



The search for the neural correlate of consciousness: Progress and challenges

Alex Lepauvre^a  (alex.lepauvre@ae.mpg.de)

Lucia Melloni^b  (Lucia.melloni@ae.mpg.de)

Abstract

Twenty years ago, Thomas Metzinger published the book *The Neural Correlates of Consciousness* amassing the state of knowledge in the field of consciousness studies at the time from philosophical and empirical perspectives. On the occasion of the 20th anniversary of this impactful publication, we review the progress the field has made since then and the important methodological challenges it faces. A tremendous number of empirical studies have been conducted, which has led to the identification of many candidate neural correlates of consciousness. Yet, this tremendous amount of work has not unraveled a consensual account of consciousness as of now. Many questions, some already raised twenty years ago, remain unanswered, and an enormous proliferation of theories sharply contrasts with the scarcity of compelling data and methodological challenges. The contrastive method, the foundational method used to study the neural correlate of consciousness (NCC), has also been called into question. And while awareness in the community of its shortcomings is widespread, few concrete attempts have been made to go beyond it and/or to revise existing theories. We propose several methodological shifts that we believe may help to advance the quest of the NCC program, while remaining uncommitted to any specific theory: (1) the currently prevalent “contrastive method” should lose its monopoly in favor of methods that attempt to explain the phenomenology of experience; (2) experimental data should be shared in centralized, multi-methods databases, transcending the limitations of individual experiments by granting granularity and power to generalize findings and “distill” the NCC proper; (3) the explanatory power of theories should be directly pitted against each other to overcome the non-productive fractioning of the field into insular camps seeking confirmatory evidence for their theories. We predict these innovations might enable the field to progress towards the goal of explaining consciousness.

Keywords

Adversarial collaboration · Big data · Consciousness · Contrastive method · Neural correlate of consciousness · Phenomenology

This article is part of a special issue on “The Neural Correlates of Consciousness,” edited by Sascha Benjamin Fink and Ying-Tung Lin.

^aDepartment of Neuroscience, Max Planck Institute for Empirical Aesthetics, Frankfurt am Main

^bDepartment of Neurology, NYU Grossmann School of Medicine



1 Introduction

Men ought to know that from the brain, and from the brain only, arise our pleasures, joys, laughter, and jests, as well as our sorrows, pains, griefs, and tears. Through it, in particular, we think, see, hear, and distinguish the ugly from the beautiful, the bad from the good, the pleasant from the unpleasant.

—Hippocrates of Cos, *The Sacred Disease*, ~400BC
(translated by W. H. S. Jones, 1959, p. 175)

Since the time of Hippocrates, mankind has attempted to understand what consciousness is, what it does, and how physical systems such as the brain, but not others such as the heart, can instantiate it. David Chalmers famously referred to the latter aspect as “The Hard Problem” (Chalmers, 1995): *why* does matter, such as the brain, give rise to perceptions and emotions that have subjective, phenomenological qualities?

As underlined by Metzinger (2000), the difficulty in studying consciousness empirically lies in the inherent subjectivity of the phenomenon. This is indeed the main reason why consciousness for a long time remained outside the scope of modern sciences, whose primary goal is to understand objective phenomena. Notwithstanding revolutionary discoveries made over the past century by studying cerebral lesions and using electrical stimulation to probe brain function, as in the pioneering work of Weiskrantz, Penfield, Libet, and Eccles among others (Eccles, 1974; Libet et al., 1967; Penfield, 1958; Weiskrantz, 1980), it was Crick and Koch’s seminal paper (1990) that provided a scientific framework for the study of the neural correlates of consciousness that set the field in motion. Crick and Koch proposed putting aside philosophical and metaphysical considerations and to focus on the “easy problem of consciousness” (Crick & Koch, 1998), i.e., to study the neural activity that *correlates* with consciousness while taking subjectivity as a dependent variable (Dehaene, 2014). The focus of the science of consciousness would be narrower, yet more tractable, clearly defined, and theoretically neutral, i.e., in identifying the neural *correlates* of consciousness (NCCs): the minimal set of neural events jointly sufficient for a specific conscious experience (given the appropriate enabling conditions, Chalmers, 2000). By quantifying which neural activity best correlates with conscious experience, one could begin to narrow down the mechanisms involved in consciousness, and to eventually formulate theories to answer the hard questions of consciousness (Chalmers, 1995).

This bottom-up strategy has dominated the field for the past thirty years, further propelled by the development of ever more refined methods for investigating the human brain non-invasively, e.g., high-field imaging (7T) with laminar and columnar resolution, and the rapid development of machine-learning algorithms that have enabled researchers to track the contents of perception. Thirty years after Crick and Koch’s findings, we now face an entire cartography of brain areas and

refined spatio-temporal patterns of brain activity, or rather, an urban jungle! Yet, correlations do not entail explanation. The question is then, are we closer to finding the mechanism(s) responsible for subjective, phenomenal qualities? We will argue that despite much effort and scientific investigation, one critical component—the explanation of the phenomenological qualities—remains as elusive as it was thirty years ago, which contrasts with the wealth of empirical data that have been amassed meanwhile. We remain optimistic that such an explanation is within reach, yet it will require us, the scientific community, to embrace a number of different practices starting from acknowledging serious methodological pitfalls in the contrastive method—the method that represents the cornerstone of our scientific practice. Expanding the contrastive method will be critical to the success of a biological research program on consciousness; yet this work will not happen if it does not include the development of descriptions that address the phenomenological level of organization of consciousness. Current theorizing has for the most part largely disregarded this aspect since Revonsuo (2000) (see also Revonsuo, 2006 for a more detailed account) made a plea for it; with just a few notable exceptions, e.g., the neurophenomenology approach advocated by the late Varela (1996), and the axiomatic approach of Integrated Information Theory (Tononi, 2004) with its most recent studies on the phenomenological qualities of space (Haun & Tononi, 2019).

Beyond that, due to the inherent bottom-up approach of the NCC program, as a community we ought to find ways to aggregate data across multiple studies and experimental paradigms while also designing studies targeted to test contradictory predictions of different theories within the same experimental context. Only then, will we be on a serious track to distill the NCC proper; anything short of this will render the NCC program futile.

In what follows, we review earlier arguments that we and others have made regarding the methodologies used to study the NCC, and we advance ideas on how to bolster the NCC program. With this article, we aim to take a critical look at Crick and Koch's framework for studying consciousness, taking the progress made since then as lessons for a renewed strategy for finding the neural processes that give rise to consciousness. In this process, we remain neutral regarding the different hypotheses that have been advanced in the field and their explanatory power, while at the same we wish to speak to the methodological challenges that lie ahead of us. We hope addressing the latter will enable us to look at the data and the proposed theories through a new lens and to gain a unified, and acceptable, explanation of consciousness in the not-so-distant future.

1.1 How the contrastive methods can help in finding the NCCs

The main approach taken in finding the NCC, and recently advocated as one of the hard criteria to evaluate empirical theories of consciousness (Doerig et al., 2021), has been the contrastive method (Baars, 2005). This method consists in contrasting

conscious and unconscious states and/or contents in consciousness, while maintaining any other factors constant. The reported perceptual state of the participant is then taken as a dependent variable and the neural activity associated with the different perceptual states is contrasted to reveal the NCCs (Dehaene, 2014). As external factors are supposed to be maintained constant and only perceptual state should vary, the associated neural activity was assumed to exclusively reflect the processes involved in consciousness. This method has been employed to compare the difference between wakefulness/vigilance and non-conscious states in order to obtain the full NCC, or to compare the neural activity between different perceptual contents to yield the content-specific NCCs (Koch et al., 2016). The current perspective focuses mostly on the latter; but the principles advocated should apply to any attempt to find a neural correlate of either a state or a content of consciousness.

Several experimental paradigms for erasing stimuli from awareness have been used to study neural processes associated with both conscious and unconscious conditions (Axelrod et al., 2015; Kim & Blake, 2005). One example of such an experimental paradigm is masking, which involves the brief presentation of a critical stimulus closely preceded and/or followed by another stimulus—a so-called “mask”—that renders the critical stimulus visible or invisible depending on the interval between the mask and the critical stimulus (Dehaene et al., 2001). Participants can then report (e.g., verbally or via button press) whether they have seen a stimulus or not. The neural activity is then compared on the basis of this report. Using this experiment, combined with fMRI and EEG recordings, Dehaene and colleagues were able to show that seen trials differed from unseen trials by a widespread fronto-parietal activation (Dehaene et al., 2001). These findings were further confirmed using other experimental paradigms, such as the attentional blink, in which participants’ attention is focused on a specific task, rendering a critical stimulus invisible for the participants (Sergent et al., 2005). They also found that a late ERP component—the so called P3b—correlated best with the subjective report of visibility on a continuous perceptual awareness scale (Del Cul et al., 2007). Another technique that has been used to search for NCCs is binocular rivalry, a phenomenon in which two incongruent stimuli are presented to each eye, leading to alternating perception of exclusively one or the other. As the stimuli presented to both eyes remain constant while perception alternates, this enables researchers to compare stimuli whose status in consciousness varies while external stimulation remains constant, and thus to dissociate sensation from conscious perception. Using this method to investigate the NCC in non-human primates with single-cell recordings, Logothetis and colleagues were able to show that neural activation in the infero-temporal cortex strongly correlates with conscious perception, i.e., what the animal reports “seeing,” whereas neural activation in V1 correlates more strongly with retinal input, i.e., the two stimuli presented to each eye (Logothetis, 1998). Other studies, however, concluded that the best correlates were found in early sensory cortices, as opposed to higher-order sensory areas. Using threshold

presentation, Pins and ffytche located the best correlates of consciousness in an ERP peaking at 100ms in the occipital cortex (Pins & ffytche, 2003). Similarly, investigating patients with Charles Bonnet syndrome, they were able to show that the hallucination of percepts correlated with changes of activity in extrastriate visual areas (ffytche et al., 1998).

This initial surge in empirical data studying the neural correlate of consciousness contrasted with the drought of the decades prior to 1990. The increased availability of neural recording methods such as fMRI, PET, and MEG, together with the clear methodology proposed by the contrastive method, seemed to make the answering of the question of consciousness a goal finally achievable. Soon enough, as data accumulated, several theories emerged aiming to explain by which mechanisms subjective experience emerges from neural activity. Many of these theories are compiled in the seminal anthology of Metzinger (2000). For instance, the early dichotomy of findings between early and late correlates was logically reflected in those theories: the recurrent feedback theory proposed by Lamme (2006) stated that consciousness emerges from local recurrent feedback loop in early sensory cortex, while one of the variants of the higher-order theory (Lau & Rosenthal, 2011) stated that consciousness is the result of higher-order processes, taking place predominantly in prefrontal areas. Similarly, the global workspace theory first proposed as a cognitive architecture by Baars (1997) was then further extended by Stanislas Dehaene and colleagues (2001). They provided a biologically realistic implementation in which consciousness is thought to be the result of widespread broadcasting of information throughout an interconnected network of prefrontal-parietal areas and high-level sensory areas, making information available to a wide range of areas for further processing. Other theories attempted to explain the “unity of consciousness”—the fact that all senses come together as a unified whole into one experience—proposing neural synchronization across cortical areas, predominantly in the gamma frequency, as the critical mechanism explaining the binding of features that gives rise to a unified experience (Crick & Koch, 1990; Edelman & Tononi, 2000; Melloni, Schwiedrzik, Müller, et al., 2011; Varela et al., 2001). These initial theories were soon followed by more, to the point that the field today is practically overwhelmed by a mass of available theories trying to explain consciousness at different levels (Doerig et al., 2021).

Thirty years after the groundbreaking insights of Crick and Koch and twenty years after Metzinger’s book, the enthusiasm of those early studies’ findings has been tempered by the realization that the apparent simplicity of the empirical research on NCCs in fact suffers from major methodological challenges. The contrastive method by itself was not selective to reveal the NCC proper. In what follows, we provide a brief description of these methodological challenges, referencing a selected set of studies to demonstrate the pervasiveness of the problem. Our review of the literature will be from that vantage point and thus be inherently selective. For recent comprehensive reviews of the empirical findings on the NCC, we refer the reader to Boly et al. (2017) and Forster et al. (2020).

1.2 Methodological challenge: More than one NCC

The core assumption of the contrastive method is that when comparing between conditions in which the external input is controlled for and only perception varies, the resulting difference directly reflects neural activity involved in consciousness. We and others have, however, argued that this strategy is too simplistic (Aru, Bachmann, et al., 2012; Graaf et al., 2012). We will briefly rehearse the main argument (for a longer treatment of the issue we refer the reader to the original work Aru, Bachmann, et al., 2012; see also Miller, 2015, where this problem is referred to as the correlation/constitution distinction problem).

When comparing neural responses associated with two perceptual states, as the contrastive method does, two other families of processes co-occur with the ones directly reflecting consciousness: the NCC-precursors (NCC-pr) and the NCC-consequences (NCC-co, Aru, Bachmann, et al., 2012, see Fig 1). NCC-pr refers to processes that precede the NCC proper. They might enable a given stimulus to reach consciousness but are not conscious themselves. The NCC-co refer to processes that might follow the NCC proper. They result from consciousness but are not conscious. A typical example of NCC-pr is attention. While the role of attention in consciousness is still being debated (Koch & Tsuchiya, 2007; Koivisto & Revonsuo, 2007; Pitts et al., 2018), it is now widely agreed that they are indeed two different mechanisms (though that was not always the case, see Posner, 1994). In classical contrastive paradigms, covert shifts of attention may explain why some stimuli are seen while others are not. For example, in masking paradigms, differential engagement of attention across trials might well explain why an otherwise identical physical stimulus is perceived in some trials but missed in others. The NCC-co are processes triggered by conscious experience but not responsible for it. Some theories of consciousness assign consciousness a function. Therefore, the fact that a critical stimulus was perceived must entail that additional processes will follow it. Such processes include encoding in working memory (Soto & Silvanto, 2014) and/or episodic memory, reflecting about the perceived stimulus, and in the case of most experimental paradigms employing the contrastive method, reporting about it (Tsuchiya et al., 2015).

This methodological caveat seriously called into question the wealth of early findings, given that by themselves they do not enable the NCC-pr and NCC-co to be distinguished from the NCC proper. As such, there is no guarantee that NCC findings that attribute, for instance, the prefrontal cortex a role in consciousness may not instead be reflecting NCC-co associated with report, task performance, or encoding in working memory and/or episodic memory. The same holds true for findings that state that early sensory activity constitutes an NCC, which could instead merely reflect NCC-pr, e.g., attentional shift, expectations, etc. As findings of NCC in early contrastive studies fostered the development of several theories of consciousness, the impact of these methodological challenges and how much they affect the proposed theories is to date unclear.

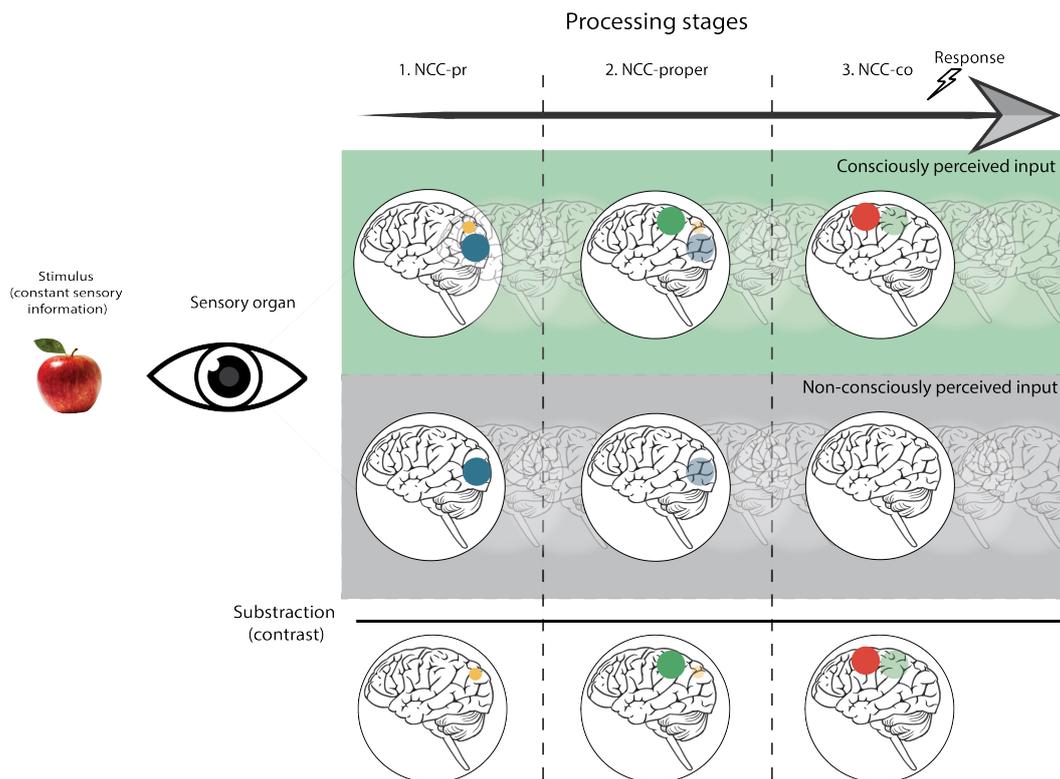


Figure 1: The contrastive method is implemented by experimentally comparing conscious versus unconscious conditions e.g., via masking, binocular rivalry, etc. The neural activations induced in the conscious vs. non-conscious condition are contrasted to reveal the NCC. While apparently straightforward, this method was argued to be insufficient to fully isolate the so-called NCC-proper (2). The results of this simple subtraction are confounded by the occurrence of temporally flanked mechanisms, termed NCC-pr (1) and NCC-co (3) (see also [Aru, Bachmann, et al., 2012](#)). The former refers to the mechanisms enabling a stimulus to reach consciousness without directly accounting for consciousness itself (yellow in the figure), while the latter refers to mechanisms that are consequences of the conscious perception of a stimulus but need not be conscious themselves (red in plot). Examples of NCC-pr are attention and prior expectations ([Aru, Axmacher, et al., 2012](#); [Aru, Bachmann, et al., 2012](#); [Melloni, Schwiedrzik, Muller, et al., 2011](#)). As most theories of consciousness assign a function to consciousness, typical NCC-co accompany it: encoding in memory, decision making to name a few. Additionally, as many contrastive studies rely on report to differentiate conscious versus non-conscious conditions, motor responses and planning, as well as self-monitoring, oftentimes confound the NCC-proper ([Tsuchiya et al., 2015](#)).

The previously accepted NCCs therefore had to be reviewed by experimentally controlling more thoroughly, ideally in the same experimental setting, for covarying processes in order to isolate the neural activity directly related to consciousness. We called this process the “distillation of the NCC,” as in effect it implies removing all confounding variables until the remaining core neural activity that correlates with consciousness is found. While many processes precede and succeed a conscious experience (and will have their own neural correlates), there is one such confound of previous studies that recently has received particular attention, namely report. This relates to the fact that participants have had to overtly report the perceptual state. Report can be performed verbally or via button press (ffytche et al., 1998; Kleinschmidt et al., 2002; Logothetis, 1998; Pitts & Britz, 2011) using either forced choice tasks (Dehaene et al., 2001; Gaillard et al., 2009; Henson et al., 2008) or the perceptual awareness scale (Melloni, Schwiedrzik, Muller, et al., 2011; Overgaard et al., 2006). When subjects report what they perceive on a trial-by-trial basis, they engage additional processes not directly reflecting conscious perception: they have to monitor their own state, make decisions about what to report depending on the afferent input, keep a template in their mind to use for reporting, and both engage in motor planning and emit a motor response (Tsuchiya et al., 2015). Introspective report has also been challenged on theoretical grounds: Ned Block has put forward the view that reports only yield information about higher-order thought processes, not about consciousness itself (Block, 2005).

To circumvent or minimize the engagement of post-perceptual processes, several paradigms have been used in which report is not directly needed in order to assess perception; these are referred to as no-report paradigms. One such example is reliance on optokinetic nystagmus (i.e., eye movements) to estimate perceptual content in binocular rivalry paradigms, which has been successfully implemented both in humans (Frässle et al., 2014) and other animals, e.g., cats (Fries et al., 1997) and monkeys (Kapoor et al., 2020). Using this methodology in combination with fMRI, Frässle et al. (2014) showed that when drifting sinusoidal gratings moving in different directions are exposed to each eye of the participant, the optokinetic nystagmus closely tracks the current percept, as validated by comparison with participant reports. When comparing the neural activation associated with perceptual changes in the absence of report (i.e., using optokinetic nystagmus measure with that observed under overt report) Frässle and colleagues observed that while parietal activation still tracked perceptual changes, frontal activation, found under reporting conditions, no longer correlated with the perceptual state (see Figure 3a and 3b in Frässle et al., 2014). Similarly, Farooqui & Manly (2018) using fMRI showed that when consciously perceived stimuli are not relevant for task performance, frontal regions get deactivated. These studies seem to indicate that prefrontal activity indexes NCC-co as opposed to the NCC proper.

In a set of experiments using EEG to investigate the electrophysiological correlates of consciousness with and without report using inattentive blindness of shapes (Pitts et al., 2012), words (Schelonka et al., 2017), and faces (Shafto & Pitts, 2015), Pitts and colleagues found P3b responses putatively indexing consciousness

only when subjects were reporting. Comparable results were obtained in forward and backward masking paradigms, whereby a robust P3b was found in the report condition, but was absent when subjects did not need to report (Cohen et al., 2020). Instead, they found that an earlier (200–300ms) negative component over posterior areas correlated with consciousness independently from report. This component, named the visual awareness negativity, has been found to correlate with consciousness across a range of studies (Koivisto & Revonsuo, 2010), making it a prime NCC candidate (but see Melloni, Schwiedrzik, Müller, et al., 2011 for a more nuanced discussion of the temporal dynamics related to the NCC). Moreover, P3b has also been shown to differ between frequent and rare stimuli in oddball paradigms, even when all the stimuli were unconsciously perceived, which calls into question P3b activation as an exclusive marker of conscious perception (Silverstein et al., 2015). Such observations are not limited to the visual domain, as Sergent et al. (2021) recently showed that the P3b correlates with awareness of auditory stimuli during the active report condition, but not during the no-report task.¹

While these studies seem to relegate late, frontal activation to post-perceptual processes only, other studies employing invasive and more sensitive methods, such as electro-corticography (ECoG) and single-unit recording, have revealed a more nuanced picture. For instance, Noy et al. (2015), using ECoG in epilepsy patients, showed that while the late-onset frontal activation was present under reporting conditions only, an initial glow spreading through all cortical areas, including the pre-frontal cortex, was observed regardless of report. Additionally, Kapoor et al. (2020), using single-unit recordings in macaque monkeys, showed that in the absence of overt report the content of consciousness can be accurately decoded from the prefrontal cortex in binocular rivalry. Thus, while some of the previously observed prefrontal activation might index confounding factors such as report, leading to an exacerbated prefrontal cortex activation, a selective portion might still correspond to the NCC proper. As such, while no-report paradigms might deal well with disregarding post-perceptual processes at the level of the experimental design, by themselves they do not resolve the inherent sensitivity issue faced by most non-invasive methodologies, e.g., fMRI, PET, M-EEG. Multimodal methodologies, and in particular a combination of invasive and non-invasive techniques, will be necessary to arbitrate on the role of prefrontal cortex in consciousness, which has been challenged by no-report paradigms.

Nevertheless, not all confounds relate to report (Melloni, Schwiedrzik, Müller, et al., 2011). Prior expectation also influences neural activations in contrastive paradigms (Aru, Axmacher, et al., 2012; Mayer et al., 2016; Melloni, Schwiedrzik, Müller, et al., 2011). Therefore, the mechanisms of expectation and prior knowledge are also confounding the search for NCC. In paradigms with passive viewing, such as in the experiment by Cohen et al. (2020), encoding in working memory occurs in the seen but not in the unseen condition, also confounding the em-

¹In that study, the best correlate of conscious access was found in a component overlapping in time with the P3-like component but with a bilateral positivity over temporal regions.

pirical findings. Attention level constitutes yet another such confounding factor. As observed by Pitts and colleagues (2014), in the case of inattention blindness paradigms, seen and unseen trials might also differ in the level of attention. Neural processes related to decision making can also obscure the NCC proper as demonstrated in a recent study by Mazzi et al. (2020). Here, while both the visual awareness negativity (VAN) and P3b (termed LP, i.e., late positivity in the text) correlated with awareness, P3b was tightly correlated with decision-making, i.e., with decision criteria. Siclari and colleagues have also recently addressed confounds present in state paradigms, i.e., those investigating neural correlates of wakefulness and sleep in REM vs. NREM states, which are confounded by differences in neuromodulation across states (Siclari et al., 2017). They have extended the methodology to a no-task, within-state paradigm, in which subjects are serially awakened and asked to report dreams both during REM and NREM, enabling investigators to contrast ‘consciousness’ to lack thereof (i.e., reported presence vs. absence of dream in both REM and non-REM sleep) while controlling for both the state and the neuromodulation milieu. Thus, there are a number of confounds, some already known, for instance: report, prior expectations, working memory, episodic memory, task demands, attention, decision making, and neuromodulatory state. This list is not exhaustive. As more studies are carried out and our understanding of the processes uniquely related to consciousness is improved, this list is likely to be expanded. Aru, Bachmann, et al. (2012) have previously proposed that, in order to isolate the NCC proper, it will first be necessary to control for all of these potential (known) confounds in order to look for the neural activity that systematically relates to consciousness. In a sense, this pursuit can be framed as a problem of generalization: the search for brain responses that systematically covary with consciousness while manipulating all other nuisance variables, e.g., stimuli, task, experimental paradigms, and other cognitive processes, etc. (Yarkoni, 2020). Certainly, no single study is expected to manipulate all these variables at once. Instead, in addition to controlling for as many confounding factors as possible within a single study, it will be necessary to perform multiple studies to cross-validate the results: a research program that will take years and significant collective effort to complete. This problem has been noted and acknowledged by the scientific community, and its impact on the theories is just now starting to be grasped (Yaron et al., 2021). For instance, there is growing empirical evidence that the P3b, a marker once thought to reflect the broadcasting of information into the global workspace (Sergent et al., 2005), might instead reflect post-perceptual processes (Forster et al., 2020). These significant developments have led to theoretical revisions of the theories—e.g., Global Neuronal Workspace theory (Sergent et al., 2021)—while at the same time driving a slow convergence on the tighter link of other earlier, ERP components, such as the VAN, with consciousness. Despite that encouraging progress, concrete actions for addressing the challenges faced by the contrastive methods have lagged behind, and most studies continue using that methodology as originally defined. Moreover, to this day, there exists no database accumulating electrophysiological

and neuroimaging data from consciousness studies available to the community, thus limiting the integration across different studies that is so dearly needed in order to disentangle the NCC proper from all potential confounds. In its absence, the performance of isolated, local studies will be highly inefficient and the search for the NCC proper could prove to be a century-long enterprise.

Far from being a gloomy depiction of the state of the field, we believe the current situation is a natural sign of its maturity. As the field has evolved, research questions have crystallized even more, and both experimental challenges and their potential solutions have manifested. In the next section we cover some proposals for addressing these challenges and more. The list is not exhaustive, yet we hope it may inspire readers to seek creative alternatives that can propel the field even further over the next twenty years.

2 Homework for the field

2.1 Development of new methods for investigating the NCC

The contrastive method, recently extended by no-report paradigms, has dominated the field of consciousness research. Even though it has been argued that the comparison of conscious vs. unconscious conditions is crucial for finding the NCC (Doerig et al., 2021), it is not the only way to better understand consciousness (Noy et al., 2015) and possibly not even the most desirable one as it suffers from a severe problem of indeterminacy. In a typical experiment, we ask subjects to report what they see, offering them binary choices, e.g., seen/unseen, or alternatives such as the perceptual awareness scale (Melloni, Schwiedrzik, Muller, et al., 2011; Overgaard et al., 2006). Yet in either case those reports tell almost nothing about what subjects have actually experienced (unless they report on their phenomenology; more on this point in the next section). When they say they have seen a stimulus: What do they mean? What have they actually experienced? On what basis are they judging the stimulus visibility? Is it the vividness of the stimulus, its contrast, motion, color, shape, etc? As experimenters, we attempt to define a “template” for the subjects to report on what they see, but we can never know what they are actually experiencing. We sometimes use tricks, for instance, in some trials we present nothing, expecting that subjects will report that they see “nothing.” If they report having seen something in those trials, we have a dilemma: either we are excited, and use those “false alarm” trials to look for the correlates of consciousness without the confounds of sensory stimulation (Gosselin & Schyns, 2003; Kim & Blake, 2005), or we discard those trials/subjects, interpreting them as featuring a “bias” (shift in criterion in signal detection theory) and not real perception (Aru, Axmacher, et al., 2012; Dehaene et al., 2001; Mazzi et al., 2020). In a nutshell, the reports that are typically obtained in most studies investigating the NCC are complicated to interpret; they lead to a number of confounds and might easily explain differences across studies, as “instructing” subjects similarly is not

easy. Complicating the matter, most studies do not even thoroughly report how subjects were instructed and what measures were taken to ensure consistent behavior across subjects. Anyone who has tried to reproduce a study in a different laboratory immediately notices how much the “context” of the research matters, including how the investigator instructed the subjects. But little attention is paid to making this information accessible in publications. What other approaches can then be taken in order to move beyond the contrastive method?

Haynes (2009) has argued for the use of multivariate analysis to decode the content of consciousness, thus providing a much finer identification of NCC than the simple study of the dichotomy between presence and absence of critical stimuli. Recent studies using this approach have shown that decodability of stimulus orientation was highest from occipital sources in early time windows (130–320ms). Furthermore, decoding performances increased with clarity of the percept as reported on a perceptual awareness scale (Andersen et al., 2016). Salti et al. (2015), on the other hand, showed decoding from frontal regions at later time points—270ms post stimulus onset—of the spatial location of perceived targets. While decoding might be an exciting approach for unravelling the neural correlate of consciousness, the results are not easily interpretable; for example, they might spuriously suggest widespread patterns that are contradicted when tested further with causal manipulations (Bouton et al., 2018). As such, multivariate decoding analysis should be considered with caution.

Other approaches relate to focusing not just on what happens following the entry of a given stimulus into consciousness, but on their maintenance and, more generally, on a defining feature of consciousness, i.e., the experience of flow and temporality. With just a few notable exceptions (Gerber et al., 2017; Golan et al., 2016), most direct empirical work has aimed at understanding the entry of stimuli into consciousness, thus obviating the questions of how and why a conscious percept is maintained in consciousness. In this regard, perceptual continuity is an intriguing concept: we perceive the world as stable despite constant sensory disruptions caused by eye movements and blinks (Golan et al., 2016). The latter, for instance, causes a massive disruption in the sensory flow as the eyelid occludes the pupil for approximately 100–150 ms, yet we do not perceive the world as blanking out. Such a natural experiment occurs approximately 1000 times an hour (Cruz et al., 2011), amounting to roughly 3–4% of our waking life spent with our eyes closed, but without having noticed that the world has disappeared. A promising and more naturalistic approach to investigating the mechanisms related to consciousness is to focus on understanding what maintains a stimulus in awareness, i.e., why we fail to update our phenomenology in light of massive changes in sensory stimulation. As we have argued elsewhere (Melloni, 2015), understanding the flow of consciousness—i.e., moving from episodes (the snapshot view that has dominated studies on consciousness) to events (integrations across moments in time) central to the experience of such “flow”—is another crucial, but hitherto neglected,

aspect of conscious perception. This resonates with early attempts by Varela (1999) to address the experience of temporality, called “specious present” by James (1890, p. 377), and with the proposition that internal brain dynamics, in particular the synchronization of widely distributed neural assemblies in the gamma band and phase transitions across states, lies at the core of time consciousness. Support for this idea comes from studies investigating binocular rivalry (Cosmelli & Thompson, 2007; Doesburg et al., 2009, among others).

Another promising approach to investigating consciousness in populations with restricted responsiveness, such as infants and patients with cerebral lesions, is the use of neural frequency tagging, which we initially developed to evaluate language capacities and access to meaning in neurotypical individuals (Ding et al., 2016). Recently, we and others have used the frequency tagging approach for both retrospective (Gui et al., 2020) and prospective (Sokoliuk et al., 2021) classification on unresponsive patients, and the results have been encouraging.

Despite the clear advantages of the contrastive method, its mostly exclusive use in the research of consciousness may have constrained the field too much, becoming an obstacle to finding NCCs. It is expected that broadening the experimental methods will help us to better understand the phenomenon.

2.2 Addressing the central problem of consciousness: Phenomenology

While the previously discussed approaches go beyond the contrastive method, they all fall short in one important aspect: the final goal of a science of consciousness should be to understand its phenomenology and relate it to physical systems. The ultimate frontier is to explain why we feel the way we do (akin to Levine’s (1983) explanatory gap): why does vision feel different from audition, why does space feel extended, why do we feel time as flowing?

Revonsuo (2000), in his visionary chapter, eloquently called for the development of a science of consciousness that systematically describes the phenomenal level of organization on its own terms—from the first-person point of view. Such descriptions should go beyond cognitive and/or computational descriptions, and they will need to be developed from within the science of consciousness. In the spirit of previous traditions in phenomenology, the goal should be to describe the structures of consciousness as experienced from the first-person point of view. Progress in this area in the past decades has been meager. Instead, the field has been taken up by a cognitivist/functionalist view that leaves little room for explanations of the phenomenal qualities.

One important conceptual disagreement in the field refers to the distinction between phenomenal (P) and access (A) consciousness (Block, 1995). Phenomenal consciousness refers to “what it is like to be in that state,” whereas access consciousness refers to the availability of information to perform cognitive operations, such as reasoning about an experience, reporting it to an external observer,

etc. Phenomenal consciousness exists in and of itself, for our own inner spectacle, while access consciousness is defined by the functions it serves. A shorthand for the distinction between “being” and “doing.” This fundamental distinction led to a clear dichotomy in the different theories of consciousness, with some assuming independence of the two and some assuming them to be one and the same.

For some theories, such as Global Neuronal Workspace, phenomenal qualities of experience are fully captured by access (see [Naccache, 2018](#)). Under this assumption, studying access is equivalent to studying its phenomenology. Other theories, in contrast, posit that P-consciousness is distinct and independent from the collection of cognitive processes related to access ([Tononi, 2004](#); [Tsuchiya et al., 2015](#)). The debate on whether phenomenology overflows access is ongoing ([Fu et al., 2021](#)). Notwithstanding the empirical challenge in arbitrating between these different conceptions, it is important to refocus the debate to the main aim of the science of consciousness, i.e., what theories of consciousness ultimately aim to explain is the phenomenological qualities of experiences, the “what it is like to be in that state” ([Revonsuo, 2006](#)).

As a matter of fact, one reason why the debate between access and phenomenal consciousness will not easily be settled empirically is that we lack clear descriptions of the phenomenological qualities of experience assumed by the different theories of consciousness. For instance, take the debate on the richness vs. sparseness of consciousness (often discussed in concert with the above-mentioned debate of A- vs. P-consciousness, see [Block, 2011](#)). One side posits that the richness of experience is an illusion and that what can actually be perceived and remembered is limited. But what does it mean for something not to be rich? For instance, if a single number, say the number 15 in red ink, is foveally presented for enough time to be perceived, say 100 ms, and at a size that makes it easy to be perceived, what do the different theories assume is consciously perceived? Just the number, just the color, just the location, or the font, or the orientation of the letters, or all of those features? This is not a naïve question, given recent work on attribute amnesia ([Chen & Wyble, 2016](#); [Wang et al., 2021](#)), which demonstrates that subjects fail to report specific attributes of stimuli, e.g., color, that are otherwise reported as perceived. Can we interpret these results as evidence that phenomenology overflows access? Not easily, as (some) theories of consciousness have not made explicit what they assume to be experienced (this is a separate question from whether P-consciousness is fully subsumed under in A-consciousness). Some could claim that subjects experience the letter but not its color ([Fu et al., 2021](#)). Thus, in the absence of explicit assumptions from the theories about the phenomenological qualities of consciousness, the debate between access and phenomenal consciousness will not easily be settled empirically.

Nevertheless, settling the P- vs. A-consciousness debate, while representing a significant advance in the field, will not provide understanding of the phenomenological qualities of experience and their generation by the brain. To that end, clear and standardized methods for studying phenomenology in cognitive neuroscience

need to be developed. Phenomenology has been extensively discussed in philosophy and efforts have been made to naturalize its investigation in the language of the cognitive neurosciences (see [Gallagher, 2012](#), for a more detailed account). Without a doubt, obtaining phenomenological insights is a challenging task ([LeVasseur, 2003](#); [Schwitzgebel, 2008](#)), but it is a task that the science of consciousness cannot shy away from inasmuch as it constitutes our very own object of study. While we wait for the perfect methodology to naturalize phenomenology, assuming that this is even achievable, there are, meanwhile, a number of areas in which progress can be made.

For a start, by making explicit the assumptions that theories of consciousness themselves have about phenomenological qualities of experience, these assumptions can be used as a source for empirical investigation. In practice, such assumptions can be tested through the front-loaded phenomenology, which consists in using phenomenological considerations to design experimental paradigms ([Gallagher, 2003](#)). By building experimental designs around phenomenological assumptions, these can be tested directly and their neural correlates investigated. Another advantage of this approach is that it makes direct comparison between theories easier since it brings the comparison with regard to the “being” into focus, suspending the question of whether “being” and “doing” are the same or different depending on the theoretical account. There are a number of interesting areas in which this approach can be tested, such as the temporality of experience, given that divergent ideas have already been formulated ([Dennett & Kinsbourne, 2011](#); [Varela, 1999](#)). Other areas concern assumptions of primary versus secondary attributes of objects. To return to our previous example regarding attribute amnesia, it is possible that theories differ in what they assume to be mandatory conscious perception of objects, e.g., an object’s color might not be considered an intrinsic attribute while others, such as its location, might be. Similarly, when it comes to the conscious perception of a face, which features are expected to be automatically available to consciousness? Could a face be perceived without experiencing its orientation or skin color, etc? Being explicit about the “grounding,” i.e., what the theories assume experiences to be, will help us not only to refocus on what we aim to understand, i.e., the phenomenal aspect of consciousness, but also to test such assumptions and thereby improve our understanding. As such, the front-loaded method is probably the one most readily compatible with cognitive neurosciences, given that it requires neither an extensive conceptual or methodological shift, nor extensive training of subjects ([Gallagher, 2003](#)).

An alternative, yet more loaded, approach could be the more extensive use of neurophenomenology introduced by the late [Varela \(1996\)](#). This approach consists in standardizing the phenomenological qualities of experience in mathematical terms, under the working hypothesis that phenomenology and cognitive science are engaged in a relation of reciprocal constraints, by which phenomenological accounts should drive the cognitive neuroscience investigation of consciousness and in return be driven by it. More specifically, this approach consists in isolating

invariants of conscious experience, formalized through mathematics to investigate the neural mechanism accounting for them (Gallagher, 2003). The search for invariants has, for example, been used to investigate the commonalities of the experience of time (Varela, 1999). This approach has been further operationalized experimentally in a number of studies by training subjects to reliably report their experiences (Bockelman et al., 2013; Lutz et al., 2002). While similar to the report in contrastive paradigms, this method differs in that it uses introspection (standardized through training) to guide analysis, rather than fitting the reports into pre-defined categories of perceptual states in order to analyze the data (Bockelman et al., 2013). Using this method, Lutz and colleagues were able to show that the inter-trial neural measurements variability correlates with reported cognitive context from trained subjects, factors otherwise very difficult to control for (Lutz et al., 2002). However, this approach presents a number of challenges already exposed by Bayne (2004); for example, it is not immediately clear how neurophenomenology naturalizes phenomenology, and this requires further conceptualization, as does the effect of training as well as the extent to which it affects the experiences themselves. These are crucial questions; and while some of them have been addressed by Antoine Lutz with his concept of generative passages (Lutz, 2002), and newer ideas have been proposed in line with a computational approach to neurophenomenology (Ramstead et al., 2021), there remains much more to be done. An interesting proposition by Fink and colleagues in this issue, akin to the description of phenomenal qualities through their invariants, puts forward that the mathematical structure of phenomenal domains ought to be preserved in the NCC proper, thus constraining the NCC proper beyond simple correlation, offering promising empirical perspectives (Fink et al., 2021).

Another approach is to take insights from phenomenology at face value and to use them to guide our understanding of whether and how a physical system might be able to express phenomenal qualities. This is the approach taken by the Integrated Information Theory (IIT), in which phenomenology is used as a basis for the theory's axioms—identified as invariants of any possible experience—which are then translated into postulates in an attempt to address the hard question of consciousness (Tononi, 2004; Tononi et al., 2016). IIT has been used principally to evaluate consciousness in unresponsive individuals, as well as to make predictions about which brain areas might participate in consciousness and which ones might not (e.g., cerebellum Tononi et al., 2016). In a recent interesting development, the machinery of IIT was used in an attempt to explain the experience of space, i.e., why space feels extended (Haun & Tononi, 2019). Extending the axiom-driven approach used in the development of the IIT, Haun and Tononi derived basic principles that must be true in order for the experience of space to feel extended. These principles were then derived into physical properties that the cause-effect power of a system must satisfy in order to account for the properties of spatial experience. Using the assumptions underlying the IIT,

they constructed a system of 8 binary units interconnected in a grid-like fashion, and they were able to demonstrate that under certain configurations the unfolded cause-effect structure of the system can account for the properties of spatial experience. Their demonstration constitutes a very interesting approach to bridging the explanatory gap (Levine, 1983) by proposing an “explanatory identity” between the phenomenology and its physical substrate.

In sum, a remaining challenge in the field is to diversify our experimental approaches, explicitly incorporating the phenomenological dimension, which includes developing a systematic framework to describe the different ways in which experience can be structured in normal and abnormal/pathological states. Only then will we be closer to finding contact points between the mind and the brain.

2.3 The need for big data

Advancing research on the neural correlates of consciousness might require not only improving in our methodology, i.e., the contrastive method, putting the focus on phenomenology, but also means for aggregating the abundance of rich multimodal data that have been (and will be) gathered. This data encompass not only correlative, invasive, and non-invasive recordings in both healthy subjects and patients, but also causal measures, such as electrical stimulation and lesion studies, in both humans and animals, e.g., macaque monkeys, mice, and, more recently, corvids (Boly et al., 2017; Nieder et al., 2020; Palmiter, 2011; Redinbaugh et al., 2020). As such, when it comes to empirical research on the neural correlate of consciousness, it is fair to say that the field has progressed tremendously (for reviews of the literature, see Forster et al., 2020; Koch et al., 2016; Mashour et al., 2020). We are far from those early days when studying consciousness within the cognitive sciences was met with hesitation (Michel, 2019). Yet most of these studies have been performed and reported in isolation, with little aggregation of data to common databases (for recent efforts to create an aggregated database on confidence, see Rahnev et al., 2020), which would enable discoveries and proper evaluation of the reliability of the findings across experimental paradigms. Most data aggregation has occurred at the level of reviews written primarily from the perspective of the proponents of the theories, without careful treatment of data contradicting a theory’s core postulates (Dehaene & Changeux, 2011; Koch et al., 2016; Lau & Rosenthal, 2011). One example of the need for data aggregation and metaanalysis relates to P3b as a marker of perceptual awareness. As previously discussed, recent evidence suggests that P3b indexes post-perceptual processes (Cohen et al., 2020; Forster et al., 2020), rather than perceptual awareness, as initially thought; yet no systematic meta-analysis of the data has so far been carried out, leaving open the question as to whether the results generalize across different set of stimuli, tasks, experimental paradigms, etc., or might instead indicate that P3b indexes post-perceptual processes under certain conditions, but not under others. A similar example concerns the role of prefrontal cortex (PFC) in consciousness.

Studies using no-report paradigms, e.g., indexing perception through eye movements (optokinetic nystagmus) and fMRI in humans, have shown that activity in PFC may reflect report, whereas other studies in non-human primates—also using no-report paradigms in a binocular rivalry paradigm, but measuring direct electrophysiological activity (single unit)—reported decoding of the content of perception from PFC, which is not explained by report per se. Aggregating data across studies in a common database would help to make sense of these disparate results. The best approximation to date is Neurosynth (<https://www.neurosynth.org/>), which provides a synthesis of human functional neuroimaging data only (Yarkoni et al., 2011), built over automated extraction of activation coordinates. While it addresses a critical need, and as such falls short of being a revolutionary resource, Neurosynth suffers from a number of shortcomings. First, it currently includes only neuroimaging data and does not include data from other modalities, such as M-EEG, ECoG, single units, or lesion/stimulation data. Second, it does not include unthresholded data, which are critical for reliable meta-analytical results, especially considering the enormous variability in analytical choices by the investigators (Botvinik-Nezer et al., 2020). Third, it currently incorporates descriptors reflecting broad cognitive states, which will need to be refined to incorporate more nuanced descriptions of the tasks and experimental paradigms.

Distilling the NCC or finding any meaningful correlate of consciousness presupposes generalization across variables such as a stimulus's parameters, tasks, research site, etc. While the scientific aim is clear, i.e., generalization over non-critical parameters, it is not done in practice, and is also not taken into account statistically. Yarkoni (2020) has recently provided the clearest discussion of this problem, calling attention to the need to align verbal descriptions (hypothesis/theories) with statistical expressions in order to avoid current problems in psychology/neuroscience, such as the replication crisis. As the number of variables to include in a statistical model is daunting, the most reasonable strategy is to aggregate across multiple studies. In principle, the prescriptive strategy seems rather simple: researchers openly share data, run a complex statistical model over the data, and voilà—the NCC are discovered. However, it is not that simple, even when leaving aside the conceptual disputes as to how to define consciousness. Meaningful data aggregation requires far richer and more detailed descriptions of the experiment and experimental setting than are commonly provided in publications (or even shared datasets). Recognizing such limitations, some researchers in the community have made efforts to provide more comprehensive descriptions of the study materials, e.g., through platforms such as *protocols.io* and/or lab protocols, study protocols recently launched at PloS One, as well as through the development of community standards, such as Neuroimaging Data Model (NIDM) experiment. The major limitation for data aggregation is the insufficient *metadata* descriptor of the study materials, without which no meaningful statistical data aggregation enabling generalization is possible. Hence, sharing data, while useful,

is an insufficient measure unless it is accompanied by the corresponding metadata. Thus, an initial, apparently trivial but quite important step towards data aggregation would be to share a clear and detailed description of the experimental materials that is aligned with standard classifications, e.g., NIDM-experiments and tasks ontologies (Poldrack & Yarkoni, 2016) when available, and, ideally, linked with a persistent identifier (DOI) to make it findable and accessible, e.g., *protocols.io*.

Expanding such standard classification and ontologies to cover domain-specific metadata for our research community will enrich the usability of these data, as information will be properly tagged and thus meaningfully digested and aggregated by our community. A good start could be to expand the suite of tasks defined in the cognitive atlas (Poldrack et al., 2011, accessible online at <http://www.cognitiveatlas.org>) to cover those tasks typically used in our community, and to define a catalogue of the cognitive processes engaged in each of them. Another challenge will be data aggregation across modalities, i.e., fMRI, M-EEG, ECoG, etc., and species, not least because these data need to be brought together into common atlases (Bjerke et al., 2018); but there are efforts in this direction. Notwithstanding the challenges that lie ahead, and while requiring development and community efforts, data aggregation across studies seems to be a tractable problem. An initial step towards this goal will simply be to aggregate the preexisting data reported in the literature into common databases, gathering data across different studies, including all research modalities, in order to create an interactive, community-augmented meta-analysis tool for the science of consciousness (for similar efforts in other communities, see <http://metalab.stanford.edu>). This effort will enable a completely new cartography of the field as well as a theoretically neutral perspective allowing us to identify tendencies, such as whether theories rely on all experimental paradigms or only a handful of them, etc (Yaron et al., 2021). The ultimate goal, however, is much more ambitious: the capacity to share raw and/or unthresholded data across investigators (including metadata). We and others (Aru, Bachmann, et al., 2012; Graaf et al., 2012) have argued that the isolation of the NCC proper from the confounds of NCC-pr and NCC-co will require the performance of several studies to account for as many confounds as possible. The existence of a centralized database would provide the community with a “mindscope,” which would greatly facilitate this enterprise, enabling analysis across studies and taking the specific experimental condition of separate studies as regressors to disentangle pre- and post-perceptual mechanisms from the NCC proper. This will, however, require a major sociological change in research culture in order to break the habit of investigators of not sharing data, to embrace team science, and to collect metadata, thus enabling data reuse (FAIR data). The slow but steady progress in the open science revolution and ongoing changes in incentives for data sharing will certainly facilitate these initiatives.

2.4 Relying on theories and their inferential power: Comparing theories

While the aggregation of data to centralized databases will be crucial to the progress of consciousness research, the technological and methodological challenges it implies are considerable, and it could take years before such resources become easily and meaningfully available. On the other hand, the generalization of the investigation of phenomenological qualities will require that the theories clearly state what they assume those qualities to be. Furthermore, the obstacles to obtaining reliable phenomenological insights attest to the difficulty of this problem. Therefore, it will take considerable time before the above-proposed development in the field of consciousness research can enable empirical progress. One approach that can be implemented in the meantime, however, is to rely on the inferential power of existing theories of consciousness, e.g., Local Recurrent Processing (LRP) theory (Lamme & Roelfsema, 2000), Global Neuronal Workspace Theory (GNWT, Baars, 2005; Dehaene & Naccache, 2001), Higher-Order Theory (HOT, Lau & Rosenthal, 2011), and Integrated Information Theory (IIT, Tononi, 2004), to name a few. Inasmuch as theories specify a neural mechanism that generates subjective experience, they all formulate explicit predictions and boundary conditions for consciousness to emerge. The comparison of the theories predictions could, therefore, serve as guiding principles for either testing theories core predictions or comparing their predictive power against other accounts, but most importantly, for guiding discoveries, given that theories should also make predictions about novel observations/experiments (comparative testing of theories predictions regarding the NCCs was extensively discussed by Fink, 2016).

The reliance on theories' inferential power is, however, more challenging than it seems. In the current state of the field many theories exist at different levels of explanation and those are rarely compared directly (Doerig et al., 2021; Signorelli et al., 2021). As discussed by Michel (2019), theories disagree not only on the mechanisms giving rise to consciousness but also on their background assumptions. A critical divergence refers to the detection rule, i.e., what is considered conscious versus not. For example, higher-order theories (HOT) assume that the failure to report a stimulus amounts to the absence of consciousness. In contrast, other theories, such as IIT, LRP, and more generally first-order theories, state that consciousness can indeed occur in the absence of report. Therefore, while the findings of no-report paradigms may challenge HOT, proponents of HOT can easily dismiss them by rejecting the background assumption that stimuli were conscious despite the absence of report, thereby leaving their theory intact.

The comparison of theories is challenging, but it is not unsolvable. Doerig et al. (2021), recently advanced four so-called hard criteria for evaluating theories of consciousness. We commend the proposal for finding agreed-upon criteria, yet some might be too restrictive, as in the case of the blind reliance on the contrastive method (e.g., criterium#1, "paradigm cases of consciousness and the unconscious

alternative”) for the reasons previously outlined, or the second criterion of the unfolding argument (e.g., criterium#2, “the unfolding argument”). In the view of Doerig et al. (2021), because any input-output function of a recurrent network can be approximated by an equivalent feedforward network, any theory of consciousness that assumes the necessity of a specific causal structure for consciousness to emerge is either false or not within the reach of science. This purely functionalist perspective, which equates consciousness with its input-output function, leaves little room for rich, phenomenological states, such as dreams. Others have argued that the treatment of the causal structure as a black box is largely unjustified (Tsuchiya et al., 2019). Overall, the reductionist strictness of the hard criteria developed by Doerig et al. (2021) runs the risk of locking down the field to a similar state as what it was in its early stages and at a time when it could benefit the most from a broadening of experimental methodologies (see Tsuchiya et al., 2019, for a more detailed argumentation against Doerig’s hard criteria for consciousness).

Until agreed-upon criteria emerge, we believe that mechanisms such as “adversarial collaboration” might offer a viable alternative for arbitrating among competing theories, as proven by the current existence of several such projects in the field (such as Melloni et al., 2021). This practice consists in bringing advocates of two or more opposing theories together to work on a common experiment or set of experiments (Cowan et al., 2020). By doing so, the different sides agree beforehand on rules that cannot be broken without direct implications for their theories. In the framework of consciousness research, for instance, higher-order and first-order theories could agree on detection rules for a given experiment (or set of experiments), such that the condition in which a stimulus is consciously perceived is agreed upon by both sides, side-stepping the problem of under-determination (Michel, 2019). As each theory has a specific set of predictions about which patterns of neural activity will lead to consciousness, the experiment(s) should be designed in such a way that direct comparison of opposing predictions between the theories is possible. Findings will thus result in the confirmation or disproof of specific aspects of a theory. While a winner-takes-all situation, in which one of the theories is fully discredited, is unlikely to occur, such projects lead to efficient identifications of core vs. auxiliary predictions (Lakatos, 1973). But perhaps more importantly, by explicitly stating these core and auxiliary predictions and pre-registering them prior to data collection (as recommended by Cowan et al., 2020), the need to patch-up theories by transforming core into auxiliary predictions of the data is made evident and is therefore tractable. Such an effort could not only aid in settling standing quarrels in the field (Michel, 2019), it could also identify the most promising theories, while sidestepping some of the challenges of the contrastive method and its indeterminacy. This is without a doubt a challenging process, yet it is a fruitful one and is already implementable. Currently, there are two consortia, one focused on testing contradictory predictions of the Global Neuronal Workspace Theory and the Integrated Information Theory, and the other pitting first-order, recurrent processing theory (RPT),

against higher-order theories of consciousness; but more are expected to come as part of the *Accelerating Research on Consciousness* program spearheaded by the Templeton World Charity Foundation. We are excited about these developments and the impact they will have on the field and the theories in the future.

3 Conclusion

We have come a long way in the research of consciousness over the past thirty years. The initial enthusiasm and the abundance of data obtained through the contrastive method were followed by the development of a number of theories attempting to explain how consciousness fits into the physical world. As the field advanced, that initial enthusiasm faded and methodological as well as theoretical and philosophical disputes clouded the horizon. At the same time, more and more innovative approaches have emerged, expanding the range of possibilities beyond the traditional use of the contrastive method. We would like to advocate for these innovative methods, since the best way through what appears to be an impasse in the field will be via methodological and theoretical shifts in current practices. We believe it will be key for future discovery to transcend the basic dichotomy of seen vs. unseen by bringing phenomenology into the center of the empirical effort. This will be accomplished, first, by expanding efforts to describe the structure of experience and using those descriptions to guide empirical research, and second, by inviting theories to make explicit their assumptions about phenomenological properties of experience so that they are then accessible to the front-loaded approach. Furthermore, we believe great advances will be facilitated by the community's active effort to gather empirical data in common databases, to enable cross-studies analyses, and to increase the spatio-temporal grain. Beyond that, the existence of common databases will help narrow the search for the NCC by disentangling the NCC-pr and NCC-co from the NCC proper. While no agreed-upon theory has so far emerged, adversarial collaboration could offer a powerful tool for testing the explanatory power of the theories, while also leveraging the theories' inferential power to fuel novel discoveries about what the function of consciousness might be and how experience arises in the soggy matter of the brain.

Acknowledgments

We would like to thank Caspar M. Schwiedrzik and the Cogitate Consortium for discussions, to Diana Koch for the figure and proofreading of earlier versions and to William Martin for the final proofreading. We would also like to thank the Max Planck Society and the Templeton World Charity Foundation, Inc. (TWCF0389) for their generous support. The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of the Templeton World Charity Foundation, Inc.

References

- Andersen, L. M., Pedersen, M. N., Sandberg, K., & Overgaard, M. (2016). Occipital MEG activity in the early time range (<300 ms) predicts graded changes in perceptual consciousness. *Cereb Cortex*, *26*(6), 2677–2688. <https://doi.org/10.1093/cercor/bhv108>
- Aru, J., Axmacher, N., Do Lam, A. T., Fell, J., Elger, C. E., Singer, W., & Melloni, L. (2012). Local category-specific gamma band responses in the visual cortex do not reflect conscious perception. *J Neurosci*, *32*(43), 14909–14914. <https://doi.org/10.1523/JNEUROSCI.2051-12.2012>
- Aru, J., Bachmann, T., Singer, W., & Melloni, L. (2012). Distilling the neural correlates of consciousness. *Neurosci Biobehav Rev*, *36*(2), 737–746. <https://doi.org/10.1016/j.neubiorev.2011.12.003>
- Axelrod, V., Bar, M., & Rees, G. (2015). Exploring the unconscious using faces. *Trends Cogn Sci*, *19*(1), 35–45. <https://doi.org/10.1016/j.tics.2014.11.003>
- Baars, B. J. (1997). In the theatre of consciousness: Global workspace theory, a rigorous scientific theory of consciousness. *Journal of Consciousness Studies*, *4*, 292–309. <https://www.ingentaconnect.com/content/imp/jcs/1997/00000004/0000004/776>
- Baars, B. J. (2005). Global workspace theory of consciousness: Toward a cognitive neuroscience of human experience. *Prog Brain Res*, *150*, 45–53. [https://doi.org/10.1016/S0079-6123\(05\)50004-9](https://doi.org/10.1016/S0079-6123(05)50004-9)
- Bayne, T. (2004). Closing the gap? Some questions for neurophenomenology. *Phenomenology and the Cognitive Sciences*, *3*(4), 349–364. <https://doi.org/10.1023/B:PHEN.0000048934.34397.ca>
- Bjerke, I. E., Ovsthus, M., Papp, E. A., Yates, S. C., Silvestri, L., Fiorilli, J., Pennartz, C. M. A., Pavone, F. S., Puchades, M. A., Leergaard, T. B., & Bjaalie, J. G. (2018). Data integration through brain atlas: Human brain project tools and strategies. *Eur Psychiatry*, *50*, 70–76. <https://doi.org/10.1016/j.eurpsy.2018.02.004>
- Block, N. (1995). On a confusion about a function of consciousness. *Behavioral and Brain Sciences*, *18*, 272. <https://doi.org/10.1017/S0140525X00038474>
- Block, N. (2005). Two neural correlates of consciousness. *Trends Cogn Sci*, *9*(2), 46–52. <https://doi.org/10.1016/j.tics.2004.12.006>
- Block, N. (2011). Perceptual consciousness overflows cognitive access. *Trends Cogn Sci*, *15*(12), 567–575. <https://doi.org/10.1016/j.tics.2011.11.001>
- Bockelman, P., Reinerman-Jones, L., & Gallagher, S. (2013). Methodological lessons in neurophenomenology: Review of a baseline study and recommendations for research approaches. *Front Hum Neurosci*, *7*, 608. <https://doi.org/10.3389/fnhum.2013.00608>
- Boly, M., Massimini, M., Tsuchiya, N., Postle, B. R., Koch, C., & Tononi, G. (2017). Are the neural correlates of consciousness in the front or in the back of the cerebral cortex? Clinical and neuroimaging evidence. *J Neurosci*, *37*(40), 9603–9613. <https://doi.org/10.1523/JNEUROSCI.3218-16.2017>
- Botvinik-Nezer, R., Holzmeister, F., Camerer, C. F., Dreber, A., Huber, J., Johannesson, M., Kirchler, M., Iwanir, R., Mumford, J. A., Adcock, R. A., Avesani, P., Baczkowski, B. M., Bajracharya, A., Bakst, L., Ball, S., Barilari, M., Bault, N., Beaton, D., Beitner, J., ... others. (2020). Variability in the analysis of a single neuroimaging dataset by many teams. *Nature*, *582*(7810), 84–88. <https://doi.org/10.1038/s41586-020-2314-9>
- Bouton, S., Chambon, V., Tyrand, R., Guggisberg, A. G., Seeck, M., Karkar, S., Ville, D. van de, & Giraud, A. L. (2018). Focal versus distributed temporal cortex activity for speech sound category assignment. *Proc Natl Acad Sci U S A*, *115*(6), E1299–E1308. <https://doi.org/10.1073/pnas.1714279115>
- Chalmers, D. J. (1995). Facing up to the hard problem of consciousness. *Journal of Consciousness Studies*, *2*(3), 200–219. <https://doi.org/10.1093/acprof:oso/9780195311105.003.0001>
- Chalmers, D. J. (2000). What is a neural correlate of consciousness? In T. Metzinger (Ed.), *Neural correlates of consciousness: Empirical and conceptual questions* (pp. 17–39). MIT Press.
- Chen, H., & Wyble, B. (2016). Attribute amnesia reflects a lack of memory consolidation for attended information. *J Exp Psychol Hum Percept Perform*, *42*(2), 225–234. <https://doi.org/10.1037/xhp0000133>
- Cohen, M. A., Ortego, K., Kyroudis, A., & Pitts, M. (2020). Distinguishing the neural correlates of perceptual awareness and postperceptual processing. *J Neurosci*, *40*(25), 4925–4935. <https://doi.org/10.1523/JNEUROSCI.0120-20.2020>
- Cosmelli, D., & Thompson, E. (2007). Mountains and valleys: Binocular rivalry and the flow of experience. *Conscious Cogn*, *16*(3), 623–41; discussion 642–4. <https://doi.org/10.1016/j.concog.2007.06.013>
- Cowan, N., Belletier, C., Doherty, J. M., Jaroslawska, A. J., Rhodes, S., Forsberg, A., Naveh-Benjamin, M., Barrouillet, P., Camos, V., & Logie, R. H. (2020). How do scientific views change? Notes from an extended adversarial collaboration. *Perspect Psychol Sci*, *15*(4), 1011–1025. <https://doi.org/10.1177/1745691620906415>
- Crick, F., & Koch, C. (1990). Towards a neurobiological theory of consciousness. *Seminars in the Neurosciences*, *2*, 203. <https://authors.library.caltech.edu/40352/>

Lepauvre A., & Melloni L. (2021). The search for the neural correlate of consciousness: Progress and challenges. *Philosophy and the Mind Sciences*, *2*, 4. <https://doi.org/10.33735/phimisci.2021.87>



- Crick, F., & Koch, C. (1998). Consciousness and neuroscience. *Cerebral Cortex*, 8(2), 97–107. <https://doi.org/10.1093/cercor/8.2.97>
- Cruz, A. A., Garcia, D. M., Pinto, C. T., & Cechetti, S. P. (2011). Spontaneous eyeblink activity. *Ocul Surf*, 9(1), 29–41. [https://doi.org/10.1016/s1542-0124\(11\)70007-6](https://doi.org/10.1016/s1542-0124(11)70007-6)
- Dehaene, S. (2014). *Consciousness and the brain: Deciphering how the brain codes our thoughts*. Penguin.
- Dehaene, S., & Changeux, J. P. (2011). Experimental and theoretical approaches to conscious processing. *Neuron*, 70(2), 200–227. <https://doi.org/10.1016/j.neuron.2011.03.018>
- Dehaene, S., & Naccache, L. (2001). Towards a cognitive neuroscience of consciousness: Basic evidence and a workspace framework. *Cognition*, 79(1-2), 1–37. [https://doi.org/10.1016/s0010-0277\(00\)00123-2](https://doi.org/10.1016/s0010-0277(00)00123-2)
- Dehaene, S., Naccache, L., Cohen, L., Bihan, D. L., Mangin, J. F., Poline, J. B., & Riviere, D. (2001). Cerebral mechanisms of word masking and unconscious repetition priming. *Nat Neurosci*, 4(7), 752–758. <https://doi.org/10.1038/89551>
- Del Cul, A., Baillet, S., & Dehaene, S. (2007). Brain dynamics underlying the nonlinear threshold for access to consciousness. *PLoS Biol*, 5(10), e260. <https://doi.org/10.1371/journal.pbio.0050260>
- Dennett, D. C., & Kinsbourne, M. (2011). Time and the observer: The where and when of consciousness in the brain. *Behavioral and Brain Sciences*, 15(2), 183–201. <https://doi.org/10.1017/s0140525x00068229>
- Ding, N., Melloni, L., Zhang, H., Tian, X., & Poeppel, D. (2016). Cortical tracking of hierarchical linguistic structures in connected speech. *Nat Neurosci*, 19(1), 158–164. <https://doi.org/10.1038/nn.4186>
- Doerig, A., Schurger, A., & Herzog, M. H. (2021). Hard criteria for empirical theories of consciousness. *Cogn Neurosci*, 12(2), 41–62. <https://doi.org/10.1080/17588928.2020.1772214>
- Doesburg, S. M., Green, J. J., McDonald, J. J., & Ward, L. M. (2009). Rhythms of consciousness: Binocular rivalry reveals large-scale oscillatory network dynamics mediating visual perception. *PLoS One*, 4(7), e6142. <https://doi.org/10.1371/journal.pone.0006142>
- Eccles, J. C. (1974). Cerebral activity and consciousness. In *Studies in the philosophy of biology* (pp. 87–107). Springer. https://doi.org/10.1007/978-1-349-01892-5_7
- Edelman, G. M., & Tononi, G. (2000). Reentry and the dynamic core: Neural correlates of conscious experience. In *Neural correlates of consciousness - empirical and conceptual questions* (pp. 139–151). MIT Press.
- Farooqui, A. A., & Manly, T. (2018). When attended and conscious perception deactivates fronto-parietal regions. *Cortex*, 107, 166–179. <https://doi.org/10.1016/j.cortex.2017.09.004>
- ffytche, D. H., Howard, R. J., Brammer, M. J., David, A., Woodruff, P., & Williams, S. (1998). The anatomy of conscious vision: An fMRI study of visual hallucinations. *Nat Neurosci*, 1(8), 738–742. <https://doi.org/10.1038/3738>
- Fink, S. B. (2016). A deeper look at the “neural correlate of consciousness.” *Frontiers in Psychology*, 7, 1044. <https://doi.org/10.3389/fpsyg.2016.01044>
- Fink, S. B., Kob, L., & Lyre, H. (2021). A structural constraint on the neural correlates of consciousness. *Philosophy and the Mind Sciences*.
- Forster, J., Koivisto, M., & Revonsuo, A. (2020). ERP and MEG correlates of visual consciousness: The second decade. *Conscious Cogn*, 80, 102917. <https://doi.org/10.1016/j.concog.2020.102917>
- Frässle, S., Sommer, J., Jansen, A., Naber, M., & Einhauser, W. (2014). Binocular rivalry: Frontal activity relates to introspection and action but not to perception. *J Neurosci*, 34(5), 1738–1747. <https://doi.org/10.1523/JNEUROSCI.4403-13.2014>
- Fries, P., Roelfsema, P. R., Engel, A. K., Konig, P., & Singer, W. (1997). Synchronization of oscillatory responses in visual cortex correlates with perception in interocular rivalry. *Proc Natl Acad Sci U S A*, 94(23), 12699–12704. <https://doi.org/10.1073/pnas.94.23.12699>
- Fu, Y., Yan, W., Shen, M., & Chen, H. (2021). Does consciousness overflow cognitive access? Novel insights from the new phenomenon of attribute amnesia. *Sci China Life Sci*, 64(6), 847–860. <https://doi.org/10.1007/s11427-020-1831-8>
- Gaillard, R., Dehaene, S., Adam, C., Clemenceau, S., Hasboun, D., Baulac, M., Cohen, L., & Naccache, L. (2009). Converging intracranial markers of conscious access. *PLoS Biol*, 7(3), e61. <https://doi.org/10.1371/journal.pbio.1000061>
- Gallagher, S. (2003). Phenomenology and experimental design: Toward a phenomenologically enlightened experimental science. *Journal of Consciousness Studies*, 10(9-10), 85–99. <https://www.ingentaconnect.com/content/imp/jcs/2003/0000010/f0020009/art00007>
- Gallagher, S. (2012). *What is phenomenology?* (pp. 7–18). Palgrave Macmillan UK. https://doi.org/10.1057/9781137283801_2
- Gerber, E. M., Golan, T., Knight, R. T., & Deouell, L. Y. (2017). Cortical representation of persistent visual stimuli. *Neuroimage*, 161, 67–79. <https://doi.org/10.1016/j.neuroimage.2017.08.028>
- Golan, T., Davidesco, I., Meshulam, M., Groppe, D. M., Megevand, P., Yeagle, E. M., Goldfinger, M. S., Harel, M., Melloni, L., Schroeder, C. E., Deouell, L. Y., Mehta, A. D., & Malach, R. (2016). Human intracranial recordings link suppressed transients rather than ‘filling-in’ to perceptual continuity across blinks. *Elife*, 5. <https://doi.org/10.7554/eLife.17243>

Lepauvre A., & Melloni L. (2021). The search for the neural correlate of consciousness: Progress and challenges. *Philosophy and the Mind Sciences*, 2, 4. <https://doi.org/10.33735/phimisci.2021.87>



- Gosselin, F., & Schyns, P. G. (2003). Superstitious perceptions reveal properties of internal representations. *Psychol Sci*, *14*(5), 505–509. <https://doi.org/10.1111/1467-9280.03452>
- Graaf, T. A. de, Hsieh, P. J., & Sack, A. T. (2012). The 'correlates' in neural correlates of consciousness. *Neurosci Biobehav Rev*, *36*(1), 191–197. <https://doi.org/10.1016/j.neubiorev.2011.05.012>
- Gui, P., Jiang, Y., Zang, D., Qi, Z., Tan, J., Tanigawa, H., Jiang, J., Wen, Y., Xu, L., Zhao, J., Mao, Y., Poo, M. M., Ding, N., Dehaene, S., Wu, X., & Wang, L. (2020). Assessing the depth of language processing in patients with disorders of consciousness. *Nat Neurosci*, *23*(6), 761–770. <https://doi.org/10.1038/s41593-020-0639-1>
- Haun, A., & Tononi, G. (2019). Why does space feel the way it does? Towards a principled account of spatial experience. *Entropy*, *21*(12), 1160. <https://doi.org/10.3390/e21121160>
- Haynes, J. D. (2009). Decoding visual consciousness from human brain signals. *Trends Cogn Sci*, *13*(5), 194–202. <https://doi.org/10.1016/j.tics.2009.02.004>
- Henson, R. N., Mouchlianitis, E., Matthews, W. J., & Kouider, S. (2008). Electrophysiological correlates of masked face priming. *Neuroimage*, *40*(2), 884–895. <https://doi.org/10.1016/j.neuroimage.2007.12.003>
- Hippocrates. (1959). *Hippocrates* (W. H. S. Jones, Trans.). Harvard Univ. Press, 1923.
- James, W., Burkhardt, F., Bowers, F., & Skrupskelis, I. K. (1890). *The principles of psychology* (Vol. 1). Macmillan London.
- Kapoor, V., Dwarakanath, A., Safavi, S., Werner, J., Besserve, M., Panagiotaropoulos, T. I., & Logothetis, N. K. (2020). Decoding the contents of consciousness from prefrontal ensembles. *bioRxiv*. <https://doi.org/10.1101/2020.01.28.921841>
- Kim, C. Y., & Blake, R. (2005). Psychophysical magic: Rendering the visible 'invisible'. *Trends Cogn Sci*, *9*(8), 381–388. <https://doi.org/10.1016/j.tics.2005.06.012>
- Kleinschmidt, A., Buchel, C., Zeki, S., & Frackowiak, R. (2002). Human brain activity during spontaneously reversing perception of ambiguous figures. *5th IEEE EMBS International Summer School on Biomedical Imaging, 2002.*, 7–pp. <https://doi.org/10.1109/SSBI.2002.1233971>
- Koch, C., Massimini, M., Boly, M., & Tononi, G. (2016). Neural correlates of consciousness: Progress and problems. *Nat Rev Neurosci*, *17*(5), 307–321. <https://doi.org/10.1038/nrn.2016.22>
- Koch, C., & Tsuchiya, N. (2007). Attention and consciousness: Two distinct brain processes. *Trends Cogn Sci*, *11*(1), 16–22. <https://doi.org/10.1016/j.tics.2006.10.012>
- Koivisto, M., & Revonsuo, A. (2007). Electrophysiological correlates of visual consciousness and selective attention. *Neuroreport*, *18*(8), 753–756. <https://doi.org/10.1097/WNR.0b013e3280c143c8>
- Koivisto, M., & Revonsuo, A. (2010). Event-related brain potential correlates of visual awareness. *Neurosci Biobehav Rev*, *34*(6), 922–934. <https://doi.org/10.1016/j.neubiorev.2009.12.002>
- Lakatos, I. (1973). Science and pseudoscience. *Philosophical Papers*, *1*, 1–7. <http://www.podstawyekonomii.pl/metodologia/files/lakatos1977.pdf>
- Lamme, V. A. (2006). Towards a true neural stance on consciousness. *Trends Cogn Sci*, *10*(11), 494–501. <https://doi.org/10.1016/j.tics.2006.09.001>
- Lamme, V. A., & Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends Neurosci*, *23*(11), 571–579. [https://doi.org/10.1016/s0166-2236\(00\)01657-x](https://doi.org/10.1016/s0166-2236(00)01657-x)
- Lau, H., & Rosenthal, D. (2011). Empirical support for higher-order theories of conscious awareness. *Trends Cogn Sci*, *15*(8), 365–373. <https://doi.org/10.1016/j.tics.2011.05.009>
- LeVasseur, J. J. (2003). The problem of bracketing in phenomenology. *Qual Health Res*, *13*(3), 408–420. <https://doi.org/10.1177/1049732302250337>
- Levine, J. (1983). Materialism and qualia: The explanatory gap. *Pacific Philosophical Quarterly*, *64*(4), 354–361. <https://doi.org/10.1111/j.1468-0114.1983.tb00207.x>
- Libet, B., Alberts, W. W., Wright, Jr., E. W., & Feinstein, B. (1967). Responses of human somatosensory cortex to stimuli below threshold for conscious sensation. *Science*, *158*(3808), 1597–1600. <https://doi.org/10.1126/science.158.3808.1597>
- Logothetis, N. K. (1998). Single units and conscious vision. *Philos Trans R Soc Lond B Biol Sci*, *353*(1377), 1801–1818. <https://doi.org/10.1098/rstb.1998.0333>
- Lutz, A. (2002). Toward a neurophenomenology as an account of generative passages: A first empirical case study. *Phenomenology and the Cognitive Sciences*, *1*(2), 133–167. <https://doi.org/10.1023/a:1020320221083>
- Lutz, A., Lachaux, J.-P., Martinerie, J., & Varela, F. J. (2002). Guiding the study of brain dynamics by using first-person data: Synchrony patterns correlate with ongoing conscious states during a simple visual task. *Proceedings of the National Academy of Sciences*, *99*(3), 1586–1591. <https://doi.org/10.1073/pnas.032658199>
- Mashour, G. A., Roelfsema, P., Changeux, J. P., & Dehaene, S. (2020). Conscious processing and the global neuronal workspace hypothesis. *Neuron*, *105*(5), 776–798. <https://doi.org/10.1016/j.neuron.2020.01.026>
- Mayer, A., Schwiedrzik, C. M., Wibral, M., Singer, W., & Melloni, L. (2016). Expecting to see a letter: Alpha oscillations as carriers of top-down sensory predictions. *Cereb Cortex*, *26*(7), 3146–3160. <https://doi.org/10.1093/cercor/bhv146>

Lepauvre A., & Melloni L. (2021). The search for the neural correlate of consciousness: Progress and challenges. *Philosophy and the Mind Sciences*, *2*, 4. <https://doi.org/10.33735/phimisci.2021.87>



- Mazzi, C., Mazzeo, G., & Savazzi, S. (2020). Late positivity does not meet the criteria to be considered a proper neural correlate of perceptual awareness. *Front Syst Neurosci*, *14*, 36. <https://doi.org/10.3389/fnsys.2020.00036>
- Melloni, L. (2015). Consciousness as inference in time. *Open MIND*, *22*, 1–13. <https://doi.org/10.15502/9783958570566>
- Melloni, L., Mudrik, L., Pitts, M., & Koch, C. (2021). Making the hard problem of consciousness easier. *Science*, *372*(6545), 911–912. <https://doi.org/10.1126/science.abj3259>
- Melloni, L., Schwiedrzik, C. M., Müller, N., Rodriguez, E., & Singer, W. (2011). Expectations change the signatures and timing of electrophysiological correlates of perceptual awareness. *J Neurosci*, *31*(4), 1386–1396. <https://doi.org/10.1523/JNEUROSCI.4570-10.2011>
- Melloni, L., Schwiedrzik, C. M., Müller, N., Rodriguez, E., & Singer, W. (2011). Expectations change the signatures and timing of electrophysiological correlates of perceptual awareness. *Journal of Neuroscience*, *31*(4), 1386–1396. <https://doi.org/10.1523/JNEUROSCI.4570-10.2011>
- Metzinger, T. (2000). *Neural correlates of consciousness: Empirical and conceptual questions*. MIT press.
- Michel, M. (2019). Consciousness science underdetermined: A short history of endless debates. *Ergo, an Open Access Journal of Philosophy*, *6*(20201214). <https://doi.org/10.3998/ergo.12405314.0006.028>
- Miller, S. M. (2015). *The constitution of phenomenal consciousness: Toward a science and theory* (Vol. 92). John Benjamins Publishing Company.
- Naccache, L. (2018). Why and how access consciousness can account for phenomenal consciousness. *Philos Trans R Soc Lond B Biol Sci*, *373*(1755), 20170357. <https://doi.org/10.1098/rstb.2017.0357>
- Nieder, A., Wagener, L., & Rinnert, P. (2020). A neural correlate of sensory consciousness in a corvid bird. *Science*, *369*(6511), 1626–1629. <https://doi.org/10.1126/science.abb1447>
- Noy, N., Bickel, S., Zion-Golumbic, E., Harel, M., Golan, T., Davidesco, I., Schevon, C. A., McKhann, G. M., Goodman, R. R., Schroeder, C. E., Mehta, A. D., & Malach, R. (2015). Ignition's glow: Ultra-fast spread of global cortical activity accompanying local "ignitions" in visual cortex during conscious visual perception. *Conscious Cogn*, *35*, 206–224. <https://doi.org/10.1016/j.concog.2015.03.006>
- Overgaard, M., Rote, J., Mouridsen, K., & Ramsøy, T. Z. (2006). Is conscious perception gradual or dichotomous? A comparison of report methodologies during a visual task. *Conscious Cogn*, *15*(4), 700–708. <https://doi.org/10.1016/j.concog.2006.04.002>
- Palmiter, R. D. (2011). Dopamine signaling as a neural correlate of consciousness. *Neuroscience*, *198*, 213–220. <https://doi.org/10.1016/j.neuroscience.2011.06.089>
- Penfield, W. (1958). Some mechanisms of consciousness discovered during electrical stimulation of the brain. *Proc Natl Acad Sci U S A*, *44*(2), 51–66. <https://doi.org/10.1073/pnas.44.2.51>
- Pins, D., & ffytche, D. (2003). The neural correlates of conscious vision. *Cereb Cortex*, *13*(5), 461–474. <https://doi.org/10.1093/cercor/13.5.461>
- Pitts, M. A., & Britz, J. (2011). Insights from intermittent binocular rivalry and EEG. *Front Hum Neurosci*, *5*, 107. <https://doi.org/10.3389/fnhum.2011.00107>
- Pitts, M. A., Lutsyshyna, L. A., & Hillyard, S. A. (2018). The relationship between attention and consciousness: An expanded taxonomy and implications for 'no-report' paradigms. *Philos Trans R Soc Lond B Biol Sci*, *373*(1755). <https://doi.org/10.1098/rstb.2017.0348>
- Pitts, M. A., Martinez, A., & Hillyard, S. A. (2012). Visual processing of contour patterns under conditions of inattentive blindness. *J Cogn Neurosci*, *24*(2), 287–303. https://doi.org/10.1162/jocn_a_00111
- Pitts, M. A., Padwal, J., Fennelly, D., Martinez, A., & Hillyard, S. A. (2014). Gamma band activity and the P3 reflect post-perceptual processes, not visual awareness. *Neuroimage*, *101*, 337–350. <https://doi.org/10.1016/j.neuroimage.2014.07.024>
- Poldrack, R. A., Kittur, A., Kalar, D., Miller, E., Seppa, C., Gil, Y., Parker, D. S., Sabb, F. W., & Bilder, R. M. (2011). The cognitive atlas: Toward a knowledge foundation for cognitive neuroscience. *Front Neuroinform*, *5*, 17. <https://doi.org/10.3389/fninf.2011.00017>
- Poldrack, R. A., & Yarkoni, T. (2016). From brain maps to cognitive ontologies: Informatics and the search for mental structure. *Annu Rev Psychol*, *67*, 587–612. <https://doi.org/10.1146/annurev-psych-122414-033729>
- Posner, M. I. (1994). Attention: The mechanisms of consciousness. *Proc Natl Acad Sci U S A*, *91*(16), 7398–7403. <https://doi.org/10.1073/pnas.91.16.7398>
- Rahnev, D., Desender, K., Lee, A. L. F., Adler, W. T., Aguilar-Lleyda, D., Akdogan, B., Arbuza, P., Atlas, L. Y., Balci, F., Bang, J. W., Begue, I., Birney, D. P., Brady, T. F., Calder-Travis, J., Chetverikov, A., Clark, T. K., Davranche, K., Denison, R. N., Dildine, T. C., ... Zylberberg, A. (2020). The confidence database. *Nat Hum Behav*, *4*(3), 317–325. <https://doi.org/10.1038/s41562-019-0813-1>

Lepauvre A., & Melloni L. (2021). The search for the neural correlate of consciousness: Progress and challenges. *Philosophy and the Mind Sciences*, *2*, 4. <https://doi.org/10.33735/phimisci.2021.87>



- Ramstead, M. J. D., Hesp, C., Sandved-smith, L., Mago, J., Lifshitz, M., Pagnoni, G., Lutz, A., & Friston, K. (2021). From generative models to generative passages: A computational approach to (neuro)phenomenology. *Preprint*, 1–29. <https://doi.org/10.31234/osf.io/k9pbn>
- Redinbaugh, M. J., Phillips, J. M., Kambi, N. A., Mohanta, S., Andryk, S., Dooley, G. L., Afrasiabi, M., Raz, A., & Saalman, Y. B. (2020). Thalamus modulates consciousness via layer-specific control of cortex. *Neuron*, *106*(1), 66–75 e12. <https://doi.org/10.1016/j.neuron.2020.01.005>
- Revonsuo, A. (2000). Prospects for a scientific research program on consciousness. In T. Metzinger (Ed.), *Neural correlates of consciousness* (pp. 57–75). The MIT press. <https://doi.org/10.7551/mitpress/4928.003.0006>
- Revonsuo, A. (2006). *Inner presence: Consciousness as a biological phenomenon*. MIT Press.
- Salti, M., Monto, S., Charles, L., King, J. R., Parkkonen, L., & Dehaene, S. (2015). Distinct cortical codes and temporal dynamics for conscious and unconscious percepts. *Elife*, *4*, 1–52. <https://doi.org/10.7554/eLife.05652>
- Schelonus, K., Grauly, C., Canseco-Gonzalez, E., & Pitts, M. A. (2017). ERP signatures of conscious and unconscious word and letter perception in an inattentive blindness paradigm. *Conscious Cogn*, *54*, 56–71. <https://doi.org/10.1016/j.concog.2017.04.009>
- Schwitzgebel, E. (2008). The unreliability of naive introspection. *Philosophical Review*, *117*(2), 245–273. <https://doi.org/10.1215/00318108-2007-037>
- Sergent, C., Baillet, S., & Dehaene, S. (2005). Timing of the brain events underlying access to consciousness during the attentional blink. *Nat Neurosci*, *8*(10), 1391–1400. <https://doi.org/10.1038/nn1549>
- Sergent, C., Corazzol, M., Labouret, G., Stockart, F., Wexler, M., King, J. R., Meyniel, F., & Pressnitzer, D. (2021). Bifurcation in brain dynamics reveals a signature of conscious processing independent of report. *Nat Commun*, *12*(1), 1149. <https://doi.org/10.1038/s41467-021-21393-z>
- Shafto, J. P., & Pitts, M. A. (2015). Neural signatures of conscious face perception in an inattentive blindness paradigm. *J Neurosci*, *35*(31), 10940–10948. <https://doi.org/10.1523/JNEUROSCI.0145-15.2015>
- Siclari, F., Baird, B., Perogamvros, L., Bernardi, G., LaRocque, J. J., Riedner, B., Boly, M., Postle, B. R., & Tononi, G. (2017). The neural correlates of dreaming. *Nat Neurosci*, *20*(6), 872–878. <https://doi.org/10.1038/nn.4545>
- Signorelli, C. M., Szcotka, J., & Prentner, R. (2021). *Explanatory profiles of models of consciousness-towards a systematic classification*. <https://doi.org/10.31234/osf.io/f5vdu>
- Silverstein, B. H., Snodgrass, M., Shevrin, H., & Kushwaha, R. (2015). P3b, consciousness, and complex unconscious processing. *Cortex*, *73*, 216–227. <https://doi.org/10.1016/j.cortex.2015.09.004>
- Sokoliuk, R., Degano, G., Banellis, L., Melloni, L., Hayton, T., Sturman, S., Veenith, T., Yakoub, K. M., Belli, A., Noppeney, U., & Cruse, D. (2021). Covert speech comprehension predicts recovery from acute unresponsive states. *Ann Neurol*, *89*(4), 646–656. <https://doi.org/10.1002/ana.25995>
- Soto, D., & Silvanto, J. (2014). Reappraising the relationship between working memory and conscious awareness. *Trends Cogn Sci*, *18*(10), 520–525. <https://doi.org/10.1016/j.tics.2014.06.005>
- Tononi, G. (2004). An information integration theory of consciousness. *BMC Neurosci*, *5*, 42. <https://doi.org/10.1186/1471-2202-5-42>
- Tononi, G., Boly, M., Massimini, M., & Koch, C. (2016). Integrated information theory: From consciousness to its physical substrate. *Nat Rev Neurosci*, *17*(7), 450–461. <https://doi.org/10.1038/nrn.2016.44>
- Tsuchiya, N., Andriillon, T., & Haun, A. (2019). *A reply to “the unfolding argument”: Beyond functionalism/behaviorism and towards a truer science of causal structural theories of consciousness*. <https://doi.org/10.31234/osf.io/a2ms9>
- Tsuchiya, N., Wilke, M., Frassle, S., & Lamme, V. A. F. (2015). No-report paradigms: Extracting the true neural correlates of consciousness. *Trends Cogn Sci*, *19*(12), 757–770. <https://doi.org/10.1016/j.tics.2015.10.002>
- Varela, F. J. (1996). Neurophenomenology: A methodological remedy for the hard problem. *Journal of Consciousness Studies*, *3*(4), 330–349. <https://www.ingentaconnect.com/content/imp/jcs/1996/00000003/00000004/718>
- Varela, F. J. (1999). The specious present: A neurophenomenology of time consciousness. In J. Petitot, F. J. Varela, J. Pachoud, & J.-M. Roy (Eds.), *Naturalizing phenomenology: Issues in contemporary phenomenology and cognitive science* (pp. 266–314). Stanford University Press.
- Varela, F., Lachaux, J. P., Rodriguez, E., & Martinerie, J. (2001). The brainweb: Phase synchronization and large-scale integration. *Nat Rev Neurosci*, *2*(4), 229–239. <https://doi.org/10.1038/35067550>
- Wang, R., Fu, Y., Chen, L., Chen, Y., Zhou, J., & Chen, H. (2021). Consciousness can overflow report: Novel evidence from attribute amnesia of a single stimulus. *Conscious Cogn*, *87*, 103052. <https://doi.org/10.1016/j.concog.2020.103052>
- Weiskrantz, L. (1980). Varieties of residual experience. *Q J Exp Psychol*, *32*(3), 365–386. <https://doi.org/10.1080/14640748008401832>
- Yarkoni, T. (2020). The generalizability crisis. *Behav Brain Sci*, 1–37. <https://doi.org/10.1017/S0140525X20001685>

Lepauvre A., & Melloni L. (2021). The search for the neural correlate of consciousness: Progress and challenges. *Philosophy and the Mind Sciences*, *2*, 4. <https://doi.org/10.33735/phimisci.2021.87>



Yarkoni, T., Poldrack, R. A., Nichols, T. E., Van Essen, D. C., & Wager, T. D. (2011). NeuroSynth: A new platform for large-scale automated synthesis of human functional neuroimaging data. *Frontiers in Neuroinformatics*, 5. <https://doi.org/10.3389/conf.fninf.2011.08.00058>

Yaron, I., Melloni, L., Pitts, M., & Mudrik, L. (2021). The consciousness theories studies (ConTraSt) database: Analyzing and comparing empirical studies of consciousness theories. *bioRxiv*. <https://doi.org/10.1101/2021.06.10.447863>

Open Access

This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

