Research Note (version 10 April 2018)

Solving Tye's 'philosophical problems of consciousness' (and some more)

J. H. van Hateren
University of Groningen, The Netherlands, j.h.van.hateren@rug.nl

Abstract A recently developed computational and neurobiological theory of phenomenal consciousness is applied to a series of persistent philosophical problems of consciousness (in recent formulations by Tye, Searle, and Chalmers). Each problem has a clear solution, as is briefly explained here.

1 Introduction

Philosophical analysis of consciousness has produced a rich literature of actual and potential problems of consciousness. Several major problems—in recent formulations by Tye 2017, Searle 2017, and Chalmers 2017—are analysed below from the perspective of the theory of consciousness proposed in van Hateren (2018). I will summarize this study in Section 2, but for fully understanding the analysis in Section 3, the reader should be familiar with the original study and its precursor (van Hateren 2015). Although the theory is conjectural and cannot be taken for granted yet, I will avoid the mays and mights that could litter the text below. Instead, I will write as if the theory is an established fact. However, I ask the reader to keep in mind that such is a stylistic choice rather than a sign of premature certainty.

A note on terminology. Below I will use the term 'intentional component', which is called X_i in van Hateren (2018). Similarly, 'fitness component' corresponds to F_i, and inverted intentional component (which is experienced) corresponds to $\bar{X}_i$. The estimate of an individual's fitness is denoted by x both here and in van Hateren (2018), but as f_{est} in van Hateren (2015), where the process X that produces x is denoted by the form of f_{est} (similarly as a mathematical function has a value and a form).

2 Summary of the theory of consciousness

A key feature of any biological organism is its evolutionary fitness, which is, in its simplest form, its propensity to survive and reproduce—'propensity' is used here to indicate that fitness is a forward-looking, predictive factor. However, fitness is often not so simple, because it can include helping related individuals (which is known as inclusive fitness), as well as social and cultural components. Organisms are likely to thrive when their fitness is high, and to wither when their fitness is low. Evolution by natural selection depends on variations in fitness.

Fitness acts continuously during the lifetime of any organism. Therefore, each organism typically strives to keep its fitness high during its lifetime, by various means. An extraordinary way to do this is when the organism implicitly utilizes an internal estimate of its own evolutionary fitness. Such an estimate turns out to be evolvable if it continually varies the internal structures of the organism in a special, random way. When fitness is low, structures should be changed with much variability ('desperate times call for desperate measures'), but little variation is required when fitness is high ('never change a winning team'). The key point here is that evolution produces estimation as a causal factor: the better the internal estimate of fitness is, the higher subsequent fitness ('fitness-to-be') will become.

Estimation does not exist as a causal factor in abiotic nature, thus it is a pure novelty. Moreover, this particular form of estimation can be shown to be an irreducible cause. This means that its causal efficacy cannot be explained by any set of micro-causes (essentially because its efficacy depends purely on modulated noise). As a result, estimation exists in a literal sense, as a distinct and irreducible entity. However, it is not well localized, because what is estimated, fitness, has widely scattered components.

A well-localized entity emerges when components of the fitness estimate (called 'intentional components' below) are prepared to be communicated to a related organism that is cooperatively

inclined. Then the inclusive fitness of the sender will increase, on average. Preparing to communicate an intentional component by a sender requires inverting it, such that it leads to a similar intentional component in the receiver (this depends on the fact that an operation followed by its inverse produces an identity operation). Inversion can be performed by the thalamocortical feedback loop in the mammalian brain, if it is used in a switched, dual-stage way. The first stage produces intentional components, whereas the second stage inverts them through feedback. Stages are switching continually, at a rate of roughly 10 Hz in the primate brain. Inverted intentional components are either communicated to a partner or are used internally as further input to the thalamocortical loop.

Inverted intentional components are causal factors that can be shown to be irreducible and spatially localized. Hence, they produce an entity that is autonomous, distinct, spatially localized to the brain, transient, and emergent. Thus, they are plausibly sensed as the feeling of consciousness. Their content equals that of the corresponding intentional components. The total content of consciousness depends on which inverted intentional components are active at any point in time. The unity of consciousness is produced by the fact that all intentional components get their causal efficacy from the causal efficacy of the overall estimate (x) of overall fitness, which are both scalars (and thus unitary).

3 Problems and Solutions

3.1 Ownership (Tye 2017, p.18)

Problem. Specific subjective experiences are necessarily owned by a specific individual. This makes them different from ordinary physical things. Such things are sometimes owned, but they still exist if they are not. In contrast, subjective experiences cannot exist in an unowned state. Thus, there is a problem if one assumes that phenomenal consciousness is wholly physical.

Solution. Although consciousness is produced by a physical system, it depends on an irreducible, emergent causal factor, estimation. Inversion of estimation produces a concrete entity (in the sense of being distinct and spatially localized) that is sensed, but which is *not* an ordinary physical thing. Hence, it is attached to the neurobiological system that produces it and it cannot exist in an unowned state.

3.2 Perspectival subjectivity (Tye 2017, p.19)

Problem. Phenomenal conscious states are perspectival, but physical states are not. Whereas the latter can be fully understood from a complete description of state and dynamics, the former can only be fully comprehended by having the proper experiential perspective (such as when having a pain, feeling a depression, and having the visual experience of red).

Solution. A conscious state (or, more appropriately, a specific conscious process) consists of components that are the inverse of specific intentional components. Thus, the content of conscious experience depends on the content of intentionality. The latter is a form of estimation produced by the brain. It concerns an estimate of components of the individual's own evolutionary fitness, produced by the individual itself. Hence, it has a subjective perspective. It is not an ordinary physical state, because it is an emergent entity.

3.3 Mechanism (Tye 2017, p.20)

Problem. What is the mechanism that produces the "what it's like" feeling? In the natural world, it seems that higher-level states or processes or properties are always grounded in—and are explained by—what is going on at lower neurophysiological or chemical or microphysical levels.

Solution. Ontological reduction may be applicable in the abiotic natural world, but not here. The key point here is evolutionary fitness, which confers causal efficacy (on fitness-to-be) to an internal fitness estimate made within the individual. Hence, estimation is a novel and emergent causal factor that has been added to abiotic nature (by being evolvable). The causal efficacy of the internal fitness estimate

depends on modulated randomness. The existence of causally efficacious randomness implies that the physical world is not completely causally closed. Moreover, describing nature completely in terms of the dynamics of instantaneous states is not correct (see Appendix D in van Hateren 2018).

3.4 Duplicates 1 (Tye 2017, p. 21)

Problem. A philosophical zombie is taken to be a perfect material duplicate of a conscious being, except that it completely lacks phenomenal consciousness. Otherwise it has identical behaviour and identical mental processes. Usually it is not claimed that such zombies are physically possible (given the features of the world that is), but rather that they are imaginable or logically possible or metaphysically possible. If they are, then consciousness seems separable from its material substrate.

Solution. Consciousness is not separable from its material substrate, and philosophical zombies are not possible—neither with 'possible' in the sense of feasible, nor in the sense of conceivable, nor in the sense of non-self-contradictory. If there is a clear explanation of how consciousness arises, as there is now, then one is not free any more to use one's imagination or logic or metaphysical assumptions in a way that conflicts with the explanation. That would amount to basing an argument on premises that are most likely false. Arguments based on philosophical zombies are now not sound any more.

3.5 Duplicates 2 (Tye 2017, p.22)

Problem. One might simulate the brain in arbitrarily fine detail in another system, such as might be realized by one billion carefully instructed people. Intuitively, one would think that such a system (as a whole) would not be conscious, even if it would perform a perfect simulation.

Solution. The main problem with this kind of simulation is that there can be no internal estimate of fitness (which is required for modulating random structural change in the system) because there is no fitness to estimate. One billion people do not survive and reproduce as a unitary entity (multiplying at once to two or three billion, or dying at once to zero). Moreover, there is no competition and cooperation with other such entities in a shared environment, nor a well-defined, unitary heredity; therefore, there is no evolution by natural selection. Without real estimation, there can be no emergent and distinct entity that is felt as consciousness.

3.6 The inverted spectrum (Tye 2017, p.23)

Problem. Suppose that Tom has a very peculiar visual system (perhaps produced by a neurosurgical rewiring at birth), such that he experiences red where others experience green, and vice versa. But nobody is aware of this difference, because otherwise Tom functions as anybody else. Thus, there is a phenomenal difference without a functional difference. More generally, one may suppose that such phenomenal inversion can occur even in microphysical duplicates.

Solution. Subjective experience is the sensed entity that is produced by inverting intentional components. Hence, the quality of the experience depends on the content of the corresponding intentional components. The content of an intentional component depends on the fitness component that it estimates. The colour red has approximately the same fitness associations in a group of culturally and functionally similar people (think of typical red things: strawberries, sunsets, fires, roses, traffic lights, socialism, blood, and so on). Therefore, their intentional components concerning red are roughly similar, and thus their phenomenal experience of red. The assumption that Tom is possible (in any sense of the term) is false.

3.7 Transparency (Tye 2017, p.24)

Problem. When attending to a visual experience, one becomes aware of what is seen (such as a particular object and its qualities), but not of the experience as such. Thus, phenomenal consciousness seems to be transparent. Why, then, is it felt?

Solution. It is felt because it equals a distinct, emergent, transient, and spatially localized entity (identical to the irreducible causes produced by inverting intentional components). The content of this entity is the content of the corresponding intentional components (pointing to a particular object and its qualities). Thus, the entity has no additional content, which may be interpreted, incorrectly, as transparency. The interpretation is incorrect, because entity and content are not separable.

3.8 Unity (Tye 2017, p.25)

Problem. There is a unity to conscious awareness. The different items that make up a specific conscious experience (e.g., the perceived objects, actions, and sensory impressions in a particular setting) are not experienced as fully separate. Rather, they are perceived as integrated in the whole. Similarly, conscious experiences stay integrated across time. How can that be?

Solution. Consciousness at any time consists of a large set of inverted intentional components. Their content corresponds to the content of the corresponding intentional components. Intentional components estimate fitness components (aspects of an individual's fitness) in such a way that together they produce a unitary (scalar) estimate of the individual's fitness, x. This estimate has irreducible causal efficacy, which is, ultimately, the reason why consciousness is felt. The intentional components (as well as their inverted versions) are automatically integrated by x. This is true at any point in time, but also across time, because X, the process that produces x, is maintained across time (even if it is nonstationary and changes slowly).

3.9 Divided consciousness (Tye 2017, p.27)

Problem. In split-brain patients the corpus callosum is cut (for medical reasons), which drastically reduces the communication between left and right half of the cortex. When conflicting information is presented to the left and right half of a patient's brain, perception seems to occur locally, not being communicated to the other half. Thus, perception is divided. Does such a patient, then, have a split consciousness too?

Solution. Consciousness is conjectured to be produced by the second stage of a dual use of the corticothalamic feedback loop (van Hateren 2018). This second stage inverts intentional components that are presumably produced by a wider loop involving thalamus, cortex, and basal ganglia, with important inputs from the upper brain stem. Together these establish x, the (distributed) estimate of an individual's fitness. Specific parts of the left or right cortex are then participating in specific intentional components, corresponding, for example, to specific perceptions. However, the unity of consciousness itself does not fully depend on the unity of left and right cortex. It also depends on the left-right unity of thalamus, basal ganglia, and upper brainstem (as these produce x too). Hence, there is no reason to assume that split-brain patients have a fully split consciousness. Moreover, they are still one individual with one fitness, thus they are likely to learn compensating strategies that repair the unity of their x, even if it were compromised initially. This may explain why split-brain patients still feel as one.

3.10 Animal consciousness (Tye 2017, p.28)

Problem. How can we decide which other creatures have consciousness?

Solution. Creatures need to have evolutionary fitness and need to make an internal estimate of that fitness (which stochastically drives structural changes in the creature). Moreover, they need to invert components of this estimate, in preparation of internal or external communication. The capability to communicate externally needs to be present in any case and must be directed at conspecifics in a cooperative setting. The latter ensures that inclusive fitness (fitness that includes helping related individuals) is increased. Then inverting estimated fitness components produces irreducible causes, which constitute the distinct, emergent entity that is felt as consciousness. In summary: in order to

have consciousness, creatures need to have evolutionary fitness, an internal fitness estimate, inversions of this estimate's components, and cooperative communications with at least some conspecifics. These conditions are necessary and sufficient. Moreover, they are amenable to empirical assessment. Note that the above assumes that the proposed mechanism is the only one capable of producing consciousness. This seems highly likely, but alternative mechanisms might exist.

3.11 Causal efficacy (Searle 2017, p.330)

Problem. One can initiate behaviour by a conscious decision. How is that possible if the brain is fully functioning through neural mechanisms?

Solution. Not all mechanisms are deterministic and reducible. Consciousness consists of sets of inverted intentional components that can be used as input to internal intentional components, which are subsequently inverted and then used as input once more, and so on. The totality of intentional components changes through time in this way, and thus the structure of the X process that produces x. The latter drives random structural changes in the brain, including ones that affect behavioural dispositions. Changes (and thus behavioural dispositions) that produce large x are sticky (because large x produces a low rate of structural change), whereas changes that produce small x are repellent (because small x produces a high rate of structural change). Which particular behavioural dispositions produce small or large x is determined by the structure of the X process, and is, thus, controlled by how consciousness proceeds. Hence, consciousness can affect which behavioural dispositions are sticky. Therefore, it can affect behavioural choices (albeit by a slow, stochastic process; instant decisions need to be prepared in advance, as stored dispositions that can be utilized nonconsciously).

3.12 Dancing qualia (Chalmers 2017, p.369)

Problem. Two functionally isomorphic systems must have the same sort of experiences. For example, a conscious biological organism may be gradually replaced, neuron by neuron and cell by cell, by silicon equivalents (this is utterly unrealistic[1], but let's suppose that it could be done). If one claims that the final, silicon version has different consciousness, or no consciousness at all, then there might be, at some point along the transition, a significant shift in experience. Moving back and forth across this point would produce dancing qualia (qualities of experience). This seems counterintuitive, thus functional isomorphism must imply equal subjective experience.

Solution. Replacing biology by silicon may not leave fitness intact, that is, the final silicon version may have lost the capability to reproduce and die. If that is so, the silicon version cannot make an internal fitness estimate (other than a fake one that would quickly fall short). Even if the silicon version had fitness (the propensity to survive and reproduce) it would not have inclusive fitness if it were the only one of its kind. Then inverting intentional components would not be sustainable (for lack of inclusive fitness), and neither would be consciousness. Assuming that fitness is indeed lost, the thought experiment would not show a sharp transition between the presence and absence of consciousness. Rather, the silicon version would gradually loose more and more of its consciousness when it senses that it is getting more and more alienated from its former conspecifics.

4 Conclusion

All problems discussed above have a clear solution. I would be happy to add—to future revisions of this note—any overlooked major problems that are brought to my attention.

[1] Biological cells work and communicate at a molecular level; it is difficult to see how this could be replaced by processes with a different material basis without having considerable consequences for fitness. What about the mass, energy requirements, and volume of the replacement? Heat dissipation? Noise? Structural changeability? Reproduction? Repairability?

References

Chalmers, D. (2017) Naturalistic dualism. In: Schneider, S., Velmans, M. (Eds.) The Blackwell Companion to Consciousness (Second Edition), pp. 363–373. Wiley-Blackwell, Chichester.

Searle, J. (2017) Biological naturalism. In: Schneider, S., Velmans, M. (Eds.) The Blackwell Companion to Consciousness (Second Edition), pp. 327–336. Wiley-Blackwell, Chichester.

Tye, M. (2017) Philosophical problems of consciousness. In: Schneider, S., Velmans, M. (Eds.) The Blackwell Companion to Consciousness (Second Edition), pp. 17–31. Wiley-Blackwell, Chichester.

van Hateren, J. H. (2015) The origin of agency, consciousness, and free will. Phenom. Cogn. Sci. 14:979-1000. Preprint at https://sites.google.com/site/jhvanhateren/home/agency/agency_web.pdf

van Hateren, J. H. (2018) A theory of consciousness: computation, algorithm, and neurobiological realization. Preprint at arXiv:1804.02952 [q-bio.NC]