Growing evidence for separate neural mechanisms for attention and consciousness

Alexander Maier 1 · Naotsugu Tsuchiya 2,3,4

Accepted: 8 September 2020 / Published online: 9 October 2020 © The Psychonomic Society, Inc. 2020

Abstract

Our conscious experience of the world seems to go in lockstep with our attentional focus: We tend to see, hear, taste, and feel what we attend to, and vice versa. This tight coupling between attention and consciousness has given rise to the idea that these two phenomena are indivisible. In the late 1950s, the honoree of this special issue, Charles Eriksen, was among a small group of early pioneers that sought to investigate whether a transient increase in overall level of attention (alertness) in response to a noxious stimulus can be decoupled from conscious perception using experimental techniques. Recent years saw a similar debate regarding whether attention and consciousness are two dissociable processes. Initial evidence that attention and consciousness are two separate processes primarily rested on behavioral data. However, the past couple of years witnessed an explosion of studies aimed at testing this conjecture using neuroscientific techniques. Here we provide an overview of these and related empirical studies on the distinction between the neuronal correlates of attention and consciousness, and detail how advancements in theory and technology can bring about a more detailed understanding of the two. We argue that the most promising approach will combine ever-evolving neurophysiological and interventionist tools with quantitative, empirically testable theories of consciousness that are grounded in a mathematically formalized understanding of phenomenology.

Keywords Attention: Neural Mechanisms · Electrophysiology · Cognitive neuroscience · Consciousness

Introduction

Consciousness and attention are both fascinating and difficult to define. The lack of widely accepted definitions stirs frequent debate in either field of study (Block, 2005; Block et al., 2014; Krauzlis, Bolimunta, Arcizet and Wang, 2014; Luo and Maunsell, 2019; Maunsell, 2015; Michel et al., 2019). For the purpose of this paper, we will define consciousness as subjective (phenomenal) experience. Note that this definition encompasses both levels and contents of consciousness (Bachmann and Hudetz, 2014; Hohwy, 2009), although we will mainly focus on the latter by examining empirical studies of perceptual experience. Moreover, our definition of consciousness does not rely on behavior, such as whether a particular experience can or does get reported (which some authors have distinguished as “access consciousness” or awareness (Block, 2005)). We will similarly bypass various distinctions of attention by collapsing them into the intuitive notion of mental focus on a particular aspect of information (a wide definition seems justified given recent evidence for a general attention factor (Huang, Mo and Li, 2012)). By relying on intuition, our definition of attention thus parallels that of the father of American psychology, William James, who famously wrote that “everyone knows what attention is” (James, 1890).

Interestingly, James went on to state that “focalization, concentration of consciousness are of its essence,” suggesting an intricate link between consciousness and attention. In this view attention is consciousness – just more of it. The notion of attention being identical to consciousness is still favored by some (O’Regan and Noe, 2001). One argument for this view is that attending something seems to amplify consciousness, such as enhancing visual contrast sensitivity and/or perceptual...
appearance (M. Carrasco, 2011). However, another possible interpretation of this observation is that attention is a different process than consciousness, but interacts with consciousness in a certain way. One such view holds that consciousness is necessary for attention in that attention is a modulation of conscious experience. Thus, some authors argue that attention depends on consciousness. To others it seems similarly intuitive that conscious experience only occurs once we attend to the world around us. In that view, consciousness depends on attention, and not the other way around (M. A. Cohen, Cavanagh, Chun and Nakayama, 2012). Finally, there is the view that attention and consciousness can be dissociated from each other (Lamme, 2004). This hypothesis is falsifiable in that it comes with several testable predictions (Koch and Tsuchiya, 2007), such as that consciousness and attention can occur independently of each other, that the effects of consciousness and attention might go in opposing directions, and that the neural substrates supporting each of the two phenomena are separable. This last point is the central focus of this review.

Eriksen’s pioneering contributions

As outlined above, the relationship between attention and consciousness has long been a subject of debate. In some ways, this discussion can be traced back to the scientist honored by this special edition, Charles Eriksen and his seminal work on “subception” in the 1950s and 60s. Eriksen was interested in studies where volunteers received electric shocks while nonsense syllables were presented for varying amounts of time. The initial data suggested that whether or not subjects consciously perceived the very briefly presented stimuli, repeated presentations of the same stimuli evoked a galvanic skin response, reminiscent of subjects anticipating the pain (Lazarus and McCleary, 1951). This phenomenon, coined “subception,” was seen as evidence that a heightened “arousal” (which, in this context, means a transient increase in alertness or unfocused attention) can be evoked in the absence of conscious perception of the stimulus that caused it (see Robbins et al. (1998), Thiele (2013), and Tsuchiya and Koch (2014) for more on the argument that arousal is closely related to attention; but also see Otazu, Tai, Yang and Zador (2009)). Eriksen was critical of this interpretation and argued that there were shortcomings of the original study (Eriksen, 1956). As we discuss below, several of Eriksen’s critiques such as that verbal reports are limited in determining consciousness (Eriksen, 1960) have re-emerged in recent years.

Behavioral evidence that attention and consciousness are dissociable

More than 40 years after Charles Eriksen’s investigation of alertness changes in the absence of conscious perception, new empirical work re-emerged that raises similar questions. In fact, the original question as to the relation of alertness and unconscious processing has been expanded and dissected in several ways: (1) Can unconscious stimuli still become the target of top-down attentional selection? (2) Can unconscious stimuli still attract attention in a bottom-up manner? And the focus of this review: (3) What are the neural correlates of unconscious stimuli that interact with attention in certain ways?

Dissociations between consciousness and attention have been demonstrated in various ways (for review, see Cohen et al., 2012; Dehaene, Changeux, Naccache, Sackur and Sergent, 2006; Koch & Tsuchiya, 2007; Tsuchiya and Koch, 2016). While manipulation of attention can be objectively verified through behavioral measures (e.g., decreased reaction time, improved performance), there is arguably no widely accepted norm for how to behaviorally assess conscious perception (Ramsey and Overgaard, 2004; Seth, Dienes, Cleeremans, Overgaard and Pessoa, 2008). While conscious experience has largely been equated with verbal reports, verbal reports are increasingly recognized as highly limited as methods for capturing richness of conscious experience (Block, 2005; Haun et al., 2017; Sperling, 1960). Recent studies even question the stringent requirement of behavioral reports, as most of our conscious experience is unreported normally and the act of reports can affect phenomenology, attention, and neural activity (Cohen, Ortego, Kyroudis and Pitts, 2020; Koch, Massimini, Boly and Tononi, 2016; Tsuchiya, Wilke, Frassle and Lamme, 2015). These issues need to be taken into account when comparing results between studies.

The null hypothesis that attention and consciousness are either identical or indivisibly coupled predicts that (1) there is no attention outside consciousness and (2) there is no consciousness without attention. Thus, any demonstration of either attention acting on nonconscious stimuli or consciousness occurring in the complete absence of attention are direct evidence for a possible dissociation between attention and consciousness.

One evidence for a complete separation between attention and consciousness would take the form of a double dissociation, where both these phenomena (i.e., attention without consciousness and consciousness without attention) are demonstrated. Such double dissociation would suggest that attention and consciousness are two separate phenomena and they can work independently. This most powerful dissociation can be achieved by manipulating conscious visibility and attentional availability in an independent manner (with a 2 x 2 design). Using such a paradigm, several studies demonstrated dissociable behavioral or neuronal effects (Brascamp, van Boxtel, Knapen and Blake, 2010; Smout and Mattingley, 2018; van Boxtel, Tsuchiya and Koch, 2010a, 2010b; J. J. A. van Boxtel, 2017; Watanabe et al., 2011). However, several of these findings have recently been challenged (Travis, Dux, & Mattingley, 2017; Yuval-Greenberg & Heeger, 2013), thus requiring further investigation.
In addition to double dissociation, there are other ways to demonstrate that attention and consciousness can be uncoupled. For example, some studies show that attention can operate outside of consciousness, or vice versa. The most robust dissociation of this kind could be summarized as “attentional effects without conscious visibility.” In these studies, sensory stimuli (typically of attentional cues) are manipulated to be either visible or invisible (Kim and Blake, 2005). Then, attentional effects are measured. If objectively measured (behaviorally or neuronally) attentional effects are constant, regardless of conscious visibility, this kind of dissociation is deemed to be established.

A more controversial dissociation is “conscious visibility regardless of attentional enhancement.” Here, attentional availability or expectation can be manipulated by dual-task designs (Matthews, Schröder, Kaunitz, Van Boxtel and Tsuchiya, 2018), rapid serial visual presentation (or attentional blink), task instruction, or task-set priming (including “inattentional blindness” or “change blindness” (Cohen, Botch and Robertson, 2020; Cohen and Rubenstein, 2020)). If conscious visibility of a stimulus does not depend on the attentional state, the dissociation is thought to be established. A part of the associated controversy stems from the fact that attentional manipulation often accompanies the changes of stimulus input per se. This leaves the possibility that attention could “spill over” to a presumably “unattended” stimulus.

Initial evidence for a dissociation between attention and consciousness was provided by a clinical study that showed that neurological patients lacking large parts of primary visual cortex (V1) respond faster when attention is drawn to a stimulus while being unable to consciously see what evoked the change in attentional state (Kertridge, Heywood and Weiskrantz, 1999, 2004; Schurer, Cowey, Cohen, Treisman and Tallon-Baudry, 2008).

A similarly stunning result was found in split-brain patients whose cerebral hemispheres were surgically separated to treat severe epilepsy (Gazzaniga, 2014). An interesting consequence of the split-brain surgery is that the operated patients seem to report that their visual perception is split in half (but see: Pinto, de Haan and Lamme, 2017; Pinto, Lamme and de Haan, 2017; Pinto et al., 2017). When these patients are asked to report their visual experience using either their left or right hand (which are controlled by the right and left hemispheres, respectively), they can only indicate whatever is shown to the left or right visual hemifields. In other words, everything that is to one side of their nose seems to be perceived by only one half of their brain and this side of their brain is completely unaware of what is shown to the other side of their nose (Volz and Gazzaniga, 2017). And yet, this is not the case for their attentional focus, which can still operate across the entire visual field (Corballis, 1995). For example, attentional cueing experiments found that when spatial cues are presented to either the left or right visual hemifield, performance is identical, suggesting that attention is a unified resource shared between cerebral hemispheres (but also see Alvarez & Cavanagh, 2005). This finding not only suggests that the neural substrates for consciousness and attention are separable, but that attention may be closely tied to subcortical structures.

This finding in human split-brain patients are also consistent with a recent surge of primate electrophysiology research, which points to non-cortical structures, such as the superior colliculus, thalamic pulvinar nucleus, and basal ganglia as critical neural substrates for attention (Arcizet and Krauzlis, 2018; Bogadhi, Bollimunta, Leopold and Krauzlis, 2018; Bollimunta, Bogadhi and Krauzlis, 2018; Fiebelkorn and Kastner, 2020; Halassa and Kastner, 2017; Krauzlis et al., 2014; Krauzlis, Bogadhi, Herman and Bollimunta, 2018; Krauzlis, Lovejoy and Zenon, 2013; Saalmann and Kastner, 2011; Saalmann, Pinsk, Wang, Li and Kastner, 2012; Zenon and Krauzlis, 2012). Attention thus might have a more expansive subcortical basis than previously appreciated. Accordingly, severing interhemispheric cortical connections in split-brain patients may have less of an impact on attention (which in some ways remains intact) than on conscious perception (which appears to be split in half, or otherwise diminished). However, note that there is also emerging evidence that certain subcortical structures, such as higher order nuclei of the thalamus, play a crucial role for both levels as well as contents of consciousness (Donoghue et al., 2019; Redinbaugh et al., 2020; Schmid and Maier, 2015; Wilke, Mueller and Leopold, 2009). Thus, while these novel findings point at dissociable neural processes for attention and consciousness, it is not as simple as attention being a subcortical phenomenon and consciousness a product of the neocortex.

Following up on these seminal studies in the blindsight and split-brain patients, a large number of studies performed on healthy observers established that attention can be captured outside of consciousness (Faiivre and Kouider, 2011; Kertridge, Nijboer and Heywood, 2008; Kiefer and Martens, 2010; Martens, Ansorge and Kiefer, 2011; Sumner, Tsai, Yu and Nachev, 2006). Based on this and other evidence (for further recent updates, see Tsuchiya & Koch, 2016), largely speaking, there is a strong agreement in the field that consciousness is not necessary for attention, but whether attention is necessary for consciousness continues to be debated (M. A. Cohen et al., 2012; Michael A Cohen, Dennett and Kanwisher, 2016; Dehaene et al., 2006; van Boxtel et al., 2010a).

Neuroscientific approaches to study the attention/consciousness divide

One of the most impactful propositions in recent years that was derived from this empirical work has been the hypothesis that neuronal mechanisms that support consciousness and
attention can work independently and be dissociated from each other (Koch & Tsuchiya, 2007; Koch and Tsuchiya, 2012; Lamme, 2003, 2004). When originally formulated, this hypothesis mostly rested on behavioral data since there were few studies at the time testing for separate neural substrates (but see Wyart and Tallon-Baudry, 2008). This situation has changed drastically during the past years, with several independent groups during investigating whether attention and consciousness can be dissociated not only in behavioral terms, but also in their underlying patterns of cortical activation (Nani et al., 2019). This empirical test is important in the attempt to falsify the original dissociation hypothesis: If there were no distinction in the brain activation patterns for attention and consciousness, it seems implausible for attention and consciousness to be supported by distinct neural mechanisms. In what follows, we review several neuroimaging and neurophysiological studies in order to determine whether or not consensus has emerged across the various experimental approaches taken to date.

**fMRI evidence for a separation between consciousness and attention**

Functional magnetic resonance imaging non-invasively yields sub-centimeter localization of brain function across the entire brain. However, this fine-grained global coverage comes at the cost of a rather poor temporal resolution of several seconds. Despite these technical limitations, recent fMRI studies have been pivotal in demonstrating that attention can increase neural response to invisible stimuli. And thanks to the spatial resolution of fMRI, such effects could be located in distinct parts of the brain, such as the visual cortex (e.g., V1) (Watanabe et al., 2011).

fMRI even produced evidence pointing at a dissociation between attention and consciousness (Kouider, Barbot, Madsen, Lehericy and Summerfield, 2016): Whether or not subjects consciously saw a stimulus yielded a larger or smaller fMRI response, respectively. Similarly, whether they attended a stimulus or not (in this particular case, whether they were trying to detect a certain stimulus type or not) produced somewhat smaller changes in fMRI response. Importantly consciousness can have that effect even when keeping attention unaltered, and vice versa (Fig. 1a – note the difference in units on the ordinate) (Kouider et al., 2016).

An even more striking dissociation was found after distinguishing between different kinds of attention. There are many different ways in which attention can be elicited, such as a subject’s behavioral goals (goal driven, or endogenous attention) or by salient stimulus events (stimulus driven, or exogenous attention). Comparing the whole-brain activation pattern evoked by endogenous attention to the brain activation linked to a threshold stimulus reaching consciousness revealed a surprisingly small overlap; the thalamus was the only part of the brain that showed a statistical interaction between these two phenomena (Chica, Bayle, Botta, Bartolomeo and Paz-Alonso, 2016).

Finally, combined with a refined behavioral control, fMRI can localize a brain region that is involved in either attention or consciousness, but not the other. For example, Webb and colleagues first established behavioral effects induced by visible and invisible attentional cues. Under these equivalent attentional conditions, they demonstrated that the brain area that remains modulated by stimulus visibility was the temporo-parietal junction (TPJ) (Webb, Igelstrom, Schurger and Graziano, 2016). While Webb’s finding is at odds with previous literature claiming that the TPJ is modulated by attention (e.g., Han and Marois, 2014), we would like to point out that most studies performed prior to 2000 rarely made any effort to keep attentional effects constant while changing consciousness. Webb et al.’s findings thus suggest that older neuroscientific studies that reached conclusions regarding either consciousness or attention without varying one of these factors while keeping the other constant, might need to be re-examined.

**EEG/MEG evidence for a separation between consciousness and attention**

In contrast to fMRI, the electro/magneto-encephalogram (E/MEG) – the other widely used non-invasive technique – does not allow for precise localization of neural activity in space. But E/MEG has an exquisite temporal resolution. It thus can provide complementary evidence to fMRI on the question of attention and consciousness operating independently of each other.

Somatosensory stimulation (touch) of fingers, for example, can be applied to participants right at the threshold of conscious perception. Such weak stimuli are consciously perceived or not, depending on each trial. Under these circumstances, physically identical stimuli evoke a greater EEG response when they produce a conscious touch sensation (Fig. 2a). Subjects then can be tasked to either attend the stimulated fingers or attend elsewhere. If the neural mechanisms supporting attention and consciousness were inseparable, there should be no effect of attention on the EEG response when the stimuli went undetected. Yet, attention increases the EEG magnitude even to unconscious stimuli (Forschack, Nierhaus, Muller and Villringer, 2017). In other words, the somatosensory EEG response follows a 2 x 2 dissociation (conscious/unconscious vs. attended/unattended) matrix (Fig. 2a), similar to what was found across behavioral studies (Tsuchiya & Koch, 2016).

Analogous work in the visual domain produced similar results (Chica, Lasaponara, Lupianez, Doricchi and
Bartolomeo, 2010; Wyart & Tallon-Baudry, 2008). In the study by Wyart and Tallon-Baudry, participants viewed faint visual patterns that hovered around the perceptual threshold, thus consciously seeing it only about half of the time. Critically, in 50% of trials subjects attended the screen location where the target images were shown and in the other 50% of trials they attended elsewhere. The authors then compared instances when subjects attended to stimuli that were either seen or unseen, and stimuli that were visible when attended versus unattended. This study used magnetoencephalography (MEG), which is believed to be more sensitive to deeper (folded) parts of the cortex than EEG (Fig. 2b). Even more noteworthy, MEG is sensitive for brain activity across a wide range of frequencies. Thus, this study was able to go beyond investigating the mere magnitude of brain responses. The authors found that consciousness was correlated with changes in low gamma activity (55–65 Hz) while attention related more to high gamma (75–90 Hz) frequencies instead (Wyart & Tallon-Baudry, 2008). A similar spectral dissociation between conscious perception and attention has been found in patients with V1 lesions (Schurger et al., 2008). What is more, there was a difference in onset of the neural activity related to

![Fig. 1](https://example.com/f1.png) fMRI and EEG evidence for separate neuronal processes underlying attention and consciousness. (a) fMRI study showing that, while attention increases neuronal response magnitudes independently of whether a stimulus reached conscious perception, the larger difference in neuronal activity is related to whether a subject was conscious or unconscious of that stimulus (note the difference in units on the y-axis). (b) EEG response magnitudes to conscious and unconscious stimuli in the presence or absence of attention, showing a similar effect to that in (a). (c) EEG responses to conscious (magenta) vs. unconscious attentional (cyan) cues. Note that even a cue that is not consciously perceived evokes an elevated (albeit somewhat smaller) EEG response, indicating attentional enhancement of neuronal responses

![Fig. 2](https://example.com/f2.png) Double dissociations between attention and consciousness demonstrated using E/MEG. (a) EEG responses to tactile stimuli that were either consciously perceived or not (y-axis) and at the same time either attended or not (x-axis, red vs. blue). Note that each of these combinations yielded a different EEG response magnitude. (b) Spectral profiles of MEG responses to consciously perceived vs. unconscious visual stimuli (left), independent of attentional state and attended vs. unattended stimuli, independent of conscious state (right). Note the difference in frequency and onset for the power maxima of each of the two contrasts
attention and the neural responses correlating with consciousness, with attention lagging conscious perception. Other studies (by the same authors) studying the magnitude of the responses rather than their spectral profile found a similar asynchrony between attention and consciousness, albeit in the reverse order (Wyart, Dehaene and Tallon-Baudry, 2011).

Attention-related EEG components were also observed both when observers consciously experienced attention-directing cues and when they failed to do so (Travis, Dux and Mattingley, 2019). However, the magnitude of this attention signal was diminished when the cue was invisible (Fig. 1c) (Travis et al., 2019). Several other studies convergingly suggest that attention can modulate EEG responses even if stimuli do not reach consciousness (Kiefer and Brendel, 2006; Koivisto and Revonsuo, 2007, 2008; Koivisto, Revonsuo and Lehtonen, 2006; Sanchez-Lopez, Savazzi, Pedersini, Cardobi and Marzi, 2019; Woodman and Luck, 2003).

On the flip side, neural correlates of consciousness have been observed independently of attentional state (Boehler, Schoenfeld, Heinze and Hopf, 2008; Koivisto et al., 2006). When attention and consciousness of stimuli are independently manipulated, both attention and consciousness of stimuli can enhance brain responses (Fig. 1b). Importantly, the general pattern emerging from these studies is that each of these response enhancements can be evoked in the absence of the other or additively superimposed (Fig. 3a).

Two noteworthy recent EEG studies are opening exciting new directions towards studying the dissociation of attention and consciousness. The first study focused on the dissociation of attention and consciousness in terms of the temporal rhythms at which they operate. Recent evidence suggests that various measures of task performance tend to oscillate at either theta- (~7Hz) or alpha- (~10Hz) rhythms, implicating a rhythmic structure to perception (Van Rullen, 2016), attention (Helfrich et al., 2018; Kienitz et al., 2018; Landau and Fries, 2012) or both. Davidson and colleagues (Matthew J Davidson, Alais, van Boxtel and Tsuchiya, 2018) investigated these rhythms in a binocular rivalry paradigm, where each eye’s view is confronted with a drastically different stimulus and perception alternates between them (Blake and Logothetis, 2002; Maier, Panagiotaropoulos, Tsuchiya and Keliris, 2012). They found that attentional cueing during binocular rivalry induces rhythmicity, around 3.5Hz or 8Hz, in terms of conscious perceptual alternations (see also Cha and Blake, 2019). Importantly, the cue was either congruent with the visible or invisible stimulus (Luonghi, Morrone and Alais, 2014). When the cue was congruent with the visible stimulus, it transiently increased 8 Hz rhythms of perceptual alternations, whereas the incongruent cue increased perceptual alternations at 3.5 Hz. To explain the latter finding, the authors hypothesized that attentional sampling at 8 Hz can be performed between conscious and nonconscious stimuli, resulting in half the attentional sampling speed. This is a novel form of dissociation between attention and consciousness that warrants further investigation.

Furthermore, Davidson et al performed simultaneous EEG recordings during the rivalry task. Their intertrial phase coherence (ITPC) analyses at 3.5 Hz and 8 Hz at the scalp EEG showed that the strength of the ITPC evoked by the attentional cues was correlated with the behaviorally observed perceptual rhythms. Specifically, parieto-occipital electrodes showed 8-Hz ITPC, which correlated with the 8-Hz rhythmic perceptual reversals, while fronto-temporal showed 3.5-Hz ITPC, which correlated with attentional effects to the invisible stimulus. This finding seems consistent with the recently proposed role of theta activity in the fronto-parietal attention network in exploitation and environmental sampling (Fiebelkorn and Kastner, 2019). To date, such dissociation between consciousness and attention in the temporal domain are still relatively scarce (but see: Naccache, Blandin and Dehaene, 2002), but this research line marks a promising future direction, particularly when utilizing the excellent temporal resolution of E/MEG.

The other notable EEG study utilized temporally modulating stimuli to evoke a steady-state visually evoked potential (SSVEP) (Norcia, Appelbaum, Ales, Cottereau and Rossion, 2015; Smout & Mattingley, 2018; Vialatte, Maurice, Dauwels and Cichocki, 2010). SSVEPs are used to “frequency-tag” stimuli by turning them on and off at distinct flicker frequencies. Since neurons are expected to respond in synchrony with the on- and offsets of flickering stimuli, this frequency tagging improves the stimulus-selectivity of EEG responses by characterizing stimulus-specific neural responses in the frequency-domain. Critically, the strength of SSVEPs has been used extensively as a correlate of both the focus of attention without overt responses (Morgan, Hansen and Hillyard, 1996; Muller et al., 2006), as well as the contents of consciousness during changes in perception (Brown and Norcia, 1997; Lansing, 1964).

Davidson and colleagues (Davidson, Mithen, Hogendoorn, van Boxtel and Tsuchiya, 2020) adapted a display where SSVEP stimuli faded from view due to visual adaptation (see also Davidson, Graafsma, Tsuchiya and van Boxtel, 2020; Schieting and Spillmann, 1987; Weil, Kilner, Haynes and Rees, 2007)). Intriguingly, this perceptual fading is known to be enhanced by attention, in that attending to flickering stimuli results in more rapid and more frequent disappearance (DeWeerd, Smith and Greenberg, 2006; Lou, 1999). This remarkable “negative” effect of attention on consciousness has also been reported for other phenomena, such as motion-induced blindness (Geng, Song, Li, Xu and Zhu, 2007; Scholvinck and Rees, 2009) and negative afterimages (J. W. Brascamp et al., 2010; Lou, 2001; Suzuki and Grabowecky, 2003; van Boxtel et al., 2010b). Davidson et al. found that perceptual fading of the flickering stimulus was accompanied by an increase in the SSVEP signal at the frequency that tagged the disappearing stimulus. At the same time, alpha-band activity (8–12Hz), whose diminishment is
commonly seen as an index of the attentional orienting (D’Andrea et al., 2019; Gould, Rushworth and Nobre, 2011; Horschig, Oosterhert, Oostenveld and Jensen, 2015; Jensen, Gips, Bergmann and Bonnefond, 2014; Spaak, Bonnefond, Maier, Leopold and Jensen, 2012), decreased at the time of target disappearance. This study raises an interesting question regarding two common assumptions about consciousness and attention. First, given that target disappearance increased the strength of stimulus-specific neural responses, the notion that consciousness is always accompanied by enhanced neuronal responses needs to be refined (see also Merten and Nieder, 2012, and next section). Second, previous SSVEP experiments that aimed to tag the neural correlates of consciousness (Brown & Norcia, 1997; Tononi, Srinivasan, Russell and Edelman, 1998) might have inadvertently tagged the neural correlates of attention instead (see also P. Zhang, Jamison, Engel, He and He, 2011).

These new advances in empirical studies are prompting a revision of the theoretical models of consciousness (see below).

Fig. 3 (a) Schematic summary of the effects of attention and conscious state on the magnitude of large-scale sensory brain responses (modified from Koivisto & Revonsuo, 2008). (b) summary of the effects of attention and conscious state on the magnitude of large-scale sensory brain responses. Specifically, results are depicted from two single-neuron recording studies that measured the effects of conscious perception (from (David A Leopold, Maier, Wilke and Logothetis, 2005), modified) and attentional selection (from (Luck et al., 1997), modified) in the visual cortex of macaque monkeys. Note the similar pattern of enhanced responses co-varying with either phenomenon. Also note that the magnitude of these enhanced responses varies considerably between cortical areas (i.e., increasing from early sensory cortex towards associative cortex). What is not shown are other changes at the single-neuron level that

have been associated with the phenomena, affecting correlated activity, changes in neural noise, intra- and inter-area coherence and changes in oscillatory activity, some of which cannot be observed when studying single neurons in isolation. Also note that these data sample neurons across entire cortical areas without regard to differences in cell type or neural circuitry (such as cortical layers) that they are embedded in (see (Maier et al., 2007; Sheinberg and Logothetis, 1997) for a sample of various types of single neuron correlate of consciousness). Thus, the neuronal concomitants of both consciousness and attention are likely more than a simple “boost” in neuronal activity, but rather constitute a complex arrangement of neural mechanisms with distinct regional characteristics

Single neuronal activity correlates with attention and consciousness in a distinct manner

All of the studies summarized so far were based on non-invasive techniques that can readily be used for human subjects, yet lag either in temporal or spatial resolution. One way to circumvent these methodological constraints is to resort to invasive techniques in animal models. Non-human primates in particular offer phylogenetic (and thus neuroanatomical) closeness (Dougherty, Schmid and Maier, 2018; Kaas, 2012a, 2012b, 2013), the ability to apply complex (non-verbal) cognitive tasks suited for the study of attention (Bosman et al., 2012; Luck, Chelazzi, Hillyard and Desimone, 1997; Maunsell, 2015; Mitchell, Sundberg and Reynolds, 2009; Pooremesaili, Poort, Thiele and Roelfsema, 2010) and consciousness (D. A. Leopold, Maier and Logothetis, 2003) as well as a greater spectrum of invasive techniques, such as the neurophysiological measurement of action potentials that can be linked back to one or more brain
cells (Roelfsema and Treue, 2014). One of the downsides of choosing this approach is the associated cost in labor-intensive work that results in long time spans from experimental planning to data collection and analysis (typically years). Perhaps as a result of this, we are not aware of any published neurophysiological studies that examined neuronal activity related to consciousness and attention while manipulating both independently of each other. There is, however, a large body of studies that examined each of these phenomena in isolation (e.g., Fig. 3b).

Interestingly, the single neuron data collected in animals parallels the human studies summarized above in that they suggest that (1) both attention and consciousness correlate with increased brain activity and (2) consciousness correlates with a bigger increase in brain activity compared to attention (Fig. 3b). Indeed, most studies studying either phenomenon in isolation found a similar general trend of enhanced neuronal responses across both cortical and subcortical structures when subjects are conscious (as opposed to unconscious) of a sensory stimulus (Maunsell, Roelfsema, Changeux and Dehaene, 2020; Noy et al., 2015; but also see Logothetis, 1998). When a stimulus is attended, similar, but more modest boosts in neuronal responses are typically observed (Maunsell, 2015). However, this is not always the case. Recent work demonstrated that under some circumstances, the exact opposite can be found (Matthew J Davidson, Mithen, et al., 2020; Merten & Nieder, 2012).

One way to explain these apparent discrepancies are regional differences in the brain. For example, while reward tends to elevate spiking responses across most of the neocortex, including most visual cortical areas, area V4 seems to show the inverse pattern. At least during some tasks, whenever subjects learn that they will be receiving reward, V4 neurons undergo profound suppression (the larger the expected reward, the deeper the suppression) (Shapcott et al., 2020). During binocular rivalry, ~20% of recorded neurons in V4 and MT also show a decrease of firing rate when the preferred stimuli are consciously experienced by the monkeys (Logothetis, 1998). This see-saw pattern of neuronal responses is also seen on a larger scale for the so-called default mode network (Lundqvist, Bastos and Miller, 2020). When a stimulus is attended, similar, but more modest boosts in neuronal responses are typically observed (Maunsell, 2015). However, this is not always the case. Recent work demonstrated that under some circumstances, the exact opposite can be found (Matthew J Davidson, Mithen, et al., 2020; Merten & Nieder, 2012).

One way to explain these apparent discrepancies are regional differences in the brain. For example, while reward tends to elevate spiking responses across most of the neocortex, including most visual cortical areas, area V4 seems to show the inverse pattern. At least during some tasks, whenever subjects learn that they will be receiving reward, V4 neurons undergo profound suppression (the larger the expected reward, the deeper the suppression) (Shapcott et al., 2020). During binocular rivalry, ~20% of recorded neurons in V4 and MT also show a decrease of firing rate when the preferred stimuli are consciously experienced by the monkeys (Logothetis, 1998). This see-saw pattern of neuronal responses is also seen on a larger scale for the so-called default mode network (Lundqvist, Bastos and Miller, 2020). When a stimulus is attended, similar, but more modest boosts in neuronal responses are typically observed (Maunsell, 2015). However, this is not always the case. Recent work demonstrated that under some circumstances, the exact opposite can be found (Matthew J Davidson, Mithen, et al., 2020; Merten & Nieder, 2012).

A final consideration is that attention-related increases in spiking magnitude are only part of the story (Luo & Maunsell, 2014) and changes in effective connectivity (van Kempen et al., 2020). While there is some paralleling evidence (Seward and Seward, 2001), less is known about the role of these phenomena for consciousness.

**Future research promises a more granular dissection of consciousness and attention**

By focusing on tractable questions, the neuroscience of consciousness has made tremendous advances during the past two decades (Boly et al., 2013; Koch et al., 2016; Mashour et al., 2020). One of these questions that seems close to consensus is whether conscious perception and selective attention are supported by two distinct brain mechanisms. Both the behavioral as well as the neuroscientific evidence, as reviewed in this article, leans increasingly towards the affirmative. The neural responses related to attention and consciousness seem distinct in magnitude, spectral properties, and at least to some degree in their spatial extent across the brain. This consensus opens more questions. Among these questions is how the neural correlates of consciousness and attention differ between different parts of the brain. Exactly which neural structures are involved in giving rise to consciousness and attention is still unclear, especially when moving from a more global perspective to the scale of neural circuits and cell types. This question is unlikely to be answered using only non-invasive techniques that are limited in either spatial or temporal resolution.

**Outlook 1: Technological developments for further refining our understanding of the neuronal substrates of consciousness and attention**

Two recent developments are particularly promising in that regard. First, thanks to the success of deep-brain stimulation, there is an ever-increasing number of people who live for extended periods of time with chronic electrode implants in their brains. This development goes hand-in-hand with invasive neurophysiological recordings that provide resolution at the level of single neurons becoming increasingly commonplace in human neurological patients (Aquino et al., 2020; Isik, Singer, Madsen, Kanwisher and Kreiman, 2018; Kreiman, Koch and Fried, 2000; Prynuk, Kfir, Gelbard-Sagiv, Fried and Paz, 2019; Rees, Kreiman and Koch, 2002; Rutishauser et al., 2013; Rutishauser et al., 2015). Given this trend, it seems increasingly likely that we are facing a future where neurophysiological implants are used in healthy individuals (Musk, 2019).

A second, related, technological breakthrough is the rapid ascent of silicone-based microelectrode arrays that provide real-time microscopic (i.e., single-neuron) resolution across...
several centimeters (Jun et al., 2017; Musk, 2019; Steinmetz, Koch, Harris and Carandini, 2018). In other words, thanks to modern electrode design and technology together with advances in computing power, big data storage and machine learning-based analyses (Pachitariu, Steinmetz, Kadir, Carandini and Harris, 2016), it has become feasible to simultaneously record thousands of neurons across a wide array of subcortical and cortical areas (Jun et al., 2017; Musk, 2019; Steinmetz et al., 2018; Steinmetz, Zatka-Haas, Carandini and Harris, 2019; Stringer, Pachitariu, Steinmetz, Carandini and Harris, 2019). It is hard to overestimate the quantum leap that has been achieved using this approach (Fig. 4).

Up until now, as we reviewed in the above sections for E/MEG and fMRI, neuroimaging techniques have sacrificed either spatial or temporal resolution when trying to cover the entire brain. Only by substantially limiting the spatial coverage, electrophysiological measures were able to get to the resolutions of single neurons. Thanks to these new electrode arrays, we can now study single neuron activity across large parts of the brain with sub-millisecond resolution. More than that, due to the large number (hundreds to thousands) and high spatial resolution (less than tens of microns) of electrodes on these silicon probes, the spiking activity of hundreds of neurons recorded simultaneously can now be sampled in a 2D spatial mosaic consisting of multiple adjacent microelectrodes. The resulting spatial profile of action potentials can be used to create an “electromorphological” fingerprint to identify and classify different neuron types using in vivo extracellular neurophysiology in a previously unprecedented way (Jia et al., 2019).

So, what do these technological advances offer for our understanding of the neural basis of consciousness and attention? The massive increase in electrode count currently underway in the field of neurophysiology is more than just a quantitative advance, but it marks the dawn of a qualitatively new advance in assessing brain activity. The implications are manifold, but immediately apparent for the question at the heart of this review. Both consciousness and attention arise from the activity of neurons that are widely dispersed across the brain. Whether neurons that support these phenomena differ biochemically (or not; Maier, Logothetis and Leopold, 2007), morphologically, or in any other physical characteristic from neighboring neurons is still unknown, as is the question of how far the neural populations supporting consciousness and attention overlap. More importantly, rather than being limited to studying isolated neurons, neurophysiology is now entering a phase where entire neural circuits, spanning cortical layers, areas and nuclei are measured without sacrificing either spatial or temporal resolution (i.e., we are now studying brain activity at the macro- and mesoscale with microscale resolution).

With this technological advance, comes increased need for a quantitative theory of consciousness and attention. We

![Fig. 4 Promising new technological developments. (a) Recent years witnessed a “quantum leap” in the count of simultaneously placed microelectrodes used in animal studies (from Steinmetz et al., 2018, modified). (b) This massive increase in the simultaneously measured activity of single neurons allows for more global (across-the-brain) measures at a sub-millisecond and sub-millimeter resolution. Given the spatio-temporal complexities associated with the neural concomitants of consciousness and attention (Fig. 3), this unprecedented combination of macroscopic scale and microscopic resolution promises a host of new insights beyond current methodological constraints from (Stringer, Pachitariu, Steinmetz, Reddy, et al., 2019) and (Jun, Steinmetz, Siegle, et al., 2017), modified](image-url)
particularly need a theory that not only manages to explain the conceptual advances outlined above, but that can also provide empirically testable and non-trivial (e.g., counterintuitive) predictions on consciousness and attention. Such theoretical guidance will promote the targeted and theory-driven data collection to specifically test hypotheses on the brain mechanisms of consciousness and attention. With a successful marriage of theory and empirical experiments, we are poised to gain unprecedented new insight in these phenomena at the neuronal circuit-level (see below).

**Outlook 2: Developments towards an empirically testable quantitative science of phenomenology**

Along with the above-mentioned technological breakthroughs, we also need to make conceptual breakthroughs to make real progress in our understanding of the neural basis of consciousness and attention.

First of all, the full potential of these new invasive neuroscientific techniques will be at least initially primarily available in animals. Therefore, we urgently need to develop improved experimental paradigms that can distinguish consciousness and attention in the absence of verbal communication, such as no-report paradigms (Frassle, Sommer, Jansen, Naber and Einhauser, 2014; D. A. Leopold et al., 2003; Tsuchiya et al., 2015; Wilke et al., 2009). Indeed, studies that eliminated overt behavioral reports during experimental investigations of conscious perception have revealed that part of what have previously been deemed the neural correlates of consciousness (NCC) may actually be brain activity related to the planning and execution of behavioral reports (Aru, Bachmann, Singer and Melloni, 2012; DeGraaf, Hsieh and Sack, 2012; Tsuchiya et al., 2015). This has been an important step in narrowing down which neural responses are truly consciousness-related, as there are arguably many additional processes that are engaged in the brain, once, and only if, consciousness has changed (attention being one of them). The no-report approach to eliminating downstream processes of the NCC thus is in some way comparable to the use of bistable (also called multistable) stimuli, which allows researchers to keep sensory stimulation constant while perception varies in order to eliminate low-level responses upstream of the NCC (J. Brascamp, Sterzer, Blake and Knapen, 2018; Frith, Perry and Lumer, 1999; Lumer, Friston and Rees, 1998).

No-report paradigms have raised interesting questions regarding the relationship between attention and consciousness (Cohen, Ortego, et al., 2020; Frassle et al., 2014; Montemayor and Haladjian, 2019; Wilke et al., 2009). For example, eliminating report also challenges whether the (low-frequency) gamma activity that has been associated with consciousness might be related to post-perceptual decision making and/or motor planning instead (Pitts, Padwal, Fennelly, Martinez and Hillyard, 2014). No-report paradigms pose certain challenges for experiments that seek to dissociate attention from consciousness in that report is often used to control attentional state. It thus has been difficult so far to combine both approaches. Future experiments will likely fill this increasingly pressing knowledge gap.

While animal studies with massive neural recordings and more sophisticated behavioral paradigms will undoubtedly play a critical role in advancing our understanding of the neural mechanisms of consciousness and attention, human studies will also need to become more quantitatively sophisticated to promote our understanding of the structure of conscious phenomenology. In general, consciousness researchers painfully know how little we understand the structure of consciousness, but this issue has been largely neglected in the last decades due to the immense advance in the side of neuroscience of consciousness. Indeed, we foresee further developments of psychophysics. For example, studies of consciousness and attention based on the phenomenon of “change blindness” can benefit from powerful sensory manipulations using virtual reality (Cohen, Botch and Robertson, 2020). The relationship between consciousness and attention in everyday life could be probed with wearable devices used for experiential sampling (Kahneman, Krueger, Schkade, Schwarz and Stone, 2004; Killingsworth and Gilbert, 2010) that are combined with wearable devices that take neural measurements (Casson, 2019). Another direction of future psychophysical work might utilize big-data analysis and massive-scale recruitment of participants online (Bridges, Pitiot, MacAskill and Peirce, 2020; Crump, McDonnell and Gureckis, 2013; Sheehan, 2018). These techniques can overcome limitations of traditional “button press” psychophysical experiments that are too restrictive to fully characterize the rich nature of conscious phenomenology (Varela, 1996; Zahavi and Gallagher, 2008). While richer descriptions of conscious phenomenology through extensive verbal reports or subjects’ drawings have been utilized in important case reports, as in neglect patients (Bisiach and Luzzatti, 1978), this approach has not been fully utilized in various research domains pertaining to consciousness and attention (Fei-Fei, Iyer, Koch and Perona, 2007; Haun et al., 2017).

A particularly exciting promising development in this regard is the rapidly emerging research program that aims to uncover and mathematically formalize isomorphisms between the structure of conscious phenomenology and the structure of information that is inherent in neuronal activity (Fig. 5). While there is a surface resemblance to studies that aim to uncover the NCC, this new line of research adds emphasis on structure (i.e., relative distances of states in both experiential and neuronal space). For example, a common approach to test whether a particular set of neurons belonged to the NCC is to keep sensory inputs constant while phenomenology is changing...
(Frith et al., 1999). This paradigm has been successfully applied to determine whether certain neurons and what aspect of neuronal responses correlate with consciousness (see D. A. Leopold and Logothetis, 1999, Michel et al., 2019, Parker and Newsome, 1998, for review). The promising next step is to focus on the structural relations that exist within phenomenology, such as the phenomenon that tones of certain frequency ratios (e.g., musical notes of the same chroma (Shepard, 1982)) sound more similar than other tones or the fact that visual stimuli are located relative to each other in visual space. This novel approach subsumes prior work on the NCC, and extends that line of research to study different sensory inputs that evoke similar phenomenology, such as short and long wavelengths of light evoking the perception of purple and red, which are phenomenologically more similar than the colors evoked by physically more similar wavelengths.

Figure 5 provides a conceptual schematic for this approach. For example, the phenomenological relationship between two types of sounds (e.g., C major chords), and the relationships among various other sounds (as well as conscious sensations in other modalities) can be quantified and mathematically expressed (e.g., in a multidimensional space through large-scale similarity rating experiments or in other ways using topology or category theory). Here, relationships among variously pitched tones are represented as a helix of musical “chroma” (Shepard, 1982), which captures the essential phenomenological feature of two tones that sound “similar but different.” At the neuronal side, connectivity and activity states of the neural networks are measured, and mathematically abstracted to propose physical structures (e.g., informational structure in IIT) that are supposed to give rise to the specific conscious experience. A search for an isomorphism starts with a unidirectional correspondence from one structure to the other (see main text). When both directions converge, we get closer to an isomorphic law that bridges between consciousness and brain activity. Attentional manipulation should preserve this isomorphism. That is, attention cannot change phenomenal structure without changing the information structure set up by neuronal activity, and vice versa. Brain icon is modified from James.mcd.nz, licensed under the Creative Commons Attribution-Share Alike 4.0 International, 3.0 Unported, 2.5 Generic, 2.0 Generic and 1.0 Generic license (https://commons.wikimedia.org/wiki/File:Brain_Surface_Gyri.SVG). The icon representing the information structure is modified from Takemori39, licensed under the Creative Commons Attribution-Share Alike 3.0 Unported license (https://commons.wikimedia.org/wiki/File:Network_self-organization_stages.png).

Fig. 5 Promising new theoretical directions. A possible research program for the search of an isomorphism between the structure of consciousness and the structure of information derived from the activity and structure of the neuronal activity. First, phenomenological structure needs to be precisely mapped out. As one possible mapping strategy, we can start with characterizing phenomenological relationships between various sounds (e.g., C major chords), as shown here, as well as conscious sensations in other modalities. Attention can be used here as a way to “perturb” phenomenology. These relationships can be quantified and mathematically expressed (e.g., in a multidimensional space through large-scale similarity rating experiments or in other ways using topology or category theory). Here, relationships among variously pitched tones are represented as a helix of musical “chroma” (Shepard, 1982), which captures the essential phenomenological feature of two tones that sound “similar but different.” At the neuronal side, connectivity and activity states of the neural networks are measured, and mathematically abstracted to propose physical structures (e.g., informational structure in IIT) that are supposed to give rise to the specific conscious experience. A search for an isomorphism starts with a unidirectional correspondence from one structure to the other (see main text). When both directions converge, we get closer to an isomorphic law that bridges between consciousness and brain activity. Attentional manipulation should preserve this isomorphism. That is, attention cannot change phenomenal structure without changing the information structure set up by neuronal activity, and vice versa. Brain icon is modified from James.mcd.nz, licensed under the Creative Commons Attribution-Share Alike 4.0 International, 3.0 Unported, 2.5 Generic, 2.0 Generic and 1.0 Generic license (https://commons.wikimedia.org/wiki/File:Brain_Surface_Gyri.SVG). The icon representing the information structure is modified from Takemori39, licensed under the Creative Commons Attribution-Share Alike 3.0 Unported license (https://commons.wikimedia.org/wiki/File:Network_self-organization_stages.png).
between the (differences in) neuronal space and the (differences in) phenomenological space — two sounds that sound similar are accompanied by similar neural activity. However, neither of these two spaces seem to be simple, Euclidian vector spaces. Phenomenological space does not seem to follow the rules of geometric space: Even just considering these similarity relations, the abstracted phenomenological structure of consciousness may not be sufficiently represented by a simple geometric model (e.g., a multi-dimensional vector space; Gärdenfors, 2000) as similarity relations are known to violate axioms of distances. One example of this phenomenon is an asymmetrical relationship in similarity judgements (Hodgetts and Hahn, 2012; Palmer, 1999; Polk, Behensky, Gonzalez and Smith, 2002; Pothos, Busemeyer and Trueblood, 2013; Tversky, 1977). Such asymmetry in relationships violates the axioms of distances, suggesting that underlying phenomenal structure cannot be captured by a simple geometric model. In that sense, the phenomenological structure mapped by this new approach might necessitate more flexible and abstract mathematical tools, such as topology and category theory (A. Haun and Tononi, 2019; Prentner, 2019; Tsuchiya and Saigo, 2020).

At the same time, the similarity in neural activation giving rise to similar experiences cannot simply be measured in physical space (including transformations such as the Talairach reference brain space): Two observers having the same conscious experience will not have the exact same 3D activation pattern in their brains (even monozygotic twins have anatomically distinct brains, with different gyri and sulci; Lohmann, Von Cramon and Steinmetz, 1999). Thus, some kind of abstraction is necessary to find the specific aspect of brain activity that is isomorphic to the structure of phenomenology. One promising idea is to mathematically extract the informational structure that is inherent in neuronal activity (Fig. 5, bottom row). A particularly promising analytical approach in this regard is provided by the Integrated Information Theory (IIT) of consciousness proposed by Tononi (Tononi, 2004). IIT rests on a small set of axioms that are based on the first-person experience of consciousness. For example, any conscious experience is integrated. That is, we have a singular experience that incorporates all senses. Further, any conscious experience is informative in that it uniquely specifies what we experience among all possible experiences. The key argument of IIT is that consciousness is integrated information. That is, any integrated information is a form of consciousness. This definition excludes most information in the universe since it is not integrated to a unified whole. In order for information to turn into consciousness, it has to become integrated by a system such as our brains. Importantly, this key aspect of integration is well defined and measurable, and doing so shows that most information-processing systems do not feature this property. Most importantly, IIT is both a mathematically formalized framework and allows for quantitative assessments. Specifically, IIT defines the amount of integrated information as the amount of causal constraints of a system that is specified by all combinations of its parts — that is, how much the whole is more than the sum of parts.

Recent advances of IIT have culminated in the development of algorithms to empirically derive (accurate proxies of) informational structures from neural recordings (Barrett and Seth, 2011; A. M. Haun et al., 2017; Leung, Cohen, Swinderen and Tsuchiya, 2020; Mayner et al., 2018; Oizumi, Albantakis and Tononi, 2014; Oizumi, Amari, Yanagawa, Fujiy and Tsuchiya, 2016; Tegmark, 2016).

IIT-based analyses rest on a solid theoretical foundation in that IIT successfully explains various properties of conscious phenomenology and known neural physiology such as why we can lose consciousness due to brain lesions, general anesthesia, certain types of epilepsy, and deep sleep while our brains remain highly active. IIT suggests that in all these cases, loss of consciousness is caused by disruptions of the structure of integrated information that is otherwise carried by the brain. IIT also explains why certain brain lesions, such as in the cerebellum, basal ganglia, and even in prefrontal cortex seem to have relatively minor impact on consciousness while those in posterior cortical areas are much more severe with regard to consciousness (Boly et al., 2017; but see Odegaard, Knight and Lau, 2017). IIT suggests that this difference is due to the causal structures that are readily available in posterior cortex in the form of lateral connections between cortical columns (A. Haun & Tononi, 2019; Tononi, Boly, Massimini and Koch, 2016).

In addition to this ad hoc explanatory power, IIT also makes various testable counterintuitive predictions about conscious phenomenology and its neuronal substrates. For example, IIT predicts that “inactivating” already “inactive” neurons in retinotopic visual areas (such as V1/V2) should completely change how we visually perceive space (A. Haun & Tononi, 2019; Song, Haun and Tononi, 2017). This prediction is based on IIT’s notion that lateral connectivity in the retinotopic areas should support topological features of space (such as inclusion, connection, and overlap). Thus, removing the possibility of causal interactions via “inactivation” is sufficient to disrupt these causal properties of the neuronal network underlying the phenomenology of visual space. Importantly, IIT predicts that the warped perception of space following such localized intervention would come about completely independent of other neural mechanisms, such as (top-down) attention. Combined with the novel, massively sampling neurophysiological techniques and sophisticated causal tools such as optogenetics (X. Han and Boyden, 2007; F. Zhang et al., 2007), neural dust (Soo et al., 2016), non-invasive ultrasound deep brain stimulation (UDBS; Folloni et al., 2019), ultrasound-based localized drug uncaging (microbubbles; McDaniel, Zhang and Vykhostdseva, 2011), and designer receptors exclusively activated by designer drugs (DREADDS; Roth, 2016), this
prediction becomes increasingly empirically testable. Future success rests on this type of strong theoretical guidance on how to link the formalized structure of phenomenology with precise whole brain-scale neurophysiology.

Lastly, when we can estimate the structure of information that is set up by neuronal activity together with the corresponding phenomenal structure, we can address the question of the exact relationship between the two. In this context, an exact isomorphism between neuronally construed information and phenomenology might prove too high a bar to be tested empirically. Rather, we propose a step-by-step one-directional test of correspondence (Tsuchiya, Taguchi and Saigo, 2016). This test distinguishes IIT’s position of assuming identity between the two abstracted structures. For example, we can ask if the structure of phenomenology (e.g., the helical chromatic structure of musical) is preserved in the structure of information derived from neural activity. In the other direction, when we manipulate the brain’s connectivity in early visual cortex (Song et al., 2017), will the spatial structure of consciousness change as predicted while maintaining the properties of the space? These are feasible tests that approach the overarching quest for isomorphism between phenomenal and neuronal information structures. Through multiple iterations of these tests, it is feasible to incrementally approach the ultimate, underlying isomorphism, which promises to bridge the gap between mind and brain.

In the current context of the division between attention and consciousness, for example, it seems feasible to manipulate attention (Marisa Carrasco, Ling and Read, 2004; Matthews et al., 2018; Prinzmetal, Nwachuku, Bodanski, Blumenfeld and Shimizu, 1997) to perturb phenomenology in an attempt to better characterize the phenomenological structure (e.g., does the helical structure of musical chroma morph or even collapse if attention is withdrawn?).

Concluding remarks

Since Eriksen’s seminal work on the impact of unconscious stimuli on heightening one’s state of vigilance, much has happened in the science of consciousness and attention. One of Eriksen’s major contributions to experimental psychology includes his extensive studies on the flanker task (also known as Eriksen’s flanker task (Eriksen and Eriksen, 1974)) as a tool to study attention. In this task, participants are asked to report the identity of a central letter (e.g., “H” or “S”) that is surrounded by task-irrelevant, congruent, or incongruent letters. The degree of interference from the task-irrelevant incongruent letters is often taken as evidence of inefficient attentional filtering (Summer et al., 2010; Ho and Cheung, 2011; (Silva et al., 2018; Wu et al., 2015), warranting further study.

Eriksen concluded from his own studies that visual stimuli that are presented too briefly to reach consciousness also fail to affect arousal or vigilance. And yet, Eriksen also found that multiple visual stimuli can evoke parallel, competitive or facilitative, influences on both neural and behavioral responses, suggesting that attention spreads beyond task-relevant items. In other words, Eriksen’s was among the first to indicate that neural processes outside the immediate focus of attention can influence neural responses to attended stimuli. The more recent research reviewed here goes even further by suggesting that attention can operate on visual stimuli that remain outside consciousness.

We are now at the edge of a new breakthrough regarding the neuronal dissection of consciousness and attention. Using methodological advances that were unavailable to Eriksen at the time, recent years witnessed a surging amount of convergent evidence that attention and consciousness are dissociable neuronal processes. At the same time, we are witnessing theoretical breakthroughs in our understanding of the formal, relational structure of consciousness. It is exciting to anticipate the insights that will be gained from symbiotically combining these theoretical and technological advances, and yet we realize how much work still needs to be done to truly understand how certain properties of our conscious experiences are supported by neuronal mechanisms.

Acknowledgements

The authors would like to thank Matthew Davidson for comments on an earlier version of the manuscript. AM is supported by a research grant from the US National Eye Institute (1R01EY027402-03). NT is supported by Australian Research Council (DP180104128 and DP180106396) and National Health and Medical Research Council (APP1183280).

References

Aquino, T. G., Minxha, J., Dunne, S., Ross, I. B., Mamelak, A. N., Rutishauser, U., & O’Doherty, J. P. (2020). Value-Related Neuronal Responses in the Human Amygdala during Observational Learning. The Journal of Neuroscience, 40(24), 4761-4772. https://doi.org/10.1523/JNEUROSCI.2897-19.2020

Arcizet, F., & Krauzlis, R. J. (2018). Covert spatial selection in primate basal ganglia. PLoS Biology, 16(10), e2005930. https://doi.org/10.1371/journal.pbio.2005930

Aru, J., Bachmann, T., Singer, W., & Melloni, L. (2012). Distilling the neural correlates of consciousness. Neuroscience and Biobehavioral Reviews, 36(2), 737-746. https://doi.org/10.1016/j.neubiorev.2011.12.003

Bachmann, T., & Hudetz, A. G. (2014). It is time to combine the two main traditions in the research on the neural correlates of consciousness: C = L x D. Frontiers in Psychology, 5, 940. https://doi.org/10.3389/fpsyg.2014.00940

Barrett, A. B., & Seth, A. K. (2011). Practical measures of integrated information for time-series data. PLoS Computational Biology, 7(1), e1001052. https://doi.org/10.1371/journal.pcbi.1001052

Bisiach, E., & Luzzatti, C. (1978). Unilateral neglect of representational space. Cortex, 14(1), 129-133.
Maier, A., Panagiotaropoulos, T. I., Tsuchiya, N., & Keliris, G. A. (2012). Introduction to research topic - binocular rivalry: A gateway to studying consciousness. Frontiers in Human Neuroscience (SEPTEMBER), 1-3. https://doi.org/10.3389/fnhum.2012.00263

Martens, U., Ansorge, U., & Kiefer, M. (2011). Controlling the unconscious: attentional task sets modulate subliminal semantic and visuomotor processes differentially. Psychological Science, 22(2), 282-291. https://doi.org/10.1177/0956797610397056

Mashour, G. A., Roelfsema, P., Changeux, J. P., & Dehaene, S. (2020). Conscious Processing and the Global Neuronal Workspace Hypothesis. Neuron, 165(5), 776-798. https://doi.org/10.1016/j.neuron.2020.01.026

Matthews, J., Schröder, P., Kaunitz, L., & Tsuchiya, N. (2018). Conscious access in the near absence of attention: critical extensions on the dual-task paradigm. Philosophical Transactions of the Royal Society, B: Biological Sciences, 373(1755), 20170352.

Maunsell, J. H. R. (2015). Neuronal Mechanisms of Visual Attention. Annual Review of Vision Science, 1, 373-391. https://doi.org/10.1146/annurev-vision-082114-035431

Mayner, W. G. P., Marshall, W., Albantakis, L., Findlay, G., Marchman, R., & Tononi, G. (2018). PyPhi: A toolbox for integrated information theory. PLoS Computational Biology, 14(7), e1006343. https://doi.org/10.1371/journal.pcbi.1006343

McDannold, N., Zhang, Y., & Vykhodtseva, N. (2011). Blood-brain barrier disruption and vascular damage induced by ultrasound bursts combined with microbubbles can be influenced by choice of anesthesia protocol. Ultrasound in Medicine & Biology, 37(8), 1259-1270. https://doi.org/10.1016/j.ultrasmedbio.2011.04.019

Merten, K., & Nieder, A. (2012). Active encoding of decisions about stimulus absence in prime prefrontal cortex neurons. Proceedings of the National Academy of Sciences, 109(16), 6289-6294.

Michel, M., Beck, D., Block, N., Blumenfeld, H., Brown, R., Carmel, D., … Yoshida, M. (2019). Opportunities and challenges for a maturing science of consciousness. Nature Human Behaviour, 3(2), 104-107. https://doi.org/10.1038/s41562-019-0531-8

Mitchell, J. F., Sundberg, K. A., & Reynolds, J. H. (2009). Spatial attention decorrelates intrinsic activity fluctuations in macaque area V4. Neuron, 65(6), 879-888. https://doi.org/10.1016/j.neuron.2009.09.013

Montemayor, C., & Haladjian, H. H. (2019). Recurrent processing theory versus general neuronal workspace theory: a comment on ‘The relationship between attention and consciousness: an expanded taxonomy and implications for ‘no-report’ paradigms’ by Pitts et al. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 374(1770), 20180517. https://doi.org/10.1098/rspb.2018.0517

Morgan, S. T., Hansen, J. C., & Hillyard, S. A. (1996). Selective attention to stimulus location modulates the steady-state visual evoked potential. Proceedings of the National Academy of Sciences of the United States of America, 93(10), 4770-4774. https://doi.org/10.1073/pnas.93.10.4770

Muller, M. M., Andersen, S., Trujillo, N. J., Valdes-Sosa, P., Malinowski, P., & Hillyard, S. A. (2006). Feature-selective attention enhances color signals in early visual areas of the human brain. Proceedings of the National Academy of Sciences of the United States of America, 103(38), 14250-14254. https://doi.org/10.1073/pnas.0606681038

Musk, E. (2019). An integrated brain-machine interface platform with thousands of channels. Journal of Medical Internet Research, 21(10), e16194.

Naccache, L., Blandin, E., & Dehaene, S. (2002). Unconscious masked priming depends on temporal attention. Psychological Science, 13(5), 416-424. https://doi.org/10.1111/1467-9280.00474

Nani, A., Manzullo, E., Mancuso, L., Liloia, D., Costa, T., & Cauda, F. (2019). The Neural Correlates of Consciousness and Attention: Two
Travis, S. L., Dux, P. E., & Mattingley, J. B. (2019). Neural correlates of conscious perception. *Nature Reviews. Neuroscience, 17*(7), 450-461. https://doi.org/10.1038/s41583-018-0217

Travis, S. L., Dux, P. E., & Mattingley, J. B. (2017). Re-examining the goal-directed enhancement and suppression of visual stimuli in the absence of conscious perception. *Attention, Perception, & Psychophysics, 81*(5), 1346-1364. https://doi.org/10.3758/s13414-018-1615-7

Tsuchiya, N., & Koch, C. (2014). On the relationship between consciousness and attention: MIT Press.

Tsuchiya, N., & Saigo, H. (2020). Applying Yoneda's lemma to consciousness research: categories of level and contents of consciousness.

van Boxtel, J. J., Tsuchiya, N., & Koch, C. (2010a). Consciousness and attention: on sufficiency and necessity. *Frontiers in Psychology, 1*, 217. https://doi.org/10.3389/fpsyg.2010.00217

van Boxtel, J. J., Tsuchiya, N., & Koch, C. (2010b). Opposing effects of attention and consciousness on afterimages. *Proceedings of the National Academy of Sciences of the United States of America, 107*(19), 8883-8888. https://doi.org/10.1073/pnas.0913292107

van Kempen, J., Gieselmann, M. A., Boyd, M., Steinmetz, N. A., Moore, T., Engel, T., & Thiele, A. (2020). Top-down coordination of local cortical state during selective attention. *bioRxiv*.

van Rullen, R. (2016). Perceptual cycles. *Trends in Cognitive Sciences, 20*(10), 723-735.

Varela, F. J. (1996). Neurophenomenology: A methodological remedy for the hard problem. *Journal of Consciousness Studies, 3*(4), 330-349.

van Boxtel J. (2017). Different Signal Enhancement Pathways of Attention and Consciousness Underlie Perception in Humans. The Journal of neuroscience : the official journal of the Society for Neuroscience, 37(24), 5912–5922. https://doi.org/10.1523/JNEUROSCI.1908-16.2017

Vialatte, F. B., Maurice, M., Daunwels, J., & Cichocki, A. (2010). Steady-state visually evoked potentials: focus on essential paradigms and future perspectives. *Progress in Neurobiology, 90*(4), 418-438. https://doi.org/10.1016/j.pneurobio.2009.11.005

Volz, L. J., & Gazzaniga, M. S. (2017). Interaction in isolation: 50 years of insights from split-brain research. *Brain, 140*(7), 2051-2060. https://doi.org/10.1093/brain/awx139

Watanabe, M., Cheng, K., Murayama, Y., Ueno, K., Asamizuya, T., Tanaka, K., & Logothetis, N. (2011). Attention but not awareness modulates the BOLD signal in the human V1 during binocular suppression. *Science, 334*(6057), 829-831. https://doi.org/10.1126/science.1203161

Webb, T. W., Igelstrom, K. M., Schurger, A., & Graziaro, M. S. (2016). Cortical networks involved in visual awareness independent of visual attention. *Proceedings of the National Academy of Sciences of the United States of America, 113*(48), 13923-13928. https://doi.org/10.1073/pnas.161505113

Weil, R. S., Kilner, J. M., Haynes, J. D., & Rees, G. (2007). Neural correlates of perceptual filling-in of an artificial scotoma in humans. *Proceedings of the National Academy of Sciences of the United States of America, 104*(12), 5211-5216. https://doi.org/10.1073/pnas.0609294104

Wilke, M., Mueller, K. M., & Leopold, D. A. (2009). Neural activity in the visual thalamus reflects perceptual suppression. *Proceedings of the National Academy of Sciences of the United States of America, 106*(23), 9465-9470. https://doi.org/10.1073/pnas.0900714106

Woodman, G. F., & Luck, S. J. (2003). Dissociations among attention, perception, and awareness during object-substitution masking. *Psychological Science, 14*(6), 605-611. https://doi.org/10.1046/j.0956-7976.2003.psci.1472.x

Wu, Q., Lo Voi, J. T., Lee, T. Y., Mackie, M.-A., Wu, Y., & Fan, J. (2015). Interocular suppression prevents interference in a flanker task. *Frontiers in Psychology, 6*, 1110.

Wyart, V., Dehaene, S., & Tallon-Baudry, C. (2011). Early dissociation between neural signatures of endogenous spatial attention and perceptual awareness during visual masking. *Frontiers in Human Neuroscience, 6*, 16. https://doi.org/10.3389/fnhum.2012.00016

Wyart, V., & Tallon-Baudry, C. (2008). Neural dissociation between visual awareness and spatial attention. *The Journal of Neuroscience, 28*(10), 2667-2679. https://doi.org/10.1523/JNEUROSCI1478-07.2008

Yuval-Greenberg, S., & Heeger, D. J. (2013). Continuous flash suppression modulates cortical activity in early visual cortex. *The Journal of neuroscience : the official journal of the Society for Neuroscience, 33*(23), 9635–9643. https://doi.org/10.1523/JNEUROSCI.4748-07.2008

Zahavi, D., & Gallagher, S. (2008). The phenomenological mind. *London: Routledge*, 244.

Zenon, A., & Krauzlis, R. J. (2012). Attention deficits without cortical neural deficits. *Nature, 499*(746), 434-437. https://doi.org/10.1038/nature12497

Zhang, F., Wang, L. P., Brauner, M., Liewald, J. F., Kay, K., Watzke, N., … Deisseroth, K. (2007). Multimodal fast optical interrogation of neural circuitry. *Nature, 446*(7136), 633-639. https://doi.org/10.1038/nature05744

Zhang, P., Jamison, K., Engel, S., He, B., & He, S. (2011). Binocular rivalry requires visual attention. *Neuron, 71*(2), 362-369. https://doi.org/10.1016/j.neuron.2011.05.035

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.