



The Origins of Consciousness or the War of the Five Dimensions

Walter Veit^{1,2}

Received: 6 January 2022 / Accepted: 1 July 2022 / Published online: 25 August 2022
© The Author(s) 2022

Abstract

The goal of this article is to break down the dimensions of consciousness, attempt to reverse engineer their evolutionary function, and make sense of the origins of consciousness by breaking off those dimensions that are more likely to have arisen later. A Darwinian approach will allow us to revise the philosopher's concept of consciousness away from a single "thing," an all-or-nothing quality, and towards a concept of phenomenological complexity that arose out of simple valenced states. Finally, I will offer support for an evaluation-first view of consciousness by drawing on recent work in experimental philosophy of mind.

Keywords Animal consciousness · Animal minds · Organisms · Origins of consciousness · Phenomenological complexity

[I]f we contemplate the subject, we shall find it difficult or impossible to imagine a form of consciousness, however dim, which does not present, in a correspondingly undeveloped condition, the capacity of preferring some of its states to others—that is, of feeling a distinction between quiescence and vague discomfort, which, with a larger accession of the mind-element, grows into the vivid contrast between a Pleasure and a Pain. I think, therefore, it is needless to say more in justification of the level on the diagram at which I have written these words.

George John Romanes (1883, p. 111)

an all-or-nothing quality that is either present or absent, and towards a recognition of multiple dimensions of what I call "phenomenological complexity" that can come in a range of shapes, varieties, and gradations (Veit 2022c, e).¹

Against the skeptics of the very viability of such a comparative study of consciousness, Birch et al. (2020) have recently called for the development of multidimensional animal consciousness profiles. Indeed, they make a convincing case that gradations and varieties can be found across at least five dimensions of self-consciousness, synchronic experience, diachronic experience, sensory experience, and evaluative experience—and these can in principle be compared and measured in different species.² Yet, their proposed tests for each dimension are derived from the study of *human* consciousness and hence should only be considered scaffolds that must eventually be overcome for a Darwinian study of consciousness. And to begin this scaffolding process by trying to understand the function of consciousness in nature we will inevitably have to address its nebulous evolutionary origins. This is the goal of the article.

In trying to reverse engineer the origins of consciousness, we are asking for its *raison d'être*; the reason for its existence.

Introduction

Any serious biological theory of consciousness must account for the full diversity and complexity of subjective experience in nature, something that won't be satisfied by a single scale or a mere recognition of levels. If consciousness is an evolved trait, we should expect as many varieties and gradations as for any other biological trait. A Darwinian approach to consciousness thus entails a partial conceptual revision of our folk notion of consciousness away from a single "thing,"

¹ It is an instance of what I have called naturalist conceptual engineering (Veit and Browning 2020).

² That is, even if the measurement of consciousness suffers from conceptual and methodological difficulties (Browning and Veit 2020; Browning 2022). While we may discover more dimensions in the future, a dimensionalist approach constitutes significant progress over monist thinking for the purposes of uncovering the origins of consciousness.

✉ Walter Veit
wvei7501@uni.sydney.edu.au

¹ School of History and Philosophy of Science, The University of Sydney, Sydney, NSW, Australia

² Department of Psychology, University of Cambridge, Cambridge, UK

As the epigraph of this article illustrates, Darwin's disciple Romanes (1883) maintained that it would be almost impossible to conceive of the dawn of subjectivity without at least a minimal sense of hedonic valence, going so far as to claim that no more justification than this would be needed. While I share Romanes's view on the origins of consciousness, a modern 21st century theory of the evolution of consciousness ought to do better than just appeal to one man's intuitions. Other possibilities need to be considered, critically evaluated, and at least argued against. And the suggested dimensions by Birch et al. (2020) give us just this possibility: five possible contenders for the crown of the most ancient kind of subjective experience. Which is the oldest in the line of succession and deserves to sit at the center of our theory of phenomenological complexity? By eliminating one dimension after another from our rich human conception of consciousness this article aims to establish the evaluative dimension as the core and origin of consciousness.

Article Outline

This article is structured as follows. In the second section, "Five Options for the Origins of Consciousness," I will narrow down our list of five contenders by arguing that both diachronic and synchronic unity are structural latecomer features of conscious experience unlikely to have been present at the earliest origins of subjective experience. In the third section, "Down to Three," I will argue against a strongly internalist view of consciousness and the centrality of self-experience as a necessary ingredient of the most basic kinds of experience. This leaves us with two dimensions: sensory and evaluative experience. In the fourth section, "Two Contenders for the Most Minimal Kinds of Subjective Experience," I will criticize the common assumption that a theory of consciousness must be built on a model of sensory experience. Worse, the hard problem may have been at least partially an artifact of a strongly externalist focus on the sensory side. To turn evaluation into *the* model for experience is a promising teleonomic alternative to a false dilemma between internalist and externalist theories of consciousness and will substantially narrow the explanatory gap between matter and mind. Finally, the fifth section, "The Spoils of War," will summarize my case for the search of the origins of consciousness in evaluation and respond to some further objections.

Five Options for the Origins of Consciousness

My treatment of consciousness as a complex multidimensional phenomenon is intended to narrow the explanatory gap by no longer demanding an explanation that requires

all of the features of consciousness to appear together as something like a "one-package deal." A new problem, however, arises in the sense that we are apparently faced with at least five different functions and origin stories for the evolution of experience, corresponding to each of the dimensions. Could self-consciousness, synchronic experience, diachronic experience, sensory experience, and evaluative experience all have their own independent origins? Or do some appear first, with others "built" on top? Depending on which dimension(s) we consider to be the most basic, we may be presented with widely different views on the place of mind in nature, perhaps even more diverse than those from theorizing about the function of the rich human consciousness we are all familiar with.

This is a bullet we ought to bite. In thinking about the function of consciousness, it would be a mistake to think that consciousness only does *one* thing. It has been an unfortunate development that much of the philosophical debate on functions has treated them in a very binary and monist fashion (see Matthewson 2020). Traits are said to either have one function *or* another, similar to how consciousness is seen as either present or not, without allowing for gradations and variations.³ But traits can be shaped by numerous selection pressures and thus have a variety of functions that can be realized to various degrees.

It is therefore unsurprising that our complex human experience may not appear to be capturable with a single function statement. Indeed, it is not uncommon among naturalists to deny that consciousness has a function over and above the cognitive mechanisms constituting it. But such thinking already implicitly presumes a model of human consciousness and complexity. In thinking about the *role* of consciousness in nature, we must address its most humble origins. And as I shall now argue, we can at least make this problem quite a bit smaller by shelving away two of the five dimensions that I group together under the *unity* of experience, i.e., diachronic and synchronic unity, or what Birch et al. (2020) call "temporality" and "unity."

Diachronic Unity of Experience

Human consciousness appears to be highly unified across time. It is no accident that the likes of William James (1890) described consciousness as a continuous stream or flow of experience, a metaphor that has become influential in work on the experience of time (Dainton 2018). Yet, out of all the five dimensions this one appears the most readily

³ One excellent exception to this trend is Matthewson (2020) who argues that a trait may more or less have any single function, urging us to accept the graded nature of natural selection into our concepts of biological functions.

conceivable as a latecomer. Unless one buys into a picture in which consciousness can *only* play a functional role if it is connected to episodic memory, or for that matter a mental simulation of the future, this dimension appears to be strictly optional. Whereas episodic memory remains contested even among mammals and birds, consciousness itself is typically granted much more liberally. This is not to deny that diachronic unity itself has important functions, but that it is not needed for consciousness to be functional, and that the burden of proof lies with those who assert that it must be present for the existence of subjective experience.

For very short timescales or what is perhaps best described as the experience of time, however, things may admittedly be less clear. The color-phi illusion experiments often used in consciousness research to investigate the threshold at which a series of pictures is experienced as motion hint at subjective experience being inherently flow-like in humans. But what may appear like continuous movement to a human may look like a fragmented set of pictures to a different animal. Indeed, some animals might even have a higher consciously experienced “frame rate” that would allow them to discriminate more individual frames per second than humans can, before it “blends” into motion for them. This could provide an adaptive benefit to animals with very quick reaction times, like a hummingbird or a fly. Broom (2014) makes just this argument, noting that “hummingbirds and mice seem to live at a much faster pace than larger, slower-moving animals such as humans” and that we shouldn’t underestimate smaller animals (p. 118). The experience of time and the speed of an organism’s behavior should, of course, be linked in a teleonomic view of mind and life.

Furthermore, there may very well be a trade-off between a higher frame rate allowing for more information being gathered per second, whereas a “flowing” mode of experience enables a better understanding of causal processes. It is therefore not clear whether a flow-like experience must be functionally “superior” to a more fragmented kind of experience. The most reasonable take on diachronic unity would be to understand it as a feature of how subjective experience can be structured, something that only comes to evolve *after* consciousness has already been set in place. While the role of metaphors in science remains contested (Veit and Ney 2021), we can use the common light metaphor to our advantage here: the “lights” at the dawn of animal consciousness might “blink” on and off, only later becoming integrated into something like a stream. Admittedly, the experience of time may be crucial to what it is like to be a typical human with long-term goals and a personal identity that is constituted by one’s experiences, but we are interested in the evolution of subjectivity, not just its supposedly most complex form.

By eliminating the need for an account of the origins of minimal consciousness to contain diachronic unity, the

explanatory gap can at least be partially reduced. It is not in the integration of experience across time that the puzzle of “qualia” will be solved nor does it appear to give rise to an additional explanatory gap. Let us thus quickly turn to the next dimension of unity.

Synchronic Unity of Experience

Whereas diachronic unity can be readily dismissed as an *unnecessary* feature that may figure in the arrangement of an organism’s subjective experience, or not, synchronic unity has long been seen as one of the most fundamental features of consciousness. To even ask the question of whether there is a subjective viewpoint appears to ask for something like an integrated picture of the world. Again, however, we ought to be careful not to confuse a certain feature of human experience with a necessary property of subjective experience in nature. The terms that have established themselves for discussions in human consciousness may not be helpful for thinking about its evolutionary origins. The “subjective” in “subjective experience” offers an unhelpful qualifier for minimal kinds of experiences that plausibly do not come in the shape of a subjective unified point of view.

Split-brain patients are here often used to study the possibility of de-integrated experience—perhaps even constituting two conscious selves. In asking for the function of unified experience, such studies of pathological cases of consciousness will doubtless be helpful, but if we want to learn about the possible advantages of a disunified experience, pathological cases *in humans* may be of little help, since our brain evolved to have a strikingly unified experience. Robbing it of this ability will doubtless lead to many dysfunctions, even if we can learn about surprising abilities that can still be performed with a split-brain. Disunified forms of experience may be easier to evolve and work “fine” if not better for many animals. It is therefore important to take a comparative approach and find natural experiments in the animal kingdom to study healthy cases of animals placed “lower” on this dimension. Note, however, that we need to be careful not to bring hierarchical thinking back into our thinking about consciousness. To have a highly unified experience will only make your experience more “point of view”-ish, not necessarily more conscious. That vast ranges of the animal tree of life do not have limited interchange between their brain halves should give us pause in thinking that unity must be a richer mode of being, rather than merely a different one.

Citing work by Güntürkün and Bugnyar (2016), Birch et al. suggest that we could rely on birds as such natural split-brain patients since they “have no structure akin to the corpus callosum connecting the two hemispheres of the dorsal pallium, which is homologous to the cortex in mammals” (2020, p. 793), a structure that was long considered

necessary for consciousness. Yet they overstate the similarity to split-brain patients since birds have unsplit lower parts of their brain, which also make up a larger fraction of the whole than the cortex does in us (Vallortigara 2000). One would therefore be too quick if one simply treated them as a case of two subjects within a single body. Non-avian reptiles or for that matter fish constitute much better neuroanatomical examples for healthy split-brain because compared to birds they have much reduced ipsilateral projections in the tectofugal and thalamofugal pathways and make it easier to infer that only one brain “half” is involved in a task (Deckel 1995, 1997; Vallortigara 2000; Sovrano et al. 2001).

In thinking about the evolutionary origins of consciousness, we must admit that a biological materialism makes the possibility of consciousness gradually becoming more unified a live possibility. Firstly, it is now widely accepted that non-avian reptiles and fish have subjective experiences, perceptual worlds, and evaluations despite coming the closest to natural split-brains. Secondly, differences in the organization of the nervous system and information processing should reflect differences in subjective experiences in a materialist picture that resists a fall back into dualism. In reverse engineering consciousness, we should see the unity of consciousness as something that gradually evolved, not just an automatic feature.

As Godfrey-Smith (2016b) once put it: “To some degree, unity is inevitable in a living agent: an animal is a whole, a physical object keeping itself alive. But in other ways, unity is optional, an achievement, an invention” (p. 87). Even the sensory experience of two eyes, seemingly central to our human subjective points of view, do not appear to be necessarily integrated in other animals. New features may be added and only come to be integrated later, and this is certainly an option that is more sensible from a gradualist perspective. Unity in a very simple mind may almost be akin to a trivial definitional affair, rather than an interesting biological property of how experience comes to be organized. And these two senses of unity between a genuine biological property and a mere inevitable feature of biological organization need to be distinguished.

If an animal’s subjective experiences are limited to a distinction between two senses, for example, hot and cold, or evaluations of good and bad, disunity is not something that can be achieved simply because there are so few distinct subjective experiences. But that should obviously not lead us to conclude that their synchronic unity is equally rich as ours. The synchronic dimension of consciousness—similar to the diachronic one—is of little relevance until there are animals with a certain degree of phenomenological complexity that allows subjective experience to be organized in interesting new ways. Until then, one can hardly speak of integrated organization at all, since disunity is not the absence of organization—it is merely a different way of being. The most

striking cases of integration and lateralization, after all, are found in vertebrates with a high degree of phenomenological complexity. And this is why we should see these various forms of organizations as later add-ons rather than original or “automatic” features of subjective experience.

Unity and Consciousness

Both diachronic unity and synchronic unity appear to be a feature of the way experiences are organized, not something we need to explain if we are merely concerned about the first sparks of experiences in their own right. They do not appear to solve the hard problem of why there are some states that *feel* like something to an organism. While those betting on unity as the fundamental dimension to solve the problem of consciousness may well turn out to be right, it appears more plausible from an evolutionary point of view to treat forms of unity as later “add-ons.” Tononi et al. (2022) argue that unity must be seen as a fundamental property of consciousness in any “satisfactory explanation of consciousness” (p. 44), but a satisfactory account of unity can explain it as a special arrangement of how consciousness is organized in humans, without thereby making it a fundamental property of consciousness.⁴ We need to be careful not to fall prey to the all-too-common confusion between human consciousness and consciousness as a phenomenon in nature. And since an “elimination” of unity from the necessary properties of consciousness substantially narrows the explanatory gap, this alone would warrant the search for a theory of consciousness in more basic experiences.

Down to Three

This leaves us with three remaining dimensions of consciousness: (1) the experience of a self, (2) sensory experience, and (3) evaluative experience. To think that these all must come together as a one-package deal for consciousness would still leave the explanatory gap incredibly wide. In thinking about the origins of minimal consciousness it is hence plausible to think that it is in one of these dimensions that consciousness arose.

Despite the fact that each dimension can provide us with a distinctive model of consciousness, however, the literature has largely focused on the first two options. As I shall argue, this has put the science of consciousness in an unfortunate dilemma between internalist and externalist approaches to consciousness, i.e., that consciousness is to be explained either through recourse of properties internal to the organism

⁴ I criticize Tononi’s integrated information theory of consciousness elsewhere in more detail (Veit 2022b).

or external to it, rather than through the dynamics and feedback between them. While these theories may well succeed at explaining parts of the phenomenon, I will make a case that there remains an explanatory leftover, i.e., precisely what the other dimension was meant to explain, and thus giving the impression of an explanatory gap.

Experience of a Self

Whereas consciousness was once seen as a higher-order form of thought or self-awareness, something that should distinguish humans from other animals, and could perhaps be found in some rudimentary fashion in our close relatives, such as chimpanzees and bonobos, or other intelligent animals such as dolphins and elephants able to recognize themselves in mirrors (Keenan et al. 2003), this kind of experience is now more typically regarded as one kind of subjective experience, i.e., self-consciousness. Indeed, the historical association of the term “consciousness” with some form of rich and exclusively human experience is precisely why many of the researchers engaged in the study of animal consciousness prefer the term “sentience” to refer to more basic kinds of experiences (see also Browning and Birch 2022).⁵

Still, it would be easy to dismiss the dimension of selfhood by appealing to the radical demands of the richest kinds within its dimension and to contrast these with a simple evaluative feeling. Just as with the dimensions of unity, we should consider the most rudimentary forms within this dimension, i.e., a minimal sense of a bodily self rather than comparatively rich self-awareness. Because capacities related to self-recognition appear to be central to animal life it is not implausible to think that the emergence of bodily self-awareness could coincide with the very emergence of consciousness as an evolutionarily ancient trait.

Frans de Waal (2006), for instance, thinks that all animals require at least a minimal form of a self-concept. But treated literally, this claim—despite being intuitive—is surely too strong when we think of the less mobile and more “plantlike” side of the animal branch of life such as corals, anemones, and sponges. This genealogical definition of animal can conflict with an older and probably still common folk understanding of animals as mobile organisms with a nervous system: a particular animal *mode of being* (Ginsburg and Jablonka 2019; Browning and Veit 2021). For these animal lifestyles, a registration of a difference between self and other appears almost necessary to enable movement and action in a complex body (see also Godfrey-Smith 2020a). After all, such animals must have some way

of distinguishing themselves from the world. As Birch et al. put it, “[a]ny complex, actively mobile animal needs a way of disentangling changes to its sensory input that are due to its own movements from changes due to events in the world” (2020, p. 797).⁶ This variety of self-consciousness may be recognized as a very minimal feature of sentience. But while there is something important to this intuition, I argue that it would be a mistake to seek this dimension in the very origins of experience.

Let us grant that a discriminatory capacity must eventually emerge in information-processing between interoceptive (sensing of internal states) and exteroceptive (sensing of external states) cognition. Once this distinction has been accomplished a basic or minimal feature of self-consciousness may be said to emerge on which more sophisticated versions can be built. But here we should think carefully about whether it makes sense to think of the distinction between interoceptive and exteroceptive experience as a necessary feature of the most minimal kinds of subjective experience. The problem with “betting” on this admittedly tempting dimension comes from a failure to meaningfully distinguish the presence of mere cognitive processes of self-recognition without any subjective “feel” from the subjective experience of a self. While the former plausibly exists even in the most basic forms of life such as bacteria, which also keep track of internal and external states, the latter ought to be a much more restrictive capacity.

Nevertheless, the idea that life, cognition, and consciousness are deeply connected is often associated with the autopoietic tradition that sees all life as being a process of cognition. A continuity thesis between mind and life can hardly be stronger than this. From here, it is only a small step towards a full embrace of biopsychism and granting consciousness to all life. This would be an odd evolutionary journey to say the least, but it is being taken seriously by the likes of Evan Thompson (2007). For a biopsychist, we take one step in the emergence of life and are suddenly presented with consciousness. But while this can be considered as a major downside of the view, we should not just reject it from the armchair. If there is something like minimal cognition present even in the earliest stages in the tree of life, might there not also be something like minimal consciousness in a form that is very different from us? Since I am committed to a strongly gradualist picture (see Veit and Huebner 2020) this view deserves at least some attention, even if we are ultimately going to reject it.

Life, the autopoietic tradition emphasizes, must be bounded and self-maintaining in order to allow its continuing existence, and these features should accordingly be the

⁵ This shift has admittedly also been motivated by ethical considerations (Browning and Veit 2022).

⁶ See also Hurley (1998); Merker (2005); Godfrey-Smith (2016b, 2020a); Trestman (2017).

center of our attention. Such a view of life emphasizing autonomy and subjecthood certainly lends itself to thinking of life as a bridge between matter and mind. Thompson (2007) maintains that consciousness can be described as “a kind of primitively self-aware liveliness or animation of the body” (p. 161) which arises from the “autopoietic identity and sense-making of living beings, but in addition it implies a feeling of self and world” (p. 221). Living systems are engaged in the “self-production of an inside that also specifies an outside to which it is normatively related” (Thompson 2007, p. 163). This emphasis on life being inherently dynamic, goal-directed, and normative is to be appreciated as a useful path to reshaping the materialist side of the mind-matter problem. But the autopoietic tradition often goes too far in various respects, overreaching in almost the opposite direction from that of the “mainstream” in philosophy of mind, a fact that is owed to its origins and conception as a “radical” challenge and alternative to the Darwinian view of life (Escobar 2012).

In opposition to the perceived excessive externalism of Darwinian views, which allegedly treat organisms as akin to passive clay molded by external forces, the autopoietic tradition has in turn overemphasized an older very internalist mode of explanation, something Godfrey-Smith (2016a) described as an unfortunate “philosophical baggage” (p. 778). In trying to motivate a view of life as active and autonomous agency, external goings-on come to be endogenized as just another kind of organismal activity, leading to a neglect of the “to-and-fro traffic characteristic of organism/environment relations” (Godfrey-Smith 2016a, p. 778). Admittedly, this neglect of feedback is also shared by mainstream representationalist views in philosophy of mind that overemphasize environmental states—whether they be outside or inside the body.⁷ In contrast, the autopoietic tradition seemingly removes the role of the environment entirely. Instead of making the externalist move of seeing internal events and processes as just another kind of environment, they neglect the role of the organism’s environment in an effort not to undermine the “autonomy” of the organism. The environment is almost seen as just another kind of internal goings-on, the organism forcing itself onto nature, without reciprocal influence.

Indeed, Thompson (2007) tellingly describes the nature of any autopoietic system as being “defined by its endogenous, self-organizing and self-controlling dynamics” which do “not have inputs and outputs in the usual sense” (p. 43). While recent work on the mechanisms of such self-organizing systems do not deny the interaction between organism

and environment—highlighting dynamic feedback, and the need of such systems to maintain organizational and operational and organizational closure in order to stave off thermodynamic processes towards entropy—the emphasis on autonomy has led, Godfrey-Smith (2016a) notes, to a continuing resistance “to the role of ecology, in a broad sense—resistant to the fact that [it] is part of the nature of life to be in ongoing interaction with an environment that is *other*” (p. 778; italics in original). It is thus perhaps not surprising that Thompson (2007) explicitly contrasts the autopoietic strategy as an explicit alternative to the Darwinian approach. In a reply to his commentators he explicitly contrasts the Darwinian reverse-engineering perspective with an autonomy perspective (Thompson 2011, p. 177).

But the “autonomy approach,” it turns out, is simply empirically inadequate, missing the reciprocal influence between organism and the world due to its emphasis on an internalist alternative to a mistaken view of the Darwinian project. Thompson’s argument is reminiscent of older vitalist debates and critics of evolutionary/functionalist approaches to consciousness for supposedly being unable to capture the most fundamental property of life and mind. It is merely asserted that autonomy is fundamental and sufficient to explain both life and mind. But the very motivation of conceiving Darwinism or for that matter adaptationism as a necessarily externalist approach is simply a failure to distinguish the externalist pre-Darwinian design thinking from the teleonomic thinking about design and normativity that conceptually re-engineered them in terms of natural selection and feedback between organism and environment.

As Godfrey-Smith (2020a) argues in his recent book *Metazoa*, I see this tradition as making consciousness too much of “an automatic feature of just being a living organism located in the world” (p. 117).⁸ It comes in a sense for free, which may be intuitively more attractive than a gradualist model, but shares some uncomfortable parallels with epiphenomenalist ideas. It does not seem to give us any purchase on its *raison d’être*, with autonomy being merely asserted. We are not given a gradualist story of how consciousness or for that matter agency emerges; rather it is something that simply comes *along* with or rather constitutes a certain kind of living activity. This makes it hard to think about things gradually becoming more agential, experiential, or—to use the language of the autopoietic tradition —“autonomous,” precisely because of their resistance to adaptationist thinking.⁹ Indeed, it is hard to make sense of the idea that a vague wash of feeling of “presence in the world” would become refined and enriched through

⁷ Elsewhere I have extended this criticism to application of the free energy principle to understanding life and mind (Veit and Browning forthcoming).

⁸ See Veit (2022f) for a review of Godfrey-Smith’s book.

⁹ See Okasha (2018); Veit (2021a, b) for treatments of the role of agency in evolutionary thinking.

the process of natural selection if these experiences represent neither useful sensory nor evaluative information. The mistaken idea to avoid the hard problem by trying to rely on a non-functionalist explanation makes it ultimately impossible to explain the gradations and variations of subjective experience across the tree of life, and this is why a search for the origins of consciousness in this dimension must ultimately fail. Even if the approach were to succeed in making sense of the qualitative experience of a self, which I doubt due to its failure to account for the role of consciousness in nature, it is unclear how the other dimensions can be built on a model of self-experience without—again—making them mere automatic features of living activity.

To take a functionalist approach to the origins of subjective experience would mean to locate the origins of selfhood in the capacity to distinguish between exteroceptive and interoceptive experience. It is this capacity that requires “disentangling” on the sensory side of consciousness, rather than giving rise to sensory experience itself. It should not be expected from the earliest organisms possessing some wash of sensory sensation and hence should be seen as a later layer. While rich forms of self-awareness are a core feature of human consciousness, we should not mistake them for its original core. What we thus need in order to uncover the ancient origins of this subjectivity is a more gradualist picture in which complex cognitive processes are recognized that, as Godfrey-Smith (2020a) argues, largely go on behind the scenes of *our* experience. In order to narrow the explanatory gap, we should treat the dimension of selfhood as something gradually built out of more basic experiences, eventually giving rise to a meaningful recognition of a subject and thus *subjective* experience, but it is in the actual building blocks of experience that we must search the origins of consciousness.

Two Contenders for the Most Minimal Kinds of Subjective Experience

This finally leaves us with two remaining options. While diachronic experience, synchronic experience, and the experience of a self can be understood as features *of* the way conscious experience can be structured as opposed to constituting it, it remains an open question whether the origins of consciousness can be found on the sensory or evaluative side. On the one hand, we have multiple sensory modalities such as seeing, touch, taste, hearing; and on the other, there are felt evaluative states such as moods and emotions, including anger, pain, grief, anxiety, fear, and more generally positive and negative feelings or “affects.”

Regardless of how difficult it may be to explain these three dimensions and how many ingenious inventions came along with them on this path, they do not seem to force us

into three further hard problems. The explanatory gap has been narrowed further and we are getting closer to the question of why the most minimal kinds of subjective experience evolved, i.e., what they were selected for. The existence of the last two remaining dimensions, however, raises the possibility that there are two explanatory gaps—one for sensing and one for evaluation.

Philosophers usually see subjective experience as something that includes a vast range of mental states, such as the sensory experience of a color like red, or an evaluative experience such as a pain or emotion. The alleged property these all have in common is a phenomenal one. Following Nagel, it is sometimes treated as a second-order property of what it’s like to be in that state (see Sytsma and Machery 2010). Godfrey-Smith (2020b) also hints at Nagel’s notion of *something it’s like* as being that property—as a possible approach to asking about the consciousness of nonhuman animals such as bees and octopuses—but suspects that it is not enough to settle the problem of disassociating the sensory from the evaluative side. He argues that Nagel’s “feature is not self-evident, such that it might, in principle, be recognized as simply present or absent across very different ways of being an animal. Something more informative is needed” (Godfrey-Smith 2020b, p. 1152). Once we take a closer look at each dimension and its relationship to consciousness, however, it appears that the idea of two explanatory gaps is mistaken.

It rests on the idea that there is a common core to both, i.e., the property of having phenomenology, which is somehow pre-theoretically obvious and clear. If this were true, we would surely expect to find something like a concept of phenomenological properties that is applied even by laypeople. But a growing evidence base coming from the work of experimental philosophers such as Justin Sytsma and Edouard Machery challenges this view (Sytsma and Machery 2009, 2010, 2012; Sytsma 2010, 2012, 2014; Machery and Sytsma 2011; Sytsma and Ozdemir 2019). Evaluation, rather than sensing, appears to be at the core of folk intuitions regarding subjective experience, and may thus constitute a better model to understand the qualitative aspects of subjective experience. But before I turn to this experimental work and evaluation as the foundation for consciousness, I will continue with the last onion layer to be shaved away: sensory experience.

Sensory Experience

A sensory explication of Nagel’s idea is widespread both among philosophers of mind and neuroscientists investigating consciousness. The sensory side, particularly studies of human vision, has often served as the model for *all* of consciousness due to its ties to both of the equally influential notions of a *point of view* and an *awareness of*. It has

been seen as the key to understanding phenomenological experience, which was only aided by the fact that human consciousness science was long deemed impossible and continues to be seen with suspicion. To a large extent, the focus on vision was an attempt to make research on consciousness as scientific as possible. Whereas experiments on vision appeared tractable, experiments on seemingly more evaluative capacities such as olfaction and interoceptive capacities took place largely elsewhere. Indeed, affective neuroscience that focuses on the moods, personality, emotions, motivations, and feelings of both humans and animals has long continued to be viewed with suspicion; not only among cognitive neuroscientists who saw this research as too closely related to “consciousness science” (see Cacioppo and Gardner 1999), but also among philosophers who saw this empirical research as largely irrelevant to their conceptual analysis of emotions (see Griffiths 1997, 2017).

At a time when scientific work on consciousness was deemed impossible and seen with high suspicion by fellow scientists, philosophers, and funding agencies alike, it may not only have been easier to begin with vision, but necessary in order to overcome the behaviorist dictum that consciousness is not investigatable. And it is of course not unreasonable to begin the scientific investigation of an incredibly complex phenomenon with an aspect of it that is the most readily operationalizable, and that can be modeled with other complexities temporarily abstracted and idealized away (Weisberg 2013; Veit 2019a, b). The problem with this approach, however, is the possibility that our early models and theories developed based on human vision may influence the way we treat all other aspects of subjective experience. Perhaps our current troubles with naturalizing consciousness are due to path dependence and we would have been in a much better place if we had begun with the study of affect and valence. But to advance such an alternative model, I will first argue against the centrality of a sensory model of all experience that treats consciousness as a mere form of representation.

One elegant argument against the centrality of a sensory model has come from Godfrey-Smith (2020a), who despite his insistence on a possible separation between the two dimensions of sensory discrimination and evaluation, has argued that a “problem with much recent work in philosophy is the idea that sensing is not only an important part of experience, but just about all that goes on there” (2020a, p. 113). Indeed, much of the recent work in philosophy of mind uses the words “sensing” or “perception” as umbrella terms to cover all forms of subjective experience. Moods, emotions, and feelings are seen through this lens as the detection of some internal phenomena such as thirst or hunger. Their only difference from olfaction and vision is a matter of perceptual direction: inward versus outward. Godfrey-Smith (2020a) uses two influential examples from different generations: Fred Dretske, who influenced Godfrey-Smith’s own

representationalist views when he was a student; and Jesse Prinz, who was his colleague at CUNY. While he doesn’t assert that all of subjective experience must be perceptual, Dretske (1993) cites the theoretical neurobiologist Bernard Baars (1988), the originator of the influential *global workspace theory* of consciousness, the English psychologist Max Velmans (1991), and neuropsychologist Nicholas Humphrey (1992) to support the view that perceptual experience and belief are taken to be the “clearest and most compelling” paradigm cases of consciousness in empirical research (Dretske 1993, p. 272). In a revealing passage, Dretske argues:

Why can’t we, following Damasio (1994), conceive of emotions, feelings, and moods as perception of chemical, hormonal, visceral, and musculoskeletal states of the body?

This way of thinking about pains, itches, tickles, and other bodily sensations puts them in exactly the same category as the experiences we have when we are made perceptually aware of our environment. The only difference is that bodily sensations are the experiences we have of *objects* in the body (the stomach, the head, the joints, etc.), not *objects* outside the body. (1999, p. 117; italics added)

But it may well be a mistake to think of these experiences as representing *objects* in the body, which makes these accounts very much feel like attempts to somehow objectively experience states of the world as described by Newtonian mechanics. The very reason these accounts haven’t satisfied proponents of the hard problem has been the lack of recognition of a *subject*. In order to address these problems, *subjectivity* needs to have a role, not merely be ignored as a mere by-product of the existence of sensory representation. But this is next to impossible due to their reliance on a strongly externalist mode of explanation.

Rather than recognizing the epistemological straight-jacket of a strongly externalist approach, representationalists tend to simply bite this bullet and declare the sense of a self or subject *unimportant*. At the end of his book on his Attended Intermediate-level Representation (AIR) theory of consciousness, Prinz (2012) confidently asserts that “[a]ll consciousness is perceptual; there is no distinctive cognitive phenomenology or any phenomenal self” despite his own concession that “[almost] all of the empirical research reviewed here comes from vision science” (p. 341). It is simply an extension of his earlier representationalist neuro-functional theory of visual consciousness (see Prinz 2000), with little to no attention paid to conflicting empirical work on other dimensions of consciousness. Such unashamed confessions of confidence in the centrality of vision as the paradigm of all of consciousness should at least raise some

worries that this path may have been somewhat of a wrong turn.

Godfrey-Smith (2020a) sees this as a general tendency of much current information-theoretic work on consciousness—such as Michael Tye’s (1995) PANIC theory of consciousness and Stanislas Dehaene’s (2014) version of the Global Workspace Theory due to Baars—to treat consciousness as a special qualitative way of information being represented in a mind (Godfrey-Smith 2020a, p. 115). These representationalist theories share too much with an older empiricist view of the mind “as merely reactive, needing to derive its patterning from elsewhere” (Godfrey-Smith 2020a, p. 188). And this is precisely what the likes of Evan Thompson rightly sought to resist.

Here, it’s important to heed Dennett’s (1991) warning not to fall back into dualist thinking and treat the mind as a mere Cartesian theater of the world, even if the new picture now includes what goes on inside the body.¹⁰ It is this aspect of the mind that many so-called illusionists of consciousness have their problems with. Godfrey-Smith (2020a) urges us to introspect whether sensory experience and belief are really the most compelling cases of consciousness. He suggests that what I have discussed under the label of affects, i.e., “emotions, willings, moods, and urges,” seem to be at least as, and in his own case “quite a bit *more* clear, as cases of conscious experience, than beliefs” (2020a, p. 114; italics in original). As a committed naturalist, I am reluctant to bring my own experience into this debate, despite sharing this intuition. We ought to resist the temptation to turn this debate into a mere pull of intuitions between different philosophers. My account of consciousness puts gradations and variations center stage, so it should not be at all surprising to expect some philosophers (who appear to be one of the most neurodiverse groups out there) to vary widely in their own subjective experiences. The different accounts may be representative of the genuinely different subjective experiences of their defenders. Some philosophers may have a much more emotionally neutral way of life (and many of us have certainly encountered people like that). But even if we grant that vision is the most paradigmatic case of human conscious experience, this may simply be an artifact of our evolutionary path, a path that has made us masters at learning about and improving our environments, thus making vision appear to represent the external world “objectively” like a mirror image presented in something like a Cartesian theater. This is a compelling intuition that would likely be quite different if we had instead evolved to have something like night vision or lacked vision and had to rely on olfaction.

Godfrey-Smith (2020a) suggests that an “alternative to [the strongly representationalist] view, rather obvious but

neglected, is the idea discussed in our section on the experience of selves that a mood is not a *presentation* of some fact or condition; it is just *the way things are* with you, at that time” (p. 114; italics in original). That was one feature of the selfhood-first view that made it attractive. However, both the autopoietic tradition and the representationalist tradition fail to recognize feedback. Treatments of consciousness by Prinz, Dretske, and others neglect the organism’s dynamic role in subjective experience, simply treating qualia as little more than a representation of an environmental state. Internal goings-on such as hunger are seen as just another kind of environment. No place is given to a subject, there is no center of agency, no feedback between the environment, action, and consciousness, making it unsurprising that such views turn the place of *subjective* experience in nature into something like a mystery. Subjective experience simply slips through the cracks in such a widening of the explanatory gap. Godfrey-Smith (2020a) once wrote that the “ancestors of qualia were born and flourished in the seventeenth, eighteenth, and nineteenth centuries” (p. 111). He described the empiricist tradition in philosophy from George Berkeley, John Locke, David Hume, and J.S. Mill in addition to the earlier work of Francis Bacon as furnishing a picture of the mind as a tabula rasa on which simple ideas and sense impressions in the form of colors, shapes, and sounds formed the theater of mind, populating it in something like a picture of mental “atoms” combining in various ways. These externalist ideas have shaped the way we think about qualia and phenomenological properties today, but despite their venerable history they may well be mistaken, similar to other strongly externalist views of life and mind.

Here, we can follow Godfrey-Smith’s suggestion to “reject the un-ecological side of the” autopoietic anti-Darwinian emphasis on autonomy and instead embrace the importance of both input and output, or perhaps more importantly the causal traffic that goes on between both sides and expresses itself in action (2016a, p. 788). Unlike Godfrey-Smith, however, I argue that this traffic important to understanding the function of consciousness is readily explicable from the evaluative side of experience, which is directly tied to action. Whereas the dimension of sensory experience emphasizes an externalist view of consciousness, the dimensions of self-experience emphasize internalism, two kinds of overreaching that make it extremely hard if not necessarily unsatisfactory to explicate either side in terms of the other, which is reminiscent of the dilemma ethologists faced between the vitalist’s internalism and the behaviourist’s externalism. Here, evaluation offers us a way out of this dilemma by being inherently dynamic and agential. What makes an external state good or bad depends on the state of the organism, and likewise whether a state of the organism is a good one depends crucially on the environment. Ecological feedback is built into this Darwinian picture of an

¹⁰ A similar warning has been made by Spinoza (see Veit 2020).

evaluation—first view from the very beginning, with action and agency being emphasized.

Evaluative Experience

What is it that unifies moods, feelings, and emotions such as pain, pleasure, anxiety, satiety, hunger, thirst, resentment, nausea, grief, pessimism, discomfort, boredom, curiosity, embarrassment, jealousy, comfort, companionship, fear, breathlessness, empathy, optimism, love, anger, among many others? Valence is usually taken to be their one common denominator and so far this is how this dimension has been treated here: emotional states are affective states. But before we can explicate a picture in which the evaluative dimension of consciousness constitutes its foundations, we need to respond to the possible objection that there is no single property shared by this dimension that gave rise to consciousness. As Browning (2020), in her attempt to naturalize a subjective notion of animal welfare, readily acknowledges: what we usually call affective states are “extremely heterogeneous states” (p. 164). They all appear to have their own unique subjective experience associated with them, which come in different varieties and gradations regarding their “intrusiveness” into our experience as a whole. But they are also heterogeneous in terms of their underlying physiology, with different brain regions being involved in different experiences and heterogeneity among both the inputs that cause said “affects” and their corresponding outputs in the form of behavior (Browning 2020). The goal of this section is thus to defend the popular idea of valence constituting an evaluative “common currency” for the comparison (or rather evaluation) of different subjective states that ties this dimension together.

That complex animals have, or perhaps even must have, a proximate common currency linked to fitness in which the values of different actions are ranked is a frequent claim in the behavioral, cognitive, and—perhaps unsurprisingly—the affective sciences (McFarland and Sibly 1975; McCleery 1977; McNamara and Houston 1986; Cabanac 1992; Shizgal and Conover 1996). Indeed, it is an influential view among those taking an explicitly evolutionary approach to consciousness, to see a currency of evaluation as something quite central in making sense of subjective experience in nature (Merker 2005, 2007; Morsella 2005; Cabanac et al. 2009; Ginsburg and Jablonka 2019). Unfortunately, philosophers—with the exception of Spurrett (2014) and Browning (2020)—have spent very little time engaging with these views.¹¹ Browning (2020), who has influenced

my own thinking here, endorses a common currency view of pleasure and pain in order to naturalize the notion of animal welfare as a single integrated state. Both are “unorthodox” philosophers *in science* because their work is motivated from a strongly naturalist stance to *contribute to* the debates in the sciences themselves, rather than just write *about* it in the confines of philosophy journals.¹² Their respective calls for philosophers to become interested in common currency claims and animal welfare science will be taken up here, although this article is likewise a contribution to science as much as it is to philosophy.

One problem with the line of argument taken here is that even if there is a common currency in humans, this doesn’t necessarily mean that such a currency was present at the very origins of subjective experience. If it did in fact come first, then the heterogeneity of affective states we are all familiar with ought to be a latecomer—not something that must arise first for a common currency to evolve. Indeed, the very notion of a “common currency” may not be useful for thinking about the origins of valence because it has the existence of multiple affective states already built in. Talk of a currency makes it seem, or at least tempting to think about, valence as having evolved to make the different aforementioned states comparable, similar to how real monetary currencies evolved only after there were goods that could be traded in a more efficient way through the implementation of a common currency. Valence could then be seen as having evolved in response to a certain kind of phenomenological complexity on the sensory side. The notion of “affective states,” on the other hand, seemingly presupposes a form of evaluation and begs the question as to whether such states could exist prior to the existence of any form of subjective valence. What makes them functional is precisely their valence.

Consider a general distinction that is often drawn between emotions as short-term states brought about through the presence or anticipation of appropriate stimuli, with a motivational and rewarding role (e.g., food, play, prey, or procreation opportunities) or a punishing one that causes avoidance (e.g., disgust, predation, and other potentially noxious stimuli such as poison or fire),¹³ and moods as comparatively long-term experiences. Some emotions last longer than others and so do moods. Here it may be more useful to think about these various states along a Darwinian continuum from very instantaneous valenced feelings like a sudden pain or cold to very general and long-lasting ones such as a general fatigue or depression. Naturally, we

¹¹ One of Spurrett’s (2014) papers is literally titled “Philosophers should be interested in ‘common currency’ claims in the cognitive and behavioural sciences.”

¹² For a defense of this “active” role for philosophers of science see Ankeny et al. (2011); Khelifaoui et al. (2021); Pradeu et al. (2022).

¹³ See LeDoux (2012); Carver (2001); Rolls (2005); Crump et al. (2020).

should not put too much weight on current terminology. The terms we use to describe the evaluative side of experience may make it hard to think about its possible evolutionary origins and which capacities come first. A radical alternative to the one defended here would be to think that some of the ancient emotions and moods occurred first and separately, and only later came to be integrated into a common currency—that is, if such a common currency of evaluation even exists, which itself can be contested. In defending a view that treats valence as the origin of subjective experience we must respond to the possibility that there is no true consciously experienced evaluative common currency.

After all, introspection into our own experience reveals a complexity of subjective experiences consisting of different states that make up our own phenomenological complexity. Browning (2020) puts this elegantly: “I consider myself right now and the combination of states I am experiencing—mild hunger, physical comfort in my office chair, slight head pain from a lingering cold, anticipation of my upcoming lunch, some intellectual discomfort from trying to write this chapter, among other states” (p. 169). While these appear to be strikingly different kinds of experiences, there does appear to be an intuitive sense in which they become integrated on something like a single scale. They have a felt evaluative sensation that becomes integrated into a single state and enables us to compare competing interests and motivations. This comparison does not take place in the cognitive or representationalist sense of a calculation, but rather as an instantaneous general feeling of one’s state—a total state of momentary *feeling* in just that sense of the word. They present a nexus of evaluation that enables us to engage in efficient decision-making under the conflicts of different needs even in the absence of felt representations. Valence came plausibly into existence with a basic feeling of good and bad, without any *felt* sensory richness. As long as the first sentient beings continued to engage in a behavior that caused them a “plus sign” and changed it up once it became a “minus,” there would have been no need for the presence of the other dimensions. Nevertheless, it is straightforward to see how the other dimensions could be built on top of such a capacity to enrich an evaluative mode of being that describes so much of animal life with further gains in phenomenological complexity being united under a “common currency” of valence. And this is ultimately what motivates me to take a valence-first view of the evolution of consciousness: unlike the other dimensions, there is no explanatory leftover of how the other mysterious properties of consciousness arise in an evaluative model of consciousness. Whereas the strongly externalist and representationalist view of consciousness based on a model of visual capacities fails to account for selfhood and evaluation, the strongly internalist model of consciousness as self-awareness fails to account for the functional capacities of representation and evaluation. Both fail

to put at center stage the sensory-evaluative-motor feedback that is so central to a teleonomic Darwinian view of life. Such feedback ought to be at the center of any theory of consciousness as a way of engaging the world as a vulnerable organism with a complex lifestyle that requires evaluation. There is one further line of evidence, however, that we can draw on to support the idea that evaluation constitutes the most basal kind of subjective experience.

Notably, this economic sense of evaluative agency can be seen as a refined commonsense view capturing much of our thinking about pleasure and pain and constitutes a folk model of decision-making. The pre-philosophical intuitions of many people appear to positively “insist on the absurdity of supposing that anyone can ultimately be motivated otherwise than by a concern for their own happiness, understood hedonistically” (Sprigge 1999, p. 314). So it is unsurprising that many engaged in the study of feelings, emotions, and moods assume an evaluative common currency, and one may even consider it a starting premise from which much other work follows. But whether this dimension is really the most central property of consciousness is something that ought to receive more empirical investigation. In my discussion of sensory experience I have mentioned that a naturalist approach to consciousness should avoid a strict dependence on the intuition of philosophers regarding which dimensions of consciousness *they* find to be most basic, basal, simple, or however else we may wish to describe the most minimal sense of experience. Too much has been made out of the apparent “obviousness” that colors and other pure qualia on the sensory side must be the core of explaining consciousness. This is not to say, however, that intuitions do not matter at all, especially when philosophers claim that theirs are representative of the population at large and used to draw conclusions about the alleged impossibility of solving the hard problem of consciousness.

Here, we can rely on a rapidly growing literature that supports the larger project I am engaged in, yet has received scant attention from philosophy of mind at large, i.e., experimental philosophy of mind. Recent work on the intuitions of the public regarding their conscious experiences appears to offer a strikingly different picture from the major views within philosophy of mind and one that fits strongly with the evaluation-first view defended here. In an influential study, Sytsma and Machery (2010) show that the public doesn’t share the philosophical consensus of consciousness as being characterized by its “phenomenality.” While there is much debate on how this notion ought to be conceived, there is broad consensus in the literature that “subjectively experienced mental states have phenomenal properties: There is something it is like to see red, smell banana, feel anger, and be in pain” (2010, p. 324). As they show, however, lay-people do not appear to recognize a commonality between all of these states. When asked whether a relatively simple

robot can smell various entities or be angry, non-philosophers appear to be unsure, suggesting that they do not share the philosophical conception of consciousness and subjective experience (Machery and Sytsma 2011). They do not define consciousness like philosophers vis-a-vis their “felt” or “phenomenal” properties. But if they do not rely on the philosophers’ “consensus” view, this raises the question of how they think about subjective experience.

The findings of Sytsma and Machery suggest that the ascription of conscious mental states to a robot rests on a hedonic evaluation, i.e., whether something feels good or bad. Laypeople appeared to distinguish those states with valence or affect such as pain, moods, and emotions and those without such as a pure sensation of red or the smell of isoamyl acetate (Sytsma and Machery 2010). Ordinary people seem to perceive consciousness as an evaluative experience. This would be a radical inversion of the way philosophers have thought about the problem: evaluation taking a much more center stage. Mere sensory discrimination did not appear to be particularly controversial when attributed to a robot, which is strikingly different from Dennett’s (1996) assertion that the robot Cog³ “cannot yet see or hear or feel at all” (p. 16). And yet, laypeople seemed resistant to assign a robot the ability to perceive “familiar smells associated with either positive or negative valence” (Sytsma and Machery 2010, p. 318). The common folk-concept of subjective experience thus appears to group “different types of perceptual experiences, bodily sensations, and felt emotions depending on their valence” (Sytsma and Machery 2010, p. 318). These results are very interesting because they may undermine the very foundation for the philosophical resistance of those defending the idea that there is something like an unbridgeable “explanatory gap” (Levine 1983) or a “hard problem of consciousness” (Chalmers 1995):

The hard problem is typically justified on the grounds that we are acquainted with the phenomenal properties of states such as pain and seeing red and that functional accounts of mental states fail to explain how they can have such phenomenal properties. Our findings challenge the first premise of this argument. Because people do not seem to conceptualize their subjective mental life as phenomenal, it is at least unclear that we are pretheoretically acquainted with the phenomenal properties of our conscious mental states. (Sytsma and Machery 2010, p. 324)

Philosophers have long religiously maintained that subjective experience is essentially conceived in the same way by both themselves and the public—that they are merely using the folk concept of consciousness. Sytsma and Machery (2010) list multiple high-profile examples of this widespread belief even among naturalist philosophers. Alvin Goldman (1993) comes out in support of “the basic integrity of the

folk-psychological conception of [phenomenal] consciousness and its importance in cognitive theorizing” (p. 364). Ned Block (2004), in his “Qualia” entry in the *The Oxford Companion to the Mind* similarly defends the phenomenological view of consciousness among philosophers as the common sense view of subjective experience. Patricia Churchland (1988) also accepts that the public holds such a view, but urges us to consider “outright replacement of the old folk notion of consciousness with new and better largescale concepts” (p. 302). Even Dennett (2005), who likewise wants to develop a scientific account of consciousness by more or less banishing the “phenomenality” or “qualia,” holds that these notions are part of “the lore” of our folk conception of consciousness that is picked up “in the course of our enculturation” (pp. 26–27). Moreover, the study of Sytsma and Machery (2010) also asked philosophers to make a prediction of the assessments laypeople would make, which they largely expected to be analogous to their own answers (though slightly less skeptical). But as Dennett’s assessment of Cog³ indicates, they may have been strikingly mistaken by confusing the enculturation of their own discipline with that of the public.

If this way of thinking about consciousness, however, rests in a mere artifact of philosophical training, we may have to radically revise our assessment of the common philosophical critique by the likes of Chalmers and Nagel, who have maintained that neuroscientists and psychologists who claim to have explained consciousness have naively failed to address the “obvious” hard problem of consciousness.¹⁴ If these experimental results are indicative of the way non-philosophers think about the problem of consciousness, such assertions ought to be seen as quite the uncharitable interpretation of what the scientists are doing and thinking. Sytsma and Machery (2010) posit that “it might be that like the folk, they do not conceive of subjective experience as being phenomenal, in spite of having plausibly carefully considered [“]what it is like[”] for them to see red, feel pain, and so on” (2010, p. 323). Both the public and scientists may simply not consider a further phenomenological property or “qualia” that needs to be addressed, which they think is in line with their own experiences, where “many ordinary people either don’t understand or don’t take seriously the philosophical concept of phenomenal consciousness even after a lengthy explanation” (2010, p. 323). If the hard problem is only something to be recognized once one has come to be “indoctrinated” by the orthodoxy in philosophy of mind, this may explain why some scientists, after passionately endorsing the possibility of and need for a science of consciousness, have later taken on the view that the hard

¹⁴ See Chalmers (1995) and Nagel (1974).

problem cannot or may not yet be solved, and that they are instead focusing on the “easy” problems of consciousness.

This may in some cases be merely a strategic choice to avoid the charges of naivety hurled against those scientists claiming to provide explanatory sketches and theories of consciousness. But if so, it would be an unfortunate one since it in turn legitimizes talk of the hard problem as something that cannot be overcome by ordinary science. Once a naturalistically inclined scientist interested in developing a science of the mind gets too close to what Dennett (2017) called the “Cartesian gravity” of Descartes’s dualistic way of thinking, it becomes all but impossible to escape the gravity of “Planet Descartes” (p. 20). And this is what the very mention of the hard problem unfortunately enables. If the experimental results reveal a vision-centric bias in our thinking about consciousness, we may wish to return to an earlier evolutionary view suggested by Dawkins (1998) in which “[t]he key to the origin of consciousness itself may lie in the emotional experience of suffering” (p. 324). This is a view I similarly defend here: evaluative experience as the most basic and original kind of subjective experience.

Ultimately, Sytsma and Machery (2010) have provided us with a beautiful case for the usefulness and perhaps even necessity of experimental philosophy to progress in philosophical debates muddled by appeals to intuition. Though this is not to say that their case for two different concepts of consciousness in the public and among philosophers has gone unchallenged.¹⁵ Godfrey-Smith (2020a), while accepting that, “Sytsma and Machery may be right about the everyday conception of experience,” nevertheless maintains that “everyday thinking may also be mistaken” (p. 311). Indeed, while I consider their results to provide striking support for an evaluation-first view of consciousness, I am reluctant to go as far as Sytsma and Machery, who see these results as undermining the *very* idea that there is anything like a hard problem or explanatory gap. My views are thus somewhat closer to Robbins and Jack (2006), who while similarly maintaining that valence and affect are the basis for the common sense concept of consciousness, see it as the underlying *cause* of the philosophical notion of phenomenology and the hard problem. They maintain that “the gap intuition is psychologically real *and deep*” (p. 60; italics in original) which Sytsma and Machery deny. But Robbins and Jack (2006) nevertheless think that the problem is an illusion and I think that deflates the problem too much.

Unlike Sytsma and Machery, I see the move towards understanding the biological basis of valence and affects as a naturalization of the vexing notion of “qualia” through

an alternative non-vision-centric model of consciousness. While Sytsma and Machery raise the possibility that the valence of consciousness is a problem akin to the phenomenological version of the hard problem, they resist the notion that we couldn’t explain it in virtue of functional and mechanistic explanations. They consider it to be straightforward to understand that the “hedonic value of a stimulus or a bodily state seems to be an evaluation of its expected value to the organism” (2010, p. 322). There doesn’t appear to be an additional problem of *why* there is valence. This makes the evaluative side of experience a compelling target for an attempt to bridge the gap between matter and mind. To have a phenomenological experience *is* to have an evaluative experience. To naturalize the puzzling notion of “qualia” is simply to explain how and why organisms have such an evaluation. Phenomenal states simply *are* explained within the context of an affect-based model of *phenomenological experience*, a concept Sytsma and Machery seemingly would prefer to eliminate, akin to the fate of the vitalist’s concept of life (2010, p. 322).

I have some sympathies for this view. There is often only a fine distinction between those who try to naturalize a concept and those who seek to eliminate and replace it. Perhaps the only difference is a degree of sympathy shared with the foregoing work of both philosophers and scientists studying the mind and in particular our phenomenological experience. Whether my problem with using the notion of valence to naturalize the phenomenological complexity we see in nature should be considered equivalent to the “hard” problem of consciousness or not, I will attempt to make progress on it here. Inevitably, this will involve some reshaping of how our ordinary concepts of both “mind” and “matter” conceive of what goes on in organisms. But that is simply what it means to find the place of consciousness in nature. Phenomenological complexity can be naturalized on a model of evaluative experience.

One final objection would be to insist that even if we think that the aforementioned points support the existence of a common currency of valence, our own human experience may strikingly differ from other animals, who may lack such a common currency even if it exists in us—or some of us. To this, of course, one can respond in a manner both Romanes and Darwin likely would have, i.e., to simply respond that pleasure—unlike language or higher-order symbolic thought—does not seem like something restricted to us. Indeed, it may even be more important for animals to possess a common currency of pleasure and pain, since they cannot engage in the same symbolic cognitive processing as us. In humans with the rare inability to feel pain, i.e., congenital analgesia, sometimes referred to as congenital insensitivity to pain, early death is common due to a neglect or inability to detect injuries and diseases (Thrush 1973; Nagasako et al. 2003; Cox et al. 2006). The experience of

¹⁵ See Huebner (2012); Talbot (2012); Peressini (2014); Chalmers (2020) for challenges to their results, and a recent empirically supported defense of their view (Sytsma and Ozdemir 2019).

pain is crucial to developing a concept of self and one's body in relation to its environment. Those with congenital analgesia, however, must exert cognitive effort to think about or rather actively represent the potential dangers to their body since they lack a system of "punishment" that would teach them from their childhood onwards. While this is certainly not easy, it can be done. Animals, however, are unlikely to even make it that far without the fast decision making and learning of important associations enabled through pain. They would not be able to think about the likely "harms" of particular behavior without pain to make these connections. Such evidence is compelling and our intuitive thinking about pleasure and pain appears to at least plausibly be on the right track in regards to a potentially much more ancient felt experience. Common sense usage and introspection do make it seem like pleasure and pain constitute something like a common currency, making it unsurprising that philosophers since antiquity have held similar views about the importance of pleasure for much of life. And it is this evaluative dimension that ultimately won the war of the five dimensions.

The Spoils of War

The goal of this war between the five dimensions was to crown the core of consciousness. While we may be lacking direct paleobiological data that could indicate which dimension has the most direct line of descent to the dawn of consciousness, this article has attempted to reverse engineer the origins of consciousness in evaluation. Let us summarize which dimensions lost and why.

I have argued that diachronic experience can be most readily dismissed as a necessary component of subjective experience since it is largely agreed to be absent in some animals, all the while granting them subjective experiences (of some kind). It can be seen as a higher-order feature of consciousness, not something that is likely to be present at its very origin.

Synchronic experience, while more often seen as a necessary component, must not be present at the very origins of consciousness either, since other animals without a strong connection between both hemispheres are likely to have a more disunified experience. Both dimensions of unity appear to be features of the way experience can be organized, rather than what makes it qualitative to begin with.

The experience of a self, I have argued, is something that is hard to disassociate from our thinking of consciousness, due to its centrality in our own conscious experience. But in a strongly gradualist picture it is unlikely to have been present at the very origins of consciousness, instead *built by* more basic kinds of subjective experience such as a distinction between exteroception and interoception in the dimension of sensory experience.

This left us with two contenders: the sensory and the evaluative side of consciousness. Sensory experience has long dominated much of our thinking about subjective experience in philosophy and science. Yet what this dimension fails to account for is the very subject that is so central to consciousness. To consider an alternative way of thinking to solve the problems of an old paradigm is a natural scientific move. By drawing on recent work in experimental philosophy of mind I have undermined the centrality of the sensory side in thinking about the most evident cases of experience. The *feelings side* which includes moods, emotions, and hedonic evaluations appears to be what drives the thinking of laypeople about consciousness and thus provides additional support for the evaluation-first view.

How animal consciousness should be studied has remained a controversial question (Birch et al. 2022). But as I hope to have shown in this article, there is great untapped potential for building an alternative model of consciousness based on what are sometimes perceived to be the background features of ordinary human conscious experience: moods, pains, evaluations—features that usually only come into the center when opportunities arise or things go badly. This alternative way of thinking has unfortunately been largely resisted, yet has striking support from and has been partially developed across economics, animal welfare science, neuroscience, behavioral ecology, animal consciousness science, and ethology. Even if a theory based on the evaluative side of conscious experience will eventually turn out to be false, we are likely to make much greater progress by developing a picture that has received scant attention, yet ought to be considered as *at least* an equal competitor to the sensory-first views that have been modeled on the human phenomenon of visual experience. Unlike rich human-like vision, after all, a basic sense of evaluation may have been present long before animals had any rich capacities to discriminate states of the world, and it is here that the hard problem challenge and explanatory gap appear to be significantly weakened. But to unearth the evolution of this evaluative spark of experience will be the target of a follow-up paper (Veit 2022a) and book (Veit 2022d).

Acknowledgments My thanks go out to Peter Godfrey-Smith, Paul Griffiths, Marian Dawkins, Jonathan Birch, Kristin Andrews, and Heather Browning for their feedback. Furthermore, I would like to thank the Animal Sentience lab at the London School of Economics and the Comparative Cognition Lab at the University of Cambridge for helpful discussions. Lastly, I would like to thank Stuart Newman, Deborah Klosky, and the reviewers at *Biological Theory* for helping me to polish the manuscript further.

Funding Open Access funding enabled and organized by CAUL and its Member Institutions. WV's research was supported under Australian Research Council's Discovery Projects funding scheme (Project Number FL170100160).

Declarations

Competing interests None.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Ankeny R, Chang H, Boumans M, Boon M (2011) Introduction: philosophy of science in practice. *Eur J Philos Sci* 1(3):303
- Baars B (1988) *A cognitive theory of consciousness*. Cambridge University Press, Cambridge
- Birch J, Broom DM, Browning H, Crump A, Ginsburg S, Halina M et al (2022) How should we study animal consciousness scientifically? *J Conscious Stud* 29(3–4):8–28
- Birch J, Schnell AK, Clayton NS (2020) Dimensions of animal consciousness. *Trends in Cognitive Sciences* 24(10):789–801
- Block N (2004) Qualia. In: Gregory RL (ed) *Oxford companion to the mind*. Oxford University Press, pp 785–789
- Broom DM (2014) *Sentience and animal welfare*. CABI, Wallingford
- Browning H (2020) *If I could talk to the animals: measuring subjective animal welfare*. PhD Thesis, Australian National University. <https://doi.org/10.25911/5f1572fb1b5be>
- Browning H (2022) The measurability of subjective animal welfare. *J Conscious Stud* 29(3–4):150–179
- Browning H, Birch J (2022) Animal sentience. *Philos Compass* 17(5):e12822. <https://doi.org/10.1111/phc3.12822>
- Browning H, Veit W (2020) The measurement problem of consciousness. *Philos Top* 48(1):85–108
- Browning H, Veit W (2021) Evolutionary biology meets consciousness: essay review of Simona Ginsburg and Eva Jablonka's *The Evolution of the Sensitive Soul*. *Biol Philos* 36(5):1–11
- Browning H, Veit W (2022) The sentience shift in animal research. *New Bioethics*. <https://doi.org/10.1080/20502877.2022.2077681>
- Cabanac M (1992) Pleasure: the common currency. *J Theor Biol* 155(2):173–200
- Cabanac M, Cabanac AJ, Parent A (2009) The emergence of consciousness in phylogeny. *Behav Brain Res* 198(2):267–272
- Cacioppo JT, Gardner WL (1999) Emotion. *Annu Rev Psychol* 50(1):191–214
- Carver CS (2001) Affect and the functional bases of behavior: on the dimensional structure of affective experience. *Pers Soc Psychol Rev* 5(4):345–356
- Chalmers DJ (1995) Facing up to the problem of consciousness. *J Conscious Stud* 2(3):200–219
- Chalmers DJ (2020) Is the hard problem of consciousness universal? *J Conscious Stud* 27(5–6):227–257
- Churchland PS (1988) Reduction and the neurobiological basis of consciousness. In: Marcel AJ, Bisiach E (eds) *Consciousness in contemporary science*. Oxford University Press, Oxford, pp 273–304
- Cox JJ, Reimann F, Nicholas AK, Thornton G, Roberts E, Springell K et al (2006) An SCN9A channelopathy causes congenital inability to experience pain. *Nature* 444(7121):894–898
- Crump A, Bethell EJ, Earley R, Lee VE, Mendl M, Oldham L et al (2020) Emotion in animal contests. *Proc R Soc B* 287(1939):20201715
- Dainton B (2018) Temporal consciousness. In: Zalta EN (ed) *The Stanford encyclopedia of philosophy*. Metaphysics Research Lab, Stanford University. <https://plato.stanford.edu/entries/consciousness-ess-temporal/>
- Damasio A (1994) *Descartes' error: emotion, reason, and the human brain*. Avon Books, New York
- Dawkins MS (1998) Evolution and animal welfare. *Q Rev Biol* 73(3):305–328
- de Waal F (2006) *Primates and philosophers: how morality evolved*. Princeton University Press, Princeton
- Deckel AW (1995) Laterality of aggressive responses in *Anolis*. *J Exp Zool* 272(3):194–200
- Deckel AW (1997) Effects of alcohol consumption on lateralized aggression in *Anolis carolinensis*. *Brain Res* 756(1–2):96–105
- Dehaene S (2014) *Consciousness and the brain: deciphering how the brain codes our thoughts*. Viking, New York
- Dennett D (1991) *Consciousness explained*. Little, Brown, New York
- Dennett DC (1996) *Kinds of minds: toward an understanding of consciousness*. MIT Press, Cambridge
- Dennett DC (2005) *Sweet dreams: philosophical obstacles to a science of consciousness*. MIT Press, Cambridge
- Dennett DC (2017) *From bacteria to Bach and back: the evolution of minds*. Norton, New York
- Dretske F (1993) Conscious experience. *Mind* 102(406):263–283
- Dretske F (1999) The mind's awareness of itself. *Philos Stud* 95(1/2):103–124
- Escobar JM (2012) Autopoiesis and darwinism. *Synthese* 185(1):53–72
- Ginsburg S, Jablonka E (2019) *The evolution of the sensitive soul: learning and the origins of consciousness*. MIT Press, Cambridge
- Godfrey-Smith P (2016a) Individuality, subjectivity, and minimal cognition. *Biol Philos* 31(6):775–796
- Godfrey-Smith P (2016b) *Other minds: the octopus, the sea, and the deep origins of consciousness*. Farrar, Straus and Giroux, New York
- Godfrey-Smith P (2020a) *Metazoa: animal minds and the birth of consciousness*. William Collins, Glasgow
- Godfrey-Smith P (2020b) Varieties of subjectivity. *Philos Sci* 87(5):1150–1159
- Goldman AI (1993) Consciousness, folk psychology, and cognitive science. *Conscious Cogn* 2(4):364–382
- Griffiths PE (1997) *What emotions really are: the problem of psychological categories*. University of Chicago Press, Chicago
- Griffiths PE (2017) Emotions. In: Bechtel W, Graham G (eds) *A companion to cognitive science*. Blackwell Publishing Ltd, pp 197–203. <https://doi.org/10.1002/9781405164535.ch11>
- Güntürkün O, Bugnyar T (2016) Cognition without cortex. *Trends Cogn Sci* 20(4):291–303
- Huebner B (2012) Reflection, reflex, and folk intuitions. *Conscious Cogn* 21(2):651–653
- Humphrey N (1992) *A history of the mind: evolution and the birth of consciousness*. Simon and Schuster, New York
- Hurley SL (1998) *Consciousness in action*. Harvard University Press, Cambridge
- James W (1890) *The principles of psychology*. Dover, New York
- Keenan JP, Gallup GG, Falk D (2003) *The face in the mirror: the search for the origins of consciousness*. HarperCollins, London
- Khelifaoui M, Gingras Y, Lemoine M, Pradeu T (2021) The visibility of philosophy of science in the sciences, 1980–2018. *Synthese* 199:6219–6249. <https://doi.org/10.1007/s11229-021-03067-x>

- LeDoux J (2012) Rethinking the emotional brain. *Neuron* 73(4):653–676
- Levine J (1983) Materialism and qualia: the explanatory gap. *Pac Philos Q* 64(4):354–361
- Machery E, Sytsma J (2011) Robot pains and corporate feelings. *Philos Mag* 52:78–82
- Matthewson J (2020) Does proper function come in degrees? *Biol Philos* 35(4):1–18
- McCleery R (1977) On satiation curves. *Anim Behav* 25:1005–1015
- McFarland D, Sibly R (1975) The behavioural final common path. *Philos Trans R Soc Lond B* 270(907):265–293
- McNamara JM, Houston AI (1986) The common currency for behavioural decisions. *Am Nat* 127(3):358–378
- Merker B (2005) The liabilities of mobility: a selection pressure for the transition to consciousness in animal evolution. *Conscious Cogn* 14(1):89–114
- Merker B (2007) Consciousness without a cerebral cortex: a challenge for neuroscience and medicine. *Behav Brain Sci* 30(1):63–81
- Morsella E (2005) The function of phenomenal states: supramodular interaction theory. *Psychol Rev* 112(4):1000
- Nagasako EM, Oaklander AL, Dworkin RH (2003) Congenital insensitivity to pain: an update. *Pain* 101(3):213–219
- Nagel T (1974) What is it like to be a bat? *Philos Rev* 83(4):435–450
- Okasha S (2018) Agents and goals in evolution. Oxford University Press, Oxford
- Peressini A (2014) Blurring two conceptions of subjective experience: folk versus philosophical phenomenality. *Philos Psychol* 27(6):862–889
- Pradeu T, Lemoine M, Khelifaoui M, Gingras Y (2022) Philosophy in science: can philosophers of science permeate through science and produce scientific knowledge? *Br J Philos Sci*. <https://doi.org/10.1086/715518>
- Prinz J (2000) A neurofunctional theory of visual consciousness. *Conscious Cogn* 9(2):243–259
- Prinz J (2012) *The conscious brain*. Oxford University Press, Oxford
- Robbins P, Jack AI (2006) The phenomenal stance. *Philos Stud* 127(1):59–85
- Rolls E (2005) *Emotion explained*. Oxford University Press, Oxford
- Romanes GJ (1883) *Mental evolution in animals, with a posthumous essay on instinct by Charles Darwin*. Kegan Paul, Trench, London
- Shizgal P, Conover K (1996) On the neural computation of utility. *Curr Dir Psychol Sci* 5(2):37–43
- Sovrano VA, Bisazza A, Vallortigara G (2001) Lateralization of response to social stimuli in fishes: a comparison between different methods and species. *Physiol Behav* 74(1–2):237–244
- Sprigge T (1999) The relation between Jeremy Bentham’s psychological, and his ethical, Hedonism. *Utilitas* 11(3):296–319
- Spurrett D (2014) Philosophers should be interested in ‘common currency’ claims in the cognitive and behavioural sciences. *S Afr J Philos* 33(2):211–221
- Sytsma J (2010) Dennett’s theory of the folk theory of consciousness. *J Conscious Stud* 17(3–4):107–130
- Sytsma J (2012) Revisiting the valence account. *Philos Top* 40(2):179–198
- Sytsma J (2014) The robots of the dawn of experimental philosophy of mind. In: Machery E, O’Neill E (eds) *Current controversies in experimental philosophy*. Routledge, New York
- Sytsma J, Machery E (2010) Two conceptions of subjective experience. *Philos Stud* 151(2):299–327
- Sytsma J, Machery E (2012) On the relevance of folk intuitions: a commentary on Talbot. *Conscious Cogn* 21(2):654–660
- Sytsma J, Ozdemir E (2019) No problem: evidence that the concept of phenomenal consciousness is not widespread. *J Conscious Stud* 26(9–10):241–256
- Sytsma JM, Machery E (2009) How to study folk intuitions about phenomenal consciousness. *Philos Psychol* 22(1):21–35
- Talbot B (2012) The irrelevance of dispositions and difficulty to intuitions about the “hard problem” of consciousness: a response to Sytsma, Machery, and Huebner. *Conscious Cogn* 21(2):661–666
- Thompson E (2007) *Mind in life*. Harvard University Press, Cambridge
- Thompson E (2011) Reply to commentaries. *J Conscious Stud* 18(5–6):176–223
- Thrush D (1973) Congenital insensitivity to pain. *Brain* 96(2):369–386
- Tononi G, Boly M, Grasso M, Hendren J, Juel BE, Mayner WG et al (2022) IIT, half masked and half disfigured. *Behav Brain Sci* 45:e60
- Trestman M (2017) Minds and bodies in animal evolution. In: Andrews K, Beck J (eds) *The Routledge handbook of animals minds*. Routledge, New York, pp 206–215
- Tye M (1995) *Ten problems of consciousness: a representational theory of the phenomenal mind*. MIT Press, Cambridge
- Vallortigara G (2000) Comparative neuropsychology of the dual brain: a stroll through animals’ left and right perceptual worlds. *Brain Lang* 73(2):189–219
- Veit W (2019a) Model pluralism. *Philos Soc Sci* 50(2):91–114. <https://doi.org/10.1177/0048393119894897>
- Veit W (2019b) Modeling morality. In: Magnani L, Nepomuceno A, Salguero F, Barés C, Fontane M (eds) *Model-based reasoning in science and technology*. Springer, London, pp 83–102
- Veit W (2020) Dennett and Spinoza. *Austral Philos Rev* 4(3):259–265
- Veit W (2021a) Agential thinking. *Synthese* 199:13393–13419
- Veit W (2021b) Samir Okasha’s philosophy. *Lato Sensu* 8(3):1–8. <https://doi.org/10.20416/LSRSPS.V8I3.1>
- Veit W (2022a) Complexity and the evolution of consciousness. Preprint. <https://philsci-archive.pitt.edu/20675>
- Veit W (2022b) Consciousness, complexity, and evolution. *Behav Brain Sci* 45:e61
- Veit W (2022c) Health, agency, and the evolution of consciousness. PhD thesis, University of Sydney. Manuscript in preparation
- Veit W (2022d) A philosophy for the science of animal consciousness. Book manuscript in preparation
- Veit W (2022e) Towards a comparative study of animal consciousness. Preprint. <https://philsci-archive.pitt.edu/20718/>
- Veit W (2022f) Review of Peter Godfrey-Smith’s metazoa: animal minds and the birth of consciousness. *Philos Sci*. <https://doi.org/10.1017/psa.2022.26>
- Veit W, Browning H (2020) Two kinds of conceptual engineering. <https://philsci-archive.pitt.edu/17452/>
- Veit W, Browning H (forthcoming) Life, mind, agency: why Markov blankets fail the test of evolution. *Behav Brain Sci*. preprint: <https://philsci-archive.pitt.edu/20126/>
- Veit W, Huebner B (2020) Drawing the boundaries of animal sentience. *Anim Sentience* 5(29):13
- Veit W, Ney M (2021) Metaphors in arts and science. *Eur J Philos Sci* 11:44. <https://doi.org/10.1007/s13194-021-00351-y>
- Velmans M (1991) Is human information processing conscious? *Behav Brain Sci* 14(4):651–669
- Weisberg M (2013) *Simulation and similarity: using models to understand the world*. Oxford University Press, New York

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.