**The Validation of Consciousness Meters: The Idiosyncratic and Intransitive Sequence of Conscious Levels**

Andrew James Latham1, 2, Cameron Ellis3, Lok-Chi Chan1 and David Braddon-Mitchell1

1Department of Philosophy, University of Sydney, Sydney, Sydney, Australia

2Brain and Mind Centre, University of Sydney, Sydney, Australia

3Department of Psychology, Princeton University, New Jersey, United States of America

**Keywords:** consciousness meter; levels of consciousness; consciousness across time; validation; integrated information theory; phi

**Correspondence:**

Andrew James Latham,

The University of Sydney,

Department of Philosophy,

School of Philosophical and Historical Inquiry,

Quadrangle A14,

Sydney, New South Wales 2006,

Email: Andrew.Latham@Sydney.edu.au

 For the purposes of this paper and some of the surrounding literature, we take ‘level of consciousness’ to mean something like the degree to which an individual is conscious. Thus a fully awake, engaged person has a high level of consciousness, a tired dazed person a moderate level, and a sleeping person a low level. Developing the tools required to determine (and to some degree operationalise) the level of consciousness of an individual would allow us to answer many important questions. It could, for instance, allow us to determine when a coma patient is aware of their surroundings, or to what extent an animal experiences pain. Thus we need an overarching theory to explain and predict the level of consciousness possessed by any organism (or system for that matter). Such a theory is sometimes called a ‘consciousness meter’, and a key example of such a 'meter’ is the perturbational complexity index (Casali et al., 2013), inspired by the integrated information theory of consciousness (for recent formulation see Oizumi, Albantakis, & Tononi, 2014). Integrated information theory suggests that consciousness exists in any system to the extent that the system integrates differentiated information. In this paper, we lay out a few interrelated issues for validating this theory (along with those like it), and the consciousness meters they inspire. First, validating any consciousness meter requires a consensus about the ordering of conscious states which is not easily achieved. This problem is particularly severe if we believe conscious states can be irreducibly smeared over time. Second, the relationship between conscious states is probably sometimes intransitive, which means levels of consciousness, will not be amenable to a single continuous measure. Third, even if we adopt a multidimensional approach to levels of consciousness in order to avoid intransitivity, we raise further problems for its validation.

 Integrated information theory (henceforth referred to as IIT) proposes phi (Φ) as a measure of the level of consciousness for any given system. While Φ is the focus of this paper, we believe the problems we raise here should generalize to validating any measure of consciousness. The value of Φ is calculated by determining the amount of information a system produces as a whole that cannot be explained by its constituent parts. Crucially, this is a single value. The exact calculation of Φ and more technical details of the IIT will not be covered here, for those details see Oizumi et al. (2014) for an overview. Whether measuring the level of integrated information measures the same thing as level of consciousness is an open question. Φ is a technical concept contained in IIT that is open to a number of philosophical interpretations. For example, it might stand for the amount of phenomenology or awareness an agent or system possesses. Alternatively, Φ might be better understood once it is deconstructed into multiple dimensions. For the purposes of this paper we set aside what is the correct interpretation of Φ, as validation of Φ as a consciousness meter faces interesting difficulties regardless of how we understand it.

This theory has been incredibly successful in making appropriate predictions of Φ for a range of different tasks and systems. Here, system simply refers to any set of elements or mechanisms that causally interact and systems can be more or less complex according to the number of elements they contain, level of integration between elements and so on. It is worth noting that for incredibly complex systems such as the human brain the calculation for Φ has not been performed[[1]](#footnote-2) and may prove to be impractical. Still, as we would predict, Φ values are lower in simple systems when compared to relatively more complex systems, and systems with higher Φ values can perform certain tasks better relative to systems with lower Φ values. Importantly though, complicated systems can still be unconscious, as Φ tracks the level of integration between parts, not just the gross number of parts or connections between them. These features are plausibly necessary conditions for validating Φ as a potential consciousness meter; a tool able to discriminate the extent to which any given system is conscious or not. We argue however that they are insufficient.

1. **Interpersonal variation**

 One major obstacle the validation of Φ faces is achieving any form of consensus in what systems ought to be considered more or less conscious. As we noted in the previous paragraph, Φ has been supported by showing higher Φ in systems that we would *intuitively* consider to be more conscious. However, in order to validate Φ successfully as a consciousness meter this needs to be true at all levels of comparison. That is, it is not sufficient to merely show intuitively correct orderings for levels of consciousness at the level of currently calculable systems; you must also show intuitively correct ordering for levels of consciousness judgments for systems such as ourselves. In addition, for these key important cases, reference can only be made to one type of datum – the report on conscious experience itself. This is an unusual level of reliance on a single source for support or confirmation (for example general relativity is supported both by converging with Newtonian physics at larger scales and by also accurately predicting the dilation of light in a gravitational field). Given this reliance on one source, we need very tight tolerances in using it, so a successful consciousness meter will require a strong consensus across reporting members of the population[[2]](#footnote-3).

In order to validate a consciousness meter there must, therefore, be a consistent correlation between the level of a consciousness produced by the meter and the order in which people rank those levels of consciousness. Of course it is not necessary for people to be able to accurately quantify the differences between any two levels of consciousness; instead, it is sufficient if they can come to consistently recognise one system or state as being higher or lower level than another. If comparisons between Φ’s and levels of consciousness consistently align, then we obtain warrant for attempts to extrapolate the levels of consciousness for systems or states that Φ can account for, but that fall outside human experience. Consciousness meters like Φ can only be validated by comparing levels of consciousness in humans and their Φ values, not through the comparison of simple systems or states that are currently calculable. However, even if the Φ values in humans could be calculated, a more pertinent problem arises. If some people report one state as being more conscious than another, while some other people report the opposite, then there is at present, no way of adjudicating who is correct. Supposedly, they both have access to the same kind of experiential subjective data but without some set of standardized criterion, there appears to be no way to reconcile the difference in their reported levels of consciousness.

Recently, Bayne, Hohwy and Owen (2016) have provided one example of such disagreement. Consider the level of consciousness associated with rapid eye movement (REM) sleep and light levels of sedation. According to the levels of consciousness approach there is a fact of the matter as to which one of these states is more or less conscious than the other. However, for Bayne and colleagues, there is no reason to suspect either is the case. Setting aside whether or not this is correct, it can be expected that when people are prompted to judge which of these gives rise to a higher level of consciousness, they will provide a diverse range of answers. In which case, there would be difficulty in validating the consciousness meter.

1. **Levels of consciousness across time**

 In line with the theme of this special issue consider what might be said about people’s levels of consciousness and Φ across time[[3]](#footnote-4). Consider the following simple example. There exists at a time, *t*1, some wakeful process *A* with an Φ of *X* which is known to underpin conscious state *A*\*. Similarly, there exists at a time, *t*2, some less-wakeful process *B* with an Φ of *Y* which is known to underpin conscious state *B*\*. Given the above it appears intuitive to say that *A* has a higher level of consciousness than *B*. In addition, so long as the Φ of *X* is greater than that of *Y* then our consciousness meter is performing successfully. However, validation of Φ becomes difficult if we believe conscious experience can be smeared across time. That is, it is possible for a single conscious experience C\* to depend on a sequence of distinct underlying states at different times, but remains unitary and irreducible to those underlying states (for example, such views are described by Dainton, 2008, Phillips 2010, and Husserl, 1893/1991). In this case it is not clear what work a consciousness meter can do to help us to make the correct judgment. Intuitively, we want to say that the answer to this question goes beyond simply locating the highest or lowest Φ values and knowing where to locate the target conscious experience.

Consider now some set of processes, *E, F*, and *G*, that exist respectively at three time points, *t1, t2,* and*t3*, and are known to underpin conscious state *D*\*. This means that there are three Φ values, the Φ of *E* at *t*1*,* the Φ *of F at t*2, and the Φof *G* at *t3*. So what is the correct Φ value of *D*\*? Is it the Φ of *E* at *t*1?The Φ *of F at t*2? The Φof *G* at *t3*? Or some totality of them? We might say that the correct Φ value of *D*\* is just whatever is the highest value, but this is just a stipulation motivated only by granting significance to the highest Φ value. For example, it is not obvious that a single conscious experience associated with many low Φ values and a single high Φ value should be considered to have a higher level of consciousness than a conscious experience with consistent, moderately high Φ values. Alternatively, perhaps *D*\* simply has three Φ values. But if *D*\* is a taken to be a single, unified and irreducible conscious experience then it should only have one Φ value. Otherwise, this would suggest the experience is reducible to its parts across time, counter our assumption. Finally, one could think that that the correct Φ value of *D*\* is some totality of the Φ values across *E, F,* and *G*. But this is still problematic. Suppose there is a conscious state *H\** that exists at a single time point and is associated with a high Φ value. Now consider a conscious experience *I\** that is smeared across numerous low Φ value time points but has a total Φ value the same as *H*\*. It is not obvious that *H*\* and *I*\* have the same level of consciousness. In sum, it is difficult to see how one could come to associate a single Φ value with irreducible conscious experiences smeared across time. As a result of this difficulty, it is unclear how a consciousness meter could be used to assess and compare such experiences if they exist.

1. **Intransitivity of levels of consciousness**

 Even assuming that levels of consciousness can be reliably compared between individuals, it is still possible that Φ remains unable to be validated because levels of consciousness do not map onto a single continuous measure. Having a single real number which is a measure of level of consciousness assumes that levels of consciousness exist on and along a single transitive dimension. For such a continuum to exist then it must be the case that all points along the continuum possess a transitive relationship with each other (Vinogradov & Hazewinkel, 2001). Without such a transitive relationship the ordering of quantities in a single dimension is not possible and thus any metric, including a consciousness meter, must order properties transitively. If an intransitive relationship exists in the intuitive ordering of conscious states a consciousness meter that encompasses all aspects of consciousness in a single dimension cannot be validated.

 Consider again the conscious states of REM sleep and light levels of sedation. We have already noted that people’s judgment for which state is the higher level of consciousness may produce a wide variety of responses. Now consider, how people may order these two states in addition to the conscious state of (some specified form of) meditation from highest to lowest level of consciousness. Some people may think that the conscious state of meditation is higher than light sedation and light sedation is higher than REM sleep, but that the conscious state of REM sleep is higher than meditation. Not everyone has to agree with the example presented here and some people may even judge that the loop proceeds in the opposite direction. What matters for our purposes is that people’s ordering judgments can be intransitive at all. If an intransitive loop is present then it may not be possible to validate Φ.

1. **Multidimensional approaches**

In order to settle disagreements regarding the ordering of conscious states and avoid intransitivity, it is natural for researchers to be attracted to multidimensional approaches to conscious states rather than a single dimensional approach (see Bayne, Hohwy, & Owen, 2016). To be more precise, the measure of consciousness like Φ is deconstructed into independent measures of multiple factors of consciousness, such as awareness, wakefulness, reasoning, action-selecting and so forth. Another possibility is to preserve the idea of Φ, but also to extract measures of independent factors out of Φ in order to be clear about the component factors in it. Limiting the comparisons between conscious states to the comparisons of some particular factors of consciousness would allow individuals to reach consensus in an easier way, and also to avoid the intransitive relationships that are caused by cases in which conscious states are having different factors that can be ordered in different ways. This will work only if the uniform source of the intransitivity’s was oscillation between different factors, but it is far from clear that this is the case. It is entirely possible that intransitivity has not been expunged by the approach and may occur in the deconstructed dimensions spelled out by the approach. Even more so with disagreement.

1. **Final remarks**

As an important aside the reader should note that this critique does not, nor does it seek to, extend to any of the excellent empirical research currently underway on IIT or related theories, but rather seeks question the theories by spelling out important possibilities that are in conflict with their ultimate goals. In addition, research that aims to identify the neural signatures associated with both conscious and unconscious states do not fall under the purview of this brief critique (e.g., Dehaene, 2014). Whilst certain boundary conditions remain unclear, such as in the case whether a presented stimuli is ‘actually unconscious’ as opposed to just ‘not reported in full’, gross comparisons between individuals who are comatose versus fully awake and lucid are unlikely to result in individual disagreement as to what state is more conscious. In addition, these comparisons are not susceptible to the problem of intransitivity as it represents only a binary comparison. If IIT restricted itself to say, trying to predict whether based on the quantity of Φ, whether or not a human is currently conscious or unconscious then this project may prove successful. However, as the project attempts to quantify consciousness in such a way that it maps direct onto levels of consciousness, it appears unlikely to succeed.

Consciousness meters, such as Φ, attempt to predict the conscious state of an individual using a single continuous metric. Unfortunately, at this time it appears not possible to validate such a measure of consciousness. Firstly, it is highly unlikely that conscious states can be ordered between individuals in a reliable fashion. This is particularly evident when we consider the possibility of ordering levels of consciousness that are smeared over time. Secondly, this ordering is highly likely to be intransitive in some instances. Any future attempts to make a consciousness meter will need to provide terms that can foster some form of consensus, even in the case where levels of consciousness is deconstructed into multiple dimensions. The benefits of developing a tool that can assess the level of consciousness in an agent that cannot report their consciousness are clear but more caution in theorising is needed to meet that end.

**References :**

Bayne. T, Hohwy, J. & Owen, A. M. (2016). Are There Levels of Consciousness? *Trends in Cognitive Sciences*. *20*(6), 405-413.

Braddon-Mitchell, D. (Manuscript). Immediacy and unitary supervenience across time.

Casali, A. G., Gosseries, O., Rosanova, M., Boly, M., Sarasso, S., Casali, K. R., et al. (2013). A Theoretically Based Index of Consciousness Independent of Sensory Processing and Behavior. *Science Translational Medicine*, 5, 198ra105.

Dainton, B (2008). The Experience of Time and Change. *Philosophy Compass*, *3*(4), 619-638.

Dehaene, S. (2014). *Consciousness and the brain: deciphering how the brain codes our thoughts*. New York, NY: Penguin.

Hurssell, E. (1893/1991). *On the Phenomenology of the Consciousness of Internal Time* (J. Brough, Trans.). Kluwer Academic Publ.

Oizumi, M., Albantakis, L., & Tononi, G. (2014). From the Phenomenology to the Mechanisms of Consciousness: Integrated Information Theory 3.0. *PLoS computational biology, 10*(5), e1003588.

Phillips, I. (2010). Perceiving temporal properties. *European Journal of Philosophy*, *18*(2), 176-202.

Vinogradov, I. M., & Hazewinkel, M. (2001). *Encyclopaedia of mathematics*: Kluwer Academic Publ.

1. Calculated exactly, it can be approximated. [↑](#footnote-ref-2)
2. This does not require consensus across *everyone’s* reports. After all, not everyone’s reports ought to be admitted or given equal weighting. Perhaps we decide only admit the reports of people found to possess some form of expertise in reporting their level of consciousness. Alternatively, we might decide to exclude all reports from people who persistently misreport their level of consciousness. For our purposes, it does not matter how this procedure is conceived as we believe there will inevitably be disagreement among the remaining reports. [↑](#footnote-ref-3)
3. The example presented here has been adapted from Braddon-Mitchell (MS). [↑](#footnote-ref-4)