



**HAL**  
open science

# Consciousness matters: phenomenal experience has functional value

Axel Cleeremans, Catherine Tallon-Baudry

► **To cite this version:**

Axel Cleeremans, Catherine Tallon-Baudry. Consciousness matters: phenomenal experience has functional value. *Neuroscience of Consciousness*, 2022, 2022 (1), pp.niac007. 10.1093/nc/niac007. hal-03874068

**HAL Id: hal-03874068**

**<https://hal.science/hal-03874068>**

Submitted on 27 Nov 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## Special Issue: Consciousness science and its theories

# Consciousness matters: phenomenal experience has functional value

Axel Cleeremans<sup>1,†,\*</sup> and Catherine Tallon-Baudry<sup>2,§</sup>

<sup>1</sup>Consciousness, Cognition & Computation Group, Center for Research in Cognition & Neuroscience, ULB Neuroscience Institute, Université libre de Bruxelles, Brussels, Belgium; <sup>2</sup>Cognitive and Computational Neuroscience Laboratory, Inserm, École Normale Supérieure—PSL University, Paris, France

<sup>†</sup>Axel Cleeremans, <http://orcid.org/0000-0002-9757-5235>

<sup>§</sup>Catherine Tallon-Baudry, <http://orcid.org/0000-0001-8480-5831>

\*Correspondence address. Center for Research in Cognition & Neurosciences, ULB Neuroscience Institute, Université Libre de Bruxelles, 50 Avenue F.-D. Roosevelt CP191, Bruxelles 1050, Belgium. Tel: +3226503296; Fax: +3226502609; E-mail: [axcleer@ulb.ac.be](mailto:axcleer@ulb.ac.be)

### Abstract

‘Why would we do anything at all if the doing was not doing something to us?’ In other words: What is consciousness good for? Here, reversing classical views, according to many of which subjective experience is a mere epiphenomenon that affords no functional advantage, we propose that subject-level experience—‘What it feels like’—is endowed with intrinsic value, and it is precisely the value agents associate with their experiences that explains why they do certain things and avoid others. Because experiences have value and guide behaviour, consciousness has a function. Under this hypothesis of ‘phenomenal worthiness’, we argue that it is only in virtue of the fact that conscious agents ‘experience’ things and ‘care’ about those experiences that they are ‘motivated’ to act in certain ways and that they ‘prefer’ some states of affairs vs. others. Overviewing how the concept of value has been approached in decision-making, emotion research and consciousness research, we argue that phenomenal consciousness has intrinsic value and conclude that if this is indeed the case, then it must have a function. Phenomenal experience might act as a mental currency of sorts, which not only endows conscious mental states with intrinsic value but also makes it possible for conscious agents to compare vastly different experiences in a common subject-centred space—a feature that readily explains the fact that consciousness is ‘unified’. The phenomenal worthiness hypothesis, in turn, makes the ‘hard problem’ of consciousness more tractable, since it can then be reduced to a problem about function.

**Keywords:** consciousness; function; phenomenal experience; hard problem; feeling

### What is the function of consciousness?

What is the function of consciousness, if any? This fundamental question has thus far been addressed mostly by philosophers (e.g. Block 1995; Chalmers 1995, 1996; Tye 1996; Seager 2001; Kriegel 2004). Surprisingly few things have been written about the functions of consciousness in the neuroscientific or psychological literature (but see Seth 2009), perhaps because it is such a thorny issue. As Seth (2009) points out, it might be that consciousness simply has no function whatsoever. This perspective is logically possible and amounts to defend either conscious ‘inessentialism’—the view that any intelligent activity that is carried out with consciousness can also be carried out without consciousness—or ‘epiphenomenalism’—the view that consciousness, while it exists, does not and cannot play any functional role. This line of thought introduces a quandary that is

at the core of the hard problem (Chalmers 1995): if phenomenal experience is but a mere epiphenomenon, a benign and causally inefficacious ‘user illusion’ (Dennett 2016, 2019; Frankish 2016), why is it there at all? Is it nothing but a mere accident of evolution—a spandrel (Robinson et al. 2015)? Conversely, if phenomenal experience itself has a function after all, what is it?

Addressing this quandary is the main goal of this article. We are convinced the time is right to do so, essentially because it seems obvious that phenomenal experience makes a difference. Indeed, why would we do anything at all if the doing was not doing something to us? Life would simply not be worth living unless we were experiencing it. Siewert (1998) asked whether one would accept zombification in exchange for a reward. Imagine, asks Siewert, that you can carry out life exactly as you do now, except that you cannot feel anything at all anymore. There would

Received: 1 May 2021; Revised: 4 February 2022; Accepted: 14 March 2022

© The Author(s) 2022. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License

(<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact [journals.permissions@oup.com](mailto:journals.permissions@oup.com)

not be anything it is like to be you anymore, yet you would remain functionally identical to who you are now (i.e. you get to keep all your non-phenomenal goods and skills, including the ability to report on your mental states). Would you accept to be zombified if offered a big enough reward? Most likely, you would refuse any such offer, except perhaps if by doing so you could save the life of a loved one, cure cancer or help address any other similar subjectively essential life-and-death situation. On the face of it, it seems rather absurd to assume that something that appears as important to us as our ability to experience things would serve no purpose at all and be a mere epiphenomenon. Most of what we do is in fact driven precisely by the experiences we seek to experience and by those we seek to avoid. In the following, we aim to leverage these powerful intuitions and to explore the possibility that phenomenal experience has a function.

In this respect, it is worth noting that modern characterizations of consciousness have often tended to strictly distinguish the putative ‘functions’ of consciousness from its ‘phenomenal’ aspects. This distinction is in fact precisely the distinction between access-consciousness and phenomenal consciousness introduced by Block (1995). Access-consciousness (A-consciousness) refers to the various ‘functions’ that consciousness enables, such as global availability, verbal report, reasoning and executive control. Phenomenal consciousness (P-consciousness), on the other hand, refers to the fact that we ‘experience’ conscious mental states: there is something it is like for us to appreciate the distinctive character of International Klein Blue or to enjoy the sweetness of a banana; there is something it is like for a bat to be chasing insects at dusk; there is nothing it is like for Alpha Go (Silver et al. 2017) to win against Lee Sedol. The distinction is also at the core of Chalmers’ ‘hard problem’ (Chalmers 1995). One can imagine, argue both Chalmers and Block, that the functions of consciousness can be wholly dissociated from its phenomenal aspects. Thus, one can imagine zombies who function exactly like we do but without experiencing anything at all. As Frankish (2021) writes, however, ‘a conceptual distinction does not necessarily mark an ontological one, of course: two distinct concepts may pick out the same thing’ (p. 53). Congruently, we should like to argue that the distinction between access-consciousness and phenomenal consciousness, as important as it has been in shaping consciousness research, is in fact misleading, precisely because it dissociates function from phenomenal experience.

Instead, we claim that phenomenal experience has a function because it has intrinsic value. And things that have value typically have a function and contribute to guiding behaviour. In the following, we explore these ideas, offering mostly discussion and (informed) speculation. We begin by discussing the surprisingly complex construct of value and review the body of theoretical and experimental arguments supporting our claim that subjective experience has intrinsic value.

## The intrinsic value of subjective experience

Consider the difference between being fed through an intravenous perfusion vs. eating a meal—even an unpalatable hospital meal. While both feeding methods serve the same physiological ‘function’, i.e., to restore appropriate blood nutrient levels, consciously experiencing the taste, smell and texture of food, as well as engaging in the act of eating, seems to be endowed with a distinctive value (Berthoud 2011). This contrast suggests that subjective experience has value in and of itself, i.e., over and above the value accrued from the caloric and nutrient contents of food. Let us consider an organism (presumably) devoid of consciousness, for

instance, a slug. The slug searches for food when hungry and uses its senses of smell and taste to detect food sources and to distinguish between and learn about edible and poisonous foods. Each food item is assigned a value corresponding to its caloric and nutrient contents, and this value is learnt or established by innate mechanisms shaped by evolution. Let us now equip this organism with the ‘subjective experience’ of food—typically, what is referred to as the ‘hedonic value’, i.e., whether the organism ‘likes’ the food or not. Now there is something if ‘feels’ like for the organism to eat this kind of food or a different one. The point is: this hedonic, subjective value can be quite different, and even wholly independent from the value associated with caloric and nutrient contents. In the light of this simple example, we propose that ‘all subjective experiences have intrinsic value’.

Before developing this central point, however, we need to consider the very notion of intrinsic value. As Lee (2019) puts it, ‘Something has intrinsic value just in case that thing is good for its own sake’. Zimmerman (2019) makes the point more precise: ‘whatever is intrinsically good is non derivatively good’, by which he means that some things, such as money, only have value because they make it possible to obtain other things that you desire. Thus, the value of money is not intrinsic because its value is only derivative with respect to something else (i.e. it has instrumental value). And indeed, what would be the point of having money at all if all you could do with it was to see numbers go up on your bank account? The same analysis applies in vastly different domains such as altruistic behaviour. For instance, as Zimmerman further discusses, you may consider that helping people who need help is good. If prompted to explain why you think doing so is good, you might say that it is good to see people’s needs be satisfied. If again asked why this is good, you might end up saying, as Zimmerman writes, ‘It just is good that people be pleased’, at which point we seem to be unable to continue this derivation and conclude that satisfying one’s needs is intrinsically good. We propose that subjective experience has, in this sense, intrinsic value. Coming back to food-related illustrations, the value of an ice cream for you is intrinsic in the sense that an ice cream is good because you enjoy the quality of the experience of consuming it, not because it restores lipid or glucose levels in your organism. The intrinsic value of subjective experience may be in line or at odds with instrumental, derivative value—you may enjoy a pistachio flavour ice cream, or dislike it, independently of whether your body’s metabolic state actually requires nutrients. The fact that the intrinsic value of subjective experience and the derivative value of extrinsic rewards may be independent of each other provides conscious beings with additional degrees of freedom.

We will elaborate somewhat on the dimension(s) of value in the following, but it is worth underlining right away that in our view, value can be either positive or negative. This important point has been debated in the philosophy of mind. Thus, in his analysis of the notion of value, Lee (2019) defends what he calls the ‘neutral view’, which he opposes to a ‘positive view’ instantiated by both Siewert’s (1998) and Seager’s (2001) arguments that consciousness is intrinsically valuable. First, says Lee, the positive view is too liberal, for it merely defines intrinsic value in terms of instances thereof: consciousness is valuable because there are experiences worth having. In addition, argues Lee, there are also instances of the ‘disvalue’ of consciousness—imagine a situation where the only experience you can have is that of constant suffering—would being conscious have any value then? Kriegel (2019) also addresses these complex philosophical issues and proposes the following thought experiment, modelled after Siewert’s: the choice you are now offered is not to be zombified (in exchange for a reward) or

not, but rather to be zombified or to continue to live a life that 'will skew constantly towards the unpleasant' (p. 10). Our intuition here is that most of us would choose the second option anyway: (moderately) unpleasant phenomenal experience is better than none whatsoever, which again suggests that experience has intrinsic value. Kriegel offers another interesting thought experiment based on Nozick's (1974) concept of an 'experience machine'. Imagine you can either enter the experience machine and continue to experience (hallucinate) life as you would were you not trapped in the machine, or to be zombified. As Kriegel puts it, it would be folly not to choose to get plugged in.

While we do not aim here to further explore the depths of the philosophical debates that involve the notion of value, it seems to us that these different arguments and intuitions are at least suggestive that phenomenal experience indeed has intrinsic value: barring exceptional circumstances, if given a choice, most of us would probably prefer to be a poor conscious agent than a rich zombie, i.e., to actually experience life vs. failing to experience anything at all. In other words, we would be quite willing to give up many non-phenomenal goods just to keep phenomenal experience, even in conditions where those phenomenal experiences might not always be experiences we want to have.

We now turn to the notion of subjective value in experimental sciences. Subjective value is a central construct in at least three different domains: perceptual consciousness, decision-making and emotion research. Nevertheless, paradigmatic phenomenal experiences, such as the taste of an apple or the warmth of an embrace, are described under different names and with emphasis on different aspects in all three domains. For thinkers interested in consciousness, such experiences constitute instances of conscious mental states characterized by their distinctive qualia, with emphasis on the subjective, modality-dependent aspects of experience rather than on its value, as discussed further below. For decision-making researchers, phenomenal experiences represent or indicate rewards to be obtained. However, while the notion of value is central to decision-making, its subjective aspects have been rather neglected in the extant literature. Finally, in affective research, phenomenal experience is discussed to the extent that emotional feelings are experienced: emotional feelings colour almost every mental state or action, with a valenced intensity reminiscent of value. Our main point here is to suggest that these different concepts are all but mere renderings of the same core assumption, namely that subjective experience has intrinsic value. In the following, we briefly overview how the notion of subjective value has been approached from these different experimental perspectives, beginning with decision-making, where the concept of value is central.

## Decision-making

In the field of decision-making, subjective value classically refers to a scalar quantity assigned to an item and that underlies choices and preference judgements: when faced with a choice, you choose whichever item has the highest value. Whether value is determined intrinsically by the participant (e.g. as when deciding whether to have red or white wine) or whether there is an objective consensus on what the best option may be (e.g. a reward of 10€ is more valuable than a reward of 1€) does not matter in most models of decision-making: value is a number on a common scale, with different values characterized by rank order and a distance metric (Levy and Glimcher 2012). Subjective value is considered subjective only inasmuch as it is subject-dependent, as in whether one prefers Bordeaux vs. Bourgogne. In other words, most of the relevant research has emphasized value in and of itself, leaving

its subjective aspect mostly unaddressed (but see Azzalini et al. 2021).

Importantly, there are ongoing debates about the very definition of value (Bartra et al. 2013), which is variously characterized as a 'common currency' scalar signal (Montague and Berns 2002; Chib et al. 2009; Levy and Glimcher 2012), as distance in a multi-dimensional quality space (O'Doherty 2014; Juechems and Summerfield 2019; Hayden and Niv 2021) or as 'whatever maximizes the evidence for the agent's model of a survivable world' (Nikolova et al. 2021) in the predictive processing literature. A recent trend in computational approaches to decision-making relates value more directly to the internal state and intrinsic goals of the subject of experience: the value of a reward is computed as the distance between the current state and the state the agent would find herself in if she obtained the reward (Keramati and Gutkin 2014; Juechems and Summerfield 2019).

However, despite massive progress in the development of computational models of decision-making, the specific computations that drive the reward function (that is, the function that associates each item relevant to a decision with value) remain largely unspecified (Juechems and Summerfield 2019). In some cases, an item might have an obvious value, as defined by evolution, to satisfy basic needs (Keramati and Gutkin 2014). But how does one compute the subjective value that one spontaneously assigns to the taste of wine, since taste, as such, does not satisfy a basic need and since the value associated with taste is distinct from the value associated with the caloric content of the wine? No rule of evolution specifies or even constrains this subjective taste value. Subjective values, and the resulting preferences, are shaped by learning, of course, but the central question is: why do we have subjective values at all? In the case of wine, we can easily imagine two wines with the same caloric content, one we like and the other we do not. What is the purpose of having different subjective values for two items that objectively fulfil the same basic need?

Current debates about reinforcement learning algorithms illustrate the gap between the rewards associated with a predefined value defined by the programmer (or by evolution in the case of living organisms) and the rewards associated with intrinsic value. Thus, all reinforcement learning algorithms are driven by 'rewards', but such rewards typically only have extrinsic value—they define what the system will do with respect to extrinsic constraints, but they have zero value 'for' the system (i.e. AlphaGO zero) itself. As Juechems and Summerfield (2019), asking where value comes from, point out, reinforcement learning algorithms embrace the 'reward hypothesis', i.e., they are based on the idea that the environment shapes the reward functions that drive learning. However, 'value must be inferred by the agent, not conferred by the world'—this they call the 'reward paradox'. Value is ontologically subjective but epistemically objective. Thus, an intravenous supply of water and nutrients has extrinsic value, and the caregiver knows that this nutrient supply is necessary to sustain his patient. However, a palatable meal has both an extrinsic, objective value, here related to caloric and nutrient content, and an intrinsic, subjective value related to its taste and smell that exists 'only because the meal is experienced'. Subjective experience would thus contribute to the computation of the reward function by creating the link between an item and its subjective, intrinsic value. In other words, the key for a reward to have 'intrinsic' value is precisely subjective experience, for in its absence, there is nobody for whom the reward means anything. How to design systems for which rewards have intrinsic value is a core challenge for future artificial intelligence research that we will not delve further into beyond pointing out that for

any system to ‘want’ anything of its own, it has to have a sense of its own existence as an intentional agent and the ability to flexibly learn about its own internal states and its possibilities of actions.

## Emotions

Emotions are an important drive for behaviour: one seeks pleasurable experiences and attempts to avoid unpleasant ones. The role of emotions in behaviour can be viewed as involving the same mechanisms as those involved in decision-making, where different emotions are associated with distinctive values that guide behaviour, leaving little room for subjective experience. Indeed, for some, emotions are only expressed through behaviour and can hence be objectively measured by the experimenter (Adolphs *et al.* 2019). Such emotional behaviours rely on survival circuits strongly constrained by evolution (Anderson and Adolphs 2014), generating a limited behavioural repertoire, such as freezing, aggression or avoidance (Livermore *et al.* 2021). However, emotions in humans are also typically associated with rich and intense subjective feelings (LeDoux and Hofmann 2018; Adolphs *et al.* 2019). Emotional feelings can be viewed as a particular instance of subjective, phenomenal experience—what it feels like to savour an excellent wine or to find pleasure—or maybe disgust—in an embrace, and so on. But what is the function of emotional feelings? Clearly, some behaviours convey an evolutionary advantage (do not be eaten, reproduce), but what is the function of joy, of remorse or of *schadenfreude*?

Whether and how emotional ‘feelings’ are related to the neural activity in survival circuits is still unclear, but there is at least some degree of independence between the two. Indeed, those two components of emotions can predict different outcomes. Let us consider the example of BASE-jumping: danger-avoidance should lead to never, ever BASE jump. On the other hand, the exhilarating feeling of flying might have a high subjective value, speaking in favour of the risky behaviour of BASE-jumping. In other words, risky behaviour can be accounted for, but only ‘if the subjective experience of risky behaviour has an intrinsic value’ that can override the value of the safe behaviour selected by evolution. Such behaviours, we argue, cannot be understood unless phenomenal states constitute their own rewards. In this sense, phenomenal states answer to a subjective rather than objective fitness function (Hesp *et al.* 2021).

While we are suggesting that an emotional subjective feeling is associated with a single value, so that it can be easily compared with something else, the characterization of subjective feelings in emotion research usually resorts to additional dimensions, from the 2D space of valence and arousal (Russell and Barrett 1999) to more complex descriptions (Cowan and Keltner 2017). How a single value is constructed from a constellation of features is an open issue, debated in both the emotion (Lindquist *et al.* 2016) and decision-making literature (Pelletier and Fellows 2021; Pessiglione and Daunizeau 2021).

## Conscious perception

In striking contrast to what is the case in the decision-making and emotion literature, the notion of value is mostly absent from the consciousness literature. Consciousness research has been largely focused on the phenomenal aspects—‘what it is like’ (Nagel 1974)—associated with certain perceptual states. Qualia refer to the specific characteristics of individual instances of conscious experience, or more prosaically, to ‘the way things seem to us’ (Dennett 1988). What is the point of seeing something consciously?

We propose that perceptual experience is always associated with an intrinsic subjective value. Intuitively, we feel that examining a painting or enjoying the song of birds in the forest has some worth—i.e., the very definition of value: the worth of something for someone. In the case of paintings, we are used to express our preferences—you might prefer a monochromatic painting by Yves Klein to one of the exquisitely detailed scenes of a Hieronymus Bosch painting, or the reverse. The argument we make here is that this applies to ‘all’ sensory conscious experiences: every act of perception—hearing, seeing, tasting or smelling something consciously—is automatically associated with some value, regardless of whether we actually consider value explicitly and regardless of whether the value itself is close to zero. This proposal echoes the ideas put forward by Barrett and colleagues (Barrett and Bar 2009; Lebrecht *et al.* 2012). In particular, Lebrecht (2012) introduced the notion of ‘micro-valence’, i.e., the idea that objects traditionally considered to be neutral in the affective neurosciences (i.e. teapots, umbrellas and telephones) nevertheless elicit automatic affective judgements. Thus, while the evidence thus far is scant, there is a small empirical literature suggesting that (conscious) perception is systematically evaluative, with object identification being automatically associated with the signalling of an object’s relevance for the perceiver, as well as its affective value.

Arguments in favour of an automatic assignment of value to arbitrary visual inputs can also be found in the experimental literature on preferences. Preferences can be measured by asking participants to rate the likeability of images, hence providing a direct measure of the subjective value associated with the stimulus, or to choose between items, allowing the experimenter to infer preference values. Importantly, the brain automatically assigns a preference value to consciously experienced visual stimuli such as faces, houses or paintings (Lebreton *et al.* 2009), even when the task (judging age) was unrelated to the preference value measured in a separate session. Such findings directly support our hypothesis, which predicts that value assignment is an intrinsic feature of consciousness and that it occurs automatically. ‘Automatic value assignment is precisely what endows conscious perceptual experience with a function.’ While the process of assigning a value to sensory inputs probably begins early in life and is fully automatized in adults, the specific value assigned to a given stimulus is likely to be modified by learning or habituation and to be dependent on context—the walk by the beach that you take every day might become less, or more, valuable with time, and be more, or less, pleasant depending on weather conditions.

Subjective value in conscious perception might be a scalar quantity, computed from a collection of features, as proposed in the field of decision-making (Pelletier and Fellows 2021; Pessiglione and Daunizeau 2021), or it might be also multi-dimensional. Despite all the work carried out on perceptual consciousness, the dimensions of conscious perceptual experience remain largely unknown and open to different conceptualizations (Bayne *et al.* 2016; Bayne and Carter 2018). One’s subjective experience of barely visible stimuli is often captured along a 4-point scale ranging from ‘nothing’ to ‘clear experience’ (Overgaard *et al.* 2006; Sandberg *et al.* 2010), suggesting a link with intensity but also with clarity. At the other end of the spectrum, some studies have attempted to offer a detailed analysis of subjective verbal reports (Petitmengin *et al.* 2019), taking advantage of recent advances in automated language processing (Bedi *et al.* 2015), but such methods have not yet reached a point where they can reliably be used in most standard experimental paradigms. Nevertheless, a notion semantically related to value—the richness of perception—has

been put forward in perceptual consciousness research. The richness of perception has initially been equated to the quantity of details that can be retrieved from an image. This view has been particularly prominent in the debate on change blindness and inattentive blindness (Simons and Levin 1997; Block 2011), where participants systematically fail to perceive otherwise prominent visual information such as a gorilla among basketball players (Simons and Chabris 1999; Simons and Ambinder 2005). People's failure to report on objectively available information in such cases stands in stark contrast with our subjective impression that visual information is rich, which has led some authors to conclude that the richness of conscious experience is illusory (Cohen and Dennett 2011). However, this conclusion only holds if one equates richness with the quantity of details that can be retrieved. A different account for the experienced richness of perception might instead be found in the impression conveyed at a glance by a visual scene (Campana and Tallon-Baudry 2013; Campana et al. 2016; Haun et al. 2017) based on the rapid extraction of summary statistics (Cohen et al. 2016; Whitney and Yamanashi Leib 2018). In any case, the notion of value we put forward here is not related to the richness brought about by details: a monochromatic blue painting by Yves Klein induces—for some people—a valuable perceptual experience despite failing to contain any detail.

In short, the notion of value, as it has been approached by the different lines of research we have briefly overviewed, is thus central in many aspects of our behaviour despite remaining surprisingly complex from a conceptual point of view. We note that similar debates around the notions of core notions value, valence and reward are unfolding in the decision-making, emotion and consciousness literature. By the same token, philosophers have thus far paid little attention to such issues, which prompted Carruthers (2018) to write: '... the correct characterization of the nature of valence becomes an urgent philosophical issue'. Strikingly, the very notion of value remains largely absent from the consciousness literature, which has thus far been mostly concerned (when it comes to describe phenomenology) with constructs such as richness or vividness (Haun et al. 2017). Despite these continuing debates, we think one thing is clear: Conscious experiences are intrinsically valuable, and conscious agents care about them in ways that non-conscious systems do not. Importantly, this does not imply that the computation of value necessarily requires awareness. For instance, food has obvious survival value for a bacteria, and the bacteria is congruently endowed with nutrient-driven (unconscious) mechanisms that drive its behaviour so as to optimize its fitness. But the bacteria, because it lacks consciousness, does not care about any of it: the bacteria is sensitive to the value of food, but remains unaware of it.

If conscious experience indeed has intrinsic value, as we suggest, then we have to consider the possibility that it has a function. We now turn to exploring this question.

### Subjective experience, once endowed with value, is functionally relevant

What is the functional advantage, if any, of having subjective experiences? Asking this question leads one directly to evolutionary considerations. While it would be too long to review the relevant literature here, we should like to emphasize a few important developments. First, Hoffman (2019) argued that perceptual systems need not capture reality 'as it is'—they need only to improve a creature's fitness, i.e., detect features of the world that are relevant to improving fitness. Thus, 'fitness trumps truth'.

At the evolutionary level, the point is simply to note that the sensory systems of different species are tailored to transduce those aspects of reality that optimally improve fitness. The evolution of sensory systems naturally leads to the emergence of inner representational spaces that are tailored to each species' ecological niche—von Uexküll's notion of 'Umwelt' (Uexküll 1957). This presumably also explains the origin of preferences, not only across species but also across individuals. In this respect, Voltaire (1962/1764), in his 'Philosophical Dictionary', wrote: 'Ask a toad what beauty is, the 'to kalon?' He will answer you that it is his toad wife with two great round eyes issuing from her little head, a wide, flat mouth, a yellow belly, a brown back.' Beauty is the eye of the beholder. What I take to be beautiful depends on preferences and biases that have accrued as a result of my experience with the world and with other people.

Second, in their hefty tome dedicated to the 'evolution of the sensitive soul', Ginsburg and Jablonka (2019), seeking to determine the biological basis of consciousness (and hence its putative functions), have argued that learning, and in particular what they call 'Universal Associative Learning' (UAL), constitutes a critical juncture in evolution. UAL is characterized by both generativity, i.e., flexibility in linking a combinatorially rich repertoire of learnable stimuli and actions, and reflectivity—an organism's ability to use the outcomes of previous learning as the basis for subsequent learning. According to Ginsburg and Jablonka, UAL leads to many functions relevant to consciousness, amongst which mention global activity and accessibility of information; binding and unification; selection, plasticity and attention; intentionality; temporal thickness; values, emotions, goals; embodiment, agency—in short, many of the features that contemporary theories (e.g. Baars 1988; Dehaene et al. 1998) associate with the functional aspects of consciousness. While shying away from articulating the difference that subjective experience in and of itself actually makes in the functional organization of an organism (see Cleeremans et al. 2020 for such an attempt), it is clear that Ginsburg and Jablonka consider that organisms so equipped are best characterized as 'experiencing subjects', the experiences of whom are no longer epiphenomenal but rather constitute the central feature that makes it possible to go beyond the pressures of natural selection. This is also, essentially, what Miller (2000) argued in his controversial book 'The Mating Mind', going as far as comparing our oversized brains with the peacock's tail: an elaborate, metabolically expensive system geared towards attracting potential sexual partners—in the case of humans, through intelligence, humour and artistic expression. It may thus be that there 'are' good evolutionary causes for the fact that we 'experience' things, but it seems impossible to go much beyond tantalizing speculations in this respect. UAL, as proposed by Ginsburg and Jablonka, does not appear to require subjective experience *per se*. As for Miller's argument, the peacock's tail is itself an illustration of how sexual selection can operate without appealing to phenomenal experience.

Be that as it may, assuming that phenomenal experience, in and of itself, has functional effects has a number of implications that we explore in the following.

### The value of subjective experience accounts for flexible, creative behaviour

Assigning value to subjective experience accounts for flexible, creative behaviour. As we have discussed, while subjective experience is connected to extrinsic rewards, some of which serve biological functions, we act mostly to find ourselves in rewarding

phenomenal states and to avoid finding ourselves in bad phenomenal states. For instance, our main ‘motivation’ to eat is to obtain the ‘reward’ of ‘feeling’ satiated, not to fulfil the biological function of nourishing our body, albeit the former undoubtedly serves the latter. In this sense, phenomenal states answer to a subjective rather than objective fitness function (Hesp et al. 2021). In many other cases, however, and crucially for this argument, we engage in behaviour that serves no biological function whatsoever and that may even be detrimental to our own existence. Thus, to return to our BASE-jumping example, people BASE jump because they enjoy the thrill of it so much that they are willing to take extreme risks to enjoy the experience itself. Strikingly, animals have been observed engaging in similarly pointless activities that appear to be pursued for their own sake. Thus, conscious agents care about their experiences in ways that non-conscious systems do not. Simple organisms can act in adaptive ways installed by evolution; algorithms can act in appropriate ways programmed by design and so on. However, the behavioural repertoire of conscious agents is much more extended. Humans create and design novel experiences, from playing music to sledging, from telling jokes to creating new cooking recipes. Further, experiences, precisely because they are non-fungible and cannot be accumulated, are now goods that can be traded, as reflected by the recent rise of ‘experiential purchasing’ and ‘experience tourism’ (Gilovich and Gallo 2019). Thus, ‘subjective experience gives value to what has no intrinsic value as defined by the rules of physics or biology’. An agent’s ability to enjoy—or to loathe—conscious experience affords considerable adaptive value—the agent can now act in ways that flexibly dodge the constraints set up by evolution as it gives it much greater freedom to act in ways it chooses for its own reasons, and even in ways that are clearly counter-adaptive, as when one commits suicide to defend an idea, or engage in otherwise self-harmful behaviour. Conscious agents act upon themselves, upon the world and upon other people because experiencing specific phenomenal states has intrinsic value.

While we argue that all phenomenal experiences have intrinsic value, it does not follow that all valuation necessarily conscious. Most of the reward-based decision-making or reinforcement learning literature successfully models behaviour using the notion of reward and does not need to refer to conscious experience. Nevertheless, in that case, the reward value is typically described from a third-person perspective and is constrained either by biological parameters (eat or die) or by the experimenter (an ‘ideal’ agent should maximize gain). Our proposal departs from such attempts because we hold that phenomenal experience itself has value, and that this value can be independent of the (potentially delayed) outcome of the behaviour associated with the phenomenal experience, as when enjoying the taste of junk food while knowing the long-term outcome of this behaviour is negative. More generally, conscious human beings often depart from ideal agents and often adopt suboptimal strategies—which echoes the concept of bounded rationality and ‘satisficing’ (Simon 1956) in economics. We argue that at least part of this suboptimal behaviour can be explained by the additional degree of freedom offered by the value of subjective experience.

### The value of subjective experience accounts for internal drive and motivation

The value of subjective experience is a powerful internal drive for behaviour. Imagine you can make money but not use it. Would you spend any effort making money just to see the numbers go up on your accounts, especially if those numbers remain unknown

to anyone but you<sup>1</sup>? Some would, but most would not, for, obviously—what would be the point? Value-driven algorithms, including reinforcement learning, work that way. The only reason they act at all is because they are programmed to maximize a reward function, but paradoxically, the reward itself is not rewarding ‘for’ them. This also holds for many biological organisms, who are programmed by evolution to survive, reproduce and otherwise achieve homeostasis. In other words, obtaining rewards is a powerful drive for behaviour. But what defines a reward, beyond basic biological needs? If conscious experience has value, then generating conscious experience is rewarding. Because the reward function is internally generated rather than imposed by biological constraints, conscious agents act for reasons of their own. ‘Seeking new conscious (hence valuable) experiences thus constitutes a drive to engage in behaviour that does not need to be justified by further reasons.’

Acting ‘just to see how it feels’, ‘for fun’—such as a child pricking a needle in her palm or a tourist trying food that smells terribly bad—does not fulfil any basic need. While such behaviours may use the same neural routes and share the same algorithms than behaviour driven by basic needs, the motivation—the internal reward function—is very different. Unconscious agents such as algorithms can act in certain ways defined by their structure, but they do not act for reasons of their own, simply because acting one way or another has no import for them. Biological systems can be driven by goals that need not be conscious—survival, reproduction and so on—but recognizing the value of subjective experience accounts for the drive of human beings to generate behaviour for reasons of their own. Conversely, zombies do not care about dying or about having an orgasm—since they lack conscious experience, nothing ever means anything to them and hence they completely lack the motivation to do anything at all.

Likewise, a central problem with Artificial Intelligence (AI) systems is their lack of internal drive and motivation. While the need to implement intrinsic motivation in artificial agents has long been recognized (Barto 2013; Juechems and Summerfield 2019), progress has been slow. Klyubin and colleagues’ concept of ‘empowerment’ (Klyubin et al. 2005, 2008), according to which agents—all other things being equal—should strive to ‘keep their options open’, i.e., to maximize the number of possible future outcomes resulting from their actions, comes close to formalize the core idea that information is intrinsically valuable. In the absence of phenomenal experience, however, it remains unclear how to quantify its value beyond basic exploration of a space of possibilities. Ultimately, the question of intrinsic motivation is connected to issues of life and death, i.e., to existence, and thus to life itself (Seth 2021).

Clinical cases further illustrate how conscious experience may act as a powerful drive to engage in behaviour. For instance, consider TN, a blind patient with bilateral damage to primary visual cortex. Despite being blind, TN could nevertheless successfully navigate around obstacles in a long corridor (de Gelder et al. 2008). Importantly, despite being able to perform well, and despite being told of his ability, TN, as any blind person, required guidance to navigate in the environment in everyday life. In other words, the lack of subjective experience deprived TN of the willingness to engage in navigation. Conversely, more intense subjective perceptual experiences, such as described in alterations of visual perception induced by psilocybin, probably contribute to subjects reporting that ‘everything looks new, fascinating and more

<sup>1</sup> There is a striking computer game where the player does just that (<https://www.decisionproblem.com/paperclips/index2.html>)

intense', even if those altered perceptions, often characterized by moving geometric shapes, have neither apparent meaning nor function (Bayne and Carter 2018; Preller and Vollenweider 2018). In other words, here, humans have a strong drive to generate novel conscious experiences even when those experiences are devoid of any other advantage beyond that of being novel.

### Endowing conscious experience with value might account for the unity of consciousness and features such as global availability

Endowing each conscious experience not only with a specific phenomenal content but also with value enables the comparison and ranking of conscious experiences. It is striking to see how humans are apt at analogy-making (French 1995), which requires mapping concepts across vastly different semantic spaces (e.g. think of a pen as a weapon). Likewise, most of us readily engage in seemingly complicated preference judgements such as 'would you rather go for a walk with friends or stay home and read?'. Those two experiences have little in common in terms of experiential features—one takes place outside, in a social context; the other indoors and in a solitary setting, but they can be easily compared in the common value space described in decision-making (Montague and Berns 2002; Lebreton et al. 2009; Levy and Glimcher 2012). 'Endowing conscious experience with value thus provides us with a unified space that probably contributes to the unity of consciousness'. Value coding, when taking place in phenomenal space (i.e. 'what does this do "to me"?'), is precisely the kind of mechanism one would expect to enable the emergence of a unified (phenomenal) space, essentially by creating a common, value-based currency in the brain. Pushing this reasoning further, perhaps it is in virtue of the fact that all our experiences take place in a unified phenomenal space that functional features such as global accessibility emerge, rather than the opposite.

### The experiencing subject should be at the heart of theories of consciousness but also of decision-making and emotions

Value is defined as the worth of something 'for someone', echoing the definition of consciousness by Nagel: 'an organism has conscious mental states if and only if there is something that it is like to 'be' that organism—something it is like 'for' the organism' (p. 436) (Nagel 1974). Similarly, an emotional feeling exists only if there is an organism to experience this feeling. In conscious perception research, the experimental measure of consciousness that has been mostly used is the report 'I saw the stimulus', but only few studies attempted at identifying the mechanisms underlying the 'I' part of the statement (Park et al. 2014). The experiencing subject should thus be at the heart of theories of (perceptual) consciousness but also of decision-making and emotion. The experiencing subject requires one adopts a first-person perspective (Azzalini et al. 2019), defined as a combination of a unified viewpoint (on both the external world and internal space) and intrinsic 'mineness' (Zahavi 2005).

What is the biological substrate of the experiencing subject? The simplest definition of a self is related to the organism that needs to be fed, regulated and protected. Interestingly, there is a convergence of theories pertaining to the emotional self (Damasio 2010), bodily self (Blanke and Metzinger 2009) or perceptual consciousness (Park and Tallon-Baudry 2014) to root the experiencing subject in the monitoring of bodily signals coming from joints and muscles (Blanke and Metzinger 2009) as well as from the viscera (Park and Tallon-Baudry 2014; Seth and Tsakiris 2018;

Azzalini et al. 2019). However, many feeding, regulatory or protective behaviours can be accounted for without introducing the notion of consciousness, instead relying solely on circuits and instrumental, derivative values selected by evolution.

Under our hypothesis, consciousness would have evolved and been selected because it adds an important degree of freedom to the machinery of reward-based behaviour: behaviour that seems purposeless from a purely functional perspective nevertheless has intrinsic value. But this, crucially, only holds when associated with conscious experience. Subjective experience hence becomes the central drive for the discovery and creation of new behaviours, leading in humans to a level of control of the environment and the ability to occupy most ecological niches unequalled by any other species. Still, under our hypothesis, observations such as crows sledging on a roof or dogs repeatedly dragging a plank up a snowy slope to enjoy the pleasure of speeding down on it thereafter, i.e., behaviours that apparently serve no function beyond that of procuring enjoyment, might be taken as a sign of phenomenal consciousness. As we elaborate below, the transition between behaviour driven mostly by circuits designed through evolution and novel, creative behaviour is likely to be a gradual, rather than an abrupt one. While our hypothesis would be in favour of animal consciousness, at least to some degree, it also suggests a close association between consciousness and life, with consequences for the possibility of artificial consciousness or of consciousness in lab-grown brains (Reardon 2020). As we have briefly discussed above, the problem with contemporary AI systems is not with further augmenting their computational prowess, but rather to design them in a way that they 'want things for reasons of their own'.

### The function of phenomenal consciousness

The distinction between access-consciousness and phenomenal consciousness has been at the core of many debates in the consciousness literature. Some have argued that there are distinct neural correlates of each (Block 2005), and others have offered arguments against that view (Baars and Laureys 2005). There has also been substantial debate about the richness of phenomenal experience, in particular about whether experience overflows access, i.e., about the extent to which we experience more than we can tell. Overflow theorists (e.g. Block 2011) have argued that it does while others have been highly critical (e.g. Phillips 2011, 2016). We will refrain from further discussion of such debates because doing so would be derivative to the two core points we wanted to emphasize with this article:

The first is that, as we hinted at earlier, the distinction between phenomenal and access-consciousness, as productive as it has been driving robust debate in consciousness research, is fundamentally misleading. On the one hand, it has led some to defend the idea that phenomenal experience is but a mere epiphenomenon that actually plays no functional role in our existence—a perspective that flies in the face of our own experience as human beings. On the other hand, it has led other authors to defend the idea that there can be phenomenal experience in the absence of access, i.e., that phenomenal experience is entirely independent of the machinery necessary for access (Fahrenfort and Lamme 2012). It is against this view that Cohen and Dennett (2011) have argued that 'the hard problem is an impossible problem', by which they mean to assert that 'dissociative' theories of consciousness, i.e., theories that distinguish between access and phenomenology, are fundamentally unfalsifiable and hence unscientific. Cohen and Dennett proposed the following 'perfect



experiment’—disconnect the entire set of brain areas responsible for colour perception (V4 in particular) and ask participants what they experience when shown an apple and asked to report on its colour. Cohen and Dennett speculate that participants would report not perceiving colour at all, despite the fact that the isolated V4 continues to function normally (Cohen and Dennett 2012), while Fahrenfort and Lamme insist that since V4 would continue to function normally, ‘the parsimonious account is to infer that perceptual states continue to exist without access’.

Here, what strikes us is how this discussion turns out to completely eschew the subject whose experiences or lack thereof form the core of the debate. What could it possibly mean for an isolated cortical region to ‘experience’ anything at all, even though its neural activity would be congruent with ‘normal’ perception? The very notion of conscious experience presupposes the existence of a subject it is the experience of. As Frege (Frege 1918) pointed out, ‘It seems absurd to us that a pain, a mood, a wish, should rove about the world without a bearer, independently. An experience is impossible without an experiencer’ (p. 299).

If phenomenal experience presupposes the existence of a subject, as we argue, it might also be necessary for access, for there would not be anything for a subject to access unless it were experienced. Congruently, Clark (2000) argued that there are cases where access implies qualia. Clark offers a contrast between two possible responses a system can give when asked to report on the differences between two visual stimuli (say, a difference in colour). The system can either, Clark argues, report that ‘the answer just comes to it’ (as in blindsight), or that it is ‘non-inferentially aware that it is using the visual vs., say, the tactile modality to make its judgement, which is what normal agents do. But in that latter case, says Clark, the system “must say that there is something it is like to ‘see’ the difference rather than e.g. to smell it.” In other words, Clark continues, at least in such cases,’ certain facts about patterns of access (the ‘easy’ stuff) can actually imply facts about phenomenality and ‘what it’s like’. Clark concludes that this excludes the possibility of zombies, for ‘honest reports of genuine direct, non-inferential access to acts of perceptual difference detection imply the presence of genuine phenomenal differences’ (p. 33).

And thus, our second core point, which admittedly goes much beyond the debates about the relationships between access-consciousness and phenomenal consciousness: we surmise that because phenomenal experience is endowed with intrinsic value, consciousness, reduced to its core definition—the ability to have phenomenal experiences—has a function. This perspective places the experiencing ‘subject’ at the heart of consciousness research and can perhaps result in a dissolution of the hard problem: phenomenal states, rather than being functionless—a mysterious ingredient added to a cognitive soup—instead constitute the very basis upon which we do everything that we do. And conversely, we probably would not do much at all were it not for the fact that we experience things.

In this, our perspective echoes Solms (2021), who writes: ‘What is the point of becoming aware of physical processes if your awareness has no influence upon those processes? [...] How can the function of feeling go on “in the dark”, without any feeling? We can legitimately ask why vision is accompanied by conscious experience. Vision does not require consciousness, and neither does any cognitive process. But feeling does. [...] There can be no such thing as a feeling that you do not feel. ‘Unconscious feeling’ is an oxymoron’ (p. 265).

To quote Dennett (personal communication): ‘How do we go from doing things for reasons to having reasons for doing things?’.

This points, in our view, at a critical juncture in evolution. All organisms behave in the ways they do for causes that have been shaped by evolution: they do the things they do for good ‘reasons’. But those reasons, crucially, are not the organism’s own reasons: they correspond to external fitness functions that follow natural selection. Humans, and probably some other animals, however, also have their ‘own’ reasons to act in the way they do, i.e., reasons that are independent of and that can actually go against the constraints set up by evolution. ‘Having reasons’ for doing things is precisely what consciousness enables and this crucially requires phenomenal experience.

## Conclusion and open questions

In closing, we propose that phenomenal experience, and hence consciousness, has a function. An organism’s ability to experience things turns out to be advantageous because phenomenal experience has intrinsic value. Importantly, the value associated with conscious phenomenal states lies in the phenomenal experience itself, not in the delayed rewards that the associated behaviour might produce. For instance, exploring the environment is typically considered to be valuable because it might lead to the discovery of new food sources. Because food is rewarding, exploration becomes rewarding and hence selected by evolution. While this may be correct, here, we propose instead that the act of exploration is rewarding in of itself for conscious agents, because exploration leads to novel phenomenal experiences and because those phenomenal experiences have intrinsic (positive or negative) value. In this respect, recent experiments in the decision-making literature (e.g. Cogliati-Dezza et al. 2020) are suggestive that in settings where agents have a choice between exploring the environment or exploiting a known source of reward (the ‘exploration-exploitation’ dilemma), the information gained through exploration can itself be rewarding.

Our proposal readily accounts for the drive to explore and generate new behaviours, including typically human achievements devoid of obvious evolutionary advantages such as artistic expression, which has been linked to evolutionary processes of sexual selection by Miller (2000). Our proposal further accounts for the unprecedented degree of mastery of the environment achieved by the human species, which results from this enhanced behavioural repertoire. The human ability to occupy all sorts of ecological niches obviously conveys an advantage over other species—but it is probably too early, in the time course of evolution, to find out whether this human ability to radically modify its own environment at the expense of other species is a curse or a blessing. In addition, being equipped with valuable phenomenal experiences is a potentially powerful drive for each individual to go beyond merely being alive. Nevertheless, our proposal does not readily assign a satisfactory role to some negative experiences such as chronic pain.

The problem of consciousness has been called a hard problem by Chalmers (1995) ‘precisely because it is not a problem about the performance of functions. The problem persists even when all the performance of all the relevant functions is explained ... What makes the hard problem hard and almost unique is that it goes ‘beyond’ problems about the performance of functions’. Because we assign a function to phenomenal consciousness, the hard problem of consciousness becomes more tractable. The precise mechanisms generating phenomenal experience and its associated value remain to be precisely identified, but taking it as a starting point that phenomenal experience has a function means that the concepts and tools of cognitive and computational neuroscience

and psychology can readily apply to our understanding of it. Thus, we believe our proposal opens a promising new venue to better identify the machinery generating consciousness, by appealing to three ingredients: the formalism associated with models of value-based decision-making and learning, the identification of the core dimensions of emotional feelings and qualia, and the embodied experiencing subject identified in consciousness research.

Finally, our hypothesis of 'phenomenal worthiness' might lead some to wonder whether we think phenomenal experience can have causal effects that go beyond its own biological causes. In other words: Are we defending a dualist position? The short answer is an emphatic 'no'. Here, we assume 'phenomenal realism', i.e., 'the view that that the phenomenal concept of consciousness is coherent and identifies something real' (Frankish 2021), and 'phenomenal efficacy', i.e., the thesis that phenomenal states are causally efficacious. In this, we follow List (2019), who developed a novel level-of-description analysis of causality and determinism to argue in favour of compatibilism: mental states are entirely caused by neural activity, but, contra Kim (2005), this does not mean they are epiphenomenal, essentially because a single mental state may be caused by different patterns of neural activity. In other words, mental states are (to some extent) realization insensitive. In such cases, List (2019) shows, rather convincingly, that Kim's causal exclusion argument does not hold, which then deflates (or rather, eliminates) epiphenomenalist perspectives on mental states and on mental causation. List writes: 'When a higher-level difference-making relation is realization insensitive, it is not accompanied by a matching difference-making relation at the lower level' (p. 137). Further: 'It is a person's intentional mental states that are normally the difference-making causes of the person's actions, not the underlying physical states of the brain and body. And this is entirely compatible with recognizing that mental states are physically realized at the level of the brain. It is just that the realizing brain states do not themselves qualify as difference-making causes of the resulting actions' (p. 138). Therefore, based on this analysis, we take it that conscious mental states, as global, high-level states of an organism, are causally efficacious in a way that cannot be wholly reduced to their biological foundations. At the same time, such states are entirely caused by and supervene on the biological activity of the brain. Consciousness matters.

## Acknowledgements

We thank the referees and the editors for abundant and often pugnacious comments that helped greatly improve the manuscript. A.C. is a research director with the F.R.S.-FNRS (Belgium). Both A.C. and C.T.B. are fellows of the Canadian Institute for Advanced Research (CIFAR) program 'Brain, Mind and Consciousness'.

## Funding

C.T.B. is supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement no 670325, Advanced grant BRAVIUS) and ANR-17-EURE-0017.

## Conflict of interest statement

None declared.

## References

Adolphs R, Mlodinow L, Barrett LF. What is an emotion? *Curr Biol* 2019;**29**:R1060–4.

- Anderson DJ, Adolphs R. A framework for studying emotions across species. *Cell* 2014;**157**:187–200.
- Azzalini D, Buot A, Palminteri S et al. Responses to heartbeats in ventro-medial prefrontal cortex contribute to subjective preference-based decisions. *J Neurosci* 2021;**41**:5102–14.
- Azzalini D, Rebollo I, Tallon-Baudry C. Visceral signals shape brain dynamics and cognition. *Trends Cogn Sci* 2019;**23**:488–509.
- Baars BJ. *A Cognitive Theory of Consciousness*. Cambridge: Cambridge University Press, 1988.
- Baars BJ, Laureys S. One, not two neural correlates of consciousness. *Trends Cogn Sci* 2005;**9**:269.
- Barrett LF, Bar M. See it with feeling: affective predictions during object perception. *Philos Trans R Soc Lond B Biol Sci* 2009;**364**:1325–34.
- Barto AG. Intrinsic motivation and reinforcement learning. In: Baldassarre G, Moiralli M (eds), *Intrinsically Motivated Learning in Natural and Artificial Systems*. Berlin: Springer, 2013, 17–47.
- Bartra O, McGuire JT, Kable JW. The valuation system: a coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value. *Neuroimage* 2013;**76**:412–27.
- Bayne T, Carter O. Dimensions of consciousness and the psychedelic state. *Neurosci Conscious* 2018;**2018**:1–8.
- Bayne T, Hohwy J, Owen AM. Are there levels of consciousness? *Trends Cogn Sci* 2016;**20**:405–13.
- Bedi G, Carrillo F, Cecchi GA et al. Automated analysis of free speech predicts psychosis onset in high-risk youths. *NPJ Schizophr* 2015;**1**:15030.
- Berthoud HR. Metabolic and hedonic drives in the neural control of appetite: who is the boss? *Curr Opin Neurobiol* 2011;**21**:888–96.
- Blanke O, Metzinger T. Full-body illusions and minimal phenomenal selfhood. *Trends Cogn Sci* 2009;**13**:7–13.
- Block N. On a confusion about a function of consciousness. *Behav Brain Sci* 1995;**18**:227–87.
- Block N. Two neural correlates of consciousness. *Trends Cogn Sci* 2005;**9**:46–52.
- Block N. Perceptual consciousness overflows cognitive access. *Trends Cogn Sci* 2011;**15**:567–75.
- Campana F, Rebollo I, Urai A et al. Conscious vision proceeds from global to local content in goal-directed tasks and spontaneous vision. *J Neurosci* 2016;**36**:5200–13.
- Campana F, Tallon-Baudry C. Anchoring visual subjective experience in a neural model: the coarse vividness hypothesis. *Neuropsychologia* 2013;**51**:1050–60.
- Carruthers P. Valence and value. *Philos Phenomenol Res* 2018;**XCVII**:658–80.
- Chalmers DJ. Facing up to the problem of consciousness. *J Conscious Stud* 1995;**2**:200–19.
- Chalmers DJ. *The Conscious Mind: In Search of a Fundamental Theory*. New York: Oxford University Press, 1996.
- Chib VS, Rangel A, Shimojo S et al. Evidence for a common representation of decision values for dissimilar goods in human ventromedial prefrontal cortex. *J Neurosci* 2009;**29**:12315–20.
- Clark A. A case where access implies qualia? *Analysis* 2000;**60**:30–7.
- Cleeremans A, Achoui D, Beauny A et al. Learning to be conscious. *Trends Cogn Sci* 2020;**24**:112–23.
- Cogliati-Dezza I, Cleeremans A, Alexander W. Independent and interacting value systems for reward and information in the human brain. *bioRxiv* 2020:1–48.
- Cohen MA, Dennett DC. Consciousness cannot be separated from function. *Trends Cogn Sci* 2011;**15**:358–64.
- Cohen MA, Dennett DC. Response to Fahrenfort and Lamme: defining reportability, accessibility, and sufficiency in conscious awareness. *Trends Cogn Sci* 2012;**16**:139–40.

- Cohen MA, Dennett DC, Kanwisher N. What is the bandwidth of perceptual experience? *Trends Cogn Sci* 2016;**20**:324–35.
- Cowen AS, Keltner D. Self-report captures 27 distinct categories of emotion bridged by continuous gradients. *Proc Natl Acad Sci U S A* 2017;**114**:E7900–09.
- Damasio A. *Self Comes to Mind: Constructing the Conscious Brain*. New York: Pantheon Books, 2010.
- de Gelder B, Tamietto M, van Boxtel G et al. Intact navigation skills after bilateral loss of striate cortex. *Curr Biol* 2008;**18**:R1128–9.
- Dehaene S, Kerszberg M, Changeux JP. A neuronal model of a global workspace in effortful cognitive tasks. *Proc Natl Acad Sci U S A* 1998;**95**:14529–34.
- Dennett DC. Quining qualia. In: Marcel AJ, Bisiach E (eds), *Consciousness in Modern Science*. Oxford, UK: Oxford University Press, 1988, 42–7.
- Dennett DC. Illusionism as the obvious default theory of consciousness. *J Conscious Stud* 2016;**23**:65–72.
- Dennett DC. Welcome to strong illusionism. *J Conscious Stud* 2019;**26**:48–58.
- Fahrenfort JJ, Lamme VAF. A true science of consciousness explains phenomenology: comment on Cohen and Dennett. *Trends Cogn Sci* 2012;**16**:138–9.
- Frankish K. Illusionism as a theory of consciousness. *J Conscious Stud* 2016;**23**:11–39.
- Frankish K. Panpsychism and the depsychologization of consciousness. *Aristotelian Soc Suppl Vol* 2021;**95**:51–70.
- Frege G. The thought: a logical enquiry. *Mind* 1918;**65**:289–311.
- French RM. *The Subtlety of Sameness: A Theory and Computer Model of Analogy-making*. Boston, MA: MIT Press, 1995.
- Gilovich T, Gallo I. Consumer's pursuit of material and experiential purchases: a review. *Consum Psychol Rev* 2019;**3**:20–33.
- Ginsburg S, Jablonka E. *The Evolution of the Sensitive Soul: Learning and the Origins of Consciousness*. Boston, MA: MIT Press, 2019.
- Haun AM, Tononi G, Koch C et al. Are we underestimating the richness of visual experience? *Neurosci Conscious* 2017;**2017**:niw023.
- Hayden BY, Niv Y. The case against economic values in the orbitofrontal cortex (or anywhere else in the brain). *Behav Neurosci* 2021;**135**:192–201.
- Hesp C, Smith R, Parr T et al. Deeply felt affect: the emergence of valence in deep inference. *Neural Comput* 2021;**33**:398–446.
- Hoffman DD. *The Case against Reality: How Evolution Hid the Truth from Our Eyes*. New York: W.W. Norton, 2019.
- Juechems K, Summerfield C. Where does value come from? *Trends Cogn Sci* 2019;**23**:836–50.
- Keramati M, Gutkin B. Homeostatic reinforcement learning for integrating reward collection and physiological stability. *Elife* 2014;**3**:1–26.
- Kim J. *Physicalism, or Something near Enough*. Princeton, NJ: Princeton University Press, 2005.
- Klyubin AS, Polani D, Nehaniv CL (2005). *All else being equal, be empowered*. Paper presented at the European Conference on Artificial Life, Canterbury, UK.
- Klyubin AS, Polani D, Nehaniv CL. Keep your options open: a information-based driving principle for sensorimotor systems. *PLoS One* 2008;**3**:e4018.
- Kriegel U. The functional role of consciousness: a phenomenological approach. *Phenomenol Cogn Sci* 2004;**3**:171–94.
- Kriegel U. The value of consciousness. *Analysis* 2019;**79**:503–20.
- Lebrecht S. "Micro-valences": affective valence in "neutral" everyday objects. Ph.D., Brown University, 2012.
- Lebrecht S, Bar M, Barrett LF et al. Micro-valences: perceiving affective valence in everyday objects. *Front Psychol* 2012;**3**:1–5.
- Lebreton M, Jorge S, Michel V et al. An automatic valuation system in the human brain: evidence from functional neuroimaging. *Neuron* 2009;**64**:431–9.
- LeDoux JE, Hofmann SG. The subjective experience of emotion: a fearful view. *Curr Opin Behav Sci* 2018;**19**:67–72.
- Lee AY. Is consciousness intrinsically valuable? *Philos Stud* 2019;**175**:1–17.
- Levy DJ, Glimcher PW. The root of all value: a neural common currency for choice. *Curr Opin Neurobiol* 2012;**22**:1027–38.
- Lindquist KA, Satpute AB, Wager TD et al. The brain basis of positive and negative affect: evidence from a meta-analysis of the human neuroimaging literature. *Cereb Cortex* 2016;**26**:1910–22.
- List C. *Why Free Will Is Real*. London: Harvard University Press, 2019.
- Livemore JJA, Klaassen FH, Bramson B et al. Approach-avoidance decisions under threat: the role of autonomic psychophysiological states. *Front Neurosci* 2021;**15**:1–12.
- Miller G. *The Mating Mind: How Sexual Choice Shaped the Evolution of Human Nature*. New York: Doubleday, 2000.
- Montague LM, Berns GS. Neural economics and the biological substrates of valuation. *Neuron* 2002;**36**:265–84.
- Nagel T. What is it like to be a bat? *Philos Rev* 1974;**83**:435–50.
- Nikolova N, Waade PT, Friston K et al. What might interoceptive inference reveal about consciousness? *Rev Philos Psychol* 2021:1–28.
- Nozick R. *Anarchy, State, and Utopia*. New York: Basic Books, 1974.
- O'Doherty JP. The problem with value. *Neurosci Biobehav Rev* 2014;**43**:259–68.
- Overgaard M, Rote J, Mouridsen K et al. Is conscious perception gradual or dichotomous? A comparison of report methodologies during a visual task. *Conscious Cogn* 2006;**15**:700–8.
- Park HD, Correia S, Ducorps A et al. Spontaneous fluctuations in neural responses to heartbeats predict visual detection. *Nat Neurosci* 2014;**17**:612–8.
- Park HD, Tallon-Baudry C. The neural subjective frame: from bodily signals to perceptual consciousness. *Philos Trans R Soc Lond B Biol Sci* 2014;**369**:20130208.
- Pelletier G, Fellows LK. Viewing orbitofrontal cortex contributions to decision-making through the lens of object recognition. *Behav Neurosci* 2021;**135**:182–91.
- Pessiglione M, Daunizeau J. Bridging across functional models: the OFC as a value-making neural network. *Behav Neurosci* 2021;**135**:277–90.
- Petitmengin C, van Beek M, Bitbol M et al. Studying the experience of meditation through micro-phenomenology. *Curr Opin Psychol* 2019;**28**:54–9.
- Phillips I. Perception and iconic memory. *Mind Lang* 2011;**26**:381–411.
- Phillips I. No watershed for overflow: recent work on the richness of consciousness. *Philos Psychol* 2016;**29**:236–49.
- Preller KH, Vollenweider FX. Phenomenology, structure and dynamics of psychedelic state. *Curr Top Behav Neurosci* 2018;**36**:221–56.
- Reardon S. Can lab-grown brains become conscious? *Nature* 2020;**586**:658–61.
- Robinson Z, Maley CJ, Piccinini G. Is consciousness a spandrel? *J Am Philos Assoc* 2015;**1**:365–83.
- Russell JA, Barrett LF. Core affect, prototypical emotional episodes, and other things called emotion: dissecting the elephant. *J Pers Soc Psychol* 1999;**76**:805–19.
- Sandberg K, Timmermans B, Overgaard M et al. Measuring consciousness: is one measure better than the other? *Conscious Cogn* 2010;**19**:1069–78.
- Seager WE. Consciousness, value and functionalism. *Psychol* 2001;**7**:20.
- Seth A. Functions of consciousness. In: Banks WP (ed.), *Encyclopedia of Consciousness*. Vol. 1. Amsterdam: Elsevier, 2009, 279–93.

- Seth A. *Being You*. London: Faber, 2021.
- Seth A, Tsakiris M. Being a beast machine: the somatic basis of selfhood. *Trends Cogn Sci* 2018;**22**:969–81.
- Siewert C. *The Significance of Consciousness*. Princeton, NJ: Princeton University Press, 1998.
- Silver D, Schrittwieser J, Simonyan K et al. Mastering the game of Go without human knowledge. *Nature* 2017;**550**:354.
- Simon HA. Rational choice and the structure of the environment. *Psychol Rev* 1956;**63**:129–38.
- Simons DJ, Ambinder MS. Change blindness: theory and consequences. *Curr Dir Psychol Sci* 2005;**14**:44–8.
- Simons DJ, Chabris CF. Gorillas in our midst: sustained inattentive blindness for dynamic events. *Perception* 1999;**28**:1059–94.
- Simons DJ, Levin DT. Change blindness. *Trends Cogn Sci* 1997;**1**:261–7.
- Solms M. *The Hidden Spring*. London: Profile Books, 2021.
- Tye M. The function of consciousness. *Noûs* 1996;**30**:287–305.
- Uexküll J. A stroll through the worlds of animals and men: a picture book of invisible worlds. In: Schiller CH (ed.), *Instinctive Behavior: The Development of a Modern Concept*. New York: International Universities Press, 1957, 5–80.
- Voltaire. *A Philosophical Dictionary*. New York: Basic Books, 1962/1764.
- Whitney D, Yamanashi Leib A. Ensemble perception. *Annu Rev Psychol* 2018;**69**:105–29.
- Zahavi D. *Subjectivity and Selfhood: Investigating the First-person Perspective*. Cambridge, MA: MIT Press, 2005.
- Zimmerman MJ. Intrinsic vs. extrinsic value. In: Zalta N (ed.), *The Stanford Encyclopedia of Philosophy*. Stanford, CA: Metaphysics Research Lab, 2019.