Psychology Faculty Articles and Research   Psychology

11-29-2020

Response to Commentaries on 'Hard Criteria for Empirical Theories of Consciousness'

Adrian Doerig
EPFL – École polytechnique fédérale de Lausanne

Aaron Schurger
Chapman University, schurger@chapman.edu

Michael H. Herzog
EPFL – École polytechnique fédérale de Lausanne

Recommended Citation

Response to Commentaries on 'Hard Criteria for Empirical Theories of Consciousness'

Response to Commentaries on "Hard Criteria for Empirical Theories of Consciousness"

We are glad to see the many responses to our contribution, as our main motivation was to foster conversation on how to compare Theories of Consciousness (ToCs). The comments vary greatly but can be roughly assigned to two camps.

First, there are those who try to explain how their ToC can cope with the criteria we proposed (Hameroff, Grossberg, Graziano, Rosenthal, Michel & Lau, Baars, O'Regan, Naccache, Damasio). The large differences between these theories highlight once more the need for stringent criteria to compare them in a common framework. In this respect, we would like to reiterate that our analysis was primarily intended to highlight general criteria that we think ToCs should address. The criteria are not seen as pass/no-pass exams, but as challenges to be addressed, invitations to take sides, make assumptions explicit, derive predictions to pit ToCs against each other and either prune some or unify them. For this reason, we will not comment on the detailed responses defending particular ToCs but leave it to the readers and authors to make up their own minds.

Experts from the second camp commented on the validity of the criteria. Most liked at least some of them, some added further ones (Fleming) or proposed the need for a general theory *for* consciousness science (Seth & Hohwy). Klein proposed to create an online platform, which we think is a good idea. Others questioned certain criteria (Fahrenfhort & van Gaal, Pinto & Stein, Rosenthal, Haun & Tsuchiya). In general, many authors seem to agree that there is an unwieldy plethora of ToCs. In this respect, criterion 1 about the unconscious alternative seems to be widely accepted, (but see Grossberg who seems to suggest that any data on perceptual processes are data about consciousness, even if there is no unconscious alternative). We, and as it seems most commentators, think that demanding conscious and unconscious alternatives is crucial for consciousness science, unless good reasons are given not to.

The other criteria were more controversial. From some comments we learned that we did not always make our points sufficiently clear and this led to misunderstandings. For example, Fahrenfort & van Gaal interpret all our criteria, except criterion 1, as relying on functionalism. However, this is only true for the unfolding argument, if at all. The small and the large network argument and the other-systems argument do not rely on functionalism, but apply to all ToCs, including those that reject functionalism. Once more, these criteria are not intended to cull the herd of ToCs straightaway, but rather to encourage proponents to make explicit which stance they take, for example, if they embrace panpsychism, whether the proposed mechanism is necessary and/or sufficient and if only necessary, to what extent it contributes to consciousness (is it the key factor for consciousness or just one factor amongst many), etc.…

Another misunderstanding occurred with the other systems argument (see comment by Rosenthal). We did not mean to imply that a ToC must hand down a verdict on whether or not any given non-human creature is conscious. With this criterion, we aimed to encourage researchers to be explicit as to whether or not a given ToC applies to humans only. If so, it needs to be made clear why this is the case, avoiding conclusions based on mere associations. For example, the capacity for language may always seem to come with consciousness but this may simply be because language is extended in time. Simply alluding to a coincidence of consciousness and X is not sufficient. The other systems argument turned out to be quite

useful because it helped to contrast computational ToCs, with their specific challenges, with cognitive ToCs, with quite different challenges. To be clear, a theory that does not address the other systems argument might still be right about human consciousness, but that does put the theory in a different category from other theories that do specify which systems in general are conscious and which are not and why.

The most controversial criterion turned out to be the unfolding argument. The unfolding argument was strongly criticized for discarding "purely" 1st person data, which are claimed to contain information not available from 3rd person reports. We realized from discussions and the comments that this criterion is tightly linked to the question of whether consciousness research should be considered as a "normal" empirical science based on 3rd person data such as reports, or if we also need "pure" 1st person data (Pinto & Stein, Haun & Tsuchiya, Fahrenfort & van Gaal). In this respect the unfolding argument can be seen as splitting ToCs in two camps by showing that causal structure theories necessarily rely heavily on "purely" 1st person data that is not accessible from the 3rd person, unlike GWT or HOT for example. However, what are 1st person data if not reports? And reports, whether verbal or otherwise, are 3rd person data. One cannot collect 3rd person reports and then call them 1st person data. There is clearly a conceptual problem here, and not just a semantic one. We would argue that if you want "pure" 1st person data then you had better keep it to yourself, because as soon as you make a verbal report or press a button, these become 3rd person data. Future discussions are needed to clarify how causal structure theories can nevertheless rely on empirical results and avoid non-interactionist dualism. Theories focusing on 3rd person data need to show why they are about consciousness and not merely about the reports themselves.

There were other instances where we feel that our view was misunderstood or misrepresented. For example, Rosenthal stated that we "believe that scientific findings about consciousness are now rich enough to pare down the many current theories of consciousness." But we certainly do not hold this view, and in fact what we contended was that there are currently too few constraints. Regarding the need for a definition of consciousness, we fully agree that we should do the necessary conceptual work (Pinto & Stein, Rosenthal) and we also agree that we should aim for a definition (Rosenthal). We certainly do not advocate for a "shut up and measure" approach to consciousness research (Pinto & Stein). On the contrary, we put forward criteria to encourage people to consider what they are claiming and demand principled reasons for why X should be identified with consciousness. Even if you leave the hard problem aside, all ToCs face significant challenges. There were also misunderstandings about falsifiability (Seth & Hohwy). While we agree that testability is important, it is not sufficient. A theory can sometimes make correct predictions for the wrong reasons, and can continue to do so with impunity if not falsifiable. A theory must be falsifiable.

In summary and to reiterate, the primary goal of our contribution was not to knock down specifics ToCs but rather to make their assumptions and implications explicit and highlight their pros and cons in a structured manner. Our criteria might serve as a first step towards a theory *for* consciousness science, as proposed by Seth & Hohwy. Rather than committing to a specific ToC, we think it is more promising to spell out the various meta-alternatives and see if we can address the pros and cons common to sets of ToCs, which eventually will allow us to compare ToCs, unify them, and test them in a consistent theoretical framework that researchers can agree on.