

# Piecing together the puzzle of emotional consciousness

Tahnée Engelen<sup>1,\*</sup> and Rocco Mennella<sup>2,\*</sup>

<sup>1</sup>Cognitive and Computational Neuroscience Laboratory (LNC<sup>2</sup>), Inserm U960, Department of Cognitive Studies, École Normale Supérieure, PSL University, 29 rue d'Ulm, Paris 75005, France; <sup>2</sup>Laboratoire sur les Interactions Cognition-Action-Émotion (LICAÉ), UFR-STAPS, Université Paris Nanterre, 200 Avenue de la République, Nanterre 92001, France

\*Corresponding authors. Cognitive and Computational Neuroscience Laboratory (LNC<sup>2</sup>), Inserm U960, Department of Cognitive Studies, École Normale Supérieure, PSL University, 29 rue d'Ulm, Paris 75005, France. E-mail: [tahnee.engelen@ens.fr](mailto:tahnee.engelen@ens.fr) Laboratoire sur les Interactions Cognition-Action-Émotion (LICAÉ), UFR-STAPS, Université Paris Nanterre, 200 Avenue de la République, Nanterre 92001, France. E-mail: [rmennella@parisnanterre.fr](mailto:rmennella@parisnanterre.fr)

## Abstract

The search for neural correlates of emotional consciousness has gained momentum in the last decades. Nonetheless, disagreements concerning the mechanisms that determine the experiential qualities of emotional consciousness—the “what is it like” to feel an emotion—as well as on their neural correlates have far-reaching consequences on how researchers study and measure emotion, sometimes leading to seemingly irresolvable impasses. The current paper lays out in a balanced way the viewpoint of both cognitive and precognitive approaches to emotional consciousness on the basis of commonalities and differences between the claims of some relevant theories of emotions. We examine the sufficiency of the existing evidence in support of the proposed theories of emotional consciousness by going through the methodological specificity of the study of emotional consciousness and its unique challenges and highlighting what can and cannot be imported by advances in research on perceptual consciousness. We propose that there are three key experimental contrasts that are each equally necessary in the search for the neural correlates of emotional consciousness and each contrast alone coming with its own limitations. We conclude by acknowledging some of the most promising avenues in the field, which may help go beyond the current limitations and collaboratively piece together the puzzle of emotional consciousness.

**Keywords:** emotion; consciousness; theories of emotional consciousness; neural correlates of consciousness

## Background and scopes

Understanding the phenomenal dimension of consciousness, the “what it is like” to have a specific conscious experience (Nagel 1974), has long been considered a “hard” scientific problem (Chalmers 1995). In the last decades, neuroscientists have proposed an increasing number of theories of consciousness (see Seth and Bayne 2022 for a review) to try to close (or at least reduce) the “explanatory gap” that we intuitively feel (but see Dennett 2019) between conscious experience and its neural determinants, the so-called neural correlates of consciousness (NCC; Koch et al. 2016). In recent years, the development of NCC theories has run in parallel with, and sometimes inspired, a particularly intense debate in the field of affective neuroscience around what emotions are and how to study them (Anderson and Adolphs 2014, Barrett 2017b, 2017c, LeDoux and Brown 2017, Panksepp et al. 2017, Berridge 2018, Fanselow and Pennington 2018, Adolphs et al. 2019, LeDoux 2019, 2020b, 2021, Mobbs et al. 2019, Taschereau-Dumouchel et al. 2022). In fact, most of the controversies around how to define and measure emotion stem from a disagreement concerning the mechanisms that determine the experiential qualities of emotional consciousness—the “what is it like” to feel an emotion—as well as on their neural correlates.

Despite the uncertainty around definitions and measures of emotion (Dukes et al. 2021), several authors have brought forward theories of the neural correlates of emotional consciousness (from now on NCEC), which—not surprisingly—are often in contradiction with one another. In our opinion, the strength of the conviction by which, for instance, some theories define core concepts (e.g. “emotion”) is at present only partially justified by empirical data, mostly due to the unique methodological challenges in the study of the NCEC, which we will detail in the following sections. As previously pointed out by others (Pessoa 2019), tight definitions can slow down scientific progress by hampering novel research ideas, which violate theoretical boundaries, rather than pushing them forward. This can create “tunnel vision” (e.g. Paré and Quirk 2017), resulting in difficulty in studying the construct of interest in its full complexity, as well as in making comparisons between theories. Finally, this can lead to aprioristically valuing or devaluing specific measures (e.g. self-reports, physiology, and behavior) or fields (e.g. human vs. animal research) in providing insightful information concerning the NCEC (for a discussion, see e.g. Panksepp et al. 2017). By no means do we question the necessity of theory building for proper confirmatory hypothesis testing, as it is evident from the contrastive approach that we propose in the second part of the present work. Nonetheless, we may be missing

Received 8 July 2022; Revised 21 February 2023; Accepted 20 March 2023

© The Author(s) 2023. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

crucial knowledge about the elements that are needed for deriving hypotheses from theories, namely concept definition, agreement on valid and reliable measures, definition of the relationships between concepts, boundary conditions, auxiliary assumptions, and statistical predictions (see [Scheel et al. 2021](#)), which sometimes leads to seemingly irresolvable impasses. The present paper aims at reflecting upon possible strategies to build this knowledge, in order to strengthen the theories of NCeC and move the debate forward. After reviewing the existing main approaches to emotional consciousness, we will try to challenge some of their fundamental assumptions, often by referring to advancements in research on the neural correlates of perceptual consciousness. The main aim of the present paper is to outline what we think are the theoretical and methodological specificities of the study of the NCeC. We argue that some of these limitations are unique to the study of the NCeC; i.e. they are in part not shared with the study of other contents of consciousness. In this spirit, we will discuss the main experimental contrasts to be used in the search for NCeC, as well as their interpretative value and limitations. In the final section, we will reflect upon the possible strategies to strengthen NCeC theories.

## The contemporary debate on the NCeC

The present section does not mean to provide a full picture of the existing theories of emotion (for comprehensive reviews, see [Lange et al. 2020](#) and [Moors 2022](#)). We will instead focus on the place of consciousness in what are, in our view, some of the most influential recent perspectives. For the purpose of this paper, theoretical propositions on emotional consciousness will be divided into two approaches, which we call “cognitive” vs. “precognitive” approaches to NCeC. The cognitive vs. precognitive distinction refers here, respectively, to whether emotional consciousness is proposed to require cognitive processing [e.g. working memory (WM)] to arise or to whether it is thought to precede it. Other dichotomies have previously been put forward, such as cognitive vs. perceptual ([Block 2019](#)), higher vs. first order ([LeDoux and Brown 2017](#)), or access vs. phenomenal consciousness theories ([Block 1995, 2005](#)), but we felt that the cognitive vs. precognitive distinction is more generalizable to the different subtheories within each category, as it will be clear in the next sections. Within each approach, we will highlight important differences between theoretical perspectives, without aiming at presenting the full complexity of each author’s theory of emotion. Of note, the topic of the interaction between affect and consciousness, using a comparative science perspective, has also recently been the object of an excellent review, to which we address the interested reader ([Paul et al. 2020](#)).

### Cognitive approach to emotional consciousness

Overall, emotional theories that we include in the “cognitive approach” to NCeC propose that, while humans and other animals share evolutionarily conserved subcortical circuits to respond to threat, face aggression, or seek rewards ([LeDoux and Daw 2018](#)), the activation of none of these circuits alone is sufficient to cause a subjective emotional experience as humans conceive it. Theories within this approach agree that, in order for the individual to experience emotion, additional top-down processes supported by neocortical associative areas are needed (e.g. [Seth 2013](#), [Smith and Lane 2015](#), [LeDoux and Brown 2017](#), [Barrett 2017c](#)). The cognitive approach does not deny the existence of behavioral and visceral reactions to salient environmental stimuli but does not support a reliable and specific causal relationship between these reactions and an instance of what we call emotion ([LeDoux and Brown 2017](#),

[Barrett 2017c](#)). Therefore, in this view, the subcortical circuits do not qualify as NCeC.

The precise way through which neocortical areas sustain emotional consciousness varies from one cognitive theory to another. Among the most influential ones, LeDoux’s “higher-order” theory of emotional consciousness (HOTEC; e.g. [LeDoux and Pine 2016](#), [LeDoux and Brown 2017](#), [LeDoux and Hofmann 2018](#), [LeDoux 2019, 2020a, 2020b, 2021](#), [Taschereau-Dumouchel et al. 2022](#)) combined the existing higher-order theories of consciousness (e.g. [Rosenthal 2005](#), [Brown 2015](#)) with the previous work on fear and threat responses (for a summary, see [LeDoux 2012](#)). For the HOTEC, for an emotional experience to arise, higher-order representations (HORs) of the stimulus, as opposed to the first order—sensorial or conceptual ones—need to enter WM, via the brain’s “general networks of cognition”, which include the lateral and medial prefrontal cortex, insular cortex, and posterior parietal cortex. For instance, for experiencing fear, a threatening stimulus (e.g. a snake) needs first to be represented in visual areas (first-order representation) and integrated with long-term semantic memories—stored in medial temporal and other cortical areas—in a common unconscious WM representation (the first HOR). This unconscious HOR can further be re-represented in WM, including autobiographic (self-related) memories (self-HOROR), becoming conscious. Specifically, for this self-HOROR to be an “emotion” (emotional self-HOROR), the consequences of the stimulus-evoked activation of the defensive circuits, including cortical arousal, behavioral, and physiological reactions, need to be included in the representation in WM as well ([LeDoux and Brown 2017](#)). The HOTEC proposes that our “emotion schemas,” i.e. what we know about emotions, which are part of the unconscious HORs, refine with experience, rendering the emotional experience more and more differentiated throughout development. Overall, the theory implies that emotions coincide with subjective feelings (i.e. conscious emotional experiences) and therefore, since emotions can never be unconscious, individuals cannot be mistaken about the emotion that they are feeling ([LeDoux and Brown 2017](#)). Hence, introspection and subjective reports, despite not being a one-to-one readout of the experience, are the gold standard for studying emotion ([LeDoux and Hofmann 2018](#)).

Within the cognitive approach, most authors agree in identifying WM as a pivotal process to maintain feelings active over short time periods in order to guide decisional and goal-directed behavior ([Mikels and Reuter-Lorenz 2019](#)). Nonetheless, not every author agrees that HORs in WM are needed for emotional consciousness to arise. For instance, [Smith and Lane \(2015, 2016\)](#), by combining Prinz’s perceptual model of emotional consciousness ([Prinz 2004, 2008](#)) with cognitive theories of consciousness based on a “global neuronal workspace (GNW)” model ([Dehaene and Naccache 2001](#), [Baars 2005](#)), proposed that the content of first-order representations of our own bodily reactions to the emotional situation can directly become conscious. This happens when it is globally broadcasted across a network of frontoparietal regions, which renders the first-order representations available in WM, to guide goal-directed behavior. More specifically, emotional episodes are proposed to involve a hierarchical, and iterative, sequence of appraisals of salient stimuli/situations’ representations. The hierarchy runs upstream in the brain from automatic appraisals of basic characteristics (i.e. novelty and concern relevance), which primarily involve the amygdala (AMY) and the hippocampus, to more sophisticated ones (i.e. goal congruence, agency, value compatibility, and affect meaning), subtended by temporal and prefrontal cortical regions. This hierarchy of appraisals triggers

and continuously refines bodily reactions (together with cognitive changes), which, as in Damasio's "as-if loops" (Damasio 1994), can sometimes only be represented and not embodied. In this view, emotions depend on our own perception of the above-mentioned bodily changes, again in a hierarchical fashion, from the perception of discrete body features (Stage 1; brainstem), to whole body patterns—which produce phenomenologically distinguishable bodily feelings (Stage 2; anterior insula)—and finally to emotional concepts (Stage 3; rostral anterior cingulate cortex), which categorize bodily feelings in an integrative manner (e.g. different bodily feelings can activate the same emotional concept). Unconscious emotional representations in Stages 2/3 are thought to compete to gain access to consciousness and, if selected via the global broadcasting mechanism in frontoparietal areas, are rendered available to guide deliberative thoughts and action. In line with HOTE, the GNW theory of emotional consciousness proposes that the NCC are cortical, largely overlap with NCC of other contents of consciousness (Smith et al. 2018b), and rely on shared cognitive mechanisms (WM and attention). However, differently from HOTE, GNW theory proposes that emotional representations can be both conscious and unconscious and that our own conscious and reportable emotional experience can sometimes not be aligned with our behavioral and physiological emotional reactions, depending on how the probabilistic competition between active emotional concepts is resolved (e.g. one can act angry without feeling angry; Smith et al. 2018a).

The focus on embodiment and on the probabilistic attribution of emotional concepts to bodily changes in specific situations is also at the core of cognitive theories centered around the notion of "predictive coding" (Seth 2013, Seth and Critchley 2013, Barrett 2017b, 2017c). In this view, the brain is specifically conceived as a Bayesian prediction machine (Clark 2013, Seth 2015), which continuously runs and updates an internal model of the body in the world, in order to perform allostasis, i.e. the regulation of the body based on metabolic cost and benefits (Barrett 2017c, see also Damasio 1994, 1998, 1999, Parvizi and Damasio 2001). Internal models contain (i) sensory predictions concerning the most probable next stimulation, (ii) motor predictions relative to the most appropriate actions to take, and (iii) visceromotor predictions regarding anticipated consequences for allostasis. The interoception of internal sensations (Craig 2009) is thought to produce the lower dimensional feelings of affect (i.e. the so-called "core affect")—valence and arousal—that are basic features of consciousness (Barrett 2017c). Crucially, although core affects are probably shared in humans and other animals (Barrett et al. 2007, Barrett 2017b), they are not emotions, nor are they specific to emotional episodes. While receiving sensory information both through exteroception and interoception, the brain computes a prediction error between the model and the actual state of the body in the world. The internal model that has the best fit, i.e. minimizes the prediction error, constitutes our present perception, namely our "experience." From this constructionist point of view, emotional experiences (i.e. fear) are concepts (internal models like any other), built upon experience, that best explain the present state of the body in the world and direct action (Barrett 2017c). With regard to the neural correlates, the same reported emotion might depend on the activation of different neural correlates under different contexts, as it has been suggested that there is a many-to-many mapping between emotional categories and the combinations of somato-visceral/cognitive reactions and situational patterns they map to (Barrett 2017c). Nonetheless, some authors insist on the importance of the salience network for running an internal model of the body in the world, centered on the anterior insula and on

the ACC (Seth 2013), while others propose a more complex interaction among brain intrinsic networks (Barrett and Satpute 2013, Barrett 2017c).

HOTE, GNW, and predictive processing theories agree that emotional concepts are learned progressively through each individual's development, ultimately allowing for the attribution of meaning to our experiences in the world (Barrett 2006, 2017c, Lane et al. 2015, LeDoux and Brown 2017). This implies that each person has their own unique emotional experiences, as the way humans mentally represent, perceive, recognize, and express emotions is profoundly shaped by life history, sociocultural influences, and language (Jack et al. 2012, Crivelli and Fridlund 2018, 2019, Barrett et al. 2019, Jackson et al. 2019). Individuals differ in their ability to conceptualize and understand their own affective states and responses—i.e. "emotional awareness" (Lane and Smith 2021). Multiple possible neurocomputational mechanisms (Smith et al. 2019a, 2019b) explain why some individuals might, for instance, be able to feel and report valenced perceptual experiences of physiological changes (e.g. the distress associated with an increase in heart rate) but fail to be aware of and report a specific emotional category (e.g. fear). The ability to "put feelings into words," reporting separate and detailed emotional experiences—the so-called "emotional differentiation" or "emotional granularity"—has proven adaptive for both psychological and social well-being and contributes to the heterogeneity of emotional experiences and reports (Kashdan et al. 2015). The GNW theory in particular does not exclude that perceptually based appraisals, preconceptual ones, can trigger bodily changes and simply felt emotions in the absence of further cognitive processing or via subcortical stimulation (Smith and Lane 2015, 2016). Nonetheless, it is safe to say that cognitive theories converge in proposing that, while it is not impossible that other animals experience some form of affect, this is hardly comparable to the complex emotional experience available to humans and will hardly be directly measurable (LeDoux and Pine 2016, Barrett 2017b, LeDoux 2021). In fact, it has been proposed that unjustified inference of subjective emotion from objective behavioral and physiological measures has led animal studies to produce scarce results in developing new pharmacological treatments for mental disorders in humans, such as depression and anxiety (LeDoux and Pine 2016, Taschereau-Dumouchel et al. 2022).

### **Precognitive approach to emotional consciousness**

The precognitive approach to NCC claims either that emotions can be conscious in a way that is inaccessible to introspective scrutiny (preconscious) or that they can remain fully unconscious, in both cases having core neural determinants in specific evolutionary old brain structures, most of which are subcortical (Izard 2007b, Adolphs and Anderson 2018, Berridge 2018, Fanselow and Pennington 2018). Indeed, some authors within this approach align with theories identifying two types of consciousness (e.g. Damasio 1994, 1998, Block 1995, Merker 2007): a "phenomenal" (or "core") consciousness, which corresponds to the experience, the "what it is like" to be in a particular state from our unique first-person perspective, and an "access" (or "extended") consciousness, which makes contents of experience available for further cognitive elaboration (e.g. subjective report). As two types of consciousness exist, two different NCC are postulated (Block 2005). Notably, it has been proposed that phenomenal consciousness emerged relatively early in evolution, as a core mechanism of integration of bodily signals and environmental objects, to guide action, and therefore strongly depends on the activation of subcortical brain structures, notably brainstem structures, such as the

superior colliculus and its connections with the thalamus (Parvizi and Damasio 2001, Merker 2007, 2013, Shine 2022). Participating in the ongoing debate around the phenomenal vs. access distinction (Block 2007, Naccache 2018), several authors within the precognitive approach explicitly adhere to this distinction (Panksepp 1998b, 2007, Izard 2007a, 2007b, Panksepp et al. 2017, Lieberman 2019), proposing that additional subcortical structures specifically determine the content of emotional phenomenal experience (e.g. the periaqueductal gray; Panksepp 1998a, 1998b); other authors propose instead that emotions can be either conscious or unconscious (and not phenomenally preconscious), still largely attributing an important role to subcortical structure for unconscious emotions, for instance to the SC and the pulvinar (Celegnin et al. 2015, Méndez et al. 2022), the AMY (Anderson and Adolphs 2014, Fanselow 2018), and the striatum (Winkelman and Berridge 2004).

Despite these differences concerning the existence of preconscious vs. fully unconscious emotional states, the underlying theory of emotion is largely shared between the authors within the precognitive approach and therefore more easily summarized compared to the cognitive approach. Overall, this approach proposes that the conscious representation that allows humans to produce verbal reports about what they feel (e.g. “I feel angry”) has no special status in defining what an emotion is. In fact, reportable subjective feelings in humans are only one of the components of a functionally organized set of behavioral, physiological, and cognitive responses to environmental challenges, caused by “central generators” or “central states” of emotions (Anderson and Adolphs 2014, Fanselow 2018). Such central generators would (at least in part) map onto dedicated neural systems (Adolphs 2013), mostly subcortical (Panksepp 2007), which have been shaped by evolution. Accordingly, these ancient circuits control the execution of typical physiological and behavioral responses, such as freezing (Fanselow and Lester 1988, Fanselow 1994), “wanting” (Berridge et al. 1989, Treit and Berridge 1990, Berridge and Valenstein 1991), or RAGE (Panksepp 1998a), which have been conserved since they have proven useful to adapt to environmental challenges. This set of physiological and behavioral responses is therefore a measurable and objective indicator of the activation of the central emotion generator. Since, for instance, the direct stimulation of dedicated emotional neural circuits has reinforcing effects on behavior (i.e. animals actively try to prolong or terminate the stimulation), these circuits are inferred to be responsible for a phenomenal form of emotional consciousness (Panksepp 1998a, 1998b), which is more differentiated than a simple core affect experience of something good or bad (as in Barrett 2017b, 2017c). Other authors argue instead that, since such reinforcing effects on behavior, as well as typical emotion-related neural and physiological activations, can be elicited even in response to emotional stimuli that are not consciously detected and in the absence of changes in subjective reports, emotions can be fully unconscious (Winkelman and Berridge 2004, Winkelman et al. 2005, Celegnin et al. 2015, Berridge 2018). Depending on the author, the functional properties of evolutionarily conserved emotional circuits are thought to be partially (Adolphs 2017) or almost entirely (Panksepp 1998a) conserved between humans and other animals. The precognitive approach largely acknowledges that humans can further elaborate, reappraise, and access core emotional reactions to report them verbally. Some authors even made the distinction between “basic emotions”—the evolutionarily conserved emotional responses—and the associations between such emotions and cognitive responses, acquired through learning, called “emotion schemas” (Ekman 1992, Izard 2007a). In this view, animal studies, as well as research on infants

(Izard 1991), are often thought to represent the gold standard for understanding the emotion primitives, non-contaminated by subsequent cognitive elaboration (Panksepp et al. 2017, Berridge 2018, Fanselow and Pennington 2018). Concerning animal research, this also stems from the possibility of directly stimulating/interfering with the central brain state with invasive techniques and measuring effects on behavior, as well as provoking strong emotional states, which would be unethical in humans (Panksepp et al. 2017). Overall, it is expected that animal studies will be crucial (and for the precognitive approach, they already have been) for gaining a better understanding of psychopathology and how to treat it (Panksepp et al. 2017, Berridge 2018, Fanselow 2018, Fanselow and Pennington 2018).

### Implications of the disagreements between approaches to studying the NCeC

The two approaches have some important points of disagreement. The primary theoretical point on the nature of emotional consciousness is whether the experience of emotion requires (or not) cognitive elaboration to take place. The “cognitive” approach proposes that it does, either in the form of a hierarchy of higher-order representations in WM (e.g. LeDoux and Brown 2017), or via the competition of first-order representations for the global broadcasting in WM (e.g. Smith and Lane 2015), or through the top-down categorization of bodily states in the world via emotional concepts (e.g. Barrett 2017c). The “preconscious” approach proposes that cognition is not needed, either because the phenomenal preconscious experience is automatically elicited by the activation of central emotion generators in the brain and differs from the ability to reflect upon this experience (access consciousness) (e.g. Ekman 1992, Panksepp 2007, Izard 2007b) or because emotions can be fully unconscious when the stimulation of these circuits is not strong enough to produce conscious emotional experiences (Winkelman and Berridge 2004, Anderson and Adolphs 2014, Fanselow 2018). Notably, this distinction can also be seen from the perspective of whether emotional consciousness and its neural correlates are largely shared with other contents of consciousness and mostly cortical (cognitive approach) or are specific to emotion and mostly subcortical (preconscious approach).

Overall, such disagreements have far-reaching consequences concerning the study of emotion (for recent discussions, see Panksepp et al. 2017, Mobbs et al. 2019, Paul et al. 2020), such as on (i) the neuroimaging techniques that have the capacity of providing evidence for the neural correlates of emotion (non-invasive techniques, such as electro/magnetoencephalography (E/MEG), functional magnetic resonance imaging (fMRI), near-infrared spectroscopy, transcranial magnetic stimulation, transcranial direct current stimulation, and transcranial alternating current stimulation, which have no or limited access to subcortical structures, vs. invasive techniques, such as intracranial EEG, intracranial stimulation and lesion studies, which can access these structures); (ii) the measures that best capture what we mean by emotion (subjective reports vs. behavioral/physiological responses); (iii) the pertinence of developmental studies, studies on patients with extended cognitive deficits, and animal studies; and (iv) the potential for developing treatment for mental disorders, such as depression and anxiety. Due to their extensive implications, the consequences of building theories of emotional consciousness on preliminary evidence can lead to particularly disturbing closed-ended impasses. This is well exemplified when comparing strong statements by authors on both sides, which are in explicit contradiction. For instance, for LeDoux and Hofmann (2018, p. 67), “the most direct way to assess conscious emotional

feelings is through verbal self-report.” In the very same year, [Adolphs and Anderson \(2018\)](#) stated that “A science of emotion should, in the first instance, use behavior, cognition, and neurobiology in its vocabulary. It should not be based on self-report of feelings in people” (p. 51). Of note, not all authors who adopt a cognitive vs. precognitive approach to NCEC would fully commit to these statements, and we think that many bridges between approaches exist. To give an example, some of the theories that we assigned to either the cognitive (e.g. GNW; [Smith and Lane 2016](#)) or the precognitive (e.g. [Winkelman and Berridge 2004](#), [Celegnin et al. 2015](#)) approach, which disagree on the centrality of cognition in emotional consciousness, agree on the other hand on the existence of both conscious and unconscious emotions and do not fully disregard some measures over others. Here, we propose possible ways to cross bridges even further by (i) identifying the methodological specificities of the study of NCEC compared to other NCC, (ii) highlighting the limitations to some of the claims of existing theories of NCEC, and (iii) exploring the methodological solutions for testing specific hypotheses to build theories on a more solid ground in the future.

## Are times mature to build theories of emotional consciousness and its neural correlates?

### Are separate theories of emotional consciousness and its neural correlates needed?

Let us start with a global reflection concerning theories of emotional consciousness. A principle of parsimony would suggest not to bother building separate theories of emotional consciousness and its neural correlates, unless evidence supports the notion that consciously experiencing an emotion differs in some way from other forms of experience, such as seeing red, feeling a fatigued muscle after exercise, or feeling thirsty. Intuition-wise, emotional phenomenal experience might seem to some of us qualitatively different from other conscious experiences, simply because watching our children making their first steps feels different from watching anybody else walking. Following this intuition makes the “hard problem” of consciousness ([Chalmers 1995](#)) even harder when it comes to emotional experience, as there seems to be a “something extra” to emotional consciousness that needs to be explained. As we have seen, across approaches, several authors provide an explanation for this “something extra,” arguing, e.g. that it is the activation of a number of so-called “limbic” areas and connections, together with its cognitive, physiological, and behavioral consequences, that is specific to the emergence of the conscious experience of emotion, irrespective of the differences on whether these areas fully qualify as NCEC ([Winkelman and Berridge 2004](#), [Anderson and Adolphs 2014](#), [Berridge 2018](#), [Fanselow 2018](#), [LeDoux 2020a](#)). Other authors insist more on the activation of brain structures that allow for our own perception of changes in bodily states in a given situation as the unique mechanism that separates emotional from other forms of consciousness (e.g. [Seth 2013](#), [Smith and Lane 2015](#)). Nonetheless, it is noteworthy that other theorists explicitly reject both the idea that there are dedicated emotional circuits in the brain and that feelings coming from the perception of bodily changes in the world are specific to emotion ([Barrett 2017b](#)), aligning with theories that renounced at distinguishing emotional from other forms of experience at the mechanistic level ([Russell 2003](#), [Moors 2022](#)), thus apparently questioning the necessity of separate theories of NCEC.

Another argument that is brought in favor of a special relationship between emotion and consciousness is that emotions are often thought to be deeply connected with unconscious aspects

of our mental life. As we have briefly introduced, some research showed that, even when not consciously perceived, emotional stimuli can elicit physiological responses and bias behavior, similarly (but not identically) to when the stimulus is consciously perceived. In a phenomenon called affective blindsight, cortically blind patients can identify the emotional expressions of faces above chance level but are incapable of doing the same for facial attributes unrelated to the emotional expression, such as identity ([Rossion et al. 2000](#)), potentially arguing for a special status for non-conscious emotional perception ([Tamietto and de Gelder 2010](#)). In healthy participants, such findings are typically supported by adapting psychophysical paradigms commonly used in the study of visual consciousness, such as backward masking, binocular rivalry, or continuous flash suppression ([Kim and Blake 2005](#)). Results from these paradigms, which manipulate conscious access to emotional stimuli, show a “preferential access to awareness” of emotional stimuli. For instance, there is a dominant viewing time for fearful faces in binocular rivalry ([Amting et al. 2010](#)) and times at which fearful faces break through binocular suppression are shortened ([Yang et al. 2007](#)). Once again, it is noteworthy that these findings are not unchallenged. Methodologically, it has been shown that differences between paradigms in the way stimulus awareness is suppressed influence emotion priming effects ([Faivre et al. 2012](#)). Furthermore, some have shown that, when stimulus awareness is not inferred by stimulus duration, but it is based on actual subjects’ report, affective categorization of emotional stimuli is not better than chance in the absence of stimulus awareness ([Lähteenmäki et al. 2015](#)) and AMY’s activation does not differ for the presentation of fearful and neutral faces ([Pessoa et al. 2006](#)). By considering the reported awareness, it has also been shown that only some physiological systems respond to unaware emotional vs. neutral stimuli in a continuous flash-suppressing paradigm ([Tooley et al. 2017](#)). Other authors more generally argued that when fully controlling for stimulus awareness, neither behavioral nor physiological responses to emotional vs. neutral stimuli are observed (for a review, see [Tsikandilakis et al. 2021](#)). Mixed findings and opposing conclusions concerning unconscious emotional processing ([Mertens and Engelhard 2020](#), [Rohr and Wentura 2021](#)) indicate that these mechanisms are still partially undetermined.

Definitive evidence either in support of or against the necessity of building separate theories of emotional consciousness is still lacking, as the debate on the “something extra” to emotional consciousness, as well as on the specificity of unconscious emotional processing, wages on. However, and most importantly for the scope of the present paper, we will argue that irrespective of one’s opinion on the matter, the study of emotional consciousness comes with unique methodological challenges, which we will further elaborate in the following sections. More than the proven necessity of a separate theory of NCEC, it is the specificity of the methodological challenges to the study of emotional consciousness that justifies, in our opinion, that this research gets its own empirical and theoretical attention before it can be clustered together with the existing theories of consciousness.

### Can content-specific NCEC be separated from neural prerequisites and consequences?

Across both approaches, a number of theories of emotional consciousness were predominantly built on evidence coming from research on single emotions, mostly on fear (e.g. [Anderson and Adolphs 2014](#), [Celegnin et al. 2015](#), [LeDoux and Brown 2017](#), [Fanselow 2018](#)), despite this not being the case for all theories ([Winkelman and Berridge 2004](#), [Panksepp 2007](#), [Izard 2007b](#), [Seth](#)

and Critchley 2013, Smith and Lane 2015, Barrett 2017b). This emphasis on fear research is understandable, as it derives from the robustness and reliability across species of fear/threat-related paradigms, such as fear/threat conditioning (e.g. Büchel and Dolan 2000, Delgado et al. 2006, Fullana et al. 2020). Nonetheless, this solution comes with an important limitation, namely the possibility to make a distinction between the content-specific NCC and the full NCC (Koch 2004, Koch et al. 2016). Content-specific NCC are defined as the neural substrates of specific phenomenal characteristics within an experience, e.g. the unique experience of seeing a face. Thus, the content-specific NCC differ from the so-called full NCC, i.e. the neural substrates of consciousness experience in its entirety, irrespective of the specific content of experience, meaning the combination of content-specific NCC for all possible contents of experience (Koch et al. 2016). To isolate content-specific NCC, typically, in the perceptual consciousness literature, contrastive approaches are used, for instance comparing brain activation when consciously perceiving a face vs. not consciously perceiving it, relying on the experimental paradigms introduced in the previous section. Importantly, a known shortcoming of such a contrastive approach is that it reveals not only content-specific NCC for perceiving a face but also all neural substrates preceding and following conscious perception, the so-called neural prerequisites and neural consequences of consciousness (Aru et al. 2012, de Graaf et al. 2012).

On the one hand, the neural prerequisites of consciousness refer to the mechanisms that are necessary for the conscious experience to arise, as, for instance, in some circumstances, the fact of directing attention toward the stimulus, but are not the determinants of the content of the phenomenal experience (de Graaf et al. 2012). On the other hand, neural consequences are the aftereffects associated with a given phenomenal experience, for instance, when an episodic memory automatically comes to mind after consciously perceiving an object, like the Proustian “madeleine” (de Graaf et al. 2012). One proposed way of separating the different types of NCC relates to the notions of content invariance and content specificity. In more detail, if a neural substrate is involved in the emergence of two distinct phenomenal experiences (content invariance), it cannot explain the subjective difference between the two (content specificity) and thus is more likely to be a prerequisite or a consequence than a content-specific NCC (de Graaf et al. 2012).

We argue that these notions directly apply to the search for NCC. For instance, it is quite undebated that global affect dimensions such as arousal, valence, and action tendencies participate to different extents in the emergence of each and every one of our emotional experiences (Lang and Bradley 2010), therefore, applying the abovementioned logic, qualifying as neural prerequisites/consequences of emotional experience, rather than as content-specific NCC. This is in line with what is proposed by predictive coding theories within the cognitive approach (Seth 2013, Seth and Critchley 2013, Barrett 2017b, 2017c), which indeed deny the existence of content-specific NCC of what we call “emotions,” as emotions are learned concepts that vary across individuals and cultures, and therefore show idiosyncratic brain activation in similar emotional circumstances (“in those theories, variability is assumed to be the norm, rather than a nuisance to be explained after the fact”; Barrett 2017a, p. 9). But even if we refer to the theories that admit the existence of specific NCC, the focus on fear over other emotions might have involuntarily produced a bias over the importance of specific neural structures over others. As an example, the amygdala and its subcortical/cortical connections are at the core of a number of theories in both

approaches, either as an important central emotion generator (Anderson and Adolphs 2014, Celeghin et al. 2015, Adolphs and Anderson 2018, Fanselow 2018, Fanselow and Pennington 2018), responsible for both conscious and unconscious fear, or at least as a necessary determinant for qualifying the conscious experience as “emotional” (see “emotional self-HOROR”; LeDoux and Brown 2017). Not surprisingly, and in our opinion partially due to the fear-centered lenses and to the methodological difficulty in distinguishing content-specific NCC from prerequisites/consequences, it is debated whether this AMY-centered network in humans is really fear/threat specific (Méndez-Bértolo et al. 2016, Burra et al. 2019, McFadyen et al. 2019) or it generally encodes affect dimensions such as arousal (Lin et al. 2020), action relevance (Guex et al. 2020), or stimulus valence (Kragel et al. 2021).

Partially supporting the confusion between content-specific NCC and its prerequisites/consequences, research focused on other phenomenal emotional experiences, such as disgust, has found quite remarkably different neural correlates (e.g. the ventral striatum and the insula; Chapman and Anderson 2012, Berridge 2018). Importantly, a similar confound applies to any research focusing on only one content of experience. Back to disgust and to the involvement of insula and generally interoceptive cortices (Chapman and Anderson 2012), many authors agree that interoception is central in a wide range of emotional experiences (e.g. Zaki et al. 2012, Pavuluri and May 2015), as we draw upon signals coming from our body to understand how we feel, and might for the very same reason as before not fully qualify as NCC. It has to be noted, to avoid oversimplification, that the idea that the entire range of emotions that we are capable of experiencing would map onto one single brain area or network is refuted by both cognitive and precognitive approaches and irrespective of whether the authors believe emotions to be discrete or continuous in nature. For example, authors neither opposing (Clark-Polner et al. 2016) nor supporting (Saarimäki et al. 2016, 2018) the existence of discrete neural signatures of emotions would argue that all kinds of emotional experience can be mapped onto a single generalized network. It is therefore at odds that approaches to emotional consciousness sometimes overlook the necessity of fully integrating research on different contents of emotional consciousness before theory building. These premises made how to exactly define what constitutes an emotional content is far from being resolved as, for instance, the debate on whether emotions are continuous or discrete in nature is still lively (Celeghin et al. 2017, Barrett 2017a). The way we define an emotional content can directly impact our methods and conclusions we draw from the results. For instance, it has been shown that assigning experimental stimuli to specific emotional categories has an impact on the performance of supervised machine learning algorithms that look for their brain correlates and that the same categories are sometimes not retrieved with non-supervised algorithms (Azari et al. 2020). The need for clarity in content definitions in the search for NCC, which we will develop further throughout the paper, calls for even more caution when constructing theories of emotional consciousness.

### Are the NCC cortical or subcortical?

When it comes to the specific claims on NCC, the cognitive approach argues that, in order for a conscious emotional experience to exist, domain-general cognitive mechanisms, such as WM and attention, need to be recruited. In support of this claim, subcortical activations are proposed to be insufficient for conscious emotional experiences to arise and cortical activations are thought to be needed, namely activation in a prefrontal-parietal

network (Smith and Lane 2015) or more precisely in specific regions of the prefrontal cortex (PFC) (Seth 2013, LeDoux and Brown 2017), which has indeed been found to correlate with self-reported emotions (Williams et al. 2006). Criticisms concerning the centrality of the prefrontal cortex in conscious experience have already been raised in the domain of perceptual consciousness, resulting from the findings obtained with the so-called no-report paradigms (Tsuchiya et al. 2015). As the name indicates, these paradigms require no subjective report from participants and rely on the idea that conscious contents can be inferred from physiological and behavioral changes, within specific experimental conditions. For instance, by inferring the perceived direction of competing moving stimuli under binocular rivalry from eye movements, rather than from subjective reports, it has been shown that frontal activations that were present during the perceptual transition in the report task were absent in the no-report variant of the task (Frässle et al. 2014). While the debate on the necessity of the prefrontal cortex in perceptual consciousness continues (Boly et al. 2017, Odegaard et al. 2017, Block 2019, Northoff and Lamme 2020), emotional theories of consciousness claiming for a role of the prefrontal cortex face on this point the same methodological challenges as their non-emotional counterparts.

However, when it comes to the involvement of the prefrontal cortex in the NCC, we argue that an additional methodological challenge exists. Namely, it is not clear yet how a no-report paradigm could be adapted to an emotional context. In the emotional domain, even if the participant is “aware” of a fearful face or of an attacking snake, one cannot unambiguously assume that they are also “experiencing” a corresponding emotion. A direct example of how no-report paradigms might be challenging to adapt to emotional stimuli comes from the work by Vetter et al. (2019). In their study, the authors presented both angry and fearful faces using a continuous flash suppression paradigm. They showed that, even in the absence of awareness of the face stimulus, eye movements were still influenced in a seemingly goal-directed fashion, notably deviating toward the fearful face and away from the angry one (Vetter et al. 2019). These findings, which align with the results for spontaneous approach/avoidance decision to task-irrelevant emotional faces (Mennella et al. 2020, 2022, Vilarem et al. 2020, Grèzes et al. 2021), show how emotional information can drive adaptive behavior in the absence of conscious experience, challenging a systematic inference of emotional states from behavior. Overall, in light of the possibility that the PFC might not be necessary for the phenomenal experience, but to the capacity to report it, we suggest that rigorous evidence in the support of the necessity of prefrontal areas to form conscious emotional experiences is lacking, in particular because no emotional counterpart to the no-report paradigm has been conceived at present.

It is important to point out that the sufficiency of subcortical structures for the emergence of emotional consciousness, as proposed by the precognitive approach, is equally debated. Possibly due to the abovementioned fear-related bias, a longstanding question in the domain has been whether conscious fear can or cannot be experienced without intact AMYs. The precognitive approach typically refers to lesion studies in which bilateral AMY damage impairs the recognition of fearful faces, the conscious experience of fear, and fear-related avoidance behaviors (Adolphs et al. 1994, Feinstein et al. 2011), as well as AMY’s stimulation studies in humans that elicited in some cases conscious experiences of fear and anxiety (e.g. Lanteaume et al. 2007, Inman et al. 2020). On the other hand, the authors within the cognitive approach more often refer to the studies in which patients with

AMY damage were still able to experience conscious fear (Anderson and Phelps 2002, Feinstein et al. 2013, 2016). Mixed findings are not surprising, however, for both methodological and theoretical reasons. Methodologically, irrespective of the value of patient and lesion data in providing causal insight into brain functioning (Adolphs 2016, Vaidya et al. 2019), such experiments come with limitations. Congenital lesions often result in plastic restructuring of the brain (Wieloch and Nikolich 2006), and stimulation studies can be difficult to generalize, as different stimulation parameters (such as polarity and intensity) create varying electric fields in different patients (Selimbeyoglu 2010). Theoretically, and more relevant for the present discussion, fear of exteroceptive threats (e.g. a predator) and of interoceptive ones (e.g. hypercapnic states due to exaggerated CO<sub>2</sub> levels), despite falling under the same “fear” label, is now known to rely on partially dissociable brain mechanisms, and AMY shows a different involvement in the two types of fear (for a recent review, see Feinstein et al. 2022), which likely explains some of the mixed findings. This exemplifies once more how ambiguity and disagreement in the definition of emotional concepts (e.g. fear) might artificially lead to conflicting and seemingly irreconcilable interpretations of the same research results.

As we have argued throughout this section, we believe that, due to (i) the lack of strong evidence against or in favor of the specificity of emotional vs. non-emotional consciousness, (ii) the present difficulty in disambiguating content-specific NCC from neural prerequisites/consequences, and (iii) unresolved issues substantiating either the predominantly cortical or subcortical nature of NCC, there is no clear need of committing to either cognitive or precognitive approaches to emotional consciousness nor to accept (or discard) one theory over another, among the proposed ones. This should, in our opinion, free researchers from some of the abovementioned impasses, which might lead to the exclusion of specific research fields, methods, or techniques in the search for NCC, pushing them to collaboratively refine concepts and find agreements on methods (see the “Conclusions and future directions” section). In the next section, we discuss three fundamental experimental contrasts that are commonly employed in the search for NCC, focusing on what can and cannot be inferred from each of them. We hope that this discussion can contribute to a shared methodological ground for future theory building and research.

## Experimental contrasts and what can(not) be inferred

We propose that there are three main experimental contrasts that are crucial for hypothesis testing on the NCC. These are widely employed in the literature, but what we can conclude from them often remains implicit. We here directly point out the interpretive value of each contrast, the extent to which it can eventually be adapted from the perceptual to the emotional consciousness field, and its specificities (and shortcomings) to the study of the NCC. We argue that, given the limitations of any of the contrasts taken alone, the combination of each of them is crucial to further our understanding of emotional consciousness and its neural correlates. Of note, for simplicity, our discussion focuses on the study of emotional consciousness in response to external stimuli, but we by no means disregard the fact that emotional responses and subjective experience might be caused by fully internal factors, both cognitive (e.g. thoughts and memories) and physiological (e.g. hormonal changes, physical fatigue, and inflammation) factors.

### Contrast 1: subliminal versus supraliminal perception of the emotional stimulus

The contrast between subliminally and supraliminally perceived stimuli is widely used in the studies of both perceptual and emotional consciousness (Mitchell and Greening 2012), which we already introduced in the previous sections. In paradigms such as masking or binocular rivalry, an emotional stimulus can alternate between being perceived by the participant (supraliminal) or not (subliminal) on a trial-by-trial basis. The report of the participant in these cases, if any is recorded, is whether they perceived the stimulus (in cases of masking or continuous flash suppression) or which stimulus they saw (in case of binocular rivalry, meaning that the non-reported stimulus was suppressed from awareness). These two types of trials are then contrasted to infer which areas were additionally activated by the conscious perception of the stimulus (Kim and Blake 2005, de Graaf et al. 2012). As highlighted earlier, this contrast alone is limited in how much insight it can give into the NCeC in two ways. First, whatever neural correlates emerge from this contrast might not be content specific and could rather be a reflection of a prerequisite or a consequence of emotional consciousness (Aru et al. 2012, de Graaf et al. 2012). Second, and more importantly, the subliminal versus supraliminal perception of the stimulus, albeit emotional, does not provide information about whether a corresponding emotion was felt. On the one hand, awareness of the stimulus and the emotion can be dissociated from one another, as not all emotional stimuli consciously perceived provoke subjective emotional experiences. On the other hand, as previously discussed, unconsciously perceived emotional stimuli might influence emotional perception, responses, and decisions (Tamietto and de Gelder 2010, Celeghin et al. 2015), including judgments of stimulus valence (Anderson et al. 2012), while possibly not conscious emotional feelings (Winkielman et al. 2005, Winkielman and Gogolushko 2018). This aspect is also intimately connected with the notion of the respective timescales of perceptual vs. emotional experiences, which adds another level of complexity. On the one hand, it is still debated whether the perceptual conscious experience of visual stimuli correlates with “early” neural activity, around 200 ms after stimulus presentation, i.e. the visual awareness negativity component of the event-related potentials (ERP), or with “late” ERP components, such as the P3 and late positivity (for a recent review, see Förster et al. 2020). On the other hand, consciousness of emotional stimuli (emotional faces) might relate in a different and specific way to the early and late ERP components and their neural substrates, as compared to neutral stimuli (e.g. Sun et al. 2023). Crucially, this whole literature, which we only briefly mention here, moderately informs regarding the timescale of the emergence of the subjective emotional experience, which seems to us much less investigated. This might be due to multiple factors, not least the fact that the timing of the conscious emotional experience after discrete stimuli is inevitably confounded with the time and the neural activity needed to report it. While some studies did investigate the neural correlates of the transition between emotional vs. neutral conscious states, via repeated emotional stimuli presentation, the time needed for the instantiation of the emotional experience, as well as its specificity compared to mood changes for instance (Eldar et al. 2021), deserves further attention. Overall, these considerations point out some of the unique limitations to the use of the “Subliminal versus supraliminal perception” contrast in the search for NCeC. Therefore, the addition of the following contrasts is necessary.

### Contrast 2: self-reported experienced (felt) vs. not experienced (unfelt/different) emotion

To be able to reveal content-specific correlates of emotional consciousness, conditions in which a stimulus elicits a reportable emotional experience need to be contrasted with conditions in which a stimulus elicits either an alternative or an absence of reportable emotional experience. Unlike the previous contrast, this relies on supraliminal presentation, while varying, e.g. stimulus content. This is a founding method in affective neuroscience, which makes use of countless databases of stimuli of different natures (e.g. pictures, sounds, and imagery scripts) that are validated for their capacity to elicit, on average, different emotional experiences in the perceiver. Let us examine the limitations inherent to this contrast for the study of the NCeC. First, if neural activation is compared following a stimulus inducing one conscious emotional experience (e.g. fear) vs. another (e.g. sadness), one can learn about different correlates of the two emotions but cannot conclude much about what determines their conscious experience (an “unfelt” control condition is lacking). Second, if neural activation is compared following a stimulus inducing one conscious emotional experience (e.g. fear) vs. a neutral state, the neural differences will contain the activity needed for the conscious emotion, but it will be confounded by the difference between stimuli, both sensorial and affective ones, such as arousal and valence (see Gasper et al. 2019 for further discussion on the use of neutral states as a baseline in contrasts). Third, for its interpretation, this contrast relies on self-report, through which participants somehow declare whether they consciously experienced an emotion and which one. Interestingly, a recent large-scale survey among researchers in the field of consciousness revealed that, although generally aware of their possible biases in measuring the content of experience, as well as of the abovementioned influence of self-reporting on neural activation, researchers overall declared subjective reports to be their preferred method to measure consciousness (Francken et al. 2022). It is true that, whereas indeed accuracy of self-report based on metacognition abilities has been shown to be poor (Nisbett and Wilson 1977), both consciousness and emotional consciousness are first-person experiences and might be quite accurately accessible through introspection (Overgaard and Sandberg 2012). Nonetheless, previous research in emotion has highlighted that the validity of self-report is influenced by the way scales are constructed (e.g. dimensional vs. discrete), by where the focus of introspective attention is placed, or by the amount of elapsed time before the experience is captured (Robinson and Clore 2002, Jack and Roepstorff 2003, Overgaard and Sandberg 2012). Furthermore, for the study of emotional consciousness, an additional level of difficulty exists, as a given stimulus might evoke not only a feeling of fear but also feelings of anger, panic, sadness, or no feeling at all. Different emotional states can coexist at the same time, a known phenomenon called “dialecticism” (Bagozzi et al. 1999, Lindquist and Barrett 2008). Therefore, even if participants were immediately asked to indicate how fearful they felt on a continuous scale, we might be missing a great deal of relevant conscious experiences. A participant might even be primed to believe that what they had felt was fear by framing the report measure in a particular way. A previous work has already shown that the mere act of asking a participant to report on their feelings can change both physiological and neural responses to a particular stimulus or task (Creswell et al. 2007, Lieberman et al. 2007, Kassam and Mendes 2013). Recent research advancements are putting forward exciting new solutions to try



to overcome some of these limitations. For instance, the use of continuous naturalistic stimuli, such as movies, has the power to profoundly modify experience (Kovarski et al. 2022), including emotion (Saarimäki 2021), which can be measured over time on multiple dimensions, without loss of complexity. Continuous behavioral reports of affective (e.g. Smirnov et al. 2019) or discrete emotional states (e.g. Hudson et al. 2020) from the participant can be used, but this comes with the neural confound associated with self-report. To circumvent this, ratings from independent samples can be collected, or automated emotion-feature extraction can be relied upon (Kragel et al. 2019). While each and every one of these methods comes with limitations (for a detailed review, see Saarimäki 2021), these advancements nicely demonstrate where the additional challenge lies when studying emotional consciousness, as well as the possible future solutions. The presented limitations for inference relative to the second contrast bring us to the third and final contrast.

### Contrast 3: presence/absence of a behavioral/physiological response

As a third point, conditions in which a stimulus elicits a given behavioral/physiological response need to be contrasted with conditions in which a stimulus elicits either an alternative or no behavioral/physiological response. As presented in the first section, the precognitive approach supports the idea that behavioral and physiological responses can be a readout of emotional experience in animals and in humans, which cannot verbally report experience (e.g. infants and patients) and, for some authors, represent the sole readout of phenomenal (as opposed to access) consciousness (Panksepp et al. 2017). For instance, as far as behavior is concerned, it has been proposed that it is possible to infer from animals' defensive behavior (e.g. freezing vs. vigorous escape attempts) their corresponding emotional experience (e.g. fear vs. panic, respectively) (The Predatory Imminence Continuum Model; Fanselow and Lester 1988, Fanselow et al. 2019, Mobbs et al. 2020). While the tight link between emotion and action is undebated, possible limitations to this kind of model are brought by converging research on defensive behavior in animals (e.g. Vale et al. 2017, Evans et al. 2019) and humans (Rotteveel and Phaf 2004, Schlund et al. 2016, Mennella et al. 2020, Vilarem et al. 2020). Such research supports the notion that defensive behavior is flexible and not stereotyped in response to threatening stimuli, resulting from a complex and still partially unexplored interaction of stimulus-driven reactions (e.g. reflexes and automated reactions) and rapid—sometimes unconscious—goal-directed responses (Moors et al. 2017, LeDoux and Daw 2018, Mendl and Paul 2020). This is corroborated by the fact that goal-directed approach/avoidance behaviors to emotional stimuli can be elicited in the absence of a reportable strategy for action (Mennella et al. 2022) and in the absence of stimulus awareness (Vetter et al. 2019). A similar discourse applies to physiological responses, which are indeed typically well correlated to subjective reports of emotion at the group level (Friedman et al. 2014, Taschereau-Dumouchel et al. 2020) but not always to a great extent at the individual level: early studies on the relationship between behavioral ratings and physiological responses showed that a correlation between zygomatic and corrugator muscle responses with the rated experienced valence reached significance in roughly 50% of the participants, and the correlations between skin conductance responses and experienced arousal reached significance in around 30% of the sample (Lang et al. 1993). Such findings, together with the fact that, as discussed earlier, consciously undetected emotional stimuli can elicit physiological responses, have raised

the question of whether subjective experience and physiological responses are subtended by the same brain mechanisms. Using multivoxel pattern analysis, Taschereau-Dumouchel et al. (2020) indeed showed that, despite a significant correlation between reported fear and skin conductance responses at the group level, some areas were differentially involved in the prediction of the two measures. Other recent work demonstrated that when modeling what contributes to subjective ratings that participants give in response to affective images, both physiological responses and neural (interoceptive) markers explain unique parts of the variance observed in self-reports (Engelen et al. 2023). This means that although physiological responses did make a significant unique contribution to self-reports, they did not account for all of the variance observed in such ratings. Overall, there is evidence that specific behavior and physiological responses can be more or less correlated with subjectively reported emotional experiences, and individual differences, as well as the intensity of the emotional situation, have been found to influence the strength of this correlation, which is named “emotional coherence” (Mauss et al. 2005). It is noteworthy that increased emotional coherence has been related to well-being (Brown et al. 2020), which aligns with the fact that emotion-related disorders, such as anxiety and depression, are undoubtedly “mental” disorders (Taschereau-Dumouchel et al. 2022) but also behavioral and physiological ones (American Psychiatric Association 2013). This is why the third contrast is necessary to complement the first two and cannot be disregarded as not pertinent for the search of NCEC.

### Conclusions and future directions

We hope to have defended the idea that researchers should feel unconstrained with respect to the existing approaches to emotional consciousness, which, despite their value in pushing toward the formation of integrated and meaningful pictures of the existing research, at present rely on preliminary findings, which do not allow to arbitrate unambiguously between them. We insist that this state of things is not necessarily specific to the NCEC literature (Oberauer and Lewandowsky 2019). Across many domains in psychology and neuroscience emerges the prevalence of a strong hypothetico-deductive method, which focuses on scientific progress as the repeated empirical test of hypotheses entailed by theories, while putting less attention upon the use of systematic and collective methods to developing theories in the first place (Borsboom et al. 2021). Another possibility is to conceive the relationship between phenomena, data, and theories as a circular one, in which, before getting to the formulation of a complex theory, several steps are taken, starting with the identification of relevant phenomena and going through the initial formulation of preliminary explanatory models (prototheories), involving a small set of general principles that putatively explain the phenomena of interest (Borsboom et al. 2021). Theory construction in this sense builds on a range of non-confirmatory activities, including descriptive research and exploratory experimentation, which are crucial to reach consensus on concept definition and measures' validity, among the elements that are necessary for theory building (Scheel et al. 2021). Back to the specific complexity of the study of NCEC, as we have detailed, methods cannot always be directly translated from the field of perceptual consciousness to the study of NCEC. For example, no-report paradigms, which have led to new insights in the search for the NCC, cannot at present lend themselves in the same way for studying emotional consciousness. Likewise, the use of subjective reports in emotional consciousness might need further methodological elaboration for the study of NCEC.

as compared to NCC, due to, among other things, the dialectical nature of emotions. Lastly, behavioral and physiological responses play a possibly privileged role in the search for NCC, as compared to NCC, as emotional coherence varies profoundly across individuals and situations.

Throughout the paper, we presented what we think are the necessary advancements to arrive to the separation of content-specific NCC. These pertain to (i) how an emotional content of consciousness is defined and (ii) how the experimental contrasts must be combined to avoid confounding factors to the identification of the NCC. Concerning concept definition, cooperative work in the form of adversarial collaborations is a way in which the energy of different laboratories (even adhering to different theories) can be combined. An adversarial collaboration is a multi-lab effort in which one prediction stemming from a particular theory is empirically tested. Such a collaboration has already been suggested (Seth and Bayne 2022) and set up in the field of consciousness (Melloni et al. 2021), where predictions stemming from GNW and integrated information theory will be directly compared against one another. Most interestingly, an adversarial collaboration has been successfully applied to the field of emotions as well to investigate how facial mimicry and voluntary facial activation can induce feelings of happiness (Coles et al. 2022). Aside from adversarial collaborations, data-driven approaches are particularly promising as they might play a part in the definition of emotion ontologies (Jack et al. 2018), thereby assisting in coming to a consensus on how to define emotional contents, in a more agnostic manner. For example, research into mapping the semantic space of emotional experience using several self-report measures in response to a wide array of emotionally evocative videos was able to identify 27 distinct categories of emotional experiences (Cowen and Keltner 2017). Data-driven approaches have likewise been used to define the characteristics that define dynamic facial and bodily movements (Jack et al. 2014, de Gelder and Poyo Solanas 2021), as well as bodily sensation maps associated with subjective feelings (Nummenmaa et al. 2014).

Concerning the proposed contrasts, individual efforts focused on combining the three experimental contrasts for uncovering the NCC can possibly benefit from some methodologies newly introduced to the study of emotion, such as computational models of behavior (Roberts and Hutcherson 2019). These methods might help give a more mechanistic account of the computational involvement of distinct brain areas and possibly in the search for NCC. For instance, recent work combining high-field fMRI, continuous flash suppression, and drift-diffusion modeling showed how the faster breaking through flash suppression of fearful faces is due to more rapid perceptual evidence accumulation, associated with activity in frontoparietal regions, occipital lobe, and AMY. Activity in other areas, such as the insula and posterior cingulate cortex, was rather correlated with a lower decision boundary (Kalhan 2022). Finally, the synthesis of existing data would make the comparison of all the different contrasts (which are unlikely to be addressed in one single study) feasible. Large-scale meta-analyses have become more and more achievable, thanks to the diffusion of open science practices, with the added advantage of increasing the statistical power and determining the consistency of effects (Yarkoni et al. 2010). Again, in the field of emotional consciousness, such an approach has already been shown to be informative, for instance, to summarize evidence about physiological responses to subliminal negative affective stimuli (van der Ploeg et al. 2017).

To draw a final parallel, in the field of perceptual consciousness, there is likewise still little agreement concerning the best

theory of consciousness, but at least there seems to be a tendency toward a consensus on which methods should be employed to study the NCC (Francken et al. 2022). Acknowledging the necessity and strengths of all different methods and paradigms at our disposal to study the NCC would be a major step forward, as the search for the NCC is likely to stagnate by being exclusionary.

## Acknowledgements

The authors would like to thank Gretty Mirdal, Julie Grèzes, Catherine Tallon-Baudry, Tom de Graaf, Minye Zhan, Klara Kovarski, Morgan Beaurenaut, Clémence Alméras, and Victor Chung for their support and valuable feedback on the manuscript. Finally, the authors acknowledge the essential contribution of the observations of Professors Anil K. Seth (editor) and Ryan Smith (reviewer), as well as an anonymous reviewer, to shape the manuscript in its present form.

## Conflict of interest:

None declared.

## Funding

T.E. was supported by funding from the European Research Council under the European Union's Horizon 2020 research and innovation program (grant agreement No 670325, advanced grant BRAVIUS). R.M. was benefited from a Gretty Mirdal Junior Fellowship at the Paris Institute for Advanced Study (France) and was supported by Fondation de France (grant number 00100076). Both authors benefited from the Agence Nationale de la Recherche grant ANR-17-EURE-0017 (EUR FrontCog).

## References

- Adolphs R. The biology of fear. *Curr Biol* 2013;**23**:R79–93.
- Adolphs R. Human lesion studies in the 21st century. *Neuron* 2016;**90**:1151–3.
- Adolphs R. How should neuroscience study emotions? By distinguishing emotion states, concepts, and experiences. *Soc Cogn Affect Neurosci* 2017;**12**:24–31.
- Adolphs R, Anderson DJ. *The Neuroscience of Emotion*. Princeton: Princeton University Press, 2018.
- Adolphs R, Mlodinow L, Barrett LF. What is an emotion? *Curr Biol* 2019;**29**:R1060–4.
- Adolphs R, Tranel D, Damasio H et al. Impaired recognition of emotion in facial expressions following bilateral damage to the human amygdala. *Nature* 1994;**372**:669–72.
- American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*, 5th edn. Washington: American Psychiatric Association, 2013.
- Amting JM, Greening SG, Mitchell DGV. Multiple mechanisms of consciousness: the neural correlates of emotional awareness. *J Neurosci* 2010;**30**:10039–47.
- Anderson DJ, Adolphs R. A framework for studying emotions across species. *Cell* 2014;**157**:187–200.
- Anderson AK, Phelps EA. Is the human amygdala critical for the subjective experience of emotion? Evidence of intact dispositional affect in patients with amygdala lesions. *J Cogn Neurosci* 2002;**14**:709–20.
- Anderson E, Siegel E, White D et al. Out of sight but not out of mind: unseen affective faces influence evaluations and social impressions. *Emotion* 2012;**12**:1210–21.

- Aru J, Bachmann T, Singer W *et al.* Distilling the neural correlates of consciousness. *Neurosci Biobehav Rev* 2012;**36**:737–46.
- Azari B, Westlin C, Satpute AB *et al.* Comparing supervised and unsupervised approaches to emotion categorization in the human brain, body, and subjective experience. *Sci Rep* 2020;**10**:20284.
- Baars BJ. Global workspace theory of consciousness: toward a cognitive neuroscience of human experience. *Prog Brain Res* 2005;**150**:45–53.
- Bagozzi RP, Wong N, Yi Y. The role of culture and gender in the relationship between positive and negative affect. *Cogn Emot* 1999;**13**:641–72.
- Barrett LF. Are emotions natural kinds? *Perspect Psychol Sci* 2006;**1**:28–58.
- Barrett LF. Categories and their role in the science of emotion. *Psychol Inq* 2017a;**28**:20–6.
- Barrett LF. *How Emotions Are Made: The Secret Life of the Brain*. Stuttgart: Pan Macmillan, 2017b.
- Barrett LF. The theory of constructed emotion: an active inference account of interoception and categorization. *Soc Cogn Affect Neurosci* 2017c;**12**:1–23.
- Barrett LF, Adolphs R, Marsella S *et al.* Emotional expressions reconsidered: challenges to inferring emotion from human facial movements. *Psychol Sci Public Interest* 2019;**20**:1–68.
- Barrett LF, Lindquist KA, Bliss-Moreau E *et al.* Of mice and men: natural kinds of emotions in the mammalian brain? A response to Panksepp and Izard. *Perspect Psychol Sci* 2007;**2**:297–312.
- Barrett LF, Satpute AB. Large-scale brain networks in affective and social neuroscience: towards an integrative functional architecture of the brain. *Curr Opin Neurobiol* 2013;**23**:361–72.
- Berridge KC. Evolving concepts of emotion and motivation. *Front Psychol* 2018;**9**:1–20.
- Berridge KC, Valenstein ES. What psychological process mediates feeding evoked by electrical stimulation of the lateral hypothalamus? *Behav Neurosci* 1991;**105**:3–14.
- Berridge KC, Venier IL, Robinson TE. Taste reactivity analysis of 6-hydroxydopamine-induced aphagia: implications for arousal and anhedonia hypotheses of dopamine function. *Behav Neurosci* 1989;**103**:36–45.
- Block N. On a confusion about a function of consciousness. *Behav Brain Sci* 1995;**18**:227–47.
- Block N. Two neural correlates of consciousness. *Trends Cogn Sci* 2005;**9**:46–52.
- Block N. Consciousness, accessibility, and the mesh between psychology and neuroscience. *Behav Brain Sci* 2007;**30**:5–6,481–99.
- Block N. What is wrong with the no-report paradigm and how to fix it. *Trends Cogn Sci* 2019;**23**:1003–13.
- Boly M, Massimini M, Tsuchiya N *et al.* Are the neural correlates of consciousness in the front or in the back of the cerebral cortex? Clinical and neuroimaging evidence. *J Neurosci* 2017;**37**:9603–13.
- Borsboom D, van der Maas HLJ, Dalege J *et al.* Theory construction methodology: a practical framework for building theories in psychology. *Perspect Psychol Sci* 2021;**16**:756–66.
- Brown R. The HOROR theory of phenomenal consciousness. *Philos Stud* 2015;**172**:1783–94.
- Brown CL, Van Doren N, Ford BQ *et al.* Coherence between subjective experience and physiology in emotion: individual differences and implications for well-being. *Emotion* 2020;**20**:818–29.
- Büchel C, Dolan RJ. Classical fear conditioning in functional neuroimaging. *Curr Opin Neurobiol* 2000;**10**:219–23.
- Burra N, Hervais-Adelman A, Celeghin A *et al.* Affective blindsight relies on low spatial frequencies. *Neuropsychologia* 2019;**128**:44–9.
- Celeghin A, de Gelder B, Tamietto M. From affective blindsight to emotional consciousness. *Conscious Cogn* 2015;**36**:414–25.
- Celeghin A, Diano M, Bagnis A *et al.* Basic emotions in human neuroscience: neuroimaging and beyond. *Front Psychol* 2017;**8**:1432.
- Chalmers D. Facing up to the problem of consciousness. *J Conscious Stud* 1995;**2**:200–19.
- Chapman HA, Anderson AK. Understanding disgust. *Ann NY Acad Sci* 2012;**1251**:62–76.
- Clark A. Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behav Brain Sci* 2013;**36**:181–204.
- Clark-Polner E, Johnson TD, Barrett LF. Multivoxel pattern analysis does not provide evidence to support the existence of basic emotions. *Cereb Cortex* 2016;**27**:1944–8.
- Coles NA, March DS, Marmolejo-Ramos F *et al.* A multi-lab test of the facial feedback hypothesis by the Many Smiles Collaboration. *Nat Hum Behav* 2022;**6**:1731–42.
- 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**:E7900–9.
- Craig AD. How do you feel—now? The anterior insula and human awareness. *Nat Rev Neurosci* 2009;**10**:59–70.
- Creswell JD, Way BM, Eisenberger NI *et al.* Neural correlates of dispositional mindfulness during affect labeling. *Psychosom Med* 2007;**69**:560–5.
- Crivelli C, Fridlund AJ. Facial displays are tools for social influence. *Trends Cogn Sci* 2018;**22**:388–99.
- Crivelli C, Fridlund AJ. Inside-out: from basic emotions theory to the behavioral ecology view. *J Nonverbal Behav* 2019;**43**:161–94.
- Damasio AR. *Descartes' Error: Emotion, Reason, and the Human Brain*. New York: Putnam, 1994.
- Damasio AR. Investigating the biology of consciousness. *Philos Trans R Soc Lond B Biol Sci* 1998;**353**:1879–82.
- Damasio AR. *The Feeling of What Happens: Body and Emotion in the Making of Consciousness*. Boston: Houghton Mifflin Harcourt, 1999.
- de Gelder B, Poyo Solanas M. A computational neuroethology perspective on body and expression perception. *Trends Cogn Sci* 2021;**25**:1–13.
- de Graaf TA, Hsieh P-J, Sack AT. The « correlates » in neural correlates of consciousness. *Neurosci Biobehav Rev* 2012;**36**:191–7.
- Dehaene S, Naccache L. Towards a cognitive neuroscience of consciousness: basic evidence and a workspace framework. *Cognition* 2001;**79**:1–37.
- Delgado MR, Olsson A, Phelps EA. Extending animal models of fear conditioning to humans. *Biol Psychol* 2006;**73**:39–48.
- Dennett DC. Welcome to strong illusionism. *J Conscious Stud* 2019;**26**:9–10,48–58.
- Dukes D, Abrams K, Adolphs R *et al.* The rise of affectivism. *Nat Hum Behav* 2021;**5**:816–20.
- Ekman P. An argument for basic emotions. *Cogn Emot* 1992;**6**:3–4,169–200.
- Eldar E, Pessiglione M, van Dillen L. Positive affect as a computational mechanism. *Curr Opin Behav Sci* 2021;**39**:52–7.
- Engelen T, Buot A, Grèzes J *et al.* Whose emotion is it? Perspective matters to understand brain-body interactions in emotions. *NeuroImage* 2023;**268**:119867.
- Evans DA, Stempel AV, Vale R *et al.* Cognitive control of escape behaviour. *Trends Cogn Sci* 2019;**23**:334–48.
- Faivre N, Berthet V, Kouider S. Nonconscious influences from emotional faces: a comparison of visual crowding, masking, and continuous flash suppression. *Front Psychol* 2012;**3**:129.
- Fanselow MS. Neural organization of the defensive behavior system responsible for fear. *Psychon Bull Rev* 1994;**1**:429–38.
- Fanselow MS. Emotion, motivation and function. *Curr Opin Behav Sci* 2018;**19**:105–9.

- Fanselow MS, Hoffman AN, Zhuravka I. Timing and the transition between modes in the defensive behavior system. *Behav Processes* 2019;**166**:103890.
- Fanselow MS, Lester LS. A functional behavioristic approach to aver- sively motivated behavior: Predatory imminence as a determi- nant of the topography of defensive behavior. In: Bolles RC and Beecher MD (ed.), *Evolution and Learning*. Hillsdale: Erlbaum, 1988, 185–212.
- Fanselow MS, Pennington ZT. A return to the psychiatric dark ages with a two-system framework for fear. *Behav Res Ther* 2018;**100**:24–9.
- Feinstein JS, Adolphs R, Damasio A et al. The human amygdala and the induction and experience of fear. *Curr Biol* 2011;**21**:34–8.
- Feinstein JS, Adolphs R, Tranel D. A tale of survival from the world of patient S. M. *Living without an Amygdala*, New York: The Guilford Press, 2016, 1–38.
- Feinstein JS, Buzza C, Hurlmann R et al. Fear and panic in humans with bilateral amygdala damage. *Nat Neurosci* 2013;**16**:270–2.
- Feinstein JS, Gould D, Khalsa SS. Amygdala-driven apnea and the chemoreceptive origin of anxiety. *Biol Psychol* 2022;**170**: 108305.
- Förster J, Koivisto M, Revonsuo A. ERP and MEG correlates of visual consciousness: the second decade. *Conscious Cogn* 2020;**80**: 102917.
- Francken JC, Beerendonk L, Molenaar D et al. An academic survey on theoretical foundations, common assumptions and the current state of consciousness science. *Neurosci Conscious* 2022;**2022**:niac011.
- Frässle S, Sommer J, Jansen A et al. Binocular Rivalry: Frontal Activity Relates to Introspection and Action But Not to Perception. *Journal of Neuroscience* 2014;**34**:1738–1747.
- Friedman BH, Stephens CL, Thayer JF. Redundancy analysis of auto- nomic and self-reported, responses to induced emotions. *Biol Psychol* 2014;**98**:19–28.
- Fullana MA, Dunsmoor JE, Schruers KRJ et al. Human fear con- ditioning: from neuroscience to the clinic. *Behav Res Ther* 2020;**124**:103528.
- Gasper K, Spencer LA, Hu D. Does neutral affect exist? How chal- lenging three beliefs about neutral affect can advance affective research. *Front Psychol* 2019;**10**:2476.
- Grèzes J, Erblang M, Vilarem E et al. Impact of total sleep deprivation and related mood changes on approach-avoidance decisions to threat-related facial displays. *Sleep* 2021;**44**:zsab186.
- Guex R, Méndez-Bértolo C, Moratti S et al. Temporal dynamics of amygdala response to emotion- and action-relevance. *Sci Rep* 2020;**10**:11138.
- Hudson M, Seppälä K, Putkinen V et al. Dissociable neural sys- tems for unconditioned acute and sustained fear. *Neuroimage* 2020;**216**:116522.
- Inman CS, Bijanki KR, Bass DI et al. Human amygdala stimula- tion effects on emotion physiology and emotional experience. *Neuropsychologia* 2020;**145**:106722.
- Izard CE. *The Psychology of Emotions*. US: Springer, 1991.
- Izard CE. Basic emotions, natural kinds, emotion schemas, and a new paradigm. *Perspect Psychol Sci* 2007a;**2**:260–80.
- Izard CE. Levels of emotion and levels of consciousness. *Behav Brain Sci* 2007b;**30**:96–8.
- Jack RE, Crivelli C, Wheatley T. Data-driven methods to diversify knowledge of human psychology. *Trends Cogn Sci* 2018;**22**:1–5.
- Jack RE, Garrod OGB, Schyns PG. Dynamic facial expressions of emo- tion transmit an evolving hierarchy of signals over time. *Curr Biol* 2014;**24**:187–92.
- Jack RE, Garrod OGB, Yu H et al. Facial expressions of emotion are not culturally universal. *Proc Natl Acad Sci U S A* 2012;**109**:7241–4.
- Jack AI, Roepstorff A. Why trust the subject? *J Conscious Stud* 2003;**10**:v–xx.
- Jackson JC, Watts J, Henry TR et al. Emotion semantics show both cul- tural variation and universal structure. *Science* 2019;**366**:1517–22.
- Kalhan S, McFadyen J, Tsuchiya N et al. Neural and computational processes of accelerated perceptual awareness and decisions: A 7T fMRI study. *Human Brain Mapping* 2022;**43**:3873–86.
- Kashdan TB, Barrett LF, McKnight PE. Unpacking emotion differenti- ation: transforming unpleasant experience by perceiving distinc- tions in negativity. *Curr Dir Psychol Sci* 2015;**24**:10–6.
- Kassam KS, Mendes WB. The effects of measuring emotion: phys- iological reactions to emotional situations depend on whether someone is asking. *PLoS One* 2013;**8**:e64959.
- Kim C-Y, Blake R. Psychophysical magic: rendering the visible ‘invis- ible’. *Trends Cogn Sci* 2005;**9**:381–8.
- Koch C. *The Quest for Consciousness: A Neurobiological Approach*. Green- wood Village: Roberts & Company Publishers, 2004.
- Koch C, Massimini M, Boly M et al. Neural correlates of consciousness: progress and problems. *Nat Rev Neurosci* 2016;**17**:307–21.
- Kovarski K, Dos Reis J, Chevais C et al. Movie editing influences spectators’ time perception. *Sci Rep* 2022;**12**:20084.
- Kragel PA, Čeko M, Theriault J et al. A human colliculus- pulvinar-amygdala pathway encodes negative emotion. *Neuron* 2021;**109**:2404–12.
- Kragel PA, Reddan MC, LaBar KS et al. Emotion schemas are embedded in the human visual system. *Sci Adv* 2019;**5**:eaaw4358.
- Lähteenmäki M, Hyönä J, Koivisto M et al. Affective processing requires awareness. *J Exp Psychol Gen* 2015;**144**:339–65.
- Lane RD, Smith R. Levels of emotional awareness: theory and mea- surement of a socio-emotional skill. *J Intell* 2021;**9**:42.
- Lane RD, Weihs KL, Herring A et al. Affective agnosia: expansion of the alexithymia construct and a new opportunity to inte- grate and extend Freud’s legacy. *Neurosci Biobehav Rev* 2015;**55**: 594–611.
- Lang PJ, Bradley MM. Emotion and the motivational brain. *Biol Psychol* 2010;**84**:437–50.
- Lange J, Dalege J, Borsboom D et al. Toward an integrative psychome- tric model of emotions. *Perspect Psychol Sci* 2020;**15**:444–68.
- Lang PJ, Greenwald MK, Bradley MM et al. Looking at pictures: affec- tive, facial, visceral, and behavioral reactions. *Psychophysiology* 1993;**30**:261–73.
- Lanteaume L, Khalfa S, Régis J et al. Emotion induction after direct intracerebral stimulations of human amygdala. *Cereb Cortex* 2007;**17**:1307–13.
- LeDoux JE. Rethinking the emotional brain. *Neuron* 2012;**73**:653–76.
- LeDoux JE. *The Deep History of Ourselves: The Four-Billion-Year Story of How We Got Conscious Brains*. New York: Penguin Publishing Group, 2019.
- LeDoux JE. How does the non-conscious become conscious? *Curr Biol* 2020a;**30**:R196–9.
- LeDoux JE. Thoughtful feelings. *Curr Biol* 2020b;**30**:R619–23.
- LeDoux JE. What emotions might be like in other animals. *Curr Biol* 2021;**31**:R824–9.
- LeDoux JE, Brown R. A higher-order theory of emotional conscious- ness. *Proc Natl Acad Sci U S A* 2017;**114**:E2016–25.
- LeDoux JE, Daw ND. Surviving threats: neural circuit and computa- tional implications of a new taxonomy of defensive behaviour. *Nat Rev Neurosci* 2018;**19**:269–82.
- LeDoux JE, Hofmann SG. The subjective experience of emotion: a fearful view. *Curr Opin Behav Sci* 2018;**19**:67–72.

- LeDoux JE, Pine DS. Using neuroscience to help understand fear and anxiety: a two-system framework. *Am J Psychiatry* 2016;**173**:1083–93.
- Lieberman MD. Boo ! The consciousness problem in emotion. *Cogn Emot* 2019;**33**:24–30.
- Lieberman MD, Eisenberger NI, Crockett MJ et al. Putting feelings into words. *Psychol Sci* 2007;**18**:421–8.
- Lindquist KA, Barrett LF. Emotional complexity. In: *Handbook of Emotions*, 3rd edn. New York: The Guilford Press, 2008, 513–30.
- Lin H, Müller-Bardorff M, Gathmann B et al. Stimulus arousal drives amygdalar responses to emotional expressions across sensory modalities. *Sci Rep* 2020;**10**:1898.
- Mauss IB, Levenson RW, McCarter L et al. The tie that binds? Coherence among emotion experience, behavior, and physiology. *Emotion* 2005;**5**:175–90.
- McFadyen J, Mattingley JB, Garrido MI. An afferent white matter pathway from the pulvinar to the amygdala facilitates fear recognition. *eLife* 2019;**8**:e40766.
- Melloni L, Mudrik L, Pitts M et al. Making the hard problem of consciousness easier. *Science* 2021;**372**:911–2.
- Méndez-Bértolo C, Moratti S, Toledano R et al. A fast pathway for fear in human amygdala. *Nat Neurosci* 2016;**19**:1041–9.
- Méndez CA, Celeghin A, Diano M et al. A deep neural network model of the primate superior colliculus for emotion recognition. *Philos Trans R Soc Lond B Biol Sci* 2022;**377**:20210512.
- Mendl M, Paul ES. Animal affect and decision-making. *Neurosci Biobehav Rev* 2020;**112**:144–63.
- Mennella R, Bavard S, Mentec I et al. Spontaneous instrumental avoidance learning in social contexts. *Sci Rep* 2022;**12**:17528.
- Mennella R, Vilarem E, Grèzes J. Rapid approach-avoidance responses to emotional displays reflect value-based decisions: neural evidence from an EEG study. *NeuroImage* 2020;**222**:117253.
- Merker B. Consciousness without a cerebral cortex: a challenge for neuroscience and medicine. *Behav Brain Sci* 2007;**30**:63–81.
- Merker B. The efference cascade, consciousness, and its self: naturalizing the first person pivot of action control. *Front Psychol* 2013;**4**:501.
- Mertens G, Engelhard IM. A systematic review and meta-analysis of the evidence for an unaware fear conditioning. *Neurosci Biobehav Rev* 2020;**108**:254–68.
- Mikels JA, Reuter-Lorenz PA. Affective working memory: an integrative psychological construct. *Perspect Psychol Sci* 2019;**14**:543–59.
- Mitchell DGV, Greening SG. Conscious perception of emotional stimuli: brain mechanisms. *Neuroscientist* 2012;**18**:386–98.
- Mobbs D, Adolphs R, Fanselow MS et al. Viewpoints: approaches to defining and investigating fear. *Nat Neurosci* 2019;**22**:1205–16.
- Mobbs D, Headley DB, Ding W et al. Space, time, and fear: survival computations along defensive circuits. *Trends Cogn Sci* 2020;**24**:228–41.
- Moors A. *Demystifying Emotions: A Typology of Theories in Psychology and Philosophy*, 1st edn. Cambridge: Cambridge University Press, 2022.
- Moors A, Boddez Y, De Houwer J. The power of goal-directed processes in the causation of emotional and other actions. *Emot Rev* 2017;**9**:310–8.
- Naccache L. Why and how access consciousness can account for phenomenal consciousness. *Philos Trans R Soc Lond B Biol Sci* 2018;**373**:20170357.
- Nagel T. What is it like to be a bat? *Philos Rev* 1974;**83**:435–50.
- Nisbett RE, Wilson TD. Telling more than we can know: verbal reports on mental processes. *Psychol Rev* 1977;**84**:231–59.
- Northoff G, Lamme V. Neural signs and mechanisms of consciousness: is there a potential convergence of theories of consciousness in sight? *Neurosci Biobehav Rev* 2020;**118**:568–87.
- Nummenmaa L, Glerean E, Hari R et al. Bodily maps of emotions. *Proc Natl Acad Sci U S A* 2014;**111**:646–51.
- Oberauer K, Lewandowsky S. Addressing the theory crisis in psychology. *Psychon Bull Rev* 2019;**26**:1596–618.
- Odegaard B, Knight RT, Lau H. Should a few null findings falsify prefrontal theories of conscious perception? *J Neurosci* 2017;**37**:9593–602.
- Overgaard M, Sandberg K. Kinds of access: different methods for report reveal different kinds of metacognitive access. *Philos Trans R Soc Lond B Biol Sci* 2012;**367**:1287–96.
- Panksepp J. *Affective Neuroscience: The Foundations of Human and Animal Emotions*. New York: Oxford University Press, 1998.
- Panksepp J. The periconscious substrates of consciousness: affective states and the evolutionary origins of the self. *J Conscious Stud* 1998b;**5**:566–82.
- Panksepp J. Emotional feelings originate below the neocortex: toward a neurobiology of the soul. *Behav Brain Sci* 2007;**30**:101–3.
- Panksepp J, Lane RD, Solms M et al. Reconciling cognitive and affective neuroscience perspectives on the brain basis of emotional experience. *Neurosci Biobehav Rev* 2017;**76**:187–215.
- Paré D, Quirk GJ. When scientific paradigms lead to tunnel vision: lessons from the study of fear. *NPJ Sci Learn* 2017;**2**:6.
- Parvizi J, Damasio AR. Consciousness and the brainstem. *Cognition* 2001;**79**:135–60.
- Paul ES, Sher S, Tamietto M et al. Towards a comparative science of emotion: affect and consciousness in humans and animals. *Neurosci Biobehav Rev* 2020;**108**:749–70.
- Pavuluri M, May A. I feel, therefore, I am: the insula and its role in human emotion, cognition and the sensory-motor system. *AIMS Neurosci* 2015;**2**:18–27.
- Pessoa L, Japee S, Sturman D et al. Target visibility and visual awareness modulate amygdala responses to fearful faces. *Cereb Cortex* 2006;**16**:366–75.
- Pessoa L. Embracing integration and complexity: Placing emotion within a science of brain and behaviour. *Cognition and Emotion* 2019;**33**:55–60.
- Prinz JJ. *Gut Reactions: A Perceptual Theory of Emotion*. Oxford: Oxford University Press, 2004.
- Prinz JJ ed. Précis of gut reactions. *Philos Phenomenol Res* 2008;**76**:707–11.
- Roberts ID, Hutcherson CA. Affect and decision making: insights and predictions from computational models. *Trends Cogn Sci* 2019;**23**:602–14.
- Robinson MD, Clore GL. Episodic and semantic knowledge in emotional self-report: evidence for two judgment processes. *J Pers Soc Psychol* 2002;**83**:198–215.
- Rohr M, Wentura D. Degree and complexity of non-conscious emotional information processing—a review of masked priming studies. *Front Hum Neurosci* 2021;**15**:689369.
- Rosenthal DM. *Consciousness and Mind*. UK: Oxford University Press, 2005.
- Rossion B, de Gelder B, Pourtois G et al. Early extrastriate activity without primary visual cortex in humans. *Neurosci Lett* 2000;**279**:25–8.
- Rotteveel M, Phaf RH. Automatic affective evaluation does not automatically predispose for arm flexion and extension. *Emotion* 2004;**4**:156–72.
- Russell JA. Core affect and the psychological construction of emotion. *Psychol Rev* 2003;**110**:145–72.
- Saarimäki H. Naturalistic stimuli in affective neuroimaging: a review. *Front Hum Neurosci* 2021;**15**:675068.
- Saarimäki H, Ejtehadian LF, Glerean E et al. Distributed affective space represents multiple emotion categories across the human brain. *Soc Cogn Affect Neurosci* 2018;**13**:471–82.

- Saarimäki H, Gotsopoulos A, Jääskeläinen IP *et al.* Discrete neural signatures of basic emotions. *Cereb Cortex* 2016;**26**:2563–73.
- Scheel AM, Tiokhin L, Isager PM *et al.* Why hypothesis testers should spend less time testing hypotheses. *Perspect Psychol Sci* 2021;**16**:744–55.
- Schlund MW, Brewer AT, Magee SK *et al.* The tipping point: value differences and parallel dorsal–ventral frontal circuits gating human approach–avoidance behavior. *NeuroImage* 2016;**136**:94–105.
- Selimbeyoglu A. Electrical stimulation of the human brain: perceptual and behavioral phenomena reported in the old and new literature. *Front Hum Neurosci* 2010;**4**:46.
- Seth AK. Interoceptive inference, emotion, and the embodied self. *Trends Cogn Sci* 2013;**17**:565–73.
- Seth AK. The Cybernetic Bayesian Brain. In: Metzinger TK and Windt JM (ed.), *Open MIND*. Frankfurt am Main: MIND Group, 2015.
- Seth AK, Bayne T. Theories of consciousness. *Nat Rev Neurosci* 2022;**23**:439–52.
- Seth AK, Critchley HD. Extending predictive processing to the body: emotion as interoceptive inference. *Behav Brain Sci* 2013;**36**:227–8.
- Shine JM. Adaptively navigating affordance landscapes: how interactions between the superior colliculus and thalamus coordinate complex, adaptive behaviour. *Neurosci Biobehav Rev* 2022;**143**:104921.
- Smirnov D, Saarimäki H, Glerean E *et al.* Emotions amplify speaker–listener neural alignment. *Hum Brain Mapp* 2019;**40**:4777–88.
- Smith R, Killgore WDS, Lane RD. The structure of emotional experience and its relation to trait emotional awareness: a theoretical review. *Emotion* 2018a;**18**:670–92.
- Smith R, Lane RD. The neural basis of one's own conscious and unconscious emotional states. *Neurosci Biobehav Rev* 2015;**57**:1–29.
- Smith R, Lane RD. Unconscious emotion: a cognitive neuroscientific perspective. *Neurosci Biobehav Rev* 2016;**69**:216–38.
- Smith R, Lane RD, Alkozei A *et al.* The role of medial prefrontal cortex in the working memory maintenance of one's own emotional responses. *Sci Rep* 2018b;**8**:1–15.
- Smith R, Lane RD, Parr T *et al.* Neurocomputational mechanisms underlying emotional awareness: insights afforded by deep active inference and their potential clinical relevance. *Neurosci Biobehav Rev* 2019a;**107**:473–91.
- Smith R, Parr T, Friston KJ. Simulating emotions: an active inference model of emotional state inference and emotion concept learning. *Front Psychol* 2019b;**10**:2844.
- Sun B, Zeng X, Chen X *et al.* Neural correlates of conscious processing of emotional faces: evidence from event-related potentials. *Neuropsychologia* 2023;**182**:108478.
- Tamietto M, de Gelder B. Neural bases of the non-conscious perception of emotional signals. *Nat Rev Neurosci* 2010;**11**:697–709.
- Taschereau-Dumouchel V, Kawato M, Lau H. Multivoxel pattern analysis reveals dissociations between subjective fear and its physiological correlates. *Mol Psychiatry* 2020;**25**:2342–54.
- Taschereau-Dumouchel V, Michel M, Lau H *et al.* Putting the “mental” back in “mental disorders”: a perspective from research on fear and anxiety. *Mol Psychiatry* 2022;**27**:1322–30.
- Tooley MD, Carmel D, Chapman A *et al.* Dissociating the physiological components of unconscious emotional responses. *Neurosci Conscious* 2017;**2017**:nix021.
- Treit D, Berridge KC. A comparison of benzodiazepine, serotonin, and dopamine agents in the taste-reactivity paradigm. *Pharmacol Biochem Behav* 1990;**37**:451–6.
- Tsikandilakis M, Bali P, Yu Z *et al.* Individual conscious and unconscious perception of emotion: theory, methodology and applications. *Conscious Cogn* 2021;**94**:103172.
- Tsuchiya N, Wilke M, Frässle S *et al.* No-report paradigms: extracting the true neural correlates of consciousness. *Trends Cogn Sci* 2015;**19**:757–70.
- Vaidya AR, Pujara MS, Petrides M *et al.* Lesion studies in contemporary neuroscience. *Trends Cogn Sci* 2019;**23**:653–71.
- Vale R, Evans DA, Branco T. Rapid spatial learning controls instinctive defensive behavior in mice. *Curr Biol* 2017;**27**:1342–9.
- van der Ploeg MM, Brosschot JF, Versluis A *et al.* Peripheral physiological responses to subliminally presented negative affective stimuli: a systematic review. *Biol Psychol* 2017;**129**:131–53.
- Vetter P, Badde S, Phelps EA *et al.* Emotional faces guide the eyes in the absence of awareness. *eLife* 2019;**8**:e43467.
- Vilarem E, Armony JL, Grèzes J. Action opportunities modulate attention allocation under social threat. *Emotion* 2020;**20**:890–903.
- Wieloch T, Nikolich K. Mechanisms of neural plasticity following brain injury. *Curr Opin Neurobiol* 2006;**16**:258–64.
- Williams LM, Liddell BJ, Kemp AH *et al.* Amygdala–prefrontal dissociation of subliminal and supraliminal fear. *Hum Brain Mapp* 2006;**27**:652–61.
- Winkielman P, Berridge KC. Unconscious emotion. *Curr Dir Psychol Sci* 2004;**13**:120–3.
- Winkielman P, Berridge KC, Wilbarger JL. Unconscious affective reactions to masked happy versus angry faces influence consumption behavior and judgments of value. *Pers Soc Psychol Bull* 2005;**31**:121–35.
- Winkielman P, Gogolushko Y. Influence of suboptimally and optimally presented affective pictures and words on consumption-related behavior. *Front Psychol* 2018;**8**:2261.
- Yang E, Zald DH, Blake R. Fearful expressions gain preferential access to awareness during continuous flash suppression. *Emotion* 2007;**7**:882–6.
- Yarkoni T, Poldrack RA, Van Essen DC *et al.* Cognitive neuroscience 2.0: building a cumulative science of human brain function. *Trends Cogn Sci* 2010;**14**:489–96.
- Zaki J, Davis JI, Ochsner KN. Overlapping activity in anterior insula during interoception and emotional experience. *NeuroImage* 2012;**62**:493–9.