are not consciousness per se. Trying to understand consciousness from this vantage point is like trying to understand the foundation of a home following years of remodeling.

For this reason, my exploration into the “why” of consciousness is guided by the evolutionary principle that the human cognitive architecture as it exists today embodies the culmination of refinements driven by problems faced by the organism's ancestors during the evolutionary past (e.g., Anderson, 1989, 1991; J. R. Anderson & Milson, 1989; Cosmides & Tooby, 1987; Klein, 2020; Klein, Cosmides, Tooby, & Chance, 2002; Sherry & Schacter, 1987). Naturally selected adaptations occur through the gradual modification of existing structures. Over time, modifications in design were likely to be incorporated to the extent they improved the functional operations of the basic architecture -- that is, increased the rate that the architecture solved adaptive problems (e.g., Cosmides & Tooby, 1987; Klein, 2020; Klein et al., 2002; Mayr, 2001; Nairne, 2005; Sherry & Schacter, 1987; G. C. Williams, 1966).

Accordingly, the most expedient approach to understanding the “why” of present-day sentient beings is to consider the adaptive challenges faced by their progenitors. In this way, we catch a glimpse of the underlying framework (the conditions that gave birth to sentience) unaccompanied by the clutter of evolutionary embellishments (processes and content subsequently conjoined with sentience to enhance its adaptive potential).

### 3. In the Beginning

In the earliest phase of organic evolution (approximately 3.5 billion years ago), life consisted primarily of micro-organisms that lived within deep-sea hydrothermal vent precipitates (e.g., Schopf, 2006). Multicellular terrestrial life appeared about 2.4 billion years later (e.g., Strother, Battison, Brasier, & Wellman, 2011).

Among the terrestrials, our interest is with the vertebrates. Vertebrates comprise all animal taxa within the subphylum Vertebrata, including mammals, birds, reptiles, amphibians, and fish. Phylogenetically, their lineage traces to the eukaryotes -- complex cells containing organelles (e.g., cell nuclei, ribosomes, mitochondria). All animals are multicellular eukaryotes (e.g., Cowen, 1995).

Vertebrates made their first appearance during what is known as the Cambrian explosion (e.g., Ginsburg & Jablonka, 2007, 2010, 2019; Zhang & Shu, 2021). This period (henceforth abbreviated CE), which spanned approximately 25 million years beginning around 545 million years ago, is considered one of the most significant transitions in evolutionary history (e.g., Marshall, 2006; Vallentine, 2004): In a relatively short time (by evolutionary standards) essentially all animal phyla first appear in the fossil records (e.g., Vallentine, 2002; Zhang & Shu, 2021).

The cause of the CE is subject to debate (e.g., changes in the oxygenation or temperature of the biotic environment; for discussion see Marshall, 2006; Vallentine, 2004; Zhang & Shu, 2021). One suggestion is that the extraordinary ecological and morphological diversification found during the CE stemmed from a genetic reorganization of the central nervous system (CNS)



that occurred in parallel among several groups of Cambrian metazoans (e.g., Ginsburg & Jablonka, 2010, 2019).

#### 4. Life on the Inside

While it is possible sentience existed prior to the appearance of vertebrates, the vertebrate lineage offers a logically defensible point of departure. Unlike sessile forms of life (organisms lacking a means of self-locomotion; e.g., coral, oysters, barnacles), vertebrates possess two features essential to the development of sentience. First, they have a CNS consisting of a brain and spinal cord. While precursors -- such as action potentials and simple nerve nets (i.e., neurons lacking a brain or cephalization) -- can be found in organisms predating the vertebrates (for example, motile, single-celled, colonial eukaryotes), appearance of the CNS dates to the first animals (e.g., Anctil, 2015).

Second, they are motile. That is, they possess the ability to change their location in the environment using energy produced by their metabolic activity. This enables them to move within, and act on, their surroundings without relying on forces originating outside the body. As I argue in section 6, these two criteria -- a central nervous system and motility -- set the stage for the emergence of sentient beings.

The functional efficacy of neurally-generated movement was greatly facilitated by evolutionary refinements of sensory organs -- i.e., biological systems used by an organism for gathering information about the world through the detection of stimuli. During sensation, these organs collect various stimuli (such as visual or tactile) for transduction (i.e., transformation into an electro-chemical discharge) that can be interpreted by the brain.

Information transmitted to the CNS via sensory organs can be used to shape and enhance the adaptive potency of an organism's interaction with its environment. While a relation between sensory organs and brain-like structures likely was in place in advance of the CE (e.g., Anctil, 2015), a notable refinement in sensory-brain interaction appears to have taken place during the CE (e.g., Feinberg, & Mallatt, 2016; Ginsburg & Jablonka, 2007, 2010, 2019). Although mental processes leave no fossil records, it can reasonably be conjectured that Cambrian vertebrates were able, in virtue of enriched sensory-brain linkage, to form more detailed neural mappings of their surroundings, enabling a host of complex adaptive behaviors (e.g., Feinberg, & Mallatt, 2016; Ginsburg & Jablonka, 2019).

The type of behaviors in evidence during the early stages of the CE likely consisted in movement occasioned by genetically acquired action schemata in concert with ontogenetic adjustments (i.e., modifications acquired in the organism's lifetime) executed in response to sensory detection of environmental stimuli. Activity for pre-sentient organisms was not an intentional effort to act one's surroundings. It was simply movement in response to a stimulus (e.g., MacMurray, 1957/1969). An example from robotics may help.

Consider the Roomba. The Roomba is a robotic vacuum cleaner that has a set of sensors which, in conjunction with robotic drives, enable it to navigate the floor area of a home. Its sensors consist in onboard mapping and navigation software that can detect the presence of obstacles. This sensory feedback is encoded in the machine's software and used to construct an electronically coded map or "floor plan", which is stored in the vacuum's central processor and updated with information about areas that already have been cleaned. Used in conjunction with the vacuum's drives, the floor plan allows it to navigate floor surfaces while avoiding obstacles.

Prior to the advent of sentience, vertebrate behavior was analogous to that of the Roomba: Animals acted in concert with their environment, based on “detection” of their surroundings. But “detection” was objective, not phenomenological. Just as a Roomba’s “detection” is nothing beyond movements made in response to internally-coded information, vertebrate “detection” was not an “awareness” of anything. Rather, it was movement, caused -- not intended -- by neurally-instantiated, cranially-located structures linking input with output.

For example, a neural system activated by detection of a stimulus will produce response (assuming the system is functioning normally) regardless of whether the neural representation resembles the physical attributes (e.g., size, color, shape, location, texture, movement) of the stimulus that set the chain of events in motion. For such organisms, the stimulus is nothing over and above the brain state(s) enabled by biological systems designed to gather information about the world (e.g., electromagnetic radiation) and translate it into neuronal spike trains (e.g., Aljadeff, Lansdell, Fairhall, & Kleinfeld, 2016; Nolan, 2011).

Thus, for pre-CE vertebrates, no meaningful distinction can be drawn between the physical world and its neural instantiation. The world is that which is in its head. This is not an argument for strong idealism (e.g., Berkeley, 1710/2003; Kastrup, 2017). Nor is it a denial of the (in my view, logically compelling; see the next section) assumption that neural mappings, to some unspecifiable degree (e.g., Forrester, 2014; Levine, 2003), capture a reality accessible to the organism’s sensory systems (the nature and existence of a reality independent of neural instantiation is topic that has been passionately debated for centuries; e.g., Berkeley, 1710/2003; Chalmers, 2019; Cornford, 1957; Furlong, 1941; Kant, 1929/1965; Kastrup, 2017; Locke, 1689-1700/1975; Moore, 1939; Nath, 2016; Rogers, 1975; Russell, 1912/1999). Regardless of where

one stands on these issues, the stubborn fact remains that the world for pre-CE vertebrates existed exclusively in terms of electro-chemical signaling among neurons.

### 5. The Problem: Adapting to a World in One's Head

Evolution trades in functionality. Accordingly, the senses and their CNS termini are not designed to provide an accurate, objective view of the physical world. Rather, they are selected on the basis of their ability to enhance their owner's chances of surviving long enough to reproduce (for discussion, see Klein, 2014b). Functionality requires only that a system work as designed, not that it remain faithful to some object or event.<sup>4</sup> Within the parameters provided by (a) environmental regularities and (b) the organisms' biological limitations, certain functional adaptations will work better than others. The ones that "work better" will be targeted by natural selection and passed to the next generation via sexual reproduction.

From a functional perspective, therefore, all that matters is that information stored in and retrieved from the brain is sufficient to meet the adaptive challenges faced by the organism. Within the (sometimes broad) constraints imposed by physical reality, neurally encoded content need not entail "precision of match" to objective reality as a criterion of success. Indeed, it is an ontological certainty that a neural representation needs bear little resemblance to the "thing-in-itself" (e.g., Hoffman, 2019; Kant, 1929/1965; Locke, 1689-1700/1975; Nolan, 2011; Plato, 2002; Russell, 1912/1999; von Uexküll, 1957).<sup>5,6</sup>

But if reality exists for the organism solely as neural representation, and the representation benefits survival, natural selection has no adaptation-driven imperative to extend reality beyond its cranial confines. For a system so designed, the physical world has no need of

observer-independent realization: It exists for the organism as it exists in the organism. A de facto solipsism characterizes non-sentient vertebrate being.

## 6. A Solution: “Being in the World”

Scholars dating back to Heraclitus have recognized that the physical world is in a state of continual flux (e.g., Brann, 2011; Cornford, 1957; Geldard, 2000; Whitehead, 1929). As environmental contingencies change, so do the challenges faced by the organism. Adaptations that facilitated survival in the evolutionary past will not necessarily be advantageous in the present. In consequence, extant adaptations come under pressure to modify their structure and operation or succumb to genomic purging. Successful modifications not only allow the organism to cope with existing contingencies -- they also present the organism new ways of “being in the world” (this phrase refers to the normal and lawful interaction between a subject and the physical environment; e.g., Binswagner, 1963; Zahorik & Jenison, 1998).

As challenges accompanying the CE diversified and intensified -- primarily in consequence of the proliferation of new, neurally sophisticated lifeforms competing for limited resources (e.g., mates, territory, food) -- the need to navigate an increasingly unfriendly landscape dramatically increased demands placed on vertebrate lifeforms (e.g., Ginsburg & Jablonka, 2010, 2019; Godfrey-Smith, 2020) . One solution to dangers posed by Cambrian existence would be for the organism to act as an agent rather than as a respondent.

### 6a. Respondent versus Agent

For the respondent, behavior is caused, not chosen. Once commenced, an act continues to completion along a pre-determined path, unaccompanied by awareness of having been issued

from a self or directed toward an other. For such creatures there is no “being in the world”; rather, “the world is in the being.”<sup>7</sup>

An agent, in contrast, behaves deliberately. Its behavior is purposefully chosen and directed toward effecting change in a world external to the agent. Prior to completion, agentic acts are subject to modification and correction based on the agent’s goals and interpretation of the situation. In short, for a respondent, the world exists in the organism; for an agent the world exists for the organism (for discussion see MacMurray, 1957/1969; Schlosser, 2019).

There are clear adaptive advantages accompanying acts fashioned by reasoned deliberation. An agent acts on its environment in virtue of being in its environment. In consequence, agentic acts are flexible -- they can be tailored to the contingencies as they present, and altered in accord with perceived changes of circumstance.

Prior to the CE, all taxa possessed of motility acted from response. However, as competition for resources intensified in consequence of the expansion of behavioral competencies, responses that could be deliberately fitted to the demands of an increasingly unpredictable world would be favored by natural selection. That is, conditions accompanying the CE provided a context in which acts issuing from agency would have adaptive priority.

#### 6b. A Stranger in a Strange Land

A necessary first step toward agentic behavior is to appreciate there is a world to behave in. For the pre-CE respondent, awareness of external reality had yet to make an appearance. Behavior originated within, and operated on, neurally-housed representations. The transition from respondent to agent required the organism to transform pre-determined, inwardly conceived

and directed acts into intentional behavior targeting objects located in a three-dimensional space outside its body. To experience the world as a phenomenological space within which the organism can move and interact, an agent must feel that its experiences are presented to, not simply present within, itself (e.g., James, 1904; Pereira, 2018; Pribram, 2004; Rudrauf, Bennequin, Granic, Landini, Friston, K., & Williford, 2017; Velmans, 2007, 2009).

To fashion a world external to the organism, neural activities must be “projected” onto the space outside the brain in which they originate (e.g., Pribram, 2004; Velmans, 2007). This process – “perceptual projection” (for discussion see (e.g., Pereira, 2018; Pribram, 2004; Velmans, 2007, 2009).<sup>8</sup> served as both the product of and occasion for observation of the physical world. To experience a world consisting of objects and their relations requires those objects be fitted with properties in virtue of which they can be individuated.

### 6c. Objects, their Properties and the Physical World

A property is a quality or characteristic (e.g., size, color, shape, texture, mass, orientation) that can be attributed to an object (for discussion see e.g., Dorr, 2019; Nolan, 2011; Orilia & Paoletti, 2022). All objects have properties. The collection of properties possessed by an object constitute and identify the object that possess them (e.g., Varzi, 2019).<sup>9</sup>

To act as an agent, an organism must differentiate the target of its behavior from non-targeted objects occupying phenomenal space. This is accomplished by conscious registration of the properties that constitute, and thus individuate, the object of interest. The evolution of sentience – i.e., the feeling of “what it is like for organism X to experience property X” (e.g., the color of an apple or the pain of a bee sting) -- is a necessary precondition for populating

phenomenal space with objects toward which an agent can direct its behavior. Absent subjective registration of object-defining properties (i.e., qualia; e.g., Shoemaker, 1990) there can be no object-oriented intentional movement (e.g., Orilia & Paoletti, 2022).

In sum, natural selection's answer to problems posed by the Cambrian explosion was to change respondents into agents. This transformation was affected, in large part, by breaking the organism free of its neural mooring and positioning it within the phenomenal space outside its brain. To enable this new way of "being in the world", external space was populated with phenomenal objects whose presence could be detected by sentient registration (i.e., "the feeling of what it is like to experience Y") of the properties of which those objects were composed.

## 7. Conclusions

In this paper I addressed the question: "What adaptive advantages does sentience confer on its possessor?" By limiting inquiry to the "why" rather than the "how" of sentience, I bypassed the seemingly intractable "hard problem" of consciousness, concentrating on issues more congruous with investigative resources.

While "investigative humility" might seem an exercise in evasion rather than explication, my position is that to best understand a construct for which even basic properties are puzzling, a "prioritize, divide and conquer" strategy is preferable to an attempt to explain the phenomenon in toto (an approach that unfortunately characterizes a sizable portion of theoretical treatments of consciousness; e.g., Dennett, 1991; Graziano, 2019; Godfrey-Smith, 2020; Kastrup, 2019).

From an evolutionary perspective, my answer to the "why" question is that to address adaptive problems posed by the Cambrian explosion (e.g., Feinberg, & Mallatt, 2016; Ginsburg



& Jablonka, 2019), evolution endowed the organism with the capacity to adopt an agentic stance toward environmental contingencies. This required the organism to project its internally-situated representation of reality into a phenomenal space existing outside its brain. Projection into an external space, in turn, required the evolution of sentience. On this view, sentience was both the consequence of agentic behavior and its *modus operandi*.

At some point after enabling the organism to take residence in the external world, sentience returned to its point of origin, empowering the organism to include mental states among its objects of consciousness. How this was affected, and what adaptive function(s) it served, are questions that are not, and need not be, addressed herein (i.e., the principle of interpretive humility). Some possibilities are considered in Humphrey (1992).

#### 7a. Blindsight and the “Why” of Sentience

Unless a conceptual meditation has consequences for how we think about a topic, it runs the risk of being seen as little more than a squabble over semantics. Consequently, I want to briefly discuss implications of my thesis for understanding one of the more puzzling phenomena in consciousness research -- blindsight.<sup>10</sup>

Blindsight is the ability of people suffering cortical blindness (i.e., loss of vision due to damage to the occipital cortex in the presence of intact sensory systems) to respond appropriately to visual stimuli they cannot consciously “see”. For example, in one study a patient was asked to locate visual stimuli presented on a screen. Because he claimed to be unable to consciously detect the presence of a stimulus, he was instructed to “guess” its location. Research showed that the patient was able to locate stimuli at levels of accuracy (often substantially) higher than would

be expected by chance (for review see Weiskrantz, 1997). The phenomenon of blindsight has led many to question whether our intuitions about the role of consciousness in visually-guided behavior needs to be revisited (for discussion see Holt, 2003).

I believe this concern may be premature. Blindsight, in light of views expressed herein, may simply be a clinically-occasioned example of agentic behavior rendered dysfunctional in consequence of clinical circumstance. Under these conditions, a response still can issue from evolutionarily more primitive, but clinically non-compromised, respondent-driven behavior.

To place this explanation in a rationally defensible context, it is necessary to recognize that the transition to agency during the CE was not accompanied by the elimination of non-agentic behavior. Agency was added to, not positioned in place of, the organism's existing ways of acting.

The coexistence respondent and agentic behavior remains true of present-day sentient beings (e.g., Milner & Goodale, 1995). Accordingly, an explanation for blindsight may be no more esoteric than the proposition that purely responsive behavior can be (and in the case of blindsight "is") executed in response to stimuli about which the organism is unable to adopt an agentic stance.

### 7c. Limitations

The argument for sentience follows from the argument for agency. Specifically, agency requires a world toward which the agent behaves. Such a world is, of bio(logical) necessity, a phenomenal projection of neurally-generated properties comprising objects positioned in space. To direct behavior toward a subset of those objects, they need to be individuated from non-

targeted objects in the same space. This is accomplished by conscious registration of their phenomenally-given properties. Hence, sentience.

It is conceivable, however, that agentic-like behavior could transpire absent the projection and detection of phenomenal properties in 3-dimensional space. Perhaps agentic-like acts can be enabled, or at least mimicked, by wholly non-conscious neural activity (e.g., the argument for philosophical zombies).

While such a proposal is (a) at odds with the definition of agency (i.e., agency consist in intentional acts directed toward the world in which the agent resides), (b) runs counter to an undeniable aspect of experience (i.e., that we are surrounded by a world outside the borders of our body), and (c) seems to violate the principle of parsimony (which is an heuristic, not a law of nature), the possibility of “non-sentient agency” (a conceptual oxymoron) cannot be categorically dismissed (cf. Hassin, 2013). Such is the nature of dealing with what arguably is among the deepest mysteries of the universe – consciousness.

### References

- Aljadeff, J., Lansdell, B. J. Fairhall, A. L., & Kleinfeld, D. (2016). Analysis of neuronal spike trains, deconstructed. Neuron 91, 221-159.
- Anctil, M. (2015). Dawn of the neuron: The early struggles to trace the origin of nervous systems. Montreal & Kingston, London, Chicago: McGill-Queen's University Press.
- Anderson, J. R. (1989). A rational analysis of human memory. In H. Roediger & F. Craik (Eds.), Varieties of memory and consciousness: Essays in honour of Endel Tulving (pp. 195–210). Hillsdale, NJ: Erlbaum.
- Anderson, J. R. (1991). Is human cognition adaptive? Behavioral and Brain Sciences, 14, 471–517.
- Anderson, J. R., & Milson, R. (1989). Human memory: An adaptive perspective. Psychological Review, 96, 703–719.
- Banick, K. (2019). What is it like to think about oneself? De se thought and phenomenal intentionality. Phenomenology and the Cognitive Sciences, 18, 919–932.
- Brann, E. (2011). The logos of Heraclitus. Philadelphia, PA: Paul Dry Books, Inc.
- Bennett, J. F. (1988). Events and their names. Indianapolis, IN: Hackett Publishing Company.
- Berkeley, G. (1710/2003). A treatise concerning the principles of human knowledge. Mineola,

Going Out of My Head

NY: Dover Publications, Inc.

Binswagner, L. (1963). Being-in-the-world; selected papers of Ludwig Binswanger. New York, NY: Basic Books. (Translated by J. Needleman).

Brentano, F. (1995). Descriptive psychology. London, United Kingdom: Routledge.

Budson, A.E ., Richman K. A., & Kensinger, E. A. (2022). Consciousness as a memory system. Cognitive Behavioral Neurology, 35, 263-297.

Carruthers, P. (2000). Phenomenal consciousness: A naturalistic theory. Cambridge, UK: Cambridge University Press.

Casati, R., & Achille, V. (2020). Events. In E. N. Zalta (Ed.), The Stanford Encyclopedia of Philosophy. URL = <https://plato.stanford.edu/archives/sum2020/entries/events/>

Chalmers, D. J. (1996). The conscious mind: In search of a fundamental theory. New York, NY: Oxford University Press.

Chalmers, D. (2019). Idealism and the mind-body problem. In W. Seager (Ed.), The Routledge Handbook of Panpsychism (pp. 353-373). New York: Routledge.

Chomsky, N. (2016). What kind of creatures are we? New York, NY: Columbia University Press.

Clement, F., & Malerstein, A. J. (2003). What is it like to be conscious? The ontogenesis of

- consciousness. Philosophical Psychology, 16, 67–85.
- Cornford, F. M. (1957). Plato and Parmenides. New York, NY: The Liberal Arts Press.
- Cosmides, L., & Tooby, J. (1987). From evolution to behavior: Evolutionary psychology as the missing link. In J. Dupré (Ed.), The latest on the best: Essays on evolution and optimality (pp. 277–306). MIT Press.
- Cowen, R. (1995). History of life (2nd edition). Blackwell Scientific Publications.
- Crane, T., & Mellor, D. H. (1990). There is no question of physicalism. Mind, 99, 185–206.
- Davidson, D. (1980). Essays on actions and events. New York, NY: Oxford University Press.
- Dennett, D. C. (1991). Consciousness explained. Boston, MA: Little, Brown and Company.
- Dorr, C. (2019). Natural Properties. In E. N. Zalta (Ed.), The Stanford Encyclopedia of Philosophy. URL =<https://plato.stanford.edu/archives/fall2019/entries/natural-properties/>
- Dretske, F. (1997). What Good Is Consciousness? Canadian Journal of Philosophy, 27, 1-15.
- Earle, W. (1955). Objectivity: An essay on phenomenological ontology. New York: The Noonday Press.
- Earle, W. E. (1972). The autobiographical consciousness. Chicago, IL: Quadrangle Books.
- Edelman, G. M. (1992). Bright air, brilliant fire: On the matter of the mind. New York, NY: Basic Books.

Farthing, W. G. (1992). The psychology of consciousness. Englewood Cliffs, N.J.: Prentice-Hall.

Feinberg, T., & Mallatt, J. (2016). The ancient origins of consciousness. Cambridge, MA: the MIT Press.

Forrester, R. (2014). Sense perception and reality. Wellington, NZ: Best Publications Limited.

Furlong, E. J. (1941). Can we prove that there is an external world? Hermathena, *57*, 107-116.

Gallagher, S., & Zahavi, D. (2008). The phenomenological mind. New York, NY: Routledge.

Geldard, R. (2000). Remembering Heraclitus. Herndon, VA: Lindisfarne Books.

Georgalis, N. (2006). The primacy of the subjective. Cambridge, MA: The MIT Press.

Ginsburg, S., & Jablonka, E. (2007). The translation to experience: Limited learning and limited experiencing. Biological Theory, *2*, 218–230.

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

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

Godfrey-Smith, P. (2020). Metazoa: Animal life and the birth of the mind. New York, NY: Farrar, Straus and Giroux.

- Graziano, M. S. A. (2019). Rethinking consciousness: A scientific theory of subjective experience. New York, NY: W. W. Norton.
- Graziano, M. S. A. (2022). A conceptual framework for consciousness. Proceedings of the National Academy of Sciences, 119, 1-9.
- Hacker, P. M. S. (2002). Is there anything it is like to be a bat? Philosophy, 77, 157–174.
- Hassin, R. R. (2013). Yes it can: On the functional abilities of the human unconscious. Perspectives on Psychological Science, 8, 195–207.
- Hoffman, D. (2019). The case against reality. New York, NY: W. W. Norton & Company.
- Holt, J. (2003). Blindsight & the nature of consciousness. Ontario, Canada: Broadview Press.
- Humphrey, N. (1992). A history of the mind: Evolution and birth of consciousness. New York, NY: HarperCollins Publishers.
- Humphrey, N. (2006). Seeing red: A study in consciousness. Cambridge, MA: Harvard University Press.
- Hurlburt, R. T., & Schwitzgebel, E. (2007). Describing inner experience? Proponent meets skeptic. Cambridge, MA: The MIT Press.
- Husserl, E. (1964). The phenomenology of internal time-consciousness. Bloomington, IN: Indiana University Press.



Jackson, F. (1982). Epiphenomenal qualia. The Philosophical Quarterly, 32, 127–136.

Jackson, F. (2003). Mind and illusion. In A. O’Hear (Ed.), Minds and persons, (pp. 251–272).

Cambridge, UK: Cambridge University Press.

James W. (1904). Does ‘consciousness’ exist? Journal of Philosophy, Psychology, and

Scientific Methods, 1, 477-491.

Jones, T. (2013). The constitution of events. The Monist, 96, 73–86.

Kant, I. (1929/1965). Critique of pure reason. Boston, MA: Bedford/St. Martin’s. (Translated

by N. K. Smith).

Kastrup, B. (2017). On the plausibility of idealism: Refuting criticisms. Disputatio, 9, 13-34.

Kirk, R. (1974). Zombies v. materialists. Proceedings of the Aristotelian Society, 48, 135–152.

Klein, S. B. (2012). The self and its brain. Social Cognition, 30, 474–516.

Klein, S. B. (2014a). The two selves: Their metaphysical commitments and functional

independence. New York, NY: Oxford University Press.

Klein, S. B. (2014b). Auto-noesis and belief in a personal past: An evolutionary theory of

episodic memory indices. Review of Philosophy and Psychology, 5, 427–447.

Klein, S. B. (2015). What memory is. Wiley Interdisciplinary Reviews: Cognitive Science, 6,

1–38.

- Klein, S. B. (2020). Thoughts on the scientific study of phenomenal consciousness. Psychology of Consciousness: Theory, Research, and Practice, 8, 74-80.
- Klein, S. B., Cosmides, L., Tooby, J., & Chance, S. (2002). Decisions and the evolution of memory: Multiple systems, multiple functions. Psychological Review, 109, 306–329.
- Kotchoubey, B. (2018). Human consciousness: Where is it from and what is it for. Frontiers in Psychology, 9, Article 567. <https://doi.org/10.3389/fpsyg.2018.00567>.
- Lehar, S. (2003). Gestalt isomorphism and the primacy of subjective conscious experience: A gestalt bubble model. Behavioral & Brain Sciences 26, 375-444.
- Levine, J. (2001). Purple haze: The puzzle of consciousness. Oxford, UK: Oxford University Press.
- Levine, J. (2003). Knowing what it is like. In B. Gertler (Ed.), Privileged access: Philosophical accounts of self-knowledge (pp. 45–53). Hampshire, UK: Ashgate.
- Locke, J. (1689-1700/1975). An essay concerning human understanding. Oxford, UK: Clarendon Press.
- MacMurray, J. (1957/1969). The self as agent. London, UK: Farber and Farber Limited.
- Marshall, C. R. (2006). Explaining the Cambrian explosion of animals. Annual Review of Earth and Planetary Sciences, 34, 355–384.

Mayr, E. (2001). What evolution is. New York, NY: Basic Books.

McGinn, C. (1991). The problem of consciousness: Essays toward a resolution. Oxford, UK: Blackwell Publishers.

Midgley, M. (2014). Are you an illusion? New York, NY: Routledge.

Milner, D. A., & Goodale, M. A. (1995). The visual brain in action. Oxford, UK: Oxford University Press.

Moore, G. E. (1939). Proof of an external world. Proceedings of the British Academy, 25, 273-300.

Nagel, T. (1974). What is it like to be a bat? Philosophical Review, 83, 435–450.

Nairne, J. S. (2005). The functionalist agenda in memory research. In A. F. Healy (Ed.), Experimental cognitive psychology and its applications: A Festschrift in honor of Lyle Bourne, Walter Kintsch and Thomas Landaur (pp. 115–126). American Psychological Association.

Nath, R. (2016). Wittgenstein on the existence of the mind in the physical World. Austrian Wittgenstein Society, 39, 181-182.

Neuhouser, F. (1990). Fichte's theory of subjectivity. New York, NY: Cambridge University Press.

Nolan, L. (2011). Primary & secondary qualities. Oxford, UK: Oxford University Press.

Orilia, F., & Paoletti, M. P. (2022). Properties. In E. N. Zalta (Ed.). The Stanford

Encyclopedia of Philosophy. URL =

<https://plato.stanford.edu/archives/spr2022/entries/properties/>.

Pereira, A. Jr. (2018). The projective theory of consciousness: From neuroscience to philosophical psychology. Trans/Form/Ação, 41, 199-232.

Plato. (2002). A Plato reader: Eight essential dialogues. Indianapolis, IN: Hackett Publishing Company, Inc.

Plonitsky, A. (2010). Epistemology and probability. New York, NY: Springer.

Rudrauf, D., Bennequin, D., Granic, I., Landini, G., Friston, K., & Williford, K. (2017). A mathematical model of embodied consciousness. Journal of Theoretical Biology, 428, 106–131.

Rogers, G. A. J. (1975). The veil of perception. Mind, 84, 210-224

Russell, B. (1912/1999). The problems of philosophy. Mineola, NY: Dover Publications.

Schwitzgebel, E. (2019). Introspection. In E. N. Zalta (Ed.). The Stanford Encyclopedia of Philosophy. URL = <https://plato.stanford.edu/archives/win2019/entries/introspection/>.

Schlosser, M. (2019). Agency. In E. N. Zalta (Ed.), The Stanford Encyclopedia of Philosophy.

URL = <https://plato.stanford.edu/archives/win2019/entries/consciousness/>.

Schopf, J. W. (2006). Fossil evidence of Archaean life. Philosophical Transactions of the Royal Society B, 361, 869–885.

Searle, J. (1997). The mystery of consciousness. New York, NY: New York Review of Books.

Sherry, D. F., & Schacter, D. L. (1987). The evolution of multiple memory systems. Psychological Review, 94, 439–454.

Shoemaker, S. (1990). Qualities and qualia: What's in the mind. Philosophy and Phenomenological Research (Supplement), 50, 109–131.

Strawson, G. (2005). The self? Oxford, UK: Blackwell Publishing.

Strawson, G. (2009). Mental reality (2nd ed.). Cambridge, MA: MIT Press.

Strother, P. K., Battison, L., Brasier, M. D., & Wellman, C. H. (2011). Earth's earliest non-marine eukaryotes. Nature, 473, 505–509.

Tsuchiya, N. (2017). “What is it like to be a bat?” A pathway to the answer from the integrated information theory. Philosophy Compass, 12407, 1–13.

Vallentine, J. W. (2002). Prelude to the Cambrian explosion. Annual Review of Earth and Planetary Sciences, 32, 285–306.

Vallentine, J. W. (2004). On the origin of phyla. Chicago, IL: University of Chicago Press.

Van Gulick, R. (2022). Consciousness. In E. N. Zalta & U. Nodelman (Eds.), The Stanford

Encyclopedia of Philosophy. URL =

<https://plato.stanford.edu/archives/win2022/entries/consciousness/>.

Varzi, A. (2019). Mereology. E. N. Zalta (Ed.), The Stanford Encyclopedia of Philosophy.

URL= <https://plato.stanford.edu/archives/spr2019/entries/mereology/>

Velmans, M. (2007). Reflexive monism. Journal of Consciousness Studies, 15, 5-50.

Velmans, M. (2009). Understanding consciousness (2<sup>nd</sup> Ed.). New York, NY: London, UK:

Routledge.

von Uexküll, J. (1957). A stroll through the worlds of animals and men: A picture book of

invisible worlds." In C. H. Schiller (Ed, & Translator), Instinctive behavior: The

development of a modern concept (pp. 5-80). New York: International Universities

Press.

Wee, C. (2002). Descartes's two proofs of the external world. Australasian Journal of

Philosophy, 80, 487-501.

Weiskrantz, L. (1997). Consciousness lost and found: A neuropsychological exploration.

Oxford, UK: Oxford University Press.

Whitehead, A. N. (1929). Process and reality. New York, NY: Macmillan.

Williams, G. C. (1966). Adaptation and natural selection: A critique of some current

evolutionary thought. Princeton, NJ: Princeton University Press.

Wittgenstein, L. (1958). The blue and brown books (R. Rees, Ed.). New York, NY: Harper and Row.

Wright, W. (2007). Explanation and the hard problem. Philosophical Studies,132, 301–330.

Zahorik, P., & Jenison, R. L. (1998). Presence as being-in-the-world. Presence, 7, 78–89.

Zhang, X., & Shu, D. (2021). Current understanding on the Cambrian Explosion: Questions and answers. Paläontologische Zeitschrift, 95, 641–660.

### Footnotes

1. In what follows I will use the terms sentience and consciousness interchangeably. While this convention is not shared by everyone (for review, see Van Gulick, 2022), my identification of consciousness with phenomenal consciousness (my reasons for so doing are presented below) explains this treatment.
2. All conscious states have content -- i.e., they are about something. Brentano (1995) called these “somethings” the intentional objects of consciousness.
3. On this view, consciousness is phenomenal consciousness (or sentience). Thus my use of the terms consciousness and sentience synonymously in the text.
4. Among philosophers, the term “event” refers to a change in an object’s properties or in its relation to other objects. Although this construal has not achieved universal consensus (for discussions see Bennett, 1988; Casati, & Achille, 2020; Davidson, 1980; Jones, 2013), for ease of exposition, in what follows I refer to “objects and events” simply as “objects”.
5. In Kantian philosophy, “things-in-themselves” refers to the status of objects as they are, independent of representation and observation (e.g., Kant, 1929/1965).
6. This is not to say that evolutionary design is unconstrained by physical reality. To effectively and reliably solve problems, a neural mapping must maintain some fidelity to that which is being mapped. Environmental regularities and the demands of reality place limits on which map-to-behavior relations will work, how well they will work, and which will fail.



7. I am aware that the word “being” in this final clause can be taken in at least two senses (i.e., existentially and organically). I leave it to the reader to determine which sense to apply. In my opinion, the two are different perspectives on a common theme.
8. Perceptual projection is a hotly debated topic in philosophical discourse (e.g., Lehar, 2003; Pereira, 2018; Pribram, 2004; Velmans, 2009). Although at present there are no adequate theoretical accounts, perceptual projection is a phenomenological fact in need of explanation, not certification.
9. Properties differ from objects in that they may belong to more than one object (e.g., baseballs and avocado pits both share the properties “round” and “hard”).
10. My proposal, if viable, has ramifications for a number of phenomena associated with sentient organisms (e.g., understanding their transition from a solipsistic existence to agentic participation in a richly configured, deeply social world). Herein, I restrict attention its implications for the phenomenon of blindsight.