Review of: "The Evolution of Consciousness Theories"

Dylan LaValley[1]

[1] University of Lethbridge

Potential competing interests: No potential competing interests to declare.

Farhadi's paper offers a descriptive review of several competing theories of consciousness that fall within the domain of the predominant neurocognitive paradigm. The function of the review seems to be to compare and contrast Farhadi's own theory of consciousness, dubbed trilogy theory, against shortcomings of existing theories. Farhadi suggests that—contrary to typical neurocognitive theories of consciousness—trilogy theory is superior in that it a) makes a clear distinction between the role of awareness and consciousness, b) incorporates volition, and c) accounts for the selective capacity of attention.

The story presented is internally valid; that is, the conclusions are not problematic, given that one accepts the foundational premises that exist in such neurocognitive discussions of consciousness. Given this, I would consider this paper to be useful to cognitivists in the following ways: first, as mentioned, the paper offers a quick description of several theories relevant to scholars of consciousness and attention; the collection of references reviewed by the author might guide readers to more thorough treatments of any given theory; and the claims made regarding the relative advantages of trilogy theory might spark discussion among consciousness scholars, which might, in turn, advance our understanding of the subject.

That said, for someone who does not subscribe to the neurocognitive framework used here (such as myself), the conclusions offered by the author might be viewed as vacuous. I will unpack this in three parts.

1. The first point to be made is that there are theories of consciousness that exist outside, or at least subvert, the neurocognitive framework on offer; and, given the theories reviewed, a more apt title for the paper might be *The Evolution of Neurocognitive Theories of Consciousness*. Obviously, I do not expect Farhadi to have formulated an exhaustive list of consciousness theories, as such an expectation would be completely unreasonable given the various constraints an author has to work within. So, here, I describe some of the alternatives. Given my own constraints, I will be *criminally* brief; and I strongly recommend readers engage with the cited material. Moreover, the alternatives below are not mutually exclusive and might be said to be branches of the same empiricist tree.

    1. Ecological conceptions of consciousness (e.g., Noë, 2009): Ecological approaches are among several that suggest (though, perhaps in not-so-many words) that neurocognitive conceptions of consciousness commit a category error (Ryle, 1949) in ascribing consciousness as a *thing* that an organism (or a machine or software) may or may not have; rather, consciousness might be better viewed as a process, analogous to something like dancing. Dancing is not a thing that an organism *has*; it is a behaviour that is distributed as part of a continuous interaction between the

organism and the environment.

   2. Behaviourist conceptions of consciousness (e.g., Skinner, 1974): Behaviourists like Skinner would agree (again, in not-so-many words) that attributions of consciousness commit the category error and, additionally, commit the nominal fallacy, and are possibly consequent affirming. We will unpack these fallacies when reviewing Farhadi's trilogy theory later on. Skinner's own terminology regarding 'conscious' and 'unconscious' processes roughly reduces to 'rule-governed behaviour' and 'contingency-shaped behaviour,' respectively. Where, given the distinction, an organism is still 'conscious' of the stimuli it responds to, in the empirical, qualitative sense of the term; however, this does not mean that it is 'conscious' of the fact that it is responding to the stimulus, in the reflective, rationalist conception of the term (e.g., LaValley, 2022).
   3. Eliminativism and Illusionism (e.g., Dennett, 1993; Frankish, 2016): Of the theories I discuss, I am least familiar with the variations of eliminativism; still, they are worth mentioning, as they too reject the premise that models of consciousness should necessarily match our intuition of the subject. Versions of eliminativism usually reject that constructs like phenomenal experience (or qualia) are exactly as they seem and may try to label them as illusory. Eliminativism is a materialist framework wherein a foundational premise is that things we think relate to consciousness will be wholly reducible to physical processes, eventually rendering the term consciousness superfluous. Of course, regardless of whether it's explicitly stated, this relates back to Skinner's work regarding the superfluousness of mental constructs in psychological explanation (e.g., 1938; 1974).

2. One of my very favourite components of Farhadi's paper is the clear rejection of the usual equivocation in consciousness studies between the terms 'consciousness' and 'awareness', going as far as to suggest that the 'hard problem of consciousness' should be re-named as 'the hard problem of awareness'. This is consistent with calls in consciousness studies to better distinguish between awareness and the awareness of one's own awareness (e.g., Frankish, 2016; LaValley, 2022), wherein one concept reflects something primary, sensory-driven, and empirical, whereas the other concept denotes something secondary, reflective, and rationalist. Consciousness scholars typically use the one word to bounce between both concepts, which seems to be no small source of confusion in the field. In my own work, I too suggest the hard problem of consciousness is actually a problem with the primary construct, not the secondary one; however, I had moved to re-name the problem 'the hard problem of perception' (LaValley, 2022, p. 531). Obviously, Farhadi and I associate different meanings with the word 'perception'; however, I sincerely believe we mean the same thing.

Still, I am not sure Farhadi goes far enough in the rejection of the typical deployment of conscious-related terminology. Discussions of consciousness typically treat the unconscious-conscious process distinction as one and the same as the subliminal-supraliminal distinction that is used in empirical tests of supposed consciousness. If an organism can differentially respond to subliminal and supraliminal stimulation, it is said to be evidence of unconscious and conscious processes, respectively; however, it is really only a test of differential responses to what does or does not exist for the organism. Responding to supraliminal stimulation need not mean the organism is conscious, in the reflective, rationalist conception of the term. This is to say, such tests do not parse unconscious processes from conscious processes; rather, they parse subliminal unconscious responses from the typical variety of unconscious responses (LaValley, 2022). As such, the term consciousness is not necessary for any description of the behaviours involved. Though, such tests usually make up the corpus of evidence on which the theories of consciousness described by Farhadi rest, despite the fact that consciousness is completely superfluous.

3. Lastly, we will briefly consider how invocations or models of consciousness, in the manner discussed within the typical neurocognitive paradigm, are explanatorily empty, in that they usually commit the nominal fallacy and are consequently affirming. We will start with an unrelated example and walk our way towards how this relates to invocations like trilogy theory.

Hypothetically, to try to account for why Megara is good at delaying gratification, whereas Marceline is not, a cognitivist might invoke a construct like executive function. If we assume that there is a thing called executive function that mediates the ability to delay gratification, then we can explain our organisms' different abilities by suggesting they have different levels of executive function. The problems with this are multifold. First, the evidence for the construct is the difference in behaviour that the construct is supposed to explain. That is, our explanatory framework is a circle. We have given a name to a difference in observable outcome, and we treat our named description as an end to enquiry. This is the nominal fallacy, and the consequence is that the actual problem just gets kicked down the road a little further. Now, to answer our original question—why Megara and Marceline have different abilities in delaying gratification—we need to answer auxiliary questions, like what *is* executive function, and *how* it mediates the delay of gratification?

Second, the invocation is consequent affirming. That is, there may be more than one explanation that fits the same description. Just because Megara and Marceline have different outcomes when it comes to delaying gratification doesn't mean the difference is a consequence of the assumed reason. For instance, the difference might be a consequence of more or less reliable schedules of reinforcement in their respective learning histories. If we can correlate our difference in outcome with verifiable antecedents, such as in this hypothetical, we render our mental construct superfluous.

Now let's consider these fallacies in relation to trilogy theory. For instance, Farhadi postulates that a key difference between artificial intelligence and the consciousness of naturally intelligent organisms is the ability to intentionally focus attention to create differences of awareness; given differences in awareness, we can make awareness-based choices, which Farhadi considers to be the function of free will in the decision-making process. First, by suggesting that organisms can intentionally focus attention, and algorithms cannot, *because* organisms have "discretionary selection for information for awareness," Farhadi provides a re-description of the original problem. We have merely invoked discretionary selection to account for differences in discretionary selection. The awareness-based choice selection module is similarly problematic within an explanatory framework. Moreover, in discussing the awareness-based choice selection mental function, Farhadi suggests a) decision making requires awareness as input and that b) this results in the emergence of true free will. The first premise seems to be falsified by the results of things like subliminal threshold studies and blindsight studies, and the conclusion does not necessarily follow, irrespective of whether the first premise is true.

A unique trait of invocations of consciousness, in contrast to the invocation of other mental constructs, is that we sometimes do not know what difference our invocations should be accounting for, anyway. Our invocations seem to be merely descriptions fit to match our intuitions about how these systems should work. That is, if we accept the neurocognitive premise that consciousness is a thing that an organism may or may not have, we should be able to answer, how do the behaviours of a conscious organism and a non-conscious organism differ? We might consider the concept of free will to be equally problematic with respect to the inability to assert what difference we should expect between the behaviour of a free organism and a determined one. This means that not only are we invoking superfluous constructs, but, in many cases, we seem to lack any reason to invoke them in the first place.

Moreover, injecting free will into the model is philosophically problematic. If awareness-based choice selection is an exercise of free will, as Farhadi suggests, we might put it to a Ryleian test (1949): were we free to choose our awareness-based choice? If we did not choose our choice, how free was it really? If we were free to choose our choice, did we get to choose to have the choice to have the choice, ad infinitum? If one tugs on either thread, they learn they both lead back to some version of a deterministic system. Moreover, if one wants to suggest that neither is true because we have discretionary selection of what we are aware of, then we have already presupposed intention in the system, and this same problem arises, only at a different module.

Of course, the problems I raise are not exclusive to the trilogy model, and if one wanted to, they could surely poke similar holes in the explanatory power of any neurocognitive invocation of consciousness. Moreover, Farhadi is explicit in stating that the discussed models do not afford prediction, calculation, or suggest a mechanism to account for how the model would function. There is an explicit awareness that the model is descriptive and is not supposed to be explanatory. This means that none of my arguments contradict the stated aim of Farhadi's paper. I raise the above arguments because I think they are important supplementary considerations in such discussions.

In sum, Farhadi has written a strong paper detailing a selection of neurocognitive theories of consciousness and has situated trilogy theory among them. I, however, find it hard to evaluate using the star rating system given here, as there are competing contingencies at play. I rather strongly oppose the neurocognitive framework (I would rate the framework ⅖ stars); and readers should be aware of some of the shortfalls of the framework and that alternative options exist, as I have tried to do in my brief review; however, regardless of my concerns about the soundness of the neurocognitive framework, the internal validity of Farhadi's paper is quite good! As mentioned, for those operating within the cognitivist framework, the paper should, at the very least, spark interesting discussion on the topic. *The only recommendations I can make are* 1) that the name of the paper is changed, as mentioned earlier, and b) that abbreviations are replaced with terms written in full so readers do not have to translate while reading abbreviations that are likely new to them (e.g., 'ABCS' remains 'awareness-based choice selection', etc.).

Sincerely, thanks for your contribution to the field, Farhadi, and thanks for the opportunity to review, Qeios.

Happy Holidays,

Dylan LaValley

University of Lethbridge

dylan.lavalley@uleth.ca

References

Dennett, D. C. (1993). *Consciousness explained*. Penguin UK.

Frankish, K. (2016). Illusionism as a theory of consciousness. *Journal of Consciousness Studies, 23*(11-12), 11-39.

LaValley, D. (2022). Equivocating on unconsciousness. *Theory & Psychology, 32*(4), 521-534.

Noë, A. (2009). *Out of our heads: Why you are not your brain, and other lessons from the*

biology of consciousness. Macmillan.

Ryle, G. (1949/2009). *The concept of mind.* Chicago University Press.

Skinner, B. F. (1938). *The behavior of organisms: An experimental analysis.* New York:

Appleton-Century-Crofts, Inc.

Skinner, B. F. (1974). *About behaviorism*. New York: Vintage.