

Aiden Karpf

May 2022

Consciousness, Evolution, and Epiphenomenalism:

Can an Evolutionary Account of Consciousness Effectively Refute Epiphenomenalism?

1. Introduction

1.1. The pursuit

Consciousness has mystified us for centuries—its ill-defined and unempirical nature makes it immensely difficult to examine and understand, and the ways that we understand it seem to change over time. When Dualist accounts of the mind were widespread, consciousness was deemed a property of a mind entirely distinct from the brain. In modern-day philosophy, however, many hold the mind and the brain to be one and the same—a view called mind-brain identity theory—which will be the fundamental assumption of this paper. If the mind is just the brain, then consciousness must be a neural phenomenon, a process that can be best understood biologically. And time has shown that biological processes are usually best understood in evolutionary terms, thus it is worth examining consciousness within an evolutionary framework.

1.2. The implications

Consciousness is vital to our experience of the world—in fact, it *is* the experience itself. Thus, understanding consciousness and its origins on a deeper level can help us to have a deeper understanding of the human experience and what makes us, us. However, beyond that, consciousness has additional massive implications for the age-old debate of free will—it *seems* to be the place of origin for our decisions, our home of control over our actions. As a result, it is

of immense importance that our conscious experience plays some causal role in producing our actions. That being said, over the years many philosophers have denied consciousness causal power—they favor a view called *epiphenomenalism*, where consciousness is said to have no causal effect on any physical events.

The beauty of evolution in the context of this discussion is that evolution is a selective process, favoring certain advantageous traits over others. Thus, if we can trace the actual phenomenon of consciousness to an evolutionary origin, it seems we can say that there is a purpose to consciousness, an actual evolutionary reason why it is better to be conscious than to not. In other words, it means that consciousness must be doing something—something causal. Uncovering this underlying evolutionary nature would combat epiphenomenalism and give more ground to free will proponents; that is, however, if we can show that the phenomenal conscious experience itself is what is being selected for, and not the underlying neural mechanisms that just happen to give rise to consciousness.

1.3. The method

To go about this investigation, I will analyze a scientific evolutionary model of the brain—Neural Darwinism—and evaluate its consistency with this idea of the evolution of consciousness. I will then shift to some more philosophical perspectives on the topic. A major aim of this paper is to evaluate the coherency of the idea (of consciousness as an adaptive trait) itself, but also to evaluate how the scientific and philosophical perspectives interact, how they talk to each other (or how they don't), and which discipline should be emphasized in continued pursuit of the broader question of consciousness's causality.

2. Scientific Models

2.1. Neural Darwinism and its consistency with consciousness

Neural Darwinism is a "large-scale selectionist theory of brain development and function that has been hypothesized to relate to consciousness" (Seth and Baars 140). Developed by Gerald Edelman in the late twentieth century, this theory outlines an evolutionary basis for neural mechanisms and proposes the idea that brain dynamics are characterized by diversity, amplification, selection, and degeneracy of neurons. If we redirect the ideas from this theory to see if they are consistent with widely recognized properties of consciousness—namely the phenomenal properties—then we can provide support for the idea of consciousness as an evolved trait. This is precisely what Anil K. Seth and Bernard J. Baars attempt to do in their paper titled *Neural Darwinism and consciousness*, and this will provide background for our investigation.

2.1.a. Neural Darwinism and subjectivity

Firstly, let's examine Neural Darwinism's (*hereafter ND*) consistency with subjectiveness itself, the experience of a private flow of events only available to the individual. To investigate subjective experience in a purely scientific, objective way initially seems quite impossible, and is referred to by many as the hard problem of consciousness. In response to this, Edelman, the original developer of ND, said the following: "A scientific theory cannot presume to replicate the experience that it describes or explains... to experience the discriminations of an individual, it is necessary to be that individual" (Edelman, 2003). Seth and Baars nicely restate this point, saying that coming up with a theory that solves the hard problem is like "developing a theory of hurricanes that is actually windy" (Seth and Baars 159). Nevertheless, we must attempt to explain this subjectivity. A question arises: how come there is a feeling in humans associated with the perception of the color red, but (presumably) not a feeling in cameras when

discriminating between red and other colors. The answer, Seth and Baars say, is likely because of the complexity involved—not the complexity of the discrimination itself, but the neural complexity involved in the discrimination (ND says we use extremely numerous, intricate sets of functional neural groups in our thalamocortical region to represent a discrimination like this). Another key component of the subjective experience is its privacy—and this seems to be explained by the correlation of consciousness to an individual brain; however, how can we explain the experience of a stream of consciousness? This seems to be tied to our experience of the self, so perhaps we should examine this as its own issue.

2.1.b. Neural Darwinism, the self, and the problem

It seems that the experience of the self can be largely correlated with several cross-region interactions—namely the connection of the posterior cortex with regions like the orbitofrontal cortex, limbic system, and brainstem (Baars, 2002). This interaction is described in many works of ND as between sensory input and value-based memory. Additionally, it seems this higher-functioning experience of the self is not limited to humans; though it has been posed to have rapidly developed (perhaps to a greater extent than in other animals) in humans in response to the connecting of linguistic areas of the brain to those tasked with concept generation. From this view, it seems that the sensation of the self may naturally arise when different areas of the brain are connected. Furthermore, it seems that consciousness may work in a largely similar fashion—arising when neural regions associated with it are connected in a specific way. However, though ND may be consistent with these correlates, does it really tell us anything? What if the neural mechanism underlying consciousness evolved but the experience itself wasn't selected for, it was just a byproduct of those underlying mechanisms that were selected for? Even given the mind and brain as identical, can we really say that both the conscious experience and the underlying

neural mechanisms are in the evolutionary feedback loop, and not just the neural mechanisms?
Could those neural functional groups be arranged in this way without giving rise to consciousness?

2.1.c. Qualia and their neural basis: A property or identity?

Seth and Baars provide some answers to these questions. They write, "Qualia are entailed by this distinctive form of neural activity in the same way that the structure of hemoglobin entails a certain spectroscopic response: One is not caused by the other, rather, one is an inevitable *property* of the other. This concept implies that neural systems underlying consciousness were selected in evolution to carry out discriminations in a high dimensional space of possible inputs, yielding adaptive advantage" (Seth and Baars 159, my emphasis). However, in their example, though the color red of hemoglobin is an inevitable property of its structure from a physics standpoint, this color was not part of the evolutionary selection process. It is not a trait with advantages, but instead the property of traits (the chemical structure, perhaps) with advantages. Thus, it seems that in this metaphor, consciousness was not in itself selected for, but was rather a *property* of the neural mechanisms that were selected for. This perspective, in the context of consciousness and philosophy of mind, is called property dualism and holds that, though both consist of physical matter, mental properties are distinct from physical properties. If property dualism proves to be true, this could spell an end to our discussion as consciousness would be distinct from its neural mechanisms (not in its material but in its properties) and thus not a part of the evolutionary feedback loop. With this possibility in mind, let's continue on and examine how Seth and Baars conclude this discussion: "According to ND, the neural systems underlying consciousness arose to enable high-order discriminations in a multidimensional space of signals, and *conscious qualia are those discriminations*" (Seth and

Baars 164, my emphasis). It appears that Seth and Baars may have contradicted themselves, calling qualia both a property of and identical to neural discriminations—and this seems to elucidate the deeper question we have been working towards: are qualia *identical* to those discriminations, or are qualia *properties* of those neural discriminations? This analysis has provided us with a working question from which we can continue our pursuit of consciousness's evolutionary trajectory and causal or acausal nature.

3. Philosophical Responses

3.1. Etiological epiphenomenal

Given consciousness as a biological trait, or a neural trait, it may seem intuitive that it (or at least its underlying neural mechanism) evolved to increase fitness. However, this is not necessarily the case. Take, for example, the human chin—a biological feature that does not serve an adaptive purpose, but rather happened to evolve in coordination with other features of the head that did have an adaptive purpose. This is an example of what Gould and Vrba call a "spandrel"—a trait without an adaptive, or etiological, function—and consciousness may be another example (Gould and Vrba, 1982).

Dreams, one type of consciousness[†], provide a good example of something that is likely a spandrel. The neurochemical brain activity that goes on during sleep clearly does have an adaptive function—it helps with cell repair, memory, hormone adjustment, *etc.*,—however the experience of the dream itself seems to be a semi-random result of brain activity that is serving these other purposes. Dreams, as described above, are an example of what is called an etiological

[†]It is important to note here that though one type of conscious state may be causal or epiphenomenal, we can't generalize that conclusion to all of consciousness.

epiphenomenon—a feature with no adaptive function. Thus, the claim that consciousness has no evolutionary purpose and does not enhance fitness can be called *etiological epiphenomenalism* (Polger and Flanagan 25).

Etiological epiphenomenalism is the type of epiphenomenalism that deals with the evolutionary analysis of consciousness we have been diving into. It is, however, by no means synonymous with general epiphenomenalism in the broad causal sense, and a trait can still be causal even if it is etiologically epiphenomenal. This is important to note in our discussion because it means that even if we show that consciousness likely serves no evolutionary function, there is still room for it to be causal. Take again dreams, for example; even though dreams may not have evolved for any adaptive purpose, they still could promote self-knowledge, or cause us to talk about them in conversation the next day, affecting our actions. Tom Polger and Owen Flanagan further support this point with their chapter in *Consciousness Evolving*. Here, Polger and Flanagan write about a dead tree with branches that no longer perform their etiological functions: "The tree is dead. I keep it because I can climb it in autumn and clear the gutters. The kids love it because of the woodpeckers. When they go to college and my knees give out, I'll hire a gutter cleaner who has a ladder. Then I'll cut the tree down and use it for firewood" (Polger and Flanagan 36). All of these are examples of non-etiological functions that still have a causal effect on the world. This idea is important because, as Polger and Flanagan put it, "even if consciousness in general were a spandrel, this would do nothing to defeat our beliefs about the nature and causal powers of consciousness" (Polger and Flanagan 37).

3.3. Susan Pockett's perspective

3.3.a. Consciousness and the self

Many people today hold the view that an individual's self can be identified with their consciousness, and that the self can cause behavior, thus consciousness can cause behavior. However, upon closer inspection, it appears that we cannot tie together consciousness and the self this closely, and thus the self does not necessarily cause behavior, providing evidence that consciousness may be epiphenomenal.

Susan Pockett makes this case in her paper *Does Consciousness Cause Behavior?* through many different pieces of evidence. One of which she focuses on is the idea of action occurring without a sense of agency or volition—for example, if certain parts of the brain are damaged, one may not have this sense of agency or volition even when they are acting (Pockett 25). Another good example of this is the phenomenon of ego-dissolution that occurs on high doses of psychedelic drugs—one can completely lose their sense of self but not their ability to act.

On the other hand, one can have a sense of agency or volition and still not have willed an action. This is shown in patients with paralysis, where subjects attempt to make movements that they are unable to make, creating a reported sensation of volition in the subject without an actual observable effect. Other experiments (Wegner and Wheatley, 1999) have caused subjects to mistakenly believe that they willed a certain action when they in fact hadn't. These examples show that our sense of self and behavior aren't necessarily as correlated as we may have thought.

This serves as an interesting side note to how our discussion has proceeded thus far because it attacks the question of epiphenomenalism from a different angle, and ultimately

comes to a different conclusion than the scientific model; it gets at the folk psychological perspectives that may underly our motivation to investigate the evolutionary perspectives in the first place. It is important to remember that not only are we investigating the models, but also the significance of the question itself, and Pockett's discussion digs into our question's basis.

3.3.b. Epiphenomenalism and science

At the end of her paper, Pockett also discusses methods to evaluate the question of consciousness's causality. She writes, "I fear that the answer to the question 'is the position that consciousness is an epiphenomenon testable empirically?' may well be 'no, it isn't'" (Pockett 38), and, as a result, she goes on to say, "so this is still just philosophy" (Pockett 38). Pockett's conclusion seems to contradict Seth's and Baars's method of investigation on the topic in their paper on Neural Determinism. They write: "Our strategy will be to emphasize empirical results and to deliberately sidestep philosophical debates" (Seth and Baars 141). So, the question is raised: can the causality of consciousness even be tested empirically at all? Should philosophy or science take the lead on this question?

4. Conclusions

4.1. Neuroscience vs philosophy

Throughout this investigation, we have taken a look at both neuroscientific and philosophical perspectives on epiphenomenalism and the evolution of consciousness, and naturally, they seem to contradict each other—namely in the way each believes we should go about investigating the question. Seth and Baars, as well as others from the realm of neuroscience, intentionally attempt to "empiricalize" the investigation of consciousness, throwing away all the philosophy involved and trying to answer the question more quantitatively

and concretely. However, in doing so, they lose out on some of the important considerations of the question and sidestep issues such as the hard problem of consciousness and property dualism; that being said, it seems that they are fully aware of what they are doing. They realize the philosophy behind consciousness can get messy and unproductive, and so they decided to attack the question in a more head-on fashion. The philosophical perspectives, like Polger, Flanagan, and Pockett, decide to take a more holistic approach to the question of epiphenomenalism, however, in doing so they lose out on the more nuanced, empirical, testable stuff. Pockett believes that this broader philosophical take is the only way to investigate the topic, as the question is just inherently unempirical. Polger and Flanagan bring up the point that an evolutionary investigation of consciousness may be completely irrelevant to epiphenomenalism. Both the scientific and philosophical approaches have their merits, and their disadvantages, and each discipline appears to be investigating the same question—but they are investigating different layers of the same question, and both the philosophers and the neuroscientists seem content with the layer that each is investigating. So, while the science and philosophy are both talking, they appear to be talking past each other, as each is investigating different sides of the same coin—and it seems that, by the nature of the question, each discipline is somewhat forced to stick to their side of the coin.

While both approaches have made some progress, the question may just be too convoluted to really attack scientifically for the time being. It seems that epiphenomenalism may be in too abstract a realm for empirical investigation, and thus, for now, we should let the philosophers take the lead—the questions of the hard problem of consciousness and property dualism are just too deep to ignore. Though Seth and Baars seem to have a coherent account of

consciousness scientifically, their investigation falls short in many regards simply by its nature, and the merit of the evolutionary lens itself should be brought into question.

4.2. Fruitfulness of the question

All in all, our analysis and pursuit of this question did not prove to be as fruitful as I originally thought—the explanatory gap in our answer seems to stem from the hard problem of consciousness, one of the most difficult questions in all of philosophy and science, and thus this investigation becomes extremely complicated and convoluted. It seems that without an answer to the hard problem, we cannot say that the conscious experience itself is a part of the evolutionary feedback loop, as opposed to being a byproduct, or *property* (see discussion in 2.1.c.) of the underlying neural mechanisms that are evolutionarily advantageous. Thus, even though we found a fairly coherent scientific model for the evolution of consciousness, it is too difficult to ensure that it is relating to the actual phenomenal experience of consciousness and not the underlying neural structures that are associated with the experience of consciousness.

On the other hand, even if we were to somehow definitively show that consciousness is *not* evolutionarily advantageous, and is not a part of the evolutionary feedback loop, that does not mean that consciousness cannot be causal, as Polger and Flanagan nicely point out (see 3.1.). This means that not only is answering this question absurdly difficult (as discussed in the paragraph above), but that one of the two possible answers (consciousness being etiologically epiphenomenal) gains us nothing definitive and does not allow us to say much about consciousness's actual causality as a whole. This provides less motivation for continued pursuit of the question and gives little merit for discussing epiphenomenalism evolutionarily at all.

4.3. A Revised Question

Through our investigation, we have determined that we may have "jumped the gun" with the question of consciousness's evolutionary origins—it appears that before we can use this lens effectively, more work must be done in the realm of the hard problem of consciousness and property dualism. It does seem that the underlying neural basis of consciousness likely evolved for some sort of etiological function (Seth and Baars), but the real question is about the relation between consciousness and its presumed neural basis. For consciousness to be an evolutionarily fitness-enhancing trait, consciousness and its underlying neural basis must be closer to each other in relation than red is to hemoglobin—they must be identical. So, the question we are left with, of which future analysis is needed, is this: is consciousness identical to the neural interactions that underlie it, or is it simply an emergent property of those neural interactions? If it is the former, it would be hard to argue against consciousness as an evolutionarily adaptive trait, and thus consciousness as having a causal role; if it is the latter, it would seem that consciousness would likely have no etiological function, existing outside of the Darwinian feedback loop—however, even if this is true, the causality of consciousness and its possible role in free will is not ruled out.

References

- Baars, B. J. (2002). The conscious access hypothesis: Origins and recent evidence. *Trends in Cognitive Science*, 6, 47– 52.
- Edelman, G. M. (2003). Naturalizing consciousness: a theoretical framework. *Proceedings of the National Academy of Sciences, USA*, 100, 5520–5524.
- Gould, S. J. and Vrba, E. 1982. Exaptation: A missing term in the science of form. *Paleo-biology*, 3: 115–151.
- Pockett, Susan (2004). Does consciousness cause behaviour? *Journal of Consciousness Studies* 11 (2):23-40.
- Polger, Tom, and Flanagan, Owen. *Consciousness Evolving*, edited by James H. Fetzer, John Benjamins 2002, pp. 21–41.
- Seth, Anil K., and Bernard J. Baars. “Neural Darwinism and Consciousness.” *Consciousness and Cognition*, vol. 14, no. 1, 2005, pp. 140–168., <https://doi.org/10.1016/j.concog.2004.08.008>.
- Wegner, D.M. and Wheatley, T. (1999), ‘Apparent mental causation: sources of the experience of will’, *American Psychologist*, 54, pp. 480–91.