

Please cite as: Veit, W. (2022). Defending the Pathological Complexity Thesis. Preprint.

Check [www.walterveit.com](http://www.walterveit.com) for citation details once published

# Defending the Pathological Complexity Thesis

Walter Veit<sup>1,2,3</sup>

<sup>1</sup>Department of Philosophy, University of Bristol

<sup>2</sup>Munich Center for Mathematical Philosophy, LMU

<sup>3</sup>Department of Biology, University of Oxford

## Abstract

In this article, I respond to commentaries by Eva Jablonka, Simona Ginsburg, and David Spurrett on my target article ‘Complexity and the Evolution of Consciousness’, in which I have offered the first extended articulation of my pathological complexity thesis as a hypothesis about the evolutionary origins and function of consciousness. My reply is structured by the arguments raised rather than by authors and will offer a more detailed explication of some aspects of the pathological complexity thesis.

*Keywords:* biological complexity; consciousness; hedonic valence; animal consciousness; teleonomy; evolution of consciousness

## Index

- 1 Introduction
- 2 What is Pathological Complexity?
- 3 From Pathological Complexity to Consciousness
- 4 Conclusion and Further Suggestions

# 1 Introduction

The goal of my target article ‘Complexity and the Evolution of Consciousness’ (Veit 2022a) was to offer a hypothesis about the evolutionary origins and function of consciousness (or for that matter sentience as the most minimal kind of consciousness). I am grateful to Spurrett (2022) and Jablonka and Ginsburg (2022) for taking the time to examine my proposal in detail and offer the criticisms I respond to here. Before we begin, however, let me briefly restate the hypothesis of interest:

*The Pathological Complexity Thesis:*

The function of consciousness is to enable the agent to respond to pathological complexity.

To summarize my thesis briefly, I have argued that a computational explosion in the pathological complexity (equivalent to life history complexity) of organisms resulting from the emergence during the Cambrian of a distinctive animal lifestyle, gave rise to the first sparks of subjective experience. Rather than locating the origins of consciousness in perceptual representations of ‘outside conditions’, I defended an evaluation-first view of consciousness, with minimal evaluative hedonic states constituting the dawn of ‘qualia’, i.e. phenomenological states. These hedonic states gave sentient animals the advantage of weighing their different demands, opportunities, and dangers against each other to effectively deal with the economic trade-offs in their decision-making (see also Veit forthcoming). While the failure to evolve such a hedonic ‘common currency’ for action selection led to the Ediacaran extinction, its later evolution led to the Cambrian explosion, allowing far more complex body-plans to be explored, that due to their high degrees of freedom were previously too costly to deal with.

Some of the problems raised for my thesis by Jablonka, Ginsburg, and Spurrett could have been resolved by a substantially longer version of my articulation of the pathological complexity thesis that was naturally beyond the length of a journal article. While some of their points will be addressed in the near future in a forthcoming book (see Veit forthcoming), I am grateful for this opportunity to address their arguments in detail here and further explicate the pathological complexity thesis.

## Article Outline

This article is organized as follows. In the second section, “What is Pathological Complexity?”, I offer further details on my notion of pathological complexity and respond to several criticisms of it. In the third section, “From Pathological Complexity to Consciousness”, I expand on and respond to criticisms to my account of how pathological complexity gives rise to sentience. Finally, the fourth section,

“and Further Suggestions”, I will conclude the discussion and offer further suggestions for future research.

## 2 What is Pathological Complexity?

What kind of complexity is relevant for the evolution of consciousness? Both sets of commentators have put pressure on my notion of ‘pathological complexity’ and it is worth looking at them in detail.

Spurrett (2022) notes that I could be clearer in specifying what pathological complexity consists in. While he acknowledges that my notion of pathological complexity is meant to offer a combination of what I perceive to be deficient externalist views of complexity (such as in Peter Godfrey-Smith’s (1996) *environmental complexity thesis*) and internalist views of complexity (such as in the *skin brain thesis* by Fred Keijzer [Keijzer et al. 2013; Keijzer 2015; Keijzer and Arnellos 2017]),<sup>1</sup> he doesn’t think my suggestion for a more dynamic view of complexity that acknowledges both internal and external sources goes much beyond his own brief proposal for a combination of those views (see Spurrett 2020), because my different descriptions of pathological complexity partially pull in different directions. This is why Spurrett criticizes that my notion of pathological complexity has not been plainly stated. Let us examine this objection more closely. Here is my canonical statement of what pathological complexity is on a fundamental ontological level:

Pathological complexity emerges dynamically from the interaction of organism and environment, as a measure of the complexity of an organism’s *life history strategy*, and will hence vary with the different “lifestyles” of different animals. It can be understood as the computational complexity of the Darwinian, or “economic,” trade-off problem faced by all biological agents as they deal with challenges and opportunities throughout their life histories in order to maximize their fitness.

– Walter Veit (2022a, p. 2)

As stated here, pathological complexity is a real biological problem faced by all living organisms (see also Veit and Browning forthcoming, 2022). Yet, in my abstract, Spurrett (2022) rightly points out, I describe this problem in terms of “having to deal

---

<sup>1</sup> Originally, the pathological complexity thesis was introduced in print as an evolutionary alternative to attempts by integrated information theory to link consciousness to the complexity of information integration (Veit 2022b).

with a complex body with high degrees of freedom” (p. 1). This may (mistakenly) suggest that pathological complexity only exists for a particular subset of organisms and we can thus easily see where Spurrett derives his confusion from. That is, it arises from a reading of a simplified and abbreviated statement of the many ideas featured in my target article, that does not capture the finer distinctions I go on to make. While I expand at length how pathological complexity is a universal obstacle to all life, I also emphasize that the Cambrian brought with it new complex bodies with high degrees of freedom that gave rise to an explosion of this type of complexity, and thus made sentience worth having. Degrees of freedom - or as I simplified them in my article, as roughly the set of alternative actions an organism can take - are the most important driver of this complexity as it relates to the origins of sentience, but importantly not the only one.

Spurrett is of course right that the degrees of freedom of an organism (the number of independent parameters that specify the possible states the organism can be in) may change without impacting their behavioural repertoire (and vice versa), and I should have given a more precise definition here to avoid confusion. While I didn't want to get too technical in a paper that offers a broad introduction to the pathological complexity thesis, Spurrett (2022) is right to insist that the distinction between behavioural repertoire and degrees of freedom is important to distinguish the parallel problems of action selection (“which potential action should be executed now?”) from the problem of action specification (“how to define potential actions and how to execute them?”) (see also Cisek 2007). I agree with all of this. Nevertheless, I will note that neither notion is intended as a definition of pathological complexity. Instead, I merely use them as important examples of how pathological complexity can increase (or for that matter decrease). Other factors that can also increase pathological complexity are the length of life, the number of life history stages, environmental heterogeneity, the presence of predators, among many others. These are all factors that can influence the *life history complexity* of an organism, which is what my notion of pathological complexity is ultimately meant to capture. This is why I offer another description of pathological complexity in terms of how it could be operationalized:

Pathological complexity can be operationalized in terms of the number of parameters and constraints in the evolutionary optimization problem studied by *state-dependent or state-based behavioral and life history theory*.

– Walter Veit (2022a, p. 2)

Rather than hinting at competing definitions that pull in different directions, I have thus offered i) a conceptual statement of the pathological complexity thesis, ii) a statement of what has led to the explosion in pathological complexity that led the evolution of sentience, and iii) a reference to state-dependent or state-based behavioral and life history theory as the means to offer us an “elegant mathematical

framework” (p. 2) for the measurement and operationalization of pathological complexity.

Here, Jablonka and Ginsburg (2022) criticize that this mathematical framework does not materialize in my target article and that they are unsure how such analyses could be undertaken. In a similar vein, Spurrett notes that if pathological complexity is a multidimensional trade-off problem, we are owed an answer as to how all these different components among which trade-offs occur could possibly be reduced to unidimensional fitness. To this, of course, we can reply that evolutionary biologists, and especially life-history theorists, recognize this trade-off complexity as a real phenomenon and routinely engage in just these kinds of calculations. Nevertheless, just as behavioural ecologists use idealizations and omissions in their models, so will we have to start with simple models to assess the pathological complexity facing different species. Such work cannot be done from the philosophical armchair alone and requires collaborations with biologists to develop better proxy measures of life-history complexity. Currently, I am working with life history researchers at the University of Oxford on developing just such measures to create just such a new research program that will help us to better understand the evolution of biological complexity.

On a more terminological level, Jablonka and Ginsburg (2022) find my term ‘pathological complexity’ confusing because their intuitive interpretation of the term is that it must have something to do with pathologies, or that this complexity is itself pathological. Nevertheless, they note that this is not what I seem to be interested in, since I do not talk much about health and disease in my target article. As I acknowledge in my target article, perhaps the terms ‘teleonomic complexity’ or ‘life history complexity’ could have been less confusing alternatives, but I chose the term ‘pathological complexity’ precisely because these others do not carry the emphasis on trade-offs that I am interested in. Jablonka and Ginsburg think that my alternatives would have been better descriptors as they do not see how health has something to do with trade-offs, but in doing so they make the mistake of thinking of health and pathology just in terms of our ordinary folk concepts based on the human case, rather than taking a broader biological notion.

Just like the notions of ‘adaptation’ and ‘design’, these concepts of health and pathology can come to be explicated in terms of natural selection. As I have argued in another paper, one that was meant as a programmatic motivation for the pathological complexity thesis, health must be understood through a Darwinian lens in order to assess one organism as being healthier than another (Veit 2023). If we compare different pathological states, such as broken bones, lesions, infections, and the like, there is simply no way of assessing these against each other without something like an ultimate ‘common currency’ - and this currency is of course fitness. Dealing better with one biological danger comes at the cost of foregoing other benefits or making one more susceptible to other dangers. Both biological design and

health are thus inherently related to the notion of trade-offs. This is why I argued that “it is only in understanding life history trade-offs that we can distinguish healthy from pathological trait variation and that includes variations of consciousness both within and across species” (Veit 2022a, p. 2). And just as health requires a common currency, so do animals with very high pathological complexity require a proximate common currency in the form of hedonic valence to deal with the trade-offs of their complex lives. This close connection is ultimately why I have titled my thesis the ‘pathological complexity thesis’ and why my dissertation was titled ‘Health, Agency, and the Evolution of Consciousness’ (Veit 2022c). Whether this view captures the ordinary folk concepts of health and pathology is not important to my project. These concepts are typically confused, vague, and indeterminate, and when I talk about them my goal is to naturalize them by drawing on the best available biology (see also Veit 2021). When Jablonka and Ginsburg describe my chosen term ‘pathological complexity’ as inadequate they are not considering that our ordinary folk concepts can and ought to be revised in the light of science.

Another criticism by Jablonka and Ginsburg (2022) is that they find my discussion of externalist and internalist alternatives to the pathological complexity thesis unnecessary and needlessly long. This is strange given how important it is to my argument that there is a requirement to develop a dynamic alternative that recognizes evolutionary feedback between organism and environment. Indeed, I am puzzled by their argument that the distinction I rely on is a meaningless straw man, maintaining that there has not been an ‘internalist’ or ‘externalist’ in biology since the nineteenth century. This is akin to saying that the distinction in political science between left- and right-wing ideologies is a meaningless strawman because no real person in the 21st century only holds political views that fall exclusively in one category or the other. Just as most distinctions in biology allow for gradualist continua, without thereby being meaningless or useless, so is the distinction between internalist and externalist views meant to be seen as a continuum. Furthermore, the authors I reference, Godfrey-Smith and Keijzer, deliberately choose the terms ‘externalist’ and ‘internalist’ to describe their views; and the pathological complexity thesis is indeed intended as something of a bridge between these views. Discussing their views and the conflicts between idealizing away important internal or external factors is not a ‘distraction’, it’s the very rationale for developing a view that emphasizes dynamic feedback in the difficult trade-off situations organisms are placed in.

I am happy to accept that there aren’t ‘true externalist’ or ‘true internalists’ in the sense that they believe internal or external factors to not matter at all for cognition, but that is simply not how the distinction is typically used in these debates.<sup>2</sup>

---

<sup>2</sup> See Godfrey-Smith (1996) for a historical discussion of the dichotomy between externalism and internalism.

Now, Jablonka and Ginsburg may think that the environmental complexity thesis and skin brain thesis are so deficient due to idealizing away important internal or external features that it wouldn't even be worth discussing them, but this would only amount to a stronger case for the pathological complexity thesis, not an objection to it.

### **3 From Pathological Complexity to Consciousness**

Jablonka, Ginsburg, and Spurrett also offer several challenges to my proposed link between pathological complexity and consciousness that I will respond to here.

#### **Evaluation-first views of consciousness**

A core motivation of the pathological complexity thesis is to emphasize preferences, motivations, and desire-like states in understanding the evolution of consciousness. This emphasis on evaluative states is meant to replace the focus on sensory representations, in order to make sense of the very origins of consciousness. Yet, in criticizing the emphasis on sensory consciousness and self-awareness within the science of consciousness, Jablonka and Ginsburg are concerned that I may give off the mistaken impression that there are only a few who acknowledge the importance of evaluation in understanding the evolution of consciousness, since I only mention Cabanac. Indeed, there are plenty of important scientists that do acknowledge the central role of evaluation (see Romanes 1883; Damasio 1999; Panksepp 2005, 2011; Morsella 2005; Merker 2005, 2007; Humphrey 2011; Ginsburg and Jablonka 2019; Solms 2021). Panksepp, for instance, once argued that “affective experience may reflect a most primitive form of consciousness [...] which may have provided an evolutionary platform for the emergence of more complex layers of consciousness” (2005, p. 32). However, while I am very happy to agree that it would be the wrong takeaway from my discussion to think that no one has defended the centrality of evaluation, I disagree with their suggestion that I should have offered a comparative analysis of all the extant approaches to consciousness that emphasize it. While this might in itself be an interesting project, I do not take it as necessary to the one I am undertaking here.

Firstly, as they themselves acknowledge, the ideas and theories of these scientists are still very heterogeneous, having only partial overlap with mine. Secondly, it is precisely because of this heterogeneity that a comparative analysis of extant views would require its own paper. Thirdly, out of all the extant views in this evaluative literature it is precisely because I see my theory as inspired by Cabanac's (or the older Benthamite idea of utility-maximizing organisms) that I emphasize his work as the closest to my own. While evaluation, preferences, desire-like states,

emotions, and action prioritizations are naturally important in the work of all these authors, they do not put as much emphasis as Cabanac did on the importance of *tradeoffs* in decision-making and the idea of a common currency. Likewise, there is little point in comparing all the theories of consciousness that emphasize sensation or self-awareness, without also clustering them into further families of theories with closer family resemblance. There are too many differences between these views, and more than Jablonka and Ginsburg acknowledge.

Most notably, I do not argue that “once an evaluation system evolved, sentience could take off” (Jablonka and Ginsburg 2022). While some of the authors above seem to endorse such a simplistic view about the relation of evaluation and sentience, I embrace a more complex picture, with plenty of unconscious evaluative processes going on. Nevertheless, I do discuss elsewhere the broader idea of linking consciousness to evaluation to highlight similarities and dissimilarities with other authors: in another paper in this journal, that was intended to motivate the pathological complexity thesis (Veit 2022d), and in my forthcoming book (Veit forthcoming). Like Solms (2021), for instance, I share the view that evaluation can make sense of why conscious states are *felt* at all. Yet, while many of these authors have similar views on the function of sentience, in the sense of conscious states involving evaluation, they do not all argue - as I do - that evaluative states are the minimal precursors of consciousness and only later became enriched to form conscious sensory representations and conscious self-awareness. As I shall shortly argue, for instance, I do not share the view of Jablonka and Ginsburg that consciousness must also involve other phenomenological states such as sensory consciousness, episodic memory, and self-awareness. Combination views that require hedonic evaluation as a part of conscious experience and other dimensions of consciousness as preconditions for hedonic feelings need to be firmly distinguished from those that see hedonic evaluation as entirely sufficient on its own.

### **Why invest in consciousness?**

Spurrett articulates the following challenge to the pathological complexity thesis: since the problem of pathological complexity involves organisms making trade-off decisions among a large set of possible actions (in addition to many possible bodily states) in order to optimize their fitness, this makes it in principle no different from problems that can be solved with a variety of different forms of unconscious reinforcement learning. Spurrett (2022) notes that this point doesn't necessarily defeat my argument that the function of consciousness is to deal with pathological complexity, but it provides a challenge to the idea that organisms with high pathological complexity can't overcome this problem with “cognitive (or computational) solutions that don't involve consciousness”. Furthermore, as Spurrett



has argued in a recent paper (Spurrett 2020), the neuroeconomic problem of efficient action selection could in principle be solved by implementing a sub-personal ranking of alternative actions that track fitness at least to a first approximation and thus could instantiate non-conscious preference orderings. So there is an open question here regarding what sentience adds to dealing with pathological complexity, that these other solutions can't. I appreciate this call by Spurrett to further expand on my view, as I will do in what follows.

Notably, I do not deny that the general problem of dealing with pathological complexity can be dealt with through non-conscious means. As I emphasized in the previous section, pathological complexity is faced by all living systems - whether microbial or multi-cellular - and the diversity of life history strategies we find in nature make it obvious that complexity can be dealt in different ways. One way to deal with pathological complexity, for instance, would be to invest in an adaptive immune system or to produce a protective shell. My argument was not that consciousness is a unique response to pathological complexity, but rather that sentience becomes worth having due to a computational explosion in pathological complexity once organisms gain greater degrees of freedom and behavioural flexibility. This, of course, is only a partial deflation of Spurrett's concerns as we may still think that basic reinforcement learning also evolved in multicellular organisms precisely to deal with the problems of efficient action control due to these factors causing an explosion in pathological complexity.

Relatedly, Spurrett raises the excellent point that there is a nearby problem about how much credit we should give to consciousness in the picture I have offered. That is, even if we accept that hedonic valence helps organisms to select fitness-maximizing actions by being compelled to pursue what feels the best (or least bad), there appears to be a lot of background work going on to produce what Spurrett (2022) describes as a "simple hedonic 'executive summary' that doesn't overwhelm the selection stage". Spurrett is correct in assuming that I do not think that this 'behind the scenes' work is done consciously. As I argue in the target article, for neuroeconomic reasons it would be overwhelming to have a conscious bottleneck at which all the information about pathological complexity trade-offs is being presented - especially when we consider the minimal kinds of consciousness at the origins of sentient *Benthamite creatures*. Indeed, it is the simplicity of the first hedonic sparks of experience that help us to bridge the explanatory gap and deny that the experience of consciousness must confirm the Cartesian intuition our own human experience tempts us towards, that conscious thought is the main player within cognitive processing. So I agree with Spurrett that what is perhaps the most impressive evolutionary accomplishment here is the design of a system in which the various dimensions of pathological complexity are being turned into hedonic feelings, rather than the role these conscious experiences play for animals and perhaps even us. As Dennett (1991, 2017, 2018) has long argued, what Chalmers (1995) describes as the

‘easy problems’ of consciousness, i.e. how the neurological, cognitive, and functional background processes of consciousness work, may really be the more interesting and complex challenges all along.

Nevertheless, what is interesting is, of course, to a large extent a matter of personal taste and while I share Spurrett’s enthusiasm for trying to understand these neuro-cognitive processes and problems of action selection and control, the public and majority of philosophers are likely to continue to think of consciousness as the philosophically more interesting phenomenon - even if the supposed hard problem turns out to be the actual easy problem: an executive summary of a lot of unconscious cognitive processing going on in the background that allows organisms to deal with their complex world in a fast and simplified manner. Consciousness thus has important roles to play, though I would agree that its importance is often overstated, with unconscious processes making up a majority of what goes on in the brain (as well as the nervous system across the body, which is doing a lot of filtering before remaining information even reaches the brain). There are multiple explanatory projects here, each interesting in their own ways; it is not necessary to choose between them.

Nevertheless, if I admit that consciousness is not as important as typically assumed and the “consciousness support team looks likely to be the real heroes of the story” (Spurrett 2022), the question arises as to why this final trade-off calculation couldn’t also happen in an unconscious manner? In principle, I am also happy to grant that evolution could have come up with different solutions to the pathological complexity challenges we associate with complex and flexible animal life and that sentience must neither be a unique nor compulsory solution. When we look at the natural world and the great diversity of life-history strategies that can be found across the animal tree of life, it is clear that similar problems can be solved in very different ways. Natural selection is more creative than any human designer so I would not want to deny this possibility. Yet, I do maintain that sentience has both an efficiency rationale and is likely to be easier to achieve when it comes to the early evolution of distinctive animal lifestyles in the Cambrian, than other potential solutions such as a representational preference ranking. It may not be the only solution available, but it was perhaps the best one for the circumstances.

Here, it is worth responding to another objection to my view by Jablonka and Ginsburg, who question my defence of hedonic valence as existing prior to complex sensory representations. To answer this, first we have to distinguish unconscious from conscious sensory representations. To deny that the origins of consciousness involved conscious sensory representations is not a denial of the existence of unconscious representations. Second, there can be successful sensory-motor information processing for action selection without necessarily involving sensory representations. Unless we treat the term ‘representation’ in a very deflationary sense, which would undermine its usage in trying to understand conscious states as special

kinds of representations, there is simply no need to see them as necessary for the existence of simple hedonic summaries of subconscious processes of the nervous system. We need to distinguish the idea that it would be *useful* for the hedonic evaluative system to become enriched with sensory representations at the final common path, such as to allow conscious associations between hedonic feelings with some actions or environmental states to enable learning, from the idea that it is *necessary* for a simpler hedonic evaluative system to have these representations in order to be useful at all. I simply do not agree with the assertion that there could not exist a prioritizing value system without an organism also investing in representing the actions and perceptions to itself - that is an additional investment that may or may not be useful to invest in. On my view, Benthamite creatures have, to borrow Dennett's (2017) slogan, *competence without comprehension*. Nevertheless, an enrichment of the representational richness of the basic hedonic system took likely place early during the Cambrian, to allow better forms of learning. Furthermore, we need to keep in mind that the plasticity/flexibility of organisms at the dawn of sentience would of course pale in comparison to that of organisms further along the evolutionary trajectory, that have benefited from further improvements of this capacity. I do not see why we should take perceptual complexity and rich memory capacities as a condition for rather than an outcome of such enrichments, that further helped to mark off the distinctive animal way of being that both Ginsburg and Jablonka (2019) and I are interested in. This is why I do not include the coevolution of sensory systems, memory systems, and learning capacities as part of my explanation. I take it that they are later features that significantly transformed consciousness but did not give rise to it.

As I see it, even the most primitive forms of sentience constituted a useful final bottleneck for dealing with competing impulses from different parts of the nervous system that require centralized processing to allow for fast and 'cheap' action selection without relying on proxies such as signal strength. As much work in AI, robotics, and cybernetics has shown, we are still unable to build robots even remotely close to solving the complexity of action selection problems real living systems have to deal with in their ordinary lives (Zhang and Mo 2021). Simulations and experiments relying on reinforcement learning still typically only deal with a very small number of variables, low degrees of freedom, and a small behavioural option space, thus giving off the impression that we are closer to understanding how living systems achieve efficient action control than we really are. Worse, computer simulations typically leave the mechanisms of learning as a black box, so that it is entirely unclear how a real biological system would implement such learning mechanistically.

Likewise, while ordinal preference orderings might seem to constitute less demanding 'system requirements' (in virtue of not assigning values to how much one action or state is to be preferred over another), it is not at all clear how a real biological

system can represent these relationships, given that organisms are constantly faced with trade-offs that instead require cardinal preference orderings (that include such comparative values) for assessing how much one action is to be preferred over another, e.g. sleeping vs. drinking. Trading this off requires a neutral indifference point that corresponds at least roughly to neutral fitness. In the real world, after all, actions are not discrete states as we might find them in a game-theoretic model. They require fine-tuning and for this a hedonic common currency of evaluation is at least one fast and efficient way that natural selection has come up with for animals to deal with the complexity of their distinctive lifestyles. Revealed preference orderings should be seen as an *outcome* of these affective processes, not a mere background ranking of actions in subpersonal states that have to be translated into affective states.

To build a general artificial intelligence/robot capable of dealing in a fast and efficient manner with the pathological complexity faced by animal life, it is not implausible that they would require something at least akin to sentience, in the sense of a hedonic common currency, which could in turn be updated with various forms of learning. As Moravec (1988) recognized early on “it is comparatively easy to make computers exhibit adult level performance on intelligence tests or playing checkers, and difficult or impossible to give them the skills of a one-year-old when it comes to perception and mobility” (p. 15). What is now often described as *Moravec’s paradox* has remained a core problem in these fields and highlights a problem that biologists have surprisingly not recognized. Sensorimotor coordination is evolutionarily a much more important and a harder problem to solve than is the abstract reasoning much of animal intelligence research has been obsessed with. The ease with which animals are able to trade off the competing demands and values of actions and situations they find themselves with has given off the mistaken impression that there is no major problem to be solved here, but that couldn’t be more wrong. Rather, it is precisely one of the features that caused the Cambrian explosion! While much remains to be done to understand the functions of affect and valence in animals, it appears that the more we learn about this dimension the more important and integral it is seen to be for animal life.

### **The Ediacaran Extinction and Cambrian Explosion**

Like several recent authors, I have argued that the origins of consciousness are to be found in the Cambrian explosion (Ginsburg and Jablonka 2007, 2010, 2019; Trestman 2013; Feinberg and Mallatt 2016; Godfrey-Smith 2016). Jablonka and Ginsburg, however, are critical of my suggestion that the failure to evolve a hedonic common currency led to the mass extinction of complex animal life during the Ediacaran Extinction. They argue that because Ediacaran animals survived for 33 million years, it’s impossible that their evaluation system was unable to cope with their bodies. This rests on a simple misunderstanding of my argument. I did not argue

that there is only a single cause for the Ediacaran extinction, only that the lack of a hedonic common currency contributed to the extinction of complex pre-Cambrian animals, so I am very happy to agree with them that neither the Ediacaran extinction nor the Cambrian explosion can be attributed to a single factor.

As I see it, the Ediacaran extinction was driven by a combination of multiple factors. The evolution of mobile complex animals in the pre-Cambrian was limited and largely restricted to ‘grazing organisms’ enabled by microbial mats that covered the seafloor and offered ample resources from which large and mobile animal bodies could benefit (see also Ginsburg and Jablonka 2019, p. 406). As this resource diminished over several million years, these animals did indeed no longer manage to survive and thrive as they once did. They had complex bodies but could effectively no longer ‘pay’ for them. But because they lacked an evaluative common currency, they were also unable to explore more complex designs in the design space, that could have provided solutions to their problem of being stuck in a lifestyle for which the ecological resource was ever more depleting.

As Spurrett (2022) recognizes, the evolution of an evaluative system enabled the “space of pathologically complex designs to be explored without sacrificing viability”. Now, Jablonka and Ginsburg (2022) themselves point out that Ediacaran animals simply did not require complex action prioritizing systems, due to the poor sensorimotor capacities and limited cognitive processing that can be found in Cambrian animals. And it is precisely because of this that I have argued that these animals were driven to extinction, against Cambrian organisms that evolved sentience as a means to handle the pathological complexity increases coming from higher degrees of freedom and a greater behavioural option space. The Avalon explosion was tied to non-sustainable ecological conditions, whereas the Cambrian explosion gave rise to animals that could flexibly respond to new challenges in virtue of possessing a hedonic common currency. While later refinements of this capacity, such as the kind of unlimited associative learning Jablonka and Ginsburg are interested in, may have sped up these evolutionary dynamics, I do not see the rationale for thinking that UAL constitutes the basis of consciousness nor that it is what initiated the Cambrian explosion. Both events can be attributed to an older evolutionary innovation that gave rise to Benthamite creatures capable of feeling pleasure and pain (in the broad sense of positive and negative valence).

### **The Pathological Complexity Thesis vs UAL**

Finally, Jablonka and Ginsburg (2022) respond to my criticism of their unlimited associative learning (UAL) framework. At this stage it is worth pointing out that I see plenty of agreement between our views, and that their search for minimal conditions and evolutionary origins of consciousness in the Cambrian has strongly influenced my own. I have even written a very positive essay-length review of their book on the

evolution of consciousness (see Ginsburg and Jablonka 2019). Most importantly perhaps, we agree that consciousness evolved in the Cambrian and that an evaluative system is necessary for consciousness. But whereas I argue that a complex evaluative system is sufficient for minimal consciousness, they believe that it is only one of several interacting capacities that need to come together for consciousness to emerge.

In my target article, I described their approach as being based on a model of human consciousness, since their list of capacities is based on properties that are seen as necessary for human consciousness. Jablonka and Ginsburg (2022) take issue with this description since their list is based on “studies of both human and animal behavior, learning and affect, which we have surveyed for over a decade with the aim of uncovering the most basic features of minimal subjective experiencing (which is why language, theory of mind, and other fancy metacognitive capacities *do not* appear in our list of characterizing minimal capacities)”. They claim that that everyone engaged in trying to understand animal consciousness is ultimately seeking a framework that is not based on the human case. With this, however, I strongly disagree. Work on animal consciousness has arisen from and has so far remained highly contingent on theories and tests for human consciousness (Browning and Birch 2022). Even in work that deliberately tries to understand the most minimal kinds of consciousness, there will be an inevitable bias towards thinking of these kinds of experiences as human-like. This is why I described their UAL approach as too demanding for a theory of minimal consciousness and instead described it as constituting a more likely contender for the understanding of a major transition in consciousness towards becoming recognizably more human-like. So I agree that my approach is too narrow to fully understand human-like consciousness, but that is simply not my primary goal here, which is why I ignore later-evolving features such as episodic memory and the refinement of this basic hedonic capacity for special affective forms of learning. I don’t deny that they are important for consciousness in almost all extant animals, but I simply do not buy into the idea that these factors must play a role in understanding of the first sparks of experience back in evolutionary history.

Asserting that consciousness is a system property - like life - rather than a functional capacity seems to me to do very little explanatory work. Seeing consciousness as simply equivalent to the operations (or summaries) of a hedonic evaluative system makes functional sense of why some states are felt and others are not, without making reference to richer capacities that are present in animals of fairly similar complexity to us. While Ginsburg and Jablonka (2019) often talk about consciousness as being equivalent to the system requirements of UAL, they also often hedge their position by claiming that they are only providing a positive *transition marker* towards definite conscious organisms, and in doing so they acknowledge that there may be sentient creatures that do not satisfy the conditions of UAL. Yet, in granting this much it seems that they are already making room for minimal sentience

to have weaker foundations than the system conditions for UAL. For an evolutionary bottom-up approach, this possibility should of course be taken very seriously.

To clarify this further, we can draw on a distinction by Birch (2020), who distinguishes between theory-heavy, theory-light, and theory-neutral approaches when it comes to animal consciousness. Whereas a theory-heavy approach attempts to first figure out how human consciousness works and then simply applies those models to non-human cases, the theory-light approach is meant to only look at experiments from the human case that provide us with evidence regarding what it is that consciousness *facilitates*. While the framework by Ginsburg and Jablonka is not human-centric in the sense of being theory-heavy, it is nevertheless a human-centric theory-light approach by arguing that consciousness facilitates unlimited associative learning. My approach in contrast is meant to be theory-neutral in the sense that we try to move away from the human case entirely – to treat it as a special case of a more general and diverse phenomenon that we can find in nature. This is not a search for what Jablonka and Ginsburg describe as a search for a “single Archimedean point”, but an attempt to develop an evolutionary bottom-up approach that focuses on the life-histories of animals and functional benefits of consciousness. This is why my forthcoming book is titled *A Philosophy for the Science of Animal Consciousness*; it is an attempt to develop an approach that will enable us to remove humans from the center of reference in this science.

Lastly, Jablonka and Ginsburg (2022) challenge me to answer the ‘who problem’ of consciousness, i.e. provide answers on its distribution across the tree of life. Here, they raise an interesting challenge. Some animals, such as ctenophores and cnidarians, have Ediacaran origins and persisted through the Cambrian. However, despite lacking a centralized nervous system and the ability to engage in associative learning, they nevertheless possess an evaluative system. Since these animals are (presumably) not sentient in my framework, there needs to be a way of offering an explanation as to why they lack sentience, whereas other animals who also have an evaluative system are sentient. Their answer is that action prioritization in ctenophores and cnidarians is “always bottom-up and based mainly on relative signal strength and suddenness” and that their nervous systems do not allow for the complexities of open-ended associative learning, lacking both cognitive and memory complexity. But they did not find an answer to this problem in my proposal.

As I mentioned before, they mistake my view as one that claims that *any* organism with an evaluative system would be conscious. But I only argue that the presence of a common currency would imply - or at least strongly indicate - the presence of sentience. Which animals have such a common evaluative currency is still an open question, which is why I do not attempt to make confident estimates regarding the spread of sentience in these animals – to get some answers, we would have to study motivational trade-offs in taxa such as cnidarians. An example of this research program can be seen in recent work on bumblebees, showing that they are

able to trade off the rewards and costs of multidimensional problems against each other (Gibbons et al. 2022) and providing strong evidence for insect sentience (though as bees are at the upper end of cognitive complexity among the insects, further comparative studies will be necessary to make the wider inference). Nevertheless, as this research makes clear, my framework provides straight-forward tests with which to assess whether a species is likely to be sentient or not, which can be compared against pathological complexity measures of different species, thus making my framework in principle ‘falsifiable’. If we find species unable to engage in such trade-offs calculations but with higher pathological complexity than other species that can, this would provide a strong case against the pathological complexity thesis. Nematodes may well turn out to be sentient following such an investigation, but as I argue in my target article, comparative neuroeconomics has unfortunately remained a very small field, so again I can only reiterate that much work remains to be done to answer the distribution question of sentience.

## 4 Conclusion and Further Suggestions

To conclude, I would like to thank Jablonka, Ginsburg, and Spurrett for their engagement with the pathological complexity thesis. Their past work has left a mark on my own thinking and it comes with a special pleasure to engage with them in this productive exchange. I hope that the clarifications and extensions of arguments in my target article I have provided here will have removed any remaining ambiguities and help anyone seeking a deeper understanding of my thesis and framework. Much work, of course, remains to be done in developing my framework further, but there are two very promising areas that I suspect will lead to immediate progress.

Firstly, the measurement of the complexity of different life-history strategies will allow us to develop a better understanding of the pathological complexity challenges animals face in their natural lives, enable a comparative study of their life histories, and to better understand the evolution of biological complexity in general. Secondly, by studying how differences in the life-histories bear out in the subjective experience of animals, we will enable a much more empirically-guided research program into the functions and roles of consciousness. Research into the phenomenological complexity of different species will allow us to make testable predictions about their life histories and research into the life histories of different species will likewise allow us to make predictions regarding their subjective experiences (see also Veit 2022f). It is this core motivation of the pathological complexity thesis - to offer a useful and progressive research program - that I see as its greatest strength compared to other theories of consciousness that have a hard time making testable predictions, especially when it comes to non-human animals (see Browning and Veit 2020).



In a previous article in *Biological Theory*, for instance, I have shown how my pathological complexity framework can be used to think about the plausible subjective experiences of arthropods and gastropods (Veit 2022e) and my forthcoming book offers similar discussions of corvids, octopuses, fish, non-avian reptiles, and humans (Veit forthcoming). As an empirical research program, my thesis and framework will inevitably undergo further refinements and modifications, but given our current knowledge and evidence base, I remain convinced that the pathological complexity thesis currently offers us the best understanding of the place of consciousness in nature.

## **Acknowledgements**

My thanks go out to Heather Browning and Samir Okasha for their feedback on this article. I would like to thank the SalGo Team at the University of Oxford for enabling me to think about the pathological complexity thesis within the context of a biology department, and in particular Samuel Gascoigne and Rob Salguero-Gómez for helpful discussions. I'd also like to thank the Animal Sentience lab at the London School of Economics and the Representing Evolution reading group at the University of Bristol for helpful discussions, in addition to audiences at the LMU's Munich Center for Mathematical Philosophy, the University of Bayreuth, the University of Canterbury, the International Society for the History, Philosophy, and Social Studies of Biology, the International Network for Economic Method, and the Australasian Association of Philosophy.

## **Competing Interests**

None.

## **Funding Information**

This paper is part of a project that has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement number 101018533).

## **References**

Birch, J. (2020). The search for invertebrate consciousness. *Noûs*.

- Browning, H. & Birch, J. (2022). Animal sentience. *Philosophy Compass*. e12822: 1-14. <https://doi.org/10.1111/phc3.12822>
- Browning, H. & Veit, W. (2020). The Measurement Problem of Consciousness. *Philosophical Topics*. 48(1), 85-108. <https://doi.org/10.5840/philtopics20204815>
- Chalmers, D. J. (1995). Facing up to the problem of consciousness. *Journal of Consciousness Studies* 2(3), 200–219.
- Cisek, P. (2007). Cortical mechanisms of action selection: the affordance competition hypothesis. *Philosophical Transactions of the Royal Society B: Biological Sciences* 362(1485), 1585–1599.
- Damasio, A. R. (1999). *The Feeling Of What Happens: Body, Emotion and the Making of Consciousness*. New York: Harcourt Brace.
- Dennett, D. (1991). *Consciousness Explained*. New York: Little, Brown and Co.
- Dennett, D. C. (2017). *From Bacteria to Bach and Back: The Evolution of Minds*. New York: WW Norton & Company.
- Dennett, D. C. (2018). Facing up to the hard question of consciousness. *Philosophical Transactions of the Royal Society B: Biological Sciences* 373(1755), 20170342.
- Feinberg, T. and J. Mallatt (2016). *The Ancient Origins of Consciousness*. Cambridge, MA: MIT press.
- Gibbons, M., E. Versace, A. Crump, B. Baran, and L. Chittka (2022). Motivational trade-offs and modulation of nociception in bumblebees. *Proceedings of the National Academy of Sciences* 119(31), e2205821119.
- Ginsburg, S. and E. Jablonka (2007). The transition to experiencing: II. The evolution of associative learning based on feelings. *Biological Theory* 2(3), 231–243.
- Ginsburg, S. and E. Jablonka (2010). The evolution of associative learning: A factor in the Cambrian explosion. *Journal of theoretical biology* 266(1), 11–20.
- Ginsburg, S. and E. Jablonka (2019). *The Evolution of the Sensitive Soul: Learning and the Origins of Consciousness*. Cambridge, MA: MIT Press.
- Godfrey-Smith, P. (1996). *Complexity and the Function of Mind in Nature*. Cambridge: Cambridge University Press.
- Godfrey-Smith, P. (2016). Animal evolution and the origins of experience. In D. Livingstone Smith (Ed.), *How Biology Shapes Philosophy: New Foundations for Naturalism*, pp. 51–71. Cambridge: Cambridge University Press.
- Humphrey, N. (2011). *Soul Dust: The Magic of Consciousness*. Princeton, NJ: Princeton University Press.
- Jablonka, E. and S. Ginsburg (2022). Sentience as a System Property: Learning Complexity and the Evolution of Consciousness. *Biological Theory*. <https://doi.org/10.1007/s13752-022-00414-0>.
- Keijzer, F. (2015). Moving and sensing without input and output: early nervous systems and the origins of the animal sensorimotor organization. *Biology & Philosophy* 30(3), 311–331.

- Keijzer, F. and A. Arnellos (2017). The animal sensorimotor organization: a challenge for the environmental complexity thesis. *Biology & philosophy* 32(3), 421–441.
- Keijzer, F., M. Van Duijn, and P. Lyon (2013). What nervous systems do: early evolution, input–output, and the skin brain thesis. *Adaptive Behavior* 21(2), 67–85.
- Merker, B. (2005). The liabilities of mobility: A selection pressure for the transition to consciousness in animal evolution. *Consciousness and Cognition* 14(1), 89–114.
- Merker, B. (2007). Consciousness without a cerebral cortex: A challenge for neuroscience and medicine. *Behavioral and Brain Sciences* 30(1), 63–81.
- Moravec, H. (1988). *Mind children: The future of robot and human intelligence*. Harvard University Press.
- Morsella, E. (2005). The function of phenomenal states: supramodular interaction theory. *Psychological Review* 112(4), 1000.
- Panksepp, J. (2005). Affective consciousness: Core emotional feelings in animals and humans. *Consciousness and Cognition* 14(1), 30–80.
- Panksepp, J. (2011). Cross-species affective neuroscience decoding of the primal affective experiences of humans and related animals. *PLoS One* 6(9), e21236.
- Romanes, G. J. (1883). *Mental Evolution in Animals*. London: Kegan Paul, Trench, & Co.
- Solms, M. (2021). *The Hidden Spring: A Journey to the Source of Consciousness*. New York: WW Norton & Company.
- Spurrett, D. (2020). The descent of preferences. *The British Journal for the Philosophy of Science*.
- Spurrett, D. (2022). Complexity, Valence and Consciousness. *Biological Theory*. <https://doi.org/10.1007/s13752-022-00415-z>.
- Trestman, M. (2013). The Cambrian explosion and the origins of embodied cognition. *Biological Theory* 8(1), 80–92.
- Veit, W. (2021). Biological normativity: a new hope for naturalism? *Medicine, Health Care and Philosophy* 24, 291–301.
- Veit, W. (2022a). Complexity and the evolution of consciousness. *Biological Theory*. <https://doi.org/10.1007/s13752-022-00407-z>.
- Veit, W. (2022b). Consciousness, complexity, and evolution. *Behavioral and Brain Sciences* 45, e61.
- Veit, W. (2022c). *Health, Agency, and the Evolution of Consciousness*. Ph.D. thesis, University of Sydney. <https://hdl.handle.net/2123/29836>
- Veit, W. (2022d). The origins of consciousness or the war of the five dimensions. *Biological Theory*. <https://doi.org/10.1007/s13752-022-00408-y>
- Veit, W. (2022e). Towards a comparative study of animal consciousness. *Biological Theory*. <https://doi.org/10.1007/s13752-022-00409-x>
- Veit, W. (2022f). Integrating Evolution into the Study of Animal Sentience. *Animal Sentience* 32(30). <https://doi.org/10.1007/10.51291/2377-7478.1765>

- Veit, W. (2023). Health, consciousness, and the evolution of subjects. *Synthese* 201(1), 1–24.
- Veit, W. (forthcoming). *A Philosophy for the Science of Animal Consciousness*. Routledge.
- Veit, W. and H. Browning (2022). Pathological complexity and the evolution of sex differences. *Behavioral and Brain Sciences* 45, e149.
- Veit, W. and H. Browning (forthcoming). Hominin life history, pathological complexity, and the evolution of anxiety. *Behavioral and Brain Sciences*.
- Zhang, T. and H. Mo (2021). Reinforcement learning for robot research: A comprehensive review and open issues. *International Journal of Advanced Robotic Systems* 18(3), 17298814211007305.