Characterising the Neural Correlates of Perceptual Experience

Benjamin Barnett

Wellcome Centre for Human Neuroimaging
Institute of Neurology
University College London

Dissertation submitted for the degree of

Doctor of Philosophy

September 2024

I, Benjamin Barnett, confirm that the work presented in my thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

Perceptual experiences are fundamental to the human condition, and yet how experiences arise in the brain remains unknown. This thesis examines various aspects of conscious awareness to better characterise how perceptual experiences are generated in the brain. In an initial study, I analyse magnetoencephalography (MEG) and functional magnetic resonance imaging (fMRI) data to test whether established properties of neural magnitude codes extend to the neural processes governing reports of awareness and vividness. Using multivariate decoding and representational similarity analysis, I report how a content-invariant code supports perceptual vividness judgements and how this code extends across visual and frontoparietal regions of the brain. The second experiment extends these findings to the concept of numerical absence. This chapter provides the first neuroscientific exploration of the number zero in the human brain and establishes zero's place on a neural number line that is independent of numerical format. This study relates to perceptual experience by providing support for an account of the relationship between basic sensory absences and more complex conceptual absences. In a third study, I analyse the fMRI data of patients with Alzheimer's disease to assess whether classic neural markers of awareness are altered in the disorder. I report that neural correlates of awareness are diminished in Alzheimer's disease and use this to argue for its characterisation as a disorder of consciousness, not just of memory. Finally, I provide the first validation of newly developed optically pumped MEG (OP-MEG) systems in a naturalistic social perception task. Across a series of analyses, I was unable to reproduce previously established neural markers of perspective-taking in a real-world task and discuss the reasons why this may have been the case. Overall, this thesis presents a number of novel insights into the neural basis of perceptual experience. It contributes empirical tests of candidate theories of consciousness, assessments of perceptual experience in disease, and, finally, ushers in a new era of ecological tests of awareness.

Impact Statement

Perceptual experiences compose the foundation of human existence. Our goals, decisions, ethics, and culture all rely on the fact we are conscious of the world around us, experiencing the sights, sounds, and smells that we encounter every day. However, the neural basis of perceptual experience remains conceptually hard to examine and is as of yet unknown. This thesis' impact lies in its multifaceted exploration of the neural basis of consciousness and its findings pertain to both academia and society at large, in both the laboratory, the clinic, and the classroom.

Across two chapters of this thesis, I have tested hypotheses regarded neural magnitudes and their relationship with the neural correlates of perceptual experience. Across different neuroimaging modalities, I have provided evidence to support the idea that a simple neural code may govern different kinds of awareness reports. Using theoretically motivated hypotheses is fundamental to scientific progress, particularly in the neuroscience of consciousness, which faces a troublesome stasis as researchers fail to move beyond simple confirmatory tests of their preferred theory. This problem has been exemplified by the introduction of adversarial collaborations in the field, representing a deliberate effort to overcome these issues. My thesis provides one example of strong, theoretically motivated science that can help progress the neuroscience of consciousness beyond its present plateau. Moreover, within this line of research, I have provided the first characterisation of the concept 'zero' in the human brain, a notoriously abstract concept. This work has the potential to impact our understanding of concept development in children, as well as providing insight into the evolutionary origins of uniquely human capacities.

In a separate study, I examined the contents of consciousness of patients suffering from Alzheimer's disease. Alzheimer's disease is a major cognitive disorder and afflicts over 50 million people globally – it is the leading cause of death in the UK. Despite its prevalence, however, no study has yet explored the possibility that patients with the disease may have an altered or degraded experience of the world. In the first study of its kind, my research reveals that classic neural markers of awareness are diminished in Alzheimer's patients, potentially highlighting the need to recharacterize the disease as a disorder of consciousness, and not simply a disorder of memory. This has the potential to impact millions of patients and caregivers alike, enabling a better understanding of what is and isn't experienced by patients, and helping to improve and develop more effective care strategies.

Finally, this thesis provides the first test of newly developed, wearable optically-pumped magnetoencephalography (OP-MEG) in a social and cognitive task. The potential for OP-MEG to revolutionise cognitive testing in real-world situations is hard to overstate. The

ability for participants to move freely whilst being scanned can enable testing of social phenomena that have so far been assessed only in contrived and artificial settings. Moreover, it is a useful clinical tool in populations who cannot remain still for extended periods of time, such as epileptic children and Parkinson's patients. My research provides a global first in attempting to validate these tools for use in social-cognitive neuroscience more broadly and provides a much-needed critical assessment of their promise.

Acknowledgements

It is hard to occupy the mind of the person who sat at this desk just three years ago. Many people have played a role in influencing, motivating, and supporting me since then, and each of them has played a unique part in helping me grow beyond who I was when I began this PhD.

Most important to my growth as a neuroscientist is my supervisor, Steve Fleming. Throughout my PhD, I have had the freedom to pursue ideas that, more than anything, have been fun to pursue, safe in the knowledge I can rely on you for expertise, guidance, and support. Most of all, I want to thank you for always congratulating – and never admonishing – me when I found a bug in my code. To my second supervisor and desk-mate, Nadine Dijkstra, thank you for answering every little question of mine, for building my independence and confidence as a scientist, and most importantly, for always caring foremost about my health and happiness. To my thesis committee, Peter Kok and Hugo Spiers, thank you for taking the time to listen to me ramble, even when I felt like I had nothing to say, and thank you for always providing insightful and productive suggestions – this thesis is all the better for our meetings. My acknowledgments would not be complete without thanking all those involved in the Ecological Brain Project, without which I would certainly not be here. In particular, thank you to Warda Sharif, for replying to all manner of anxious and confusing emails I may have sent over the years.

Completing a PhD in the Meta Lab has been a privilege, and I am endlessly appreciative of the supportive and easy-going colleagues I have the pleasure of collaborating with. In particular, thank you to Cormac, for being a kindred spirit in many ways – my PhD experience would have been very different without you. Thank you to Yongling for revealing to me the limits of my badminton prowess, to Marco for every word of encouragement, and to Astrid for being a close friend each time ASSC rolled around. Across the FIL, I have had the pleasure of working with a huge number of excellent people and scientists. To Rob Seymour, thank you for your generous supervision when collecting and analysing OP-MEG data – if anyone could have made me enjoy working with OPMs, it would have been you. To Nic Alexander, thank you for your unending patience when teaching me how to set up the scannercasts. It is no doubt a skill I will take to the grave. To Jon Huntley, for trusting me with your data and for generally being a supremely

calming presence. Thank you to the imaging support team, particularly Yasmin, Dimitra, and Dan. You took my least favourite part of the week and filled it with fun, gossip, and friendship. Thank you also to the burgeoning PhD support network: Gina, Arjun, Aaron, Sam, and many others. It's nice to know you're not alone – better late than never! Finally, thanks to Yan – my Ecobrain partner – for being interested in how things are going from the very start.

Completing a PhD can be lonely and insular. To survive it with a smile on your face requires a little help from a great many friends. Thank you to Alex and George, for being there from the very outset of our shared intrigue for consciousness science. Thank you to my housemates over the years, especially Tilly and James, for every lovely evening spent on our terrible sofa. To Astrid and everyone at AC-Aquatics, thank you for giving me goals that have nothing to do with science. To Ned and Krista, thank you for always listening and for never judging. Finally, to my school friends: Sal, George, Dom, Ed, Henry, Karum, Finn – I never feel more at home than when I'm with you.

Thank you to my parents, for giving me every opportunity to try anything in life, and for supporting me with all your might no matter what it was I tried to do. To my sister and best-friend Gemma, for helping me deal with our parents. To Dot, thank you for your ever-present support over the last two years. More than anyone, you made life at work richer and easier, and I will always be grateful for how you changed my experience of the FIL.

Finally, to Georgina, for the beautiful places we have been together, for the hours spent persevering as I try to explain phenomenal magnitude to you, and for the fiercest love imaginable.

Table of Contents

1. General Introduction

1.1 Consciousness

'Birdwatchers often keep a life list of all the species they have seen. Suppose you and I are birdwatchers, and we both hear a bird singing in the trees above our heads; I look up and say "I see it—do you?" You stare right where I am staring, and yet you say, truthfully, "No. I don't see it." I get to write this bird on my list; you do not, in spite of the fact that you may be morally certain that its image must have swum repeatedly across your foveae.

…Presumably it's not sufficient for reflected light from the object merely to enter your eyes, but what further effect must the reflected light have — what further notice must your brain take of it — for the object to pass from the ranks of the merely unconsciously responded to into … conscious experience?'

Daniel Dennett, Consciousness Explained, 1991, p. 336.

When you walk through the forest, you are confronted by a tapestry of sensations. Through widened eyes you examine the overbearing canopy, the dappled light on the forest floor, and the looming presence of the trees. You feel a dull ache in your heels as you reach the end of your walk. You hear sounds from every angle: the rustle of the leaves underfoot, and the creaking of tree trunks as they wave in the breeze. You are not, however, aware that your brain is sending signals to warm your body after hours spent out in the cold. Nor do you experience, as Daniel Dennett writes above, the bird your

partner expertly points out to you, despite the fact you set your eyes on its supposed location.

Why are we aware of the nexus of leaves and branches overhead, but remain unconscious of our own internal thermoregulatory processes? What difference is there between your partner's brain activity, which generates the breathless experience of spotting a new bird, and the neural signals inside your own brain, elicited by the foveated bird that you fail to consciously perceive? These questions form the basis of the problem of consciousness, often called the "hard problem" (Chalmers, 1995), which asks how mindless, inert electrical signals in the brain can generate the rich perceptual experiences that characterise our mental lives.

1.2 The Hard Problem and The Easy Problems

The hard problem describes the difficulty in bridging the apparent "explanatory gap" between the objective, biological workings of the human brain and the subjective, ineffable qualities of our perceptual experiences, or consciousness (Chalmers, 1995). At first pass, it seems difficult to explain our subjective phenomenology in physiological terms. Recasting the sights, sounds, and emotions that comprise our inner lives in terms of action potentials, mutual inhibition, and synaptic transmission seems implausible, as if it were some kind of category error. When viewed in this light, the hard problem demands a grand solution: how could a run-of-the-mill, neuroscientific theory bridge the gap between two worlds as conceptually distinct as the objective and the subjective? Such solutions have of course been proffered. Dualist accounts of consciousness propose that psychophysical laws – which supplement the physical laws already described by modern physics – provide the missing ingredient to bridge the explanatory gap (Chalmers, 1995). Alternatively, panpsychists propose that consciousness is a fundamental feature of the physical world, such that any form of physical matter is conscious to some extent (Goff, 2019; Strawson, 2016).

The trouble with these metaphysical accounts of consciousness is that they operate outside the realm of empirical science, and thus offer very little in our attempts to provide a scientific explanation of how conscious experiences are formed in the brain. A more promising avenue is to consider the "easy problems" of consciousness (Chalmers, 1995), originally defined as the tractable neuroscientific questions associated with perceptual experience that fall short of tackling the hard problem. The easy problems consist of phenomena such as visual detection, attentional focus, and reporting one's mental states. If you favour the idea that consciousness will eventually submit to an entirely physical explanation, the hard problem of consciousness dissolves, and the easy problems are all that remain (Dennett, 1991; Frankish, 2019). In fact, for a physicalist, the only hard problem is explaining why we think there's a problem of consciousness in the first place (Chalmers, 2018; Frankish, 2019).

Tackling so-called easy problems has already advanced our understanding of the neural mechanisms underpinning perceptual experiences. Attention has been shown to alter the appearance of simple stimuli (Carrasco, 2018; Carrasco et al., 2004), confidence in the visual periphery can be systematically inflated (Knotts et al., 2020; Odegaard et al., 2018), and the neural correlates of mental imagery are now known to overlap with those for perception (Dijkstra et al., 2020; Pearson, 2019; Siclari et al., 2017). Admittedly, these findings relate to perception and not necessarily to consciousness per se. However, with an increased focus on how perceptual systems and their computations relate to phenomenological reports, this approach promises to be the most practical and efficient means towards revealing the physical basis of consciousness (Dennett, 1991; Seth, 2021; Varela, 1996). All that would be left to explain is why these perceptual processes end up being described, by the experiencer, in experiential terms (Frankish, 2019) – a hard but empirically tractable problem. This approach bears parallels to 20th century biology. The fall of vitalism – the belief that life would not submit to physical explanation – came about once it was no longer doubted that physical mechanisms could support the varied functions of life (Chalmers, 1995). It was continued testing of empirically tractable problems in biology that eventually ended the scepticism surrounding the limits of physical

systems, and such an approach now provides the best empirical means towards elucidating the brain basis of conscious experience.

1.3 Theories of Consciousness

Although consciousness has long been subject to neuroscientific examination (LeDoux et al., 2020), the rise of neuroimaging techniques at the end of the 20th century led to a proliferation of novel theories explaining how perceptual experience might arise from brain activity (Seth & Bayne, 2022).

Theoretical work on consciousness can generally be divided into two broad frameworks. The first consists of First-Order theories of consciousness. First-Order theories propose that certain kinds of perceptual representations are sufficient for a phenomenal experience to arise. For example, Recurrent Processing Theory stipulates that top-down feedback signals in sensory cortices are the primary determinant of conscious experience (Lamme, 2006; Lamme & Roelfsema, 2000). Meanwhile, Global Workspace Theory (GWT) proposes that representation in sensory cortices alone is not sufficient for a perceptual experience. Instead, Global Workspace theorists suggest that sensory representations must be broadcast to the brain's many cognitive systems (e.g., attention, working memory, planning, action, etc.) before it can be experienced by an agent (Baars, 1993; Dehaene et al., 1998; Dehaene & Changeux, 2011). In GWT, making information available to these domains is associated with late activity in frontoparietal brain regions – the so-called Global Workspace (Dehaene & Naccache, 2001; Mashour et al., 2020).

Empirical support for GWT emerges mostly from contrastive analyses that compare experimental conditions where participants are aware of a stimulus with conditions where they are not. Such studies generally find that conscious perception is associated with activity in frontoparietal regions (Sadaghiani et al., 2009; Sanchez et al., 2020; Van Vugt et al., 2018) or at late time points with respect to stimulus presentation (Berkovitch et al., 2018; Charles et al., 2014; Noel et al., 2019). Such qualities are even observed when participants are not required to report whether they've seen a stimulus or not

(Hatamimajoumerd et al., 2022; Sergent et al., 2021). Nevertheless, concerns remain with respect to GWT, including the conflation of subjective experience with information processing (Lau, 2022), the difficulty of dissociating consciousness and cognition in such research (Block, 2019), and observations of unconscious content being represented in frontoparietal regions (Mei et al., 2022).

A second broad class of theories of consciousness are Higher-Order theories (HOT). The core claim behind HOTs is that a first-order sensory representation is not on its own sufficient for a conscious experience. Instead, the representation must be subject to re-representation or monitoring from a meta-level cognitive system (Brown et al., 2019; Lau & Rosenthal, 2011). This criterion follows from an intuitive understanding of what it means to be conscious: if a system is not aware of itself as having a particular percept, then that percept cannot be consciously experienced (Brown et al., 2019). Like GWT, HOTs rely on empirical data showing higher-order brain regions to be associated with awareness. In contrast to GWT, however, they place importance on cases where conscious and unconscious conditions exhibit the same level of task performance (Lau & Passingham, 2006; Persaud et al., 2011; Rounis et al., 2010). Such data cannot be explained by GWT since GWT would predict that the frontal activity driving awareness should also improve task performance. Additional support comes from the close theoretical correspondence between metacognition and perceptual experience (Fleming, 2020; Lau, 2022) and related empirical work showing that metacognitive confidence judgements are largely computed from the detectability of different stimuli (Cortese et al., 2016; Peters et al., 2017). There are several different HOTs, diverging from one another with respect to the character of the higher-order representations in play. ‘Rich’ HOTs propose that perceptual experience is supported by complete or near-complete re-representations of perceptual content, wherein the first-order content is recapitulated in detail in higher-order regions (Brown, 2015; Cleeremans et al., 2020). In contrast, ‘sparse’ HOTs suggest that higher-order representations only encode the precision or reliability of first-order sensory representations (Fleming, 2020; Lau, 2019).

1.4 Clinical Approaches to Consciousness

Empirical insights into the neural basis of consciousness have rightly influenced clinical settings where the presence of consciousness is doubted. Disorders of consciousness such as coma, minimally conscious states (MCS) and vegetative states (VS) are characterised by varying degrees of disruption to arousal and awareness of one's environment (Giacino et al., 2014) and correctly determining which level of awareness a patient has can be difficult given the impaired behavioural and communicative function associated with these conditions (Giacino et al., 2018). This is exemplified by the high number of erroneous diagnoses made in such situations (approximately 40%; Schnakers et al., 2009). Critically, the diagnosis given to a patient suffering from a disorder of consciousness informs potential treatments and even end-of-life decisions (Arzi et al., 2020; Thibaut et al., 2019) so it is ethically imperative that research surrounding the neural correlates of consciousness is used to inform and aid clinical decisions where doubt persists.

The principal way in which cognitive research has aided clinical neuroscience is by putting forward candidate biomarkers of conscious awareness to assist in diagnoses when patient communication is limited. As discussed earlier, GWT identifies consciousness with late-stage activity in frontoparietal networks of the brain (Dehaene and Changeux, 2011; Mashour et al., 2020). In line with this, the presence of frontal activity in certain tasks can reliably distinguish healthy controls from unresponsive patients (Bekinschtein et al., 2009). For instance, in tasks where individuals passively listen to a sequence of tones, local (within trial) irregularities elicit sensory activity in the auditory cortex of controls and patients alike, but detection of global (across trial) irregularities produced late electrophysiological responses around 400 ms after stimulus onset in healthy controls only (Bekinschtein et al., 2009; King et al., 2013). Late electrophysiological responses to global irregularities are thus one potential marker for consciousness that can be used in a clinical setting. In line with this, event-related potentials such as the P300, which occurs around 300 ms after stimulus onset, have been shown to effectively classify awareness

– as well as the capacity to regain awareness – during recovery from disorders of consciousness (Zhang et al., 2017).

Aside from GWT-inspired electrophysiological markers, measures of informational complexity such as the perturbational complexity index (PCI) have also shown great promise in detecting consciousness in non-communicative patients (Casali et al., 2013; Casarotto et al., 2016). Inspired by complexity-based theories of consciousness (Tononi, 2008; Tononi & Edelman, 1998), the PCI measures the mathematical complexity of the brain's response to perturbation from transcranial magnetic stimulation. The PCI value can readily be interpreted as providing evidence for conscious awareness if it is above a particular value (Casali et al., 2013), meaning it can easily be incorporated into clinical decision-making when behavioural assays of consciousness are insufficient. The PCI results in over 90% sensitivity when detecting awareness in MCS and VS patients (Casarotto et al., 2016) and provides an extremely rapid test of the capacity for consciousness, long before behavioural responsiveness emerges after injury.

Determining to what degree a patient is aware of themselves and their environment is vital in clinical settings, and as the above approaches make evident, the neuroscience of consciousness continues to contribute to this effort. However, the *degree* of consciousness is only one dimension of our perceptual experience (Bayne et al., 2016). An alternative aspect, which has received considerably less attention with respect to disorders of consciousness, is that of the *content* of consciousness. If PCI and GWT can determine which patients have the capacity for conscious experience, how can we tell what that experience is like for the patient? What exactly *are they aware of?* In patients with disorders of consciousness, it would be helpful to know whether interventions or therapeutic care make a difference to them, or whether they can experience the faces and voices of the clinicians and family members that they interact with. Such concerns also exist for neurological conditions that are not typically thought of as disorders of consciousness and, in this thesis, I will present analyses that raise the question of whether dementia, typically understood as a disorder of memory, may also be described as a disorder of consciousness.

1.5 Ecological Approaches to Consciousness

Although the cognitive neuroscience of consciousness has impacted the clinical domain, its relevance to common, everyday perceptual experiences has been called into question (Mudrik et al., 2024). The experimental methods that pervade consciousness science rely on strictly controlled stimuli whilst employing artificial apparatus and scenarios that rarely occur outside the laboratory. A technique called binocular rivalry, where different images are simultaneously presented to each eye (Breese, 1909; Leopold & Logothetis, 1996), provides an illustrative example. To the participant, this procedure results in a percept of one of the two images which, after a delay of a few seconds (Kim et al., 2020), switches to the alternative image. Rather than experiencing a mixture of the two images, this dynamic switching between the two competing stimuli continues over time. This procedure is useful for examining perceptual experience because it fixes the visual input while enabling the subjective experience to change. However, experiences of rivalry rarely – if ever – occur in our day to day lives (Arnold, 2011), and if they do, they are seldom noticed (O'Shea, 2011). Since the perceptual states evoked by binocular rivalry (as well as those evoked by other popular techniques such as masking (Dehaene et al., 1998) and continuous flash suppression (Tsuchiya & Koch, 2005)) can only be induced under highly artificial settings, it's unclear how much they can tell us about perceptual experiences in the real world (Mudrik et al., 2024). Indeed, previously established phenomena from other cognitive domains are either diminished or abolished completely in naturalistic settings (primacy-recency in memory: Lee & Chen, 2022; Weber's law in motor acts: Ozana & Ganel, 2019).

Attempts to introduce ecological validity into tests of perceptual experience have already extended or nuanced findings from lab-based studies. When participants were asked to explore naturalistic videos in a virtual environment, they were unable to detect drastic changes to their visual scene (Cohen et al., 2020). Specifically, a significant minority of participants failed to detect a gradual change to grayscale in the periphery, even when only 10º of viewing angle remained coloured (Cohen et al., 2020). The use of virtual reality has thus revealed a gross inability to detect substantial changes to our visual scene,

outstripping the limits of change blindness previously found in computer studies where only single items were removed or modified (Levin & Simons, 1997). Other ecological procedures such as “real-life” continuous flash suppression (Korisky et al., 2019) have nuanced our understanding of the gating mechanisms that regulate entry into conscious awareness. Use of this novel technique has indicated that real life objects are easier to detect than 2D or physically-impossible versions of the same object (Korisky & Mudrik, 2021; Suzuki et al., 2019), implying a role for affordances and agential interaction in the formation of perceptual experience.

Experiments with high ecological validity may be especially important when testing theories of consciousness that centre social computations, such as Attention Schema Theory (AST; Graziano, 2013). AST unites both our subjective experience of the world and our ability to model and predict the minds of others under one core principle: the modelling of attention (Graziano, 2019). AST suggests that consciousness of the self and the environment stems from a schematic and detail-poor model of our own attentional mechanisms (i.e., an attention schema), which facilitates the control of attention (Graziano, 2013; Webb & Graziano, 2015). In many ways, AST is a variant of a HOT: it defines subjective awareness as the modelling of low-level attentional mechanisms by higher-order processes (Graziano, 2013; Webb and Graziano, 2015).[1] Thus, the attention schema is an evolved, functional mechanism that allows simple and efficient control of our own attention, and in doing so has resulted in subjective and phenomenal experience. According to AST, however, attention schemata are equally fundamental to predicting the behaviour of other people (Graziano, 2019). If we can infer the attentional state of other individuals, so AST claims, we can make accurate predictions regarding their behaviour. Although not deemed identical, AST therefore suggests that the attributions of

[1] AST goes further than other HOTs, however, because not only does it describe the conditions necessary for conscious experience, but it also explains why higher order modelling should make experiences feel the way they do: ineffable and ethereal (Dennett, 1988). It is because the attention schema is abstracted away from physical properties like action potentials and lateral inhibition that it depicts something physically incoherent that resists description in physical terms (Webb and Graziano, 2015; Graziano, 2019). This provides a built-in solution to the question of why humans perceive there to be a hard problem (Chalmers, 2018; Frankish, 2019) and is one of the theory’s great strengths.

consciousness to self and other are grounded in the same core mechanism of modelling attentional processes (Graziano, 2019; Kelly et al., 2014).

There is empirical evidence in favour of a shared foundation uniting self-awareness and social cognition. Notably, transcranial magnetic stimulation to subject-specific regions of the right temporoparietal junction (rTPJ), which were activated in a task involving attribution of awareness to cartoon faces, also influenced performance in a detection task in the same participants (Kelly et al., 2014). Additionally, examples of 'altercentric perception' – where the presence of a partner's attentional focus spontaneously impacts the perceptual experience of the subject – give credence to the idea that neural mechanisms devoted to perceptual experience and social cognition may overlap or interact with one another (Kampis & Southgate, 2020; Seow & Fleming, 2019). For example, in a task involving the identification of rotated letters, performance was facilitated if the letter was upright from the perspective of a task-irrelevant agent (Ward et al., 2019) and, in a separate experiment, the preference of subjects' face-sensitive N170 component to inverted (over upright) faces tracked a confederate's perspective, rather than the subject's (Böckler & Zwickel, 2013). These results indicate that not only do subjects spontaneously take the perspective of others, but their inference as to the content of the other's perception becomes the input to their own perceptual system as well.

The findings presented above evince an interaction between perceptual and social processes in the brain and reinforce calls for increased ecological validity in the neuroscience of consciousness. As a final, illustrative example, consider a recent study that reported the face-selective N170 component was abolished for real (vs. 2D) faces (Sagehorn et al., 2023). With this result in mind, previous descriptions of altercentric interference to the N170 in studies with 2D faces (Böckler and Zwickel, 2013) lose some of their explanatory power because it is not clear whether such results will generalise outside of artificial settings that do not frequently occur in naturalistic social situations. Following this logic, ecological approaches towards testing social *and* non-social aspects of perceptual experience are in need of development. Only then can we be confident that

our neural theories of perceptual experience extend to the kinds of experiences we most care to explain.

1.6 Neural Magnitude Codes

To construct accurate theories of perceptual experience, it is crucial to consider the neural architectures that govern cognitive processing more broadly, such that theories of consciousness do not operate outside the bounds of how we understand the brain to work. Predictive processing, for example, is a general theory of neural computation (Friston, 2009; Rao & Ballard, 1999), and not specifically a theory regarding conscious experience. However, embedding predictive processing mechanisms within theories of perceptual experience (e.g., Fleming, 2020; Whyte & Smith, 2021) ensures a grounded approach to studying the neural basis of consciousness, since it constrains ideas about how experience is generated to computational processes already hypothesised to occur throughout the brain (Hohwy & Seth, 2020). The neuroscience of consciousness can therefore benefit from 'theory-neutral' frameworks that can help operationalise abstract concepts into established neural or computational architectures.

One organising principle that motivated work contained within this thesis is that of magnitude coding (Summerfield et al., 2020; Tsouli et al., 2021; Walsh, 2003). Magnitudes are values along a one-dimensional manifold, or line, that describe the relative position a stimulus occupies along that line. Numerosity is a characteristic magnitude domain, with numbers each having their place on a number line from small to large magnitudes (Dehaene et al., 1993; Dehaene et al., 1998; Nieder, 2016; Piazza et al., 2004). However, many other cognitive domains also rely on magnitude estimation, such as size (Harvey et al., 2015), reward (McNamee et al., 2013), brightness and loudness (Stevens & Marks, 1965), decision-making (Shenhav & Greene, 2010), and confidence judgements (Mazor et al., 2022). As with predictive processing, magnitude coding does not relate to consciousness directly, instead focusing on the neural architectures responsible for encoding magnitudes in the brain. However, predictions from theories of consciousness, particularly sparse higher-order theories (e.g., Fleming,

2020), pertain to the neural encoding of certain magnitudes and therefore naturally interface with magnitude coding systems.

A core feature of magnitude coding architectures is the tuning of neural responses to specific magnitudes (Tsouli et al., 2021). In the same way that visual neurons are tuned to different orientations (Hubel & Wiesel, 1968) and motion directions (Dubner & Zeki, 1971), neurons throughout the brain are selective for different numerical magnitudes (Nieder & Dehaene, 2009). Numerosity-selective neurons fire maximally when their preferred numerosity is presented and the firing rate decreases as the presented number shifts away from that preferred numerosity. To model this, their tuning curves are typically described using Gaussian functions on a logarithmic scale (**Figure 1.1**; Dehaene et al., 1998; Tsouli et al., 2021). This coding scheme is emphasised in functional magnetic resonance imaging (fMRI) adaptation studies, where repeated presentations of a particular numerosity (adaptor) desensitise the fMRI response to that numerosity (Piazza et al., 2004). As the distance between the test numerosity and adaptor grows, this adaptation systematically decreases (Piazza et al., 2004). This is illustrative of a distance effect, where numerosities closer together in numerical space share overlapping tuning curves and are thus represented as closer together in the brain (Dehaene, 1998). fMRI adaptation, direct electrophysiological recordings, and population receptive field modelling have all converged on this architecture, revealing numerical magnitude codes across high-level visual (Harvey & Dumoulin, 2017; Paul et al., 2022), medial temporal (Kutter et al., 2018, 2023), parietal (Piazza et al., 2004; Harvey et al., 2013) and pre-frontal regions of the brain (Nieder & Miller, 2003, 2004).

Tuned neurons are not specific to numerical magnitudes or even vision itself. Single neurons have been shown to encode specific stimulus durations (Duysens et al., 1996), line length (Tudusciuc & Nieder, 2007), object size (Harvey et al., 2015), auditory event duration (He et al., 1997), and haptic numerosity (Hofstetter et al., 2021). The fact that this coding architecture is shared across domains and modalities has inspired suggestions that there may be domain-general magnitude coding systems in the brain, which efficiently encode quantities irrespective of what kind of quantity is being processed

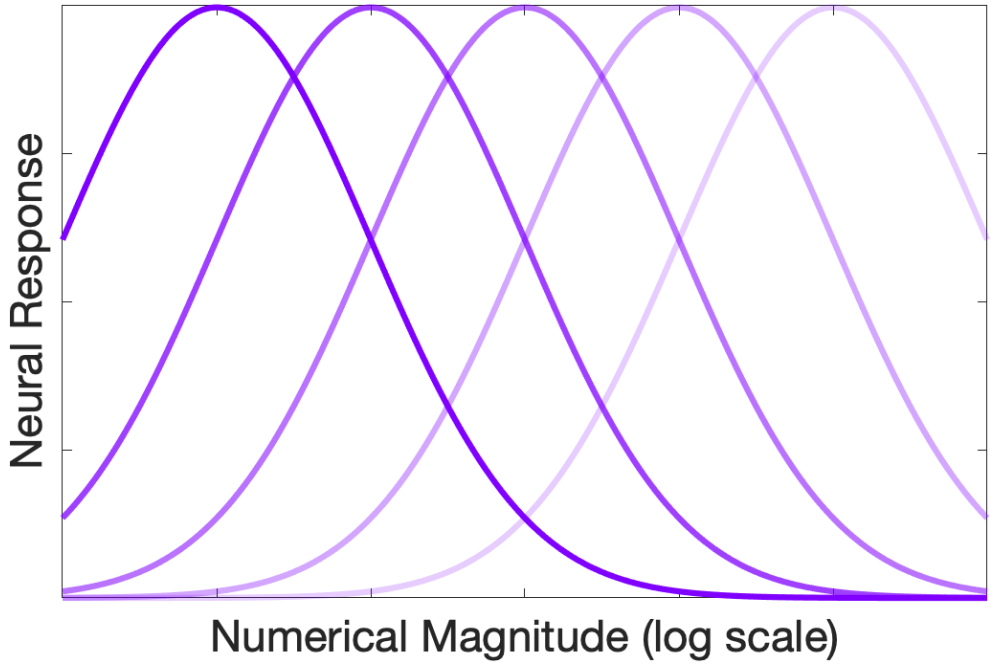


Figure 1.1. Tuning curves for numerical magnitudes. Different neurons exhibit selectivity for different magnitudes. However, tuning curves for nearby numerosities overlap, meaning tuned neurons will remain responsive to the presentation of neighbouring numerosities. This leads to the so-called distance effect. Gaussian curves are used to model neurons' tuning functions when numerical magnitude is on a logarithmic scale.

(Walsh, 2003; Summerfield et al., 2020). Evidence for a domain general magnitude coding system is typically taken from behavioural and neuroimaging studies showing interference or interaction between different magnitude domains (Walsh, 2003). The SNARC (spatial-numerical association of response codes) effect, for example, describes behavioural cases where higher numerical magnitudes are associated with the right side of visual space (Dehaene et al., 1993), a robust interaction between number and space. Neuroimaging studies in support of domain generality have shown overlapping regions of activation across different numerical formats such as dot arrays and numerals (Eger et al., 2003; Piazza et al., 2007) or representational similarity across cognitive domains (Luyckx et al., 2018). However, the greater spatial resolution afforded by single cell recordings and high field fMRI rarely show evidence for format- or domain-general magnitude encoding, instead indicating anatomically overlapping but distinct quantity

representations (Eiselt & Nieder, 2016; Harvey et al., 2015; Hofstetter et al., 2021; Kutter et al., 2018).

Sparse HOTs of consciousness, such as the Higher Order State Space (HOSS; Fleming, 2020) and Perceptual Reality Monitoring (PRM; Lau, 2019), make predictions regarding the relevance of unidimensional magnitudes in the computational architecture supporting perceptual experience. Specifically, HOSS describes a higher-order monitoring system tracking the reliability or precision of perceptual states, which can be represented on a one-dimensional axis (Fleming, 2020). PRM, on the other hand, tracks the reliability of these perceptual signals in order to classify whether or not they correspond to external events (Lau, 2019). Given the prevalence of magnitude coding systems in the brain, it is natural to use features associated with these systems, such as distance effects and format-invariance, to test hypotheses related to the relevant aspects of sparse higher-order theories of consciousness. In this way, magnitude coding provides a helpful framework within which to examine the neural correlates of perceptual experience.

1.7 Measuring Neural Responses During Perceptual Experiences

To examine the brain basis of perceptual experience, we must use tools that allow us access, either directly or indirectly, to the brain's activity. Non-invasive neuroimaging techniques have granted researchers this capability and it is no coincidence that the emergence of tools such as positive emission tomography and fMRI in the late 20th century co-occurred with the rapid rise of neural theories of consciousness (Crick and Koch, 1990).

fMRI is a fundamental technique in the field of cognitive neuroscience and a method used in Chapters 2 and 4 of this thesis. fMRI measures the blood-oxygen-level-dependent (BOLD) signal, an indirect measure of neural activity based on the differential level of oxygen supplied to the blood surrounding different brain regions. The logic goes that, if a particular area of the brain is more active, it will require a greater supply of oxygenated blood. The different magnetic properties of oxygenated and deoxygenated blood are

detectable using MRI and are therefore able to identify active brain regions during different perceptual experiences. Since the BOLD signal tracks the oxygenation of blood, it is sluggish in comparison to the neural activity it represents. As such, fMRI analyses are often limited by their lack of temporal sensitivity and are better suited to questions requiring high spatial resolution, since fMRI can typically resolve brain activity to voxels sized only a few millimetres cubed.

For questions that require greater temporal resolution, magnetoencephalography (MEG) is often used. Using sensors placed around the scalp, MEG directly measures the magnetic fields caused by the electrical activity of neurons. This procedure enables brain activity to be recorded at millisecond precision, enabling delineation of the precise onset of sensory processes. However, MEG suffers from a lack of spatial sensitivity, due to both the relatively large distance from the sensors to the scalp and the ambiguity surrounding which neural sources contribute to sensor-level data. As such, MEG can often only make broad statements regarding the localisation of neural activity. In Chapters 2 and 3 of this thesis, I use MEG to answer different questions regarding the magnitude-related properties of perceptual experience.

Together, fMRI and MEG have been central to the field of cognitive neuroscience. However, they both suffer from the same problem: ecological invalidity. In both modes of neuroimaging, participants are required to stay as still as possible: often movement of only a few millimetres introduces enough noise to render data unusable. The experience of lying or sitting down with restricted movement and interacting with a screen is not a valid model of real life, and, as I argued in Section 1.5, this limits the extent to which we can trust results from fMRI and MEG studies to generalise to real world experiences. One promising method to overcome this is optically-pumped MEG (OP-MEG; Brookes et al., 2022). OP-MEG records the same magnetic fields as conventional MEG but does so without requiring superconducting coils and the associated helium cooling system (Tierney et al., 2019). The fact that OPM sensors need not rely on cumbersome cooling systems means they need not be fixed in place and can instead be attached to the scalps of participants (Boto et al., 2018). This brings the sensors to within a few millimetres of

the scalp, improving signal to noise ratio compared to conventional MEG, where the sensors are at best a few centimetres away (Boto et al., 2018). Perhaps more importantly, attaching sensors directly to participants' heads means they are free to move whilst being scanned, since the sensors will move in tandem (Seymour et al., 2021). As illustrated in Chapter 5, this allows for the neural processes underlying real-world, ecological behaviours to be studied in a way that MEG and fMRI experiments do not permit.

Finally, once participants have been scanned, there are still several different ways to analyse their neural data. More traditional analyses are univariate in nature, typically contrasting the average response of fMRI voxels or MEG sensors across different conditions. This approach allows for conclusions to be drawn regarding the brain regions that are more active in condition A vs. B, in the case of fMRI, or which time window or frequency band differentiates experimental conditions in MEG. Univariate analyses (Chapters 3, 4, and 5) are relatively simple and easy to interpret but because they address individual voxels or sensors in isolation they come at a cost of sensitivity (Haynes, 2015). Multivariate methods (Chapters 2, 3, and 4) were developed to overcome this exact issue (Haxby et al., 2001; Haynes & Rees, 2006). Multivariate methods, including decoding (Haxby et al., 2001) and representational similarity analysis (RSA; Kriegeskorte et al., 2008), consider the covariance of activity occurring across voxels or sensors, rather than simply averaging over them. This increases power to detect differences in brain activity that may be distributed across different brain areas and allows us to reveal how particular mental contents are encoded in the brain (Haynes and Rees, 2006; Haynes, 2015).

1.8 Thesis Outline

This thesis describes various experiments pertaining to perceptual experience. In Chapter 2, I re-analysed data from MEG (Andersen et al., 2016) and fMRI (Dijkstra et al., 2021) studies to test whether the neural code underlying reports of perceptual vividness are specific to the content of perception or whether they are content-invariant. This experiment was motivated by results mentioned in Section 1.6, which suggest magnitude codes in different cognitive domains, such as number, may be independent of the

presented perceptual format. It was also driven by sparse Higher-Order models of awareness, which predict the existence of content-invariant codes for phenomenal magnitude that monitor first-order neural signals. Using RSA and decoding analyses, I provide evidence for a content-invariant neural code for perceptual vividness, which extends throughout the visual, parietal, and frontal cortex.

In Chapter 3, I characterise the neural representation of numerical absence, i.e., zero, as a means towards examining the explicit representation of absence predicted by certain HOTs. The number zero is a uniquely abstract number, and its cognitive appreciation shares several features with perceiving sensory absences. However, the representation of zero had yet to be studied in the human brain. By combining MEG, multivariate decoding, and source reconstruction, I reveal how zero is represented at the beginning of a neural number line in the posterior association cortex, and that representations of zero are shared across symbolic “0” and non-symbolic empty sets. This provides the first delineation of zero in the human brain and is a preliminary test of a shared representation between numerical absence and the perceptual experience of sensory absence.

In Chapter 4, I explored whether patients suffering from dementia may exhibit differences in the content of their perceptual experience compared to healthy controls. By analysing the fMRI responses of patients and controls in a classic visual masking paradigm, I was able to characterise the neural responses associated with visual awareness. Decoding analyses revealed diminished neural correlates of consciousness in dementia patients across visual and frontoparietal regions. Taken together with findings associating frontoparietal activity with conscious awareness, I present evidence to suggest patients suffering from dementia may exhibit differences or degradations in perceptual experiences compared to healthy populations.

Finally, in Chapter 5, I report a pilot study that aimed to develop a naturalistic perspective-taking paradigm for use in OPMs. Throughout the course of this project, I transformed an established, computerised perspective-taking task into one involving real people in a naturalistic setting. Using time-frequency analyses and source reconstruction, I analysed

OP-MEG data from three subjects in an attempt to replicate findings from conventional MEG using a naturalistic task. I failed to reproduce previous findings pertaining to perspective taking, particularly in relation to the naturalistic condition. I interpret these null findings as potentially indicating a lack of power with a small sample, but also discuss whether OP-MEG systems may currently be insensitive to high-level cognitive effects.

Overall, this thesis takes a varied approach to characterising the neural correlates of perceptual experience. It comprises theoretically motivated tests of awareness-related architectures, assessments of awareness in disease, and the evaluation of one promising methodology needed to address the lack of ecological validity in the neuroscience of consciousness.

2. Identifying content-invariant neural signatures of perceptual vividness

2.1 Introduction

Some experiences are more vivid than others. For example, seeing a bird on a clear day will be more vivid than seeing one on a foggy evening. Similarly, a car alarm outside your office can be very vivid until your attention is consumed by a task at work. The neural correlates of experience are therefore likely to involve some representation of the magnitude of perceptual vividness. While the neural basis of perceptual vividness is yet to be systematically characterised, neural codes for magnitude quantities in other cognitive domains, such as reward and numerosity, are better understood. Many neural magnitude codes exhibit a content-invariant component, where the magnitude property is represented independently of its sensory features (Chib et al., 2009; McNamee et al., 2013; Piazza et al., 2007). For instance, the number “9” is represented as larger than the number “5”, regardless of whether we are comparing 9 vs. 5 apples, oranges, or saxophones. Here, I ask whether the magnitude properties of perceptual vividness are also invariant to stimulus content: i.e., is the difference in vividness between seeing a bird on a clear day compared to a foggy evening represented in a similar manner as the difference in vividness between hearing a car alarm when we are attending to it, compared to when are distracted? I investigate this question by testing the extent to which neural signatures associated with reports of perceptual vividness are independent of perceptual content.

Content-invariance is a well-established feature of several neural magnitude codes. In the orbitofrontal cortex, for example, common representations of reward magnitude are shared across vastly different reward identities (Chib et al., 2009; Howard et al., 2015; Klein-Flügge et al., 2013; McNamee et al., 2013; Padoa-Schioppa & Assad, 2006). Furthermore, presentation of the same numerosity elicits suppression effects across symbolic (Arabic numerals) and non-symbolic (dots) stimuli in the intraparietal lobe (Piazza et al., 2007), and multivariate cross-classification has revealed common representations of numerosity across symbolic and non-symbolic formats (Teichmann et al., 2018). There is also evidence that numerical and reward magnitudes (amongst others) are encoded in a domain-general manner where, for example, higher numbers are represented similarly to highly rewarding stimuli and lower numbers are represented similarly to stimuli with low reward values, indicating a shared neural system underpinning representations of magnitude in both domains (Luyckx et al., 2019; Summerfield et al., 2020; Walsh, 2003).

Given the evidence for content-invariant neural magnitude codes in other domains, it is intriguing to ask whether invariance to perceptual content is also a feature of the neural activity covarying with the magnitude of perceptual vividness. If perceptual vividness is only encoded in a content-specific manner, our experience of a stimulus such as a red circle may become vivid through the increased firing of neural populations representing this feature (Itti & Baldi, 2009; **Figure 2.1**, left). However, if neural activity covarying with perceptual vividness also contains a content-invariant component, we should be able to find neural signatures of vividness that are independent of those covarying with sensory features. The drivers of such content-invariant signals may include changes in attention, emotion, and other cognitive factors that surpass stimulus-specific salience, but nevertheless contribute to the vividness of experience (Morales, 2018, 2021) – an idea consistent with philosophical positions that distinguish the content of percepts from their ‘force’ and ‘vivacity’ (Hume, 2000; Teng, 2022). As such, rather than being solely bound to content-specific representations, perceptual vividness might also covary with neural activity in a domain-general fashion, independently of stimulus content (**Figure 2.1**, right) (Levinson et al., 2021; Podvalny et al., 2019; Samaha et al., 2017; Sanchez et al., 2020).

Content-specific and content-invariant coding schemes should not be viewed as mutually exclusive. For instance, it is well-known that stimulus-driven aspects of perceptual salience such as stimulus contrast are reflected in modality-specific neural activity (Albrecht & Hamilton, 1982; Bartlett & Doty, 1974), and such properties in turn influence the subjective experience of vividness. Other studies have shown that activity in content-specific brain areas is associated with changes in perceptual awareness, even when holding the stimulus constant (Boehler et al., 2008; Fisch et al., 2009; Moutoussis & Zeki, 2002). Moreover, representations of magnitude in other domains such as reward or number often exhibit both content-specific and content-invariant components (Howard et al., 2015; McNamee et al., 2013; Teichmann et al., 2018). The present study aimed to investigate whether, beyond these content-specific codes, there is also evidence for content-invariant neural signatures of perceptual vividness

To test whether the neural code for perceptual vividness exhibits a content-invariant component, I reanalysed both magnetoencephalography (MEG; Andersen et al., 2016) and functional magnetic resonance imaging (fMRI; Dijkstra et al., 2021) data to investigate how perceptual vividness is represented in the human brain. The difficulties in isolating pure correlates of vividness and awareness from co-varying neural signals (e.g. those related to arousal or performance) are well known, and I did not attempt to tackle these issues here (Aru et al., 2012; Lau, 2022). Instead, I sought to determine the representational structure of awareness and visibility reports about different stimulus contents, to ask whether neural signatures covarying with vividness did so in a content-specific or content-invariant manner. To anticipate the results, I found evidence that neural representations of perceptual vividness generalize over stimulus content, exhibit a graded structure, and can be identified across visual, parietal, and frontal brain regions, consistent with signatures of magnitude codes in other cognitive domains.

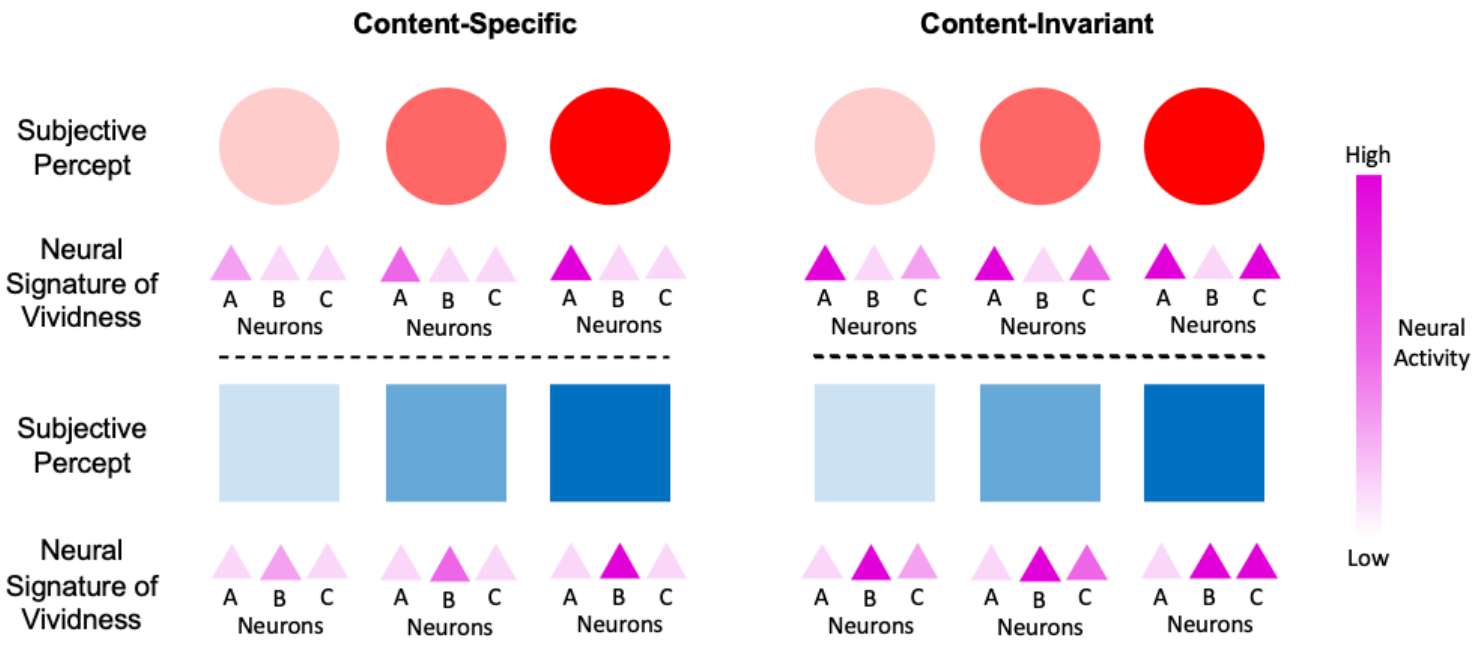


Figure 2.1. Hypothesised neural signatures of perceptual vividness. Left: Content-specific neural signatures associated with perceptual vividness. The subjective vividness of a red circle is associated with the strength of red circle-representing neurons (neuron A) while the vividness of a blue square is associated with the strength of blue square-representing neurons (neuron B). For example, as red circle-representing neurons increase their activity (top-left), the subjective percept of a red circle becomes more vivid. The neural signatures correlating with the vividness of red circles and blue squares are therefore different. Right: Content-invariant neural signatures associated with perceptual vividness. The subjective vividness of both red circles and blue squares is associated with a common neural signature (i.e., the activity of neuron C), which tracks vividness over and above any stimulus-specific neural activity (i.e., neurons A and B). Attention, emotion, and other cognitive factors may drive a content-invariant neural signal of vividness. The hypothetical coding schemes represented here are not mutually exclusive, and it is possible that a combination of both schemes underpin the vividness of perceptual experience.

2.2 Materials and Methods

2.2.1 MEG Experiment

To explore the structure and dynamics of abstract representations of awareness ratings, I re-analysed an MEG dataset previously acquired at Aarhus University (Andersen et al. 2016). The data were recorded in a magnetically shielded room using an Elekta

Neuromag Triux system with 102 magnetometers and 204 orthogonal planar gradiometers. Data were recorded at a frequency of 1000 Hz.

2.2.1.1 Participants

Nineteen participants took part in the experiment (Mean age = 26.6 years; SD = 4.4 years). Two participants were excluded from the analyses: one for failing to complete the experiment and the other for not using the 'almost clear experience' rating at all (see procedure details below).

2.2.1.2 Experimental Design and Statistical Analyses

In order to obtain a range of awareness ratings from each subject, a visual masking paradigm was used (**Figure 2.2A**). First, a fixation cross was presented for either 500, 1000, or 1500 ms, followed by a target stimulus for 33.3 ms. The target stimulus was either a square or a diamond presented in white/grey on a black (RGB value 0, 0, 0) background (**Figure 2.2A**). A static random noise mask followed the target and was presented for 2000 ms. Participants were required to identify the target during these 2000 ms, before rating their awareness of the stimulus on the perceptual awareness scale (PAS). PAS consists of four possible responses: no experience (NE), weak glimpse (WG), almost clear experience (ACE), and clear experience (CE). Following identification of the target, participants reported their awareness of the stimulus. Participants pressed the upper button of a second response box (PAS response) to cycle through the different PAS categories. The lower button was then pressed to confirm the selection. The response boxes used for target identification and awareness report were swapped between hands every 36 trials to minimise lateralised motor responses contributing to MEG activity patterns. Additionally, the temporal sequence of responses offers protection against decoding preparatory motor signals when decoding awareness ratings, since participants were first required to report the stimulus identity. More details regarding the instructions given to participants about each PAS response can be found in Andersen et al. (2016).

The experiment consisted of one practice block and 11 experimental blocks, each with 72 trials. A contrast staircase was used for the target stimuli in order to obtain a sufficient number of responses for each PAS rating. The staircase procedure had 26 contrast levels ranging from a contrast of 2% to 77%, with a step size of 3 percentage points. In the practice block and first experimental block, the staircase increased by 1 level if a participant made an incorrect judgement on the identification task, and decreased by 2 levels if a participant made 2 successive correct identification judgements. For the remainder of the blocks, the staircase was adjusted based on which PAS rating the participant had used least throughout the experiment so far. Specifically, if NE had been used the least number of times throughout a block, 3 levels were subtracted after 2 consecutive correct answers, and 1 added for a wrong answer. If WG was the least used response, 2 levels were subtracted and 1 added. For ACE, 1 level was subtracted and 2 added. For CE, 1 level was subtracted and 3 added. This staircase procedure ensured a sufficient number of responses for each rating.

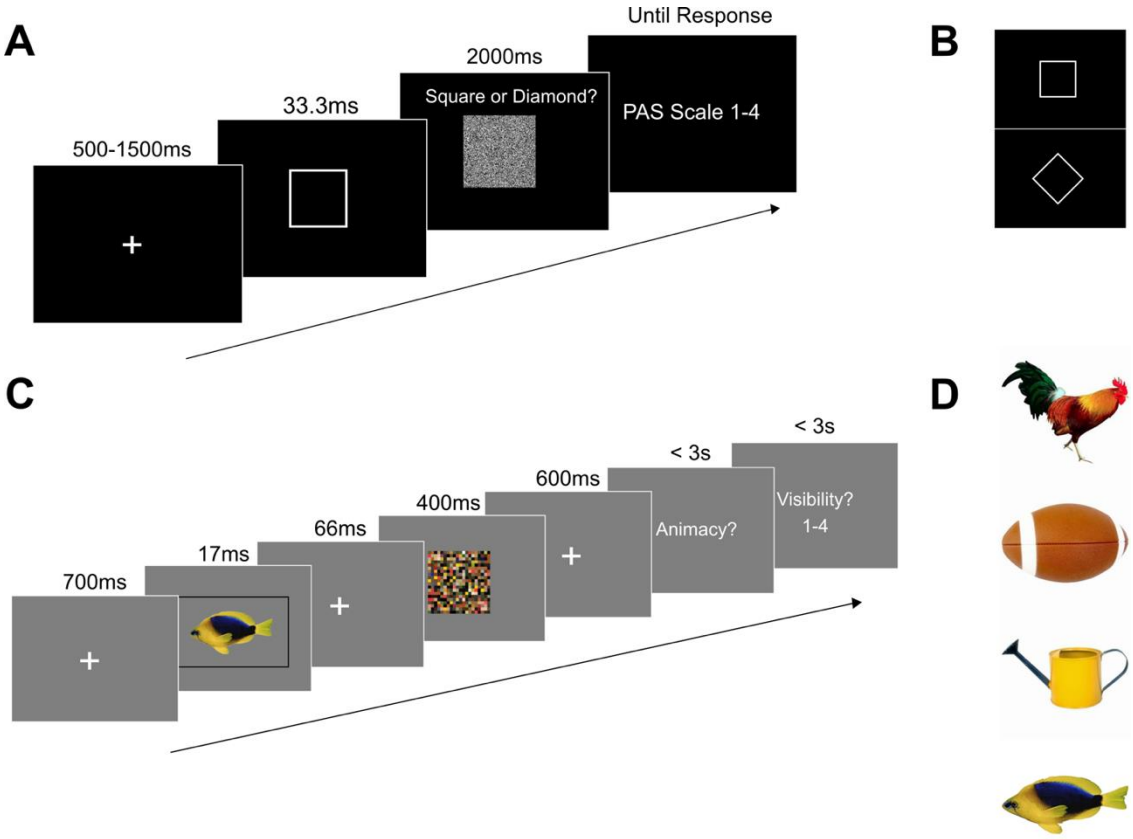


Figure 2.2. Experimental paradigms. A: Experimental paradigm for the MEG data collected by Andersen et al. (2016). First, a fixation cross was presented for 500, 1000 or 1500 ms. Then, either a square or a diamond was shown for 33.3 ms, followed by a static noise mask for 2000 ms. While the mask was shown, participants reported the identity of the target. Finally, they reported their awareness of the stimulus using the PAS scale. B: Stimuli used in Andersen et al. (2016). C: Experimental paradigm for the fMRI data collected by Dijkstra et al. (2021). A stimulus was presented for 17 ms, followed by a 66 ms ISI and a 400 ms mask. Participants then indicated whether the stimulus was animate or inanimate, and finally rated the visibility of the stimulus on a four-point scale. D: Stimuli used in Dijkstra et al. (2021).

2.2.1.3 Pre-processing

MEG data were analysed using MATLAB 2019a and FieldTrip (Oostenveld et al., 2011). The data were pre-processed with a low pass filter at 100 Hz, as well as a Discrete Fourier Transform (DFT) and bandstop filters at 50 Hz and its harmonics. The data were split into epochs of -200 to 2000 ms around stimulus onset and down-sampled to 250 Hz. For baseline correction, for each trial, activity 200 ms prior to stimulus presentation was averaged per channel and subtracted from the entire epoch. During artefact rejection, trials with high variance were visually inspected and removed if they were judged to contain excessive artefacts. This procedure was performed blind to the experimental condition to avoid experimenter bias and was completed separately for the magnetometers and gradiometers in the Elekta Neuromag Triux system. Following artefact rejection the mean number of trials per PAS rating were as follows (numbers in brackets refer to standard deviations): No Experience: 180.35 (59.64); Weak Glimpse: 168.10 (74.75); Almost Clear Experience: 186.35 (82.74); Clear Experience: 115.24 (77.13). To further remove eye-movement artefacts, an independent components analysis was carried out on the MEG data, and the components with the highest correlation with each of the electro-oculographic (EOG) signals were discarded after visual inspection. Components showing topographic and temporal signatures typically associated with heart rate artefacts were also removed by eye.

Since I re-analysed previously collected data, I was unable to fully control for neural signals that typically covary with awareness level. As such, to better characterise the contribution of these signals to ratings of awareness I created two additional analysis pipelines. First, to ensure the results were not entirely driven by the contrast level of the stimuli, I regressed stimulus contrast level on each trial out of the pre-processed MEG

data. Second, to investigate whether differences in pre-stimulus activity contributed to differences in perceptual visibility (Benwell et al., 2017; Podvalny et al., 2019; Samaha et al., 2017) I ran the data through the same pre-processing pipeline as above except for two adjustments: removing the baseline correction stage and lengthening the epochs to -450 ms to 2000 ms around stimulus presentation. The omission of baseline correction allows the analysis to be sensitive to differences in the pre-stimulus activity (in the offset or mean amplitude, for example) of trials associated with different awareness ratings which would otherwise be removed by baseline correction (the baseline correction procedure results in each trial's pre-stimulus window having a mean activity of zero across all time points for each channel, such that my RSA (Representational Similarity Analysis) and decoding analyses would be unable to detect and characterise any pre-stimulus contribution to visibility codes).

2.2.1.4 Representational similarity analysis

RSA allows one to directly compare bespoke hypotheses about the structure of neural data (Kriegeskorte & Kievit, 2013). In RSA, hypotheses are expressed as model representational dissimilarity matrices (RDMs), which define the predicted similarity of neural patterns between different conditions according to each hypothesis. Here, I defined 4 model RDMs that make different predictions about whether or not awareness ratings generalise over perceptual content, and whether or not each rating leads to a graded activation pattern partially shared by neighbouring ratings (**Figure 2.3A**).

In the Abstract-Graded RDM, I model awareness ratings as being independent of perceptual content (such that ratings of a clear experience of a square have an identical neural profile to those of a clear experience of a diamond), as well as being graded in nature (exhibiting a distance effect such that ratings of "no experience" are more similar to those of "weak glimpse", than of "almost clear experience"). In the Specific-Graded RDM, awareness ratings are modelled as being graded in the same way, but they are now represented differently depending on which specific stimulus they are related to. Conversely, the Abstract-Discrete RDM represents PAS ratings as independent of

perceptual content but with no graded structure/distance effects (such that the neural code underpinning a report of “no experience” is equally (dis)similar to the neural code reflecting either a “weak glimpse” or a “clear experience”). Finally, the Specific-Discrete RDM reflects the null model, whereby there is no observable representational similarity structure among conditions, such that neural patterns reflecting one specific awareness rating are equally dissimilar to all other awareness ratings.

RSA involves the comparison of the model RDMs with empirical RDMs constructed from neural data. To do this, I first ran a linear regression on the MEG data with dummy coded predictors for each of the eight conditions (Square trials: NE, WG, ACE, CE; Diamond trials: NE, WG, ACE, CE; trial condition coded with a 1, alternative classes coded with a 0). This resulted in coefficient weights for each condition at each time point and sensor, with the weights representing the neural response per condition, averaged over trials. I then computed the Pearson distance between each pair of condition weights over sensors, resulting in an 8 x 8 neural RDM reflecting the similarity of neural patterns across different awareness ratings and stimulus types (Luyckx et al., 2019). Neural RDMs were subsequently smoothed over time via convolution with a 60 ms uniform kernel.

I then compared this neural RDM with the model RDMs. To compare model RDMs with the neural RDM, I correlated the lower triangle of the model and neural RDMs using Kendall’s Tau rank correlation (Nili et al., 2014). I performed this procedure at every time point, resulting in a correlation value at each time point for each model. Importantly, I only correlated the lower triangle of the RDMs, excluding the diagonal to avoid spurious correlations driven by the increased similarity of on-diagonal values compared to off-diagonal values (Ritchie et al., 2017). This precluded me from directly testing the Specific-Discrete model, since it is represented by a uniform RDM, and as such would give identical rank correlation values regardless of the neural RDM it was compared to. However, since this RDM reflects the null model (i.e. that there is no observable representational structure amongst awareness ratings), this model is implicitly compared with the other model RDMs when I examine whether the correlation of these model RDMs with the neural RDM is greater than 0.

One concern with this approach is that the graded hypotheses may win due to the neural data itself being noisy. In other words, if the neural correlates of ratings are not cleanly dissociable, there will be greater overlap between all ratings in the empirical RDM, including those of close neighbours. To ensure I did not obtain spuriously high similarity with the Abstract-Graded model in virtue of this model's low-frequency content, I performed a shuffling and blending procedure. This procedure involved shuffling the lower triangle of the Abstract-Discrete RDM before apportioning neighbours of the four (shuffled) high correlation cells with graded amounts of correlation. Correlation was blurred most to immediate neighbours, and less to diagonal neighbours, matching the format of the graded RDMs (**Figure 2.3D**). I ran this procedure 1000 times per subject, resulting in 1000 Shuffled-Discrete and 1000 Shuffled-Graded RDMs. I compared all shuffled-discrete and shuffled-graded RDMs with subjects' neural data at each time point. Finally, I took the average correlation value for each time point across all permutations such that I had a Kendall's Tau value for both the Shuffled-Discrete and Shuffled-Graded RDMs across time per subject. Through this approach, I was able to compare neural data with RDMs that shared no representational similarity with the Abstract-Graded RDM while controlling for differences in variance and frequency profile.

2.2.1.5 Within-subject multivariate decoding analysis

To support and extend conclusions drawn from the RSA analyses, I ran an exploratory analysis using temporal generalisation methods (King & Dehaene, 2014) to identify content-invariant and graded representations of awareness ratings while also investigating the stability of these representations over time. In this procedure, a separate classifier is trained on each time point (from 200 ms pre-stimulus to 2000 ms post-stimulus) and tested on all other time points. This method results in a time-by-time decoding accuracy matrix indicating the extent to which neural representations are stable over time. Above chance decoding at a particular point in the decoding matrix indicates that neural representations present at the training and testing time points are similar, whilst chance decoding indicates the representations have changed.

I ran the above temporal generalisation analysis using both a within-condition and cross-condition decoding procedure. Within-condition decoding involved training and testing the decoder to classify PAS ratings on trials from one stimulus type (either squares or diamonds). In cross-condition decoding, I trained on trials from one stimulus type and tested on trials from the other (e.g. trained on square trials and tested on diamond trials, and vice-versa). In both cases, I used a 5-fold cross validation scheme, with a balanced number of trials per class within each fold. Cross-condition decoding, where a classifier trained to decode multivariate neural patterns in one class of stimuli is tested on an unseen class of stimuli, offers an empirical test of whether the neural patterns associated with each class share a similar neural code across conditions (Albers et al., 2013; Bernardi et al., 2020; Dijkstra et al., 2018). This analysis therefore complements the RDM analysis in being able to test for content-invariant perceptual visibility codes, while also providing information about their stability over time. I performed all multiclass decoding analyses with a multiclass Linear Discriminant Analysis (LDA) decoder using the MVPA-light toolbox (Treder, 2020) with FieldTrip. Each of the four PAS ratings served as classes for the decoder to classify trials into. I used L1-regularisation of the covariance matrix, with the shrinkage parameter calculated automatically using the Ledoit-Wolf formula within each training fold (Ledoit & Wolf, 2004a).

It is important to note that cross-validation is not technically necessary during cross-condition decoding because the test data is never seen by the classifier during training, so there is no risk of overfitting. However, I employed a cross-validation scheme for all the decoders so that differences in their performance would not be due to differences in training procedures (e.g. number of trials in the training or test set). Data was smoothed over 7 samples (28ms) and classification analysis was run on individual time points throughout the whole trial to characterise the temporal dynamics of the representations (-200 to 2000 ms post-stimulus).

2.2.1.6 Stimulus decoding

One difficulty with interpreting a content-invariant representation of perceptual visibility is that it may reflect a lack of power to detect content-specific differences between conditions (e.g. square vs. diamond). To control for this possibility, I sought to ensure that the resolution of the data was sufficiently fine-grained to pick up differences in the neural encoding of different stimuli. To do this, I applied a binary decoding procedure using a binary LDA decoder with the same classification parameters as above. In this analysis, the two stimulus types (squares and diamonds) were used as classes for the decoder to classify trials into. For this analysis I grouped trials into low (NE and WG) and high visibility (ACE and CE) trials to ensure sufficient power, performing the decoding analysis separately in each group. Once again, data were smoothed over 7 samples (28ms) and analysed on individual time points throughout the whole trial (-200 to 2000 ms post-stimulus).

2.2.1.7 Statistical Inference

To determine whether the RSA and decoding results were statistically significant, I used cluster-based permutation testing (Maris & Oostenveld, 2007) with 1000 permutations. For RSA, within each permutation I flipped the sign of each ranked correlation value at each time point for each participant and performed a one-sample t-test against 0. Resulting t-values associated with a p-value smaller than 0.05 were used to form clusters across the single time dimension. For each cluster, an associated cluster statistic was computed, the largest of which was stored per permutation to build a group-level null distribution. The cluster statistic computed from the observed data was then compared to this chance distribution to determine statistical significance with an alpha level of 0.05. This procedure controls for the multiple comparisons problem by only performing one comparison at the inference stage and specifically tests the null hypothesis that the observed data are exchangeable with data from the permuted (null) distribution (Maris & Oostenveld, 2007). I used the same cluster-based permutation procedure to compare how well different model RDMs predicted the neural data. In this case, I performed paired-comparisons where ranked correlation values per RDM were randomly swapped within subjects per permutation to build up a group-level null distribution.

For decoding results, I used the same cluster-forming parameters, but this time randomly flipped the sign of individual subjects' accuracy scores per permutation to build up a group-level null distribution. Additionally, I formed clusters over both time dimensions of the temporal generalisation matrices. I used the same cluster-based permutation procedure to compare performance between cross-condition and within-condition decoders.

It is important to note that the cluster-based permutation testing procedure does not allow for inference as to the exact time points at which neural representations come into existence. This is because the algorithm does not consider individual time points at the statistical inference stage, since at this point it only relies on cluster statistics, which encompass multiple time points (Sassenhagen & Draschkow, 2019). Still, as I am not interested in the precise onset of content-invariant representations of awareness ratings but rather their general temporal profile, this method is sufficient for my purposes.

2.2.2 fMRI experiment

To help localise representations of perceptual visibility in the brain, I re-analysed a previously collected fMRI dataset (Dijkstra et al., 2021). It is worth noting that, while source-space decoding in MEG is certainly possible (Andersen et al., 2016; Sandberg et al., 2013), fMRI is much better suited to answering this question at a fine spatial scale, especially as I wish to compare the (potentially fine-grained) differences and similarities in regional activity covarying with perceptual content and/or visibility.

2.2.2.1 Participants

Thirty-seven participants took part in the study. Eight participants were excluded from the analyses. One was excluded because they quit the experiment early, and another because they failed to follow task instructions. The final six subjects were excluded because they did not have at least 10 trials in each visibility rating class after the grouping

procedure (see *Within-subject multivariate searchlight decoding analysis* below). Twenty-nine subjects were thus included in the final analyses (mean age = 25.35; SD = 6.31).

2.2.2.2 Stimuli

The stimuli used were taken from the POPORO stimulus data set (Kovalenko et al., 2012). The stimuli selected were a rooster, a fish, a watering can, and a football (**Figure 2.2D**), and were selected based on familiarity and visual difference to maximise classification performance as well as both accuracy and visibility scores calculated in a pilot experiment run by Dijkstra et al. (2021). The mask was created by randomly scrambling the pixel values of all stimuli combined (**Figure 2.2C**).

2.2.2.3 Experimental Design and Statistical Analyses

The experiment consisted of two tasks: a perception task and an imagery task. Each of these tasks were executed in interleaved blocks and were counterbalanced across participants. The re-analysis used data only from the perception task and thus I omit details of the imagery component of the study. The perception task ran as follows. A stimulus was presented for 17 ms, followed by a backward mask for 400 ms. Participants then indicated whether the stimulus was animate or inanimate and rated the visibility of the stimulus on a scale from 1 (not visible at all) to 4 (perfectly clear). For both the discrimination and visibility decisions, button response mappings were randomised across trials, thus preventing preparatory motor responses from contaminating the neural signals of interest. The task was made up of visible and invisible trials. The difference between these trials was the length of the interstimulus interval (ISI) between the stimulus and the mask. In the visible trials the ISI was 66 ms, and in the invisible trials the ISI was 0 ms. In the present study, I only analysed data from invisible trials (**Figure 2.2C**) because these were associated with the variation in the visibility ratings that I am interested in. Choosing to focus on a single ISI also means that differences in visibility ratings were not driven by differences in stimulus presentation characteristics. There were 184 trials in total, with 46 repetitions per stimulus divided over 4 blocks. More detailed information regarding the study protocol can be found in (Dijkstra et al., 2021).

2.2.2.4 Acquisition

fMRI data were recorded on a Siemens 3T Skyra scanner with a Multiband 6 sequence (TR: 1 s; voxel size: 2 x 2 x 2 mm; TE: 34 ms) and a 32-channel head coil. The tilt of each participant's field of view was controlled using Siemens AutoAlign Head software, such that each participant had the same tilt relative to their head position. T1-weighted structural images (MPRAGE; voxel size: 1 x 1 x 1 mm; TR: 2.3 s) were also acquired for each participant.

2.2.2.5 Preprocessing

Data were pre-processed using SPM12. Motion correction (realignment) was performed on all functional imaging data before co-registration to the T1 structural scan. The scans were then normalised to MNI space using DARTEL normalisation and smoothed with a 6 mm Gaussian kernel, which has been shown to improve group-level decoding accuracy (Gardumi et al., 2016; Hendriks et al., 2017; Misaki et al., 2013). Slow signal drift was removed using a high pass filter of 128s.

2.2.2.6 General Linear Model

Coefficient weights were estimated per trial with a general linear model that contained a separate regressor for each trial at the onset of the stimulus convolved with the canonical HRF. Alongside nuisance regressors (average WM and CFG signals and motion parameters), the screen onset and button press of both the animacy and visibility responses were included as regressors, as well as a constant value per run to control for changes in mean signal amplitude across runs.

2.2.2.7 Within-subject multivariate searchlight decoding analysis

For decoding the fMRI data, I binarized the visibility ratings into low and high visibility classes. This is because in contrast to the MEG experiment, visibility was not staircased

per-participant, leading to a large number of participants failing to have enough trials at each of the four visibility ratings in both animate and inanimate trials. Because training a decoder on such a small number of trials would yield unreliable and noisy results, trials were therefore sorted into low visibility and high visibility classes on a subject-by-subject basis prior to analysis. This was performed as follows. The median visibility rating (from 1 to 4) was extracted from each subject and trials with a lower visibility rating than the median were classed as low visibility trials, and those with visibility ratings equal to or greater than the median were classed as high visibility trials. This procedure allowed me to control for the fact that different subjects had different distributions of visibility ratings, such that the lower 1 and 2 ratings did not always correspond to low visibility trials, and likewise the higher 3 and 4 ratings did not always correspond to high visibility trials. For instance, one subject may have used visibility ratings 2 and 3 in around 50% of trials, rating 4 on the other 50%, and not used rating 1 at all. In this case, I would label ratings 2 and 3 as low visibility, and rating 4 as high visibility.

Trials were next grouped according to whether they contained an animate or inanimate stimulus. For each participant, if there were less than 10 trials in either the low or high visibility class for either the animate or inanimate trials, the participant was removed. This was the case for 6 participants. The mean number of trials per condition following this procedure were as follows (numbers in brackets denote the standard deviation): animate-high visibility: 63.48 (11.34); animate-low visibility: 25.31 (10.38); inanimate-high visibility: 61.10 (12.24); inanimate-low visibility: 28.86 (11.48).

I used an LDA classifier on the beta estimates per trial to decode low and high visibility ratings within and across animate/inanimate stimulus conditions. Cross-condition decoding was performed by training the LDA classifier on low versus high visibility ratings in animate trials and then testing it on low versus high visibility ratings in inanimate trials, and vice versa. Cross-condition decoding was performed with the same logic as in the MEG analysis: if I train a classifier to decode visibility ratings in animate trials and use this classifier to successfully decode visibility ratings in inanimate trials, I can conclude the representations of visibility ratings are similar across different perceptual content. Once

again, I also performed within-condition decoding, where the classifier was trained on low versus high ratings in one condition (e.g., animate trials), and tested on trials in the same condition allow a direct comparison of within- and across-condition decoding performance. This comparison allowed me to determine where content-specific representations of perceptual visibility may exist in the brain. Decoding was performed with a 5-fold cross validation scheme using L1 regularisation with a shrinkage parameter of 0.2, and, similar to the MEG analysis, cross-validation was used for both within-condition and cross-condition decoding. Trials were down-sampled prior to decoding, such that there was an equal number of low and high visibility trials in each fold. To ensure that the data were sensitive enough to show content-specific codes, I additionally ran a similar analysis that sought to decode stimulus content (animate or inanimate) rather than visibility. This analysis was similar in structure except the classifier was trained to decode animate vs. inanimate trials rather than visibility level.

Decoding was performed using a searchlight method. Searchlights had a radius of 4 voxels (257 voxels per searchlight). As such, at every searchlight the classifier was trained on 257 features (one beta estimate for each voxel in the searchlight) for each trial in every fold. The searchlights moved through the brain according to the centre voxel, meaning that each voxel was entered into multiple searchlights. After decoding in each searchlight, the accuracy of the classifier was averaged across folds and this value was stored at the centre of the searchlight to produce a brain map of decoding accuracy.

2.2.2.8 Stimulus decoding in fMRI Regions of Interest (ROI)

As in the MEG analysis, I again wished to establish that findings of content-invariant awareness representations were not due to an inability to decode content itself. I tested whether I could decode perceptual content within two ROIs with successful visibility decoding from the searchlight results. To do this, I created two masks, one visual and one frontal, and then selected the 200 voxels within this mask that had the highest mean visibility decoding accuracy averaged across all four decoding directions (within animate; within inanimate; train animate-test inanimate; train inanimate-test animate). For the

frontal mask I used a connectivity-based parcellation of the orbitofrontal cingulate cortex that spanned frontal regions with successful visibility decoding. These were regions 8m (*x, y, z peak voxel coordinates per hemisphere* LH: -14.6, 33.8, 43.3; RH: 13.5, 32.3, 44) and 32d (LH: -8.7, 37.5, 23.4; RH: 12.7, 40.4, 17.5) (Neubert et al., 2015). The visual mask spanned an area with successful visibility decoding in occipital regions VO1 (LH: -27.1, -70.9, -11.3; RH: 27.5, -69.5, -10.6), VO2 (LH: -25.6, -64.3, -10.6; RH: 26.7, -59.9, -9.1), PHC1 (LH: -27.1, -54, -8.3; RH: 28.3, -53.2, -8.3), and PHC2 (LH: -28.6, -45.9, -8.3; RH: 29, -43.7, -9.8) (L. Wang et al., 2015). Coordinates for the clusters obtained within each ROI can be found in **Supplementary Table 2.1**. Using the 200 ROI voxels as features, I decoded animate (rooster and fish) versus inanimate (watering can and football) stimuli in low and high visibility trials separately using the same 5-fold cross validation procedure and LDA parameters as above, down-sampling trials prior to decoding to ensure an equal number of animate and inanimate trials in each fold.

2.2.2.9 Group-level statistical inference

Distributions of accuracy values from classification of fMRI data are often non-Gaussian and asymmetric around chance level. This means that parametric statistical comparisons, such as t-tests against chance decoding (50%), are unable to provide valid tests of whether group-level accuracy values are significant (Stelzer et al., 2013). Therefore, to determine where classifiers had performed significantly above chance, I compared mean performance across all participants with a null distribution created by first permuting the class labels 25 times prior to decoding per participant and then using bootstrapping to form a group-level null distribution of 10,000 bootstrapping samples (Stelzer et al., 2013). I did this separately for each decoding direction (within: train and test on animate; train and test on inanimate; cross: train on animate, test on inanimate; train on inanimate, test on animate). To perform statistical inference on an average cross-decoding map created by averaging the two cross-condition decoding directions, this average map was compared to a group-level null distribution formed by averaging the two null distributions created for the two separate maps. To compare within-condition and cross-condition classification performance, a group-level null distribution was formed by taking the

difference between cross and within decoding scores throughout the bootstrapping procedure. To control for multiple comparisons in the searchlight analysis, resulting p values were subsequently corrected for multiple comparisons with a false discovery rate of 0.01.

2.3 Results

2.3.1 Representational structure of perceptual visibility in whole-brain MEG data

I used RSA to test whether perceptual visibility levels (PAS ratings) correlated with MEG activity patterns independently of perceptual content (Abstract RDMs) or together with perceptual content (Specific RDMs). I additionally tested whether neural activity patterns covaried with visibility levels in a graded or discrete manner (**Figure 2.3A**). A model instantiating graded and abstract representations of awareness ratings significantly predicted the neural data throughout most of the post-stimulus period (purple line; **Figure 2.3B**). In contrast, a model with an abstract but discrete representational structure was able to predict the neural data only in an early phase of the trial between approximately 100 to 500 milliseconds after stimulus onset (green line). Paired comparisons between these two models showed that the Abstract-Graded model was significantly better at predicting the neural data than the Abstract-Discrete model throughout the majority of the trial (purple and green dots). The Specific-Graded model did not significantly predict the neural data at any point during the trial (gold line), and likewise the Abstract-Graded model was found to be significantly better at predicting the neural data than the Specific-Graded model in a direct comparison (purple and yellow dots), indicating that an abstract model of awareness ratings better described their neural representation. In line with this, multidimensional scaling of awareness ratings revealed a principal dimension encoding vividness that was shared by both square and triangle stimuli (**Figure 2.3C**). To assess the spatial distribution of Abstract-Graded signals across sensors, I repeated the analysis for frontal and occipital sensors separately (following Hu et al., 2018), finding similar

results in each case (**Supplementary Figure 2.1**). These results indicate that neural correlates of perceptual visibility generalise over perceptual content and exhibit distance effects, indicative of neural populations tuned to specific degrees of visibility with overlapping tuning curves.

To ensure that the neural data did not exhibit spuriously high similarity with the Abstract-Graded model in virtue of its increased variance and reduced frequency when compared to the Abstract-Discrete and Specific models, I performed a shuffling and blending control analysis (**Figure 2.3D**). This procedure revealed no significant prediction of the neural data for either the shuffled-discrete or shuffled-graded RDMs (**Figure 2.3E**). As such, RDMs with frequency and variance profiles matching that of the Abstract-Graded RDM, but without any relationship with awareness ratings, were not able to significantly predict neural data, in contrast to the Abstract-Graded model that captures the graded and content-invariant structure of awareness ratings. To additionally control for the possible influence of stimulus contrast on the RDM results, I confirmed that similar results were obtained when regressing out the linear component of contrast (**Supplementary Figure 2.2**). It is possible that nonlinear or multivariate effects of contrast may have still driven some of my findings. Indeed, whilst we see a clear linear trend from no experience to clear experience across the first dimension in the original data, this dimension is somewhat compressed following the removal of the linear component of stimulus contrast. Along this compressed dimension, higher ends of the scale are represented more similarly than those at the lower end. This is potentially in line with a Weber scaling law in the neural representation of perceptual vividness, as also found for other magnitude codes (e.g. the 'size effect' in numerical cognition), and also hints at a role for stimulus contrast in driving some of the difference between CE and ACE in the original analysis. However, even after removing potentially confounding effects of stimulus contrast, the difference in perceptual vividness between NE, WG and ACE/CE is clearly distinguished in **Supplementary Figure 2.3B**.

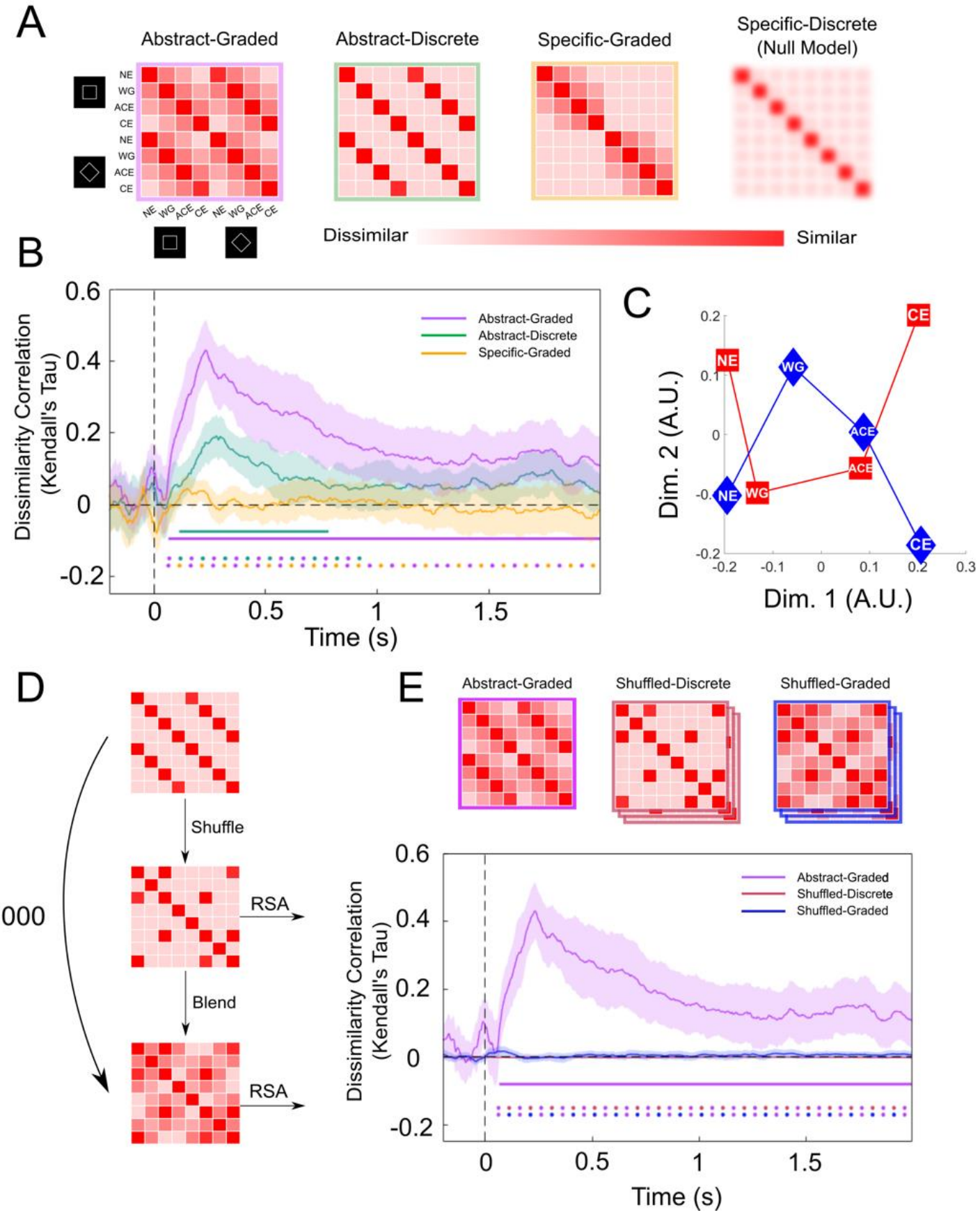

A
Abstract-Graded
Abstract-Discrete
Specific-Graded
Specific-Discrete
(Null Model)
NE
WG
ACE
CE
Dissimilar
Similar
B
Dissimilarity Correlation
(Kendall's Tau)
Abstract-Graded
Abstract-Discrete
Specific-Graded
Time (s)
C
Dim. 2 (A.U.)
Dim. 1 (A.U.)
D
Shuffle
RSA
Blend
RSA
x1000
E
Abstract-Graded
Shuffled-Discrete
Shuffled-Graded
Dissimilarity Correlation
(Kendall's Tau)
Time (s)

Figure 2.3. Neural representations of perceptual visibility are abstract and graded. A: From left to right: Abstract-Graded model where neural correlates of awareness ratings are independent of perceptual content and follow a graded structure; Abstract-Independent model where awareness ratings are independent of perceptual content but do not follow a graded structure; Specific-Graded model where awareness ratings are specific to the perceptual content to which they relate and follow a graded structure; Specific-Discrete (Null hypothesis) model where there is no observable representational structure amongst awareness ratings (PAS ratings, NE: No Experience, WG: Weak Glimpse, ACE: Almost Clear Experience, CE: Clear Experience). B: RSA reveals that the Abstract-Graded model was the best predictor of the representational structure of neural patterns in whole-brain sensor-level MEG data. Solid horizontal lines represent time points significantly different from 0 for a specific RDM at $p < .05$, corrected for multiple comparisons. Horizontal dots denote statistically significant paired comparisons between the different models at $p < .05$, corrected for multiple comparisons. I obtained similar findings across occipital (**Supplementary Figure 2.1A**) and frontal (**Supplementary Figure 2.1B**) sensors separately, as well as in datasets with stimulus contrast level regressed out (**Supplementary Figure 2.2**) and without baseline correction (**Supplementary Figure 2.5**). We also examined the pattern of classifier mistakes during cross-stimulus decoding, again revealing distance-like effects in perceptual visibility decoding (**Supplementary Figure 2.4**). C: Shuffling and blending procedure. This analysis was performed to control for naturally occurring low-frequency content in neural data. D: Results from both shuffled models reflect the average Kendall's Tau over 1000 shuffling permutations. Purple, red, and blue lines represent similarity of the Abstract-Graded, shuffled-discrete, and shuffled-graded models respectively with neural data. The shuffled-discrete line varies only slightly from 0 and is thus hard to see. The Abstract-Graded model is the only model under consideration that significantly predicted the neural data.

To further characterize the graded representational structure of perceptual visibility, I computed confusion matrices between each rating and its neighbours. By plotting the proportion of predictions for each awareness rating made by the multiclass classifier separately for trials of each rating, I can visualise when the decoder makes mistakes, and which PAS ratings it most often confuses (**Supplementary Figure 2.4**). These confusion plots confirm the distance effects identified with the RSA model comparison, in which neighbouring PAS ratings are most often confused with each other by the classifier, and more distant ratings less so, suggesting that visibility is represented in a graded, ordinal manner.

Finally, I asked whether the model RDMs could also predict pre-stimulus neural activity. If a graded, abstract structure for perceptual visibility is already evident prior to stimulus presentation, this would be indicative of trial-to-trial fluctuations in attention or arousal contributing to my ability to decode content-invariant visibility signals. Interpreting (a lack

of) pre-stimulus decoding from the previous RSA analyses is confounded by the baseline correction procedure applied during pre-processing. To address this issue, I re-ran my analysis on data that had not been baseline corrected. I found that pre-stimulus activity was not captured by any of the candidate RDMs, and that stimulus-triggered responses continued to show the same graded/abstract pattern of results as in the initial analysis (**Supplementary Figure 2.5**). Together these results indicate that pre-trial fluctuations in attention and/or arousal are unlikely to drive my results.

2.3.2 Temporal profile of perceptual visibility codes

Next, I performed a temporal generalisation analysis to further unpack the content-invariant nature of neural signatures of perceptual visibility and to characterize how and whether their patterns change from timepoint to timepoint. Off-diagonal panels in **Figure 2.4** (top right and bottom left) depict temporal generalisation matrices for both directions of cross-condition decoding (top-right: train on squares-test on diamonds; bottom-left: train on diamonds-test on squares). Within these panels, above-chance decoding on the major diagonal indicates that representations of visibility begin to show content-invariance from just after stimulus onset up until the moment of report. Contrasting cross-condition decoding with within-condition decoding resulted in no significant differences in decoding accuracy for either comparison (train on squares, test on diamonds vs. within squares: all $p > 0.89$; train on diamonds, test on squares vs. within diamonds: all $p > 0.4$). In other words, I did not find any evidence that there was content-specific visibility information available over and above content-invariant information. Furthermore, the lack of off-diagonal decoding in each temporal generalisation matrix indicates that the format of content-invariant neural signatures of visibility change rapidly over time.

I again replicated this analysis in a dataset that had not undergone baseline correction to test whether activity contributing to participants' awareness ratings could be decoded prior to stimulus presentation. In line with the RSA analysis on this dataset, I could not decode awareness ratings prior to stimulus presentation when data had not been baseline-corrected (**Supplementary Figure 2.6**).

2.3.3 Content-invariant representations of visibility are found across visual, parietal, and frontal cortex

To localise brain regions supporting content-invariant representations of perceptual visibility, I re-analysed an existing fMRI dataset (Dijkstra et al., 2021). I used a searchlight approach to identify brain regions that represent perceptual visibility in an abstract manner. Both cross-condition and within-condition decoding resulted in above chance accuracy in a number of regions across the visual, parietal, and frontal cortex (**Figure**

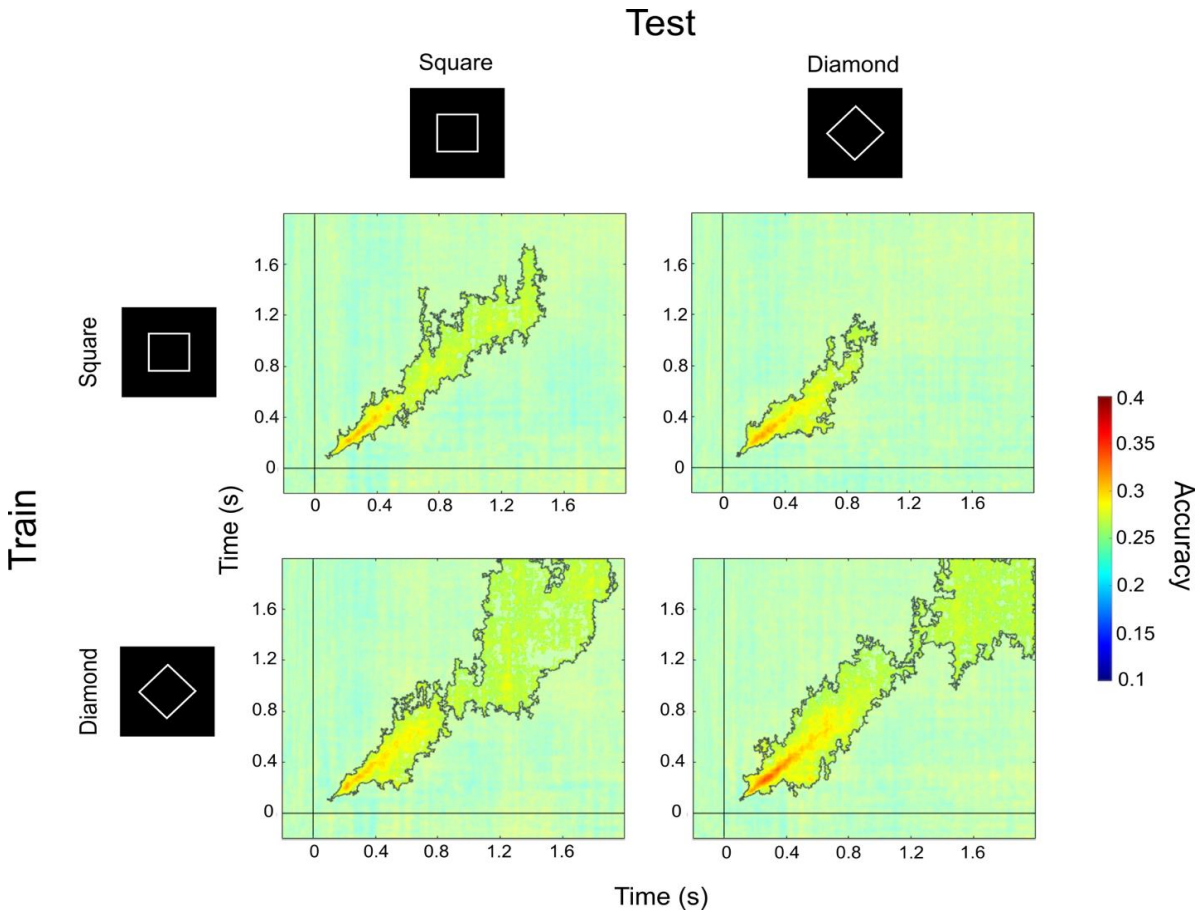


Figure 2.4. Abstract representations of perceptual visibility evolve rapidly over time. Temporal generalisation results for the classification of PAS ratings from MEG data (4 PAS responses; chance = 0.25). For each row, statistical comparisons between the two columns showed no significant differences in decoding accuracy between within- and cross-condition decoding. Non-translucent regions within solid lines highlight above chance decoding, as revealed by cluster-based permutation tests. I replicated these findings in non-baseline-corrected data (**Supplementary Figure 2.6**).

2.5). To assess whether these representations of perceptual visibility were stimulus-dependent, I compared cross-condition decoding to within-condition decoding in both animate and inanimate trials (training on animate trials vs. within animate trials; training on inanimate trials vs. within inanimate trials) and found no significant differences. In other words, I could find no evidence that stimulus-specific visibility information was present over and above stimulus-invariant visibility information. See **Supplementary Table 2.2** for details of the clusters found to be significantly above chance in both cross-condition and within-condition decoding analyses.

2.3.4 Stimulus content can be decoded from both MEG and fMRI data

I next considered the possibility that a content-invariant neural signature of visibility may be obtained because of insufficient sensitivity to perceptual content in the dataset. To address this, I sought to decode stimulus identity, rather than visibility level. Stimulus decoding was above chance in both datasets for high visibility trials. In the MEG data, I was able to decode stimulus identity (square vs. diamond) in trials in which participants used the upper two PAS ratings (ACE/CE), but not when participants used the lower two PAS ratings (NE/WG; **Figure 2.6A**). Similarly, in the fMRI data, the decoding of animate vs. inanimate stimuli was significantly above chance in a visual cortical ROI during trials reported as high visibility (mean accuracy = 0.52; $p = 0.007$) but not in trials reported as low visibility (mean accuracy = 0.496; $p = 0.655$; **Figure 2.6B**). It was not possible to decode stimulus content from a frontal cortical ROI in either low (mean accuracy = 0.5; $p = 0.406$) or high visibility trials (mean accuracy = 0.507; $p = 0.159$). Together, these analyses indicate that stimulus content could be reliably decoded in posterior visual regions.

2.3.5 Stimulus Content and Visibility are Encoded in Dissociable Brain Regions

To further probe the relationship between neural signatures of content and visibility, I ran a searchlight decoding procedure to decode animate versus inanimate stimuli in the fMRI data. Since content could only be decoded in high visibility trials in the ROI analysis (**Figure 2.6B**), I restricted the analysis to these trials. I then compared the overlap between the content-decoding searchlight and the content-invariant visibility searchlight maps. To do this I computed mean content cross-decoding accuracy averaged over the

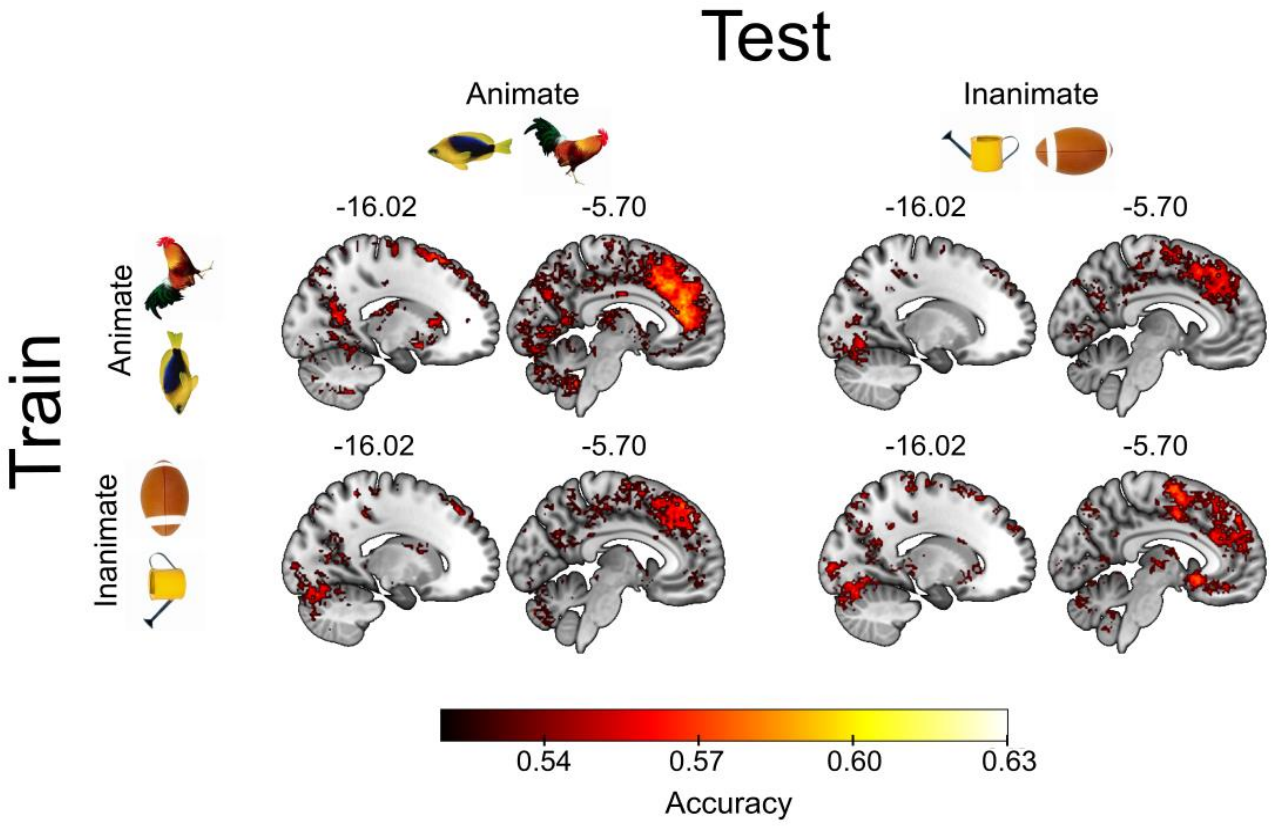


Figure 2.5. Abstract representations of perceptual visibility are found across visual, parietal, and frontal cortex. Searchlight decoding in fMRI data revealed significantly above-chance accuracy in both cross-condition (off-diagonal cells of matrix) and within-condition (on-diagonal cells) in decoding of visibility ratings. Clusters of successful cross-condition decoding were found across the frontal, parietal, and visual cortex. The statistical comparison of cross and within-condition decoding accuracy (comparing the on- and off-diagonal statistical maps) revealed no significant differences anywhere in the brain. Significance was assessed at $p < .05$, corrected for multiple comparisons with an FDR of 0.01. Clusters are reported in **Supplementary Table 2.2.**

two cross-decoding directions (train on animate, test on inanimate; train on inanimate, test on animate) prior to group-level inference (see Methods).

Overall, there was minimal overlap between representations of content and visibility (**Figure 2.6C**). Despite overlapping clusters being obtained in the superior and inferior lateral occipital cortex (see **Supplementary Table 2.3** for full list of individual and overlapping clusters), clear anatomical distinctions in occipital regions can be seen between representations of stimulus content and visibility, with the former being decoded from more lateral regions of the occipital cortex, while the latter was decoded closer to the medial surface (**Figure 2.6C**, **Supplementary Table 2.3**). Fewer clusters of above-chance stimulus content decoding were found in frontal regions, whereas content-invariant representations of visibility were more abundant in these areas (Hatamimajoumerd et al., 2022). Distinct decoding patterns for content and visibility representations further strengthens the notion that content-invariant representations of visibility exist partly independently of perceptual content, even in regions typically associated with the encoding of stimulus content such as the visual cortex (Kamitani & Tong, 2005; Kriegeskorte et al., 2008; Mazor et al., 2022).

2.4 Discussion

In this study I asked whether perceptual vividness covaries with neural activity patterns in a content-specific and/or content-invariant manner. By applying multivariate analyses to MEG and fMRI datasets in which participants rated their awareness of visual stimuli, I found that the vividness of experience is represented in a similar way across different stimulus contents and exhibits signatures of an ordered and graded magnitude code. Furthermore, neural representations of perceptual vividness were found to change rapidly over time and were localized to visual, parietal, and frontal cortices.

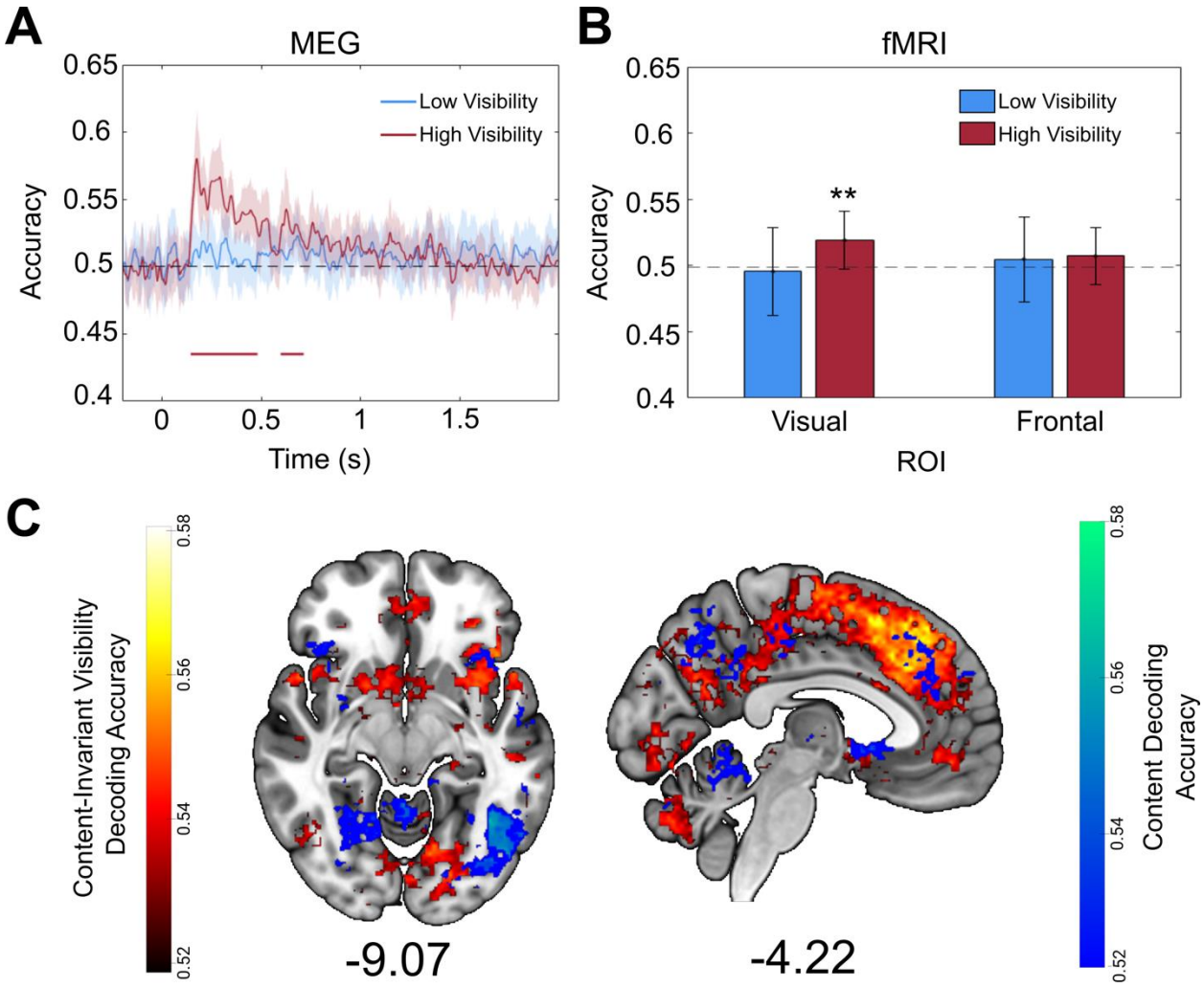


Figure 2.6 Perceptual content can be decoded in high visibility trials and shows distinct representations to visibility. A: Decoding of perceptual content on each trial (squares or diamonds) from participants' whole-brain sensor-level MEG data for low visibility (NE and WG) and high visibility (ACE and CE) trials separately. Successful decoding was possible in high visibility trials up to approximately 700ms post-stimulus onset. Lines are smoothed using a Gaussian-weighted moving average with a window of 20 ms. Shaded area denotes 95% confidence intervals. The solid horizontal line reflects above-chance decoding, as revealed by cluster-based permutation tests. B: Decoding of perceptual content on each trial (animate or inanimate) from participants' fMRI data for low visibility and high visibility trials separately. Decoding was successful in a visual ROI in high but not low visibility trials, and unsuccessful in a frontal ROI. Asterisks denote significance at $p < .01$. Error bars illustrate 95% confidence intervals. C: Searchlight decoding accuracy for content decoding in high visibility trials (blue) and for content-invariant visibility decoding (red). Clusters illustrate areas where content or content-invariant visibility could be decoded significantly above chance. Content-invariant representations of visibility were more widespread than content representations and extended into the prefrontal cortex, whereas both content and visibility could be decoded in distinct locations of the visual cortex. Significance was assessed at $p < .05$, corrected for multiple comparisons with an FDR of 0.01. Clusters are reported in **Supplementary Table 2.3.**

The identification of content-invariant representations of perceptual vividness is in line with recent work highlighting a dissociation between neural correlates of awareness and perceptual content. For example, Sanchez et al. (2020) found neural patterns that indicated whether an individual was aware of a stimulus or not, irrespective of which sensory modality it was presented in. Likewise, Mazor et al. (2022) reported that, while stimulus identity was best decoded from occipital regions, perceptual visibility (stimulus presence vs. absence) could be effectively decoded from a wider range of areas including the parietal and frontal cortex. Notably, a recent study also found that graded changes in perceptual vividness could be reliably decoded from the prefrontal cortex, even in the absence of report, consistent with a contribution to the vividness of experience (Hatamimajoumerd et al., 2022). Whilst I do not claim that representations of vividness are solely content-invariant, I build on these findings by showing that neural signals underlying graded awareness ratings – ranging from the absence of an experience of particular content, to a clear and vivid experience – exhibit a content-invariant neural signature.

Content invariant representations of vividness may also provide a new understanding of the mechanisms supporting intensity-matching in psychophysical tasks. For example, studies of cross-modal intensity matching have demonstrated that subjects can reliably match intensities across sensory domains (Marks et al., 1986, 1988; Stevens & Marks, 1965), and even provide some evidence for absolute equivalences between intensities in different modalities (Marks et al., 1986). Success in such tasks could be mediated by some form of common currency for intensity that is modality-invariant. My findings offer a potential neural framework within which to explain this capacity. Specifically, if the intensity of an experience is mapped onto a low-dimensional and content-invariant neural code for vividness, it should be possible to leverage this representation to reliably match the intensity of stimuli across sensory modalities. This is the essence of ‘mapping theory’ (Krantz, 1972), and could be directly tested by combining intensity-matching psychophysical methods with neuroimaging to examine the degree to which psychophysical estimates of cross-modal magnitudes rely on the same low dimensional neural manifolds associated with vividness observed here.

Although I find evidence for content-invariant signals underlying perceptual vividness, the mechanism by which these signals influence vividness remains to be determined. One candidate mechanism may be the top-down modulation of content-specific representations, driven by content-invariant attention signals. For example, fluctuations in the (content-invariant) degree of attention may increase the perceived contrast of stimuli (Carrasco et al., 2000, 2004), perhaps through the modulation of content-specific neuronal responses. In line with this model, there may be multiple components to a neural representation of perceptual vividness: content-invariant signals that are associated with the degree of attention or other domain-general factors, and content-specific representations modulated by such attentional signals. Such a model neatly exemplifies how content-invariant and content-specific neural signatures may together contribute to the subjective experience of perceptual vividness. Indeed, on this view, content-specific modulations may be subtle compared to changes in abstract representations determining the degree of attention, which could in turn explain why the content-specific vividness model did not provide a good fit to the neural data.

The finding that neural representations of awareness ratings display a distance effect is suggestive of perceptual vividness relying on similar schemes to those encoding magnitude in other domains such as number. Specifically, my results are consistent with the possibility that distributed populations of neurons are tuned to specific phenomenal magnitudes, in the same way that specific populations of neurons are sensitive to certain numerical magnitudes (Harvey et al., 2013; Kutter et al., 2018; Piazza et al., 2004). Such a prediction could be tested through repetition suppression experiments (Piazza et al., 2004), and/or by collecting single-unit recordings from human patients while they provide subjective awareness ratings (Pereira et al., 2021). A variety of analogue magnitudes have been shown to rely on common magnitude representations (Luyckx et al., 2019; Pinel et al., 2004; Yallak & Balcı, 2021), prompting a hypothesis that domain-general representations are responsible for encoding low-dimensional quantities in the brain (Summerfield et al., 2020; Walsh, 2003). Therefore, an intriguing possibility is that perceptual vividness is supported by similar domain-general magnitude codes. Future work could explore this hypothesis by assessing whether representations of vividness

share neural resources with other analogue magnitude codes such as those for reward or number (Luyckx et al., 2019).

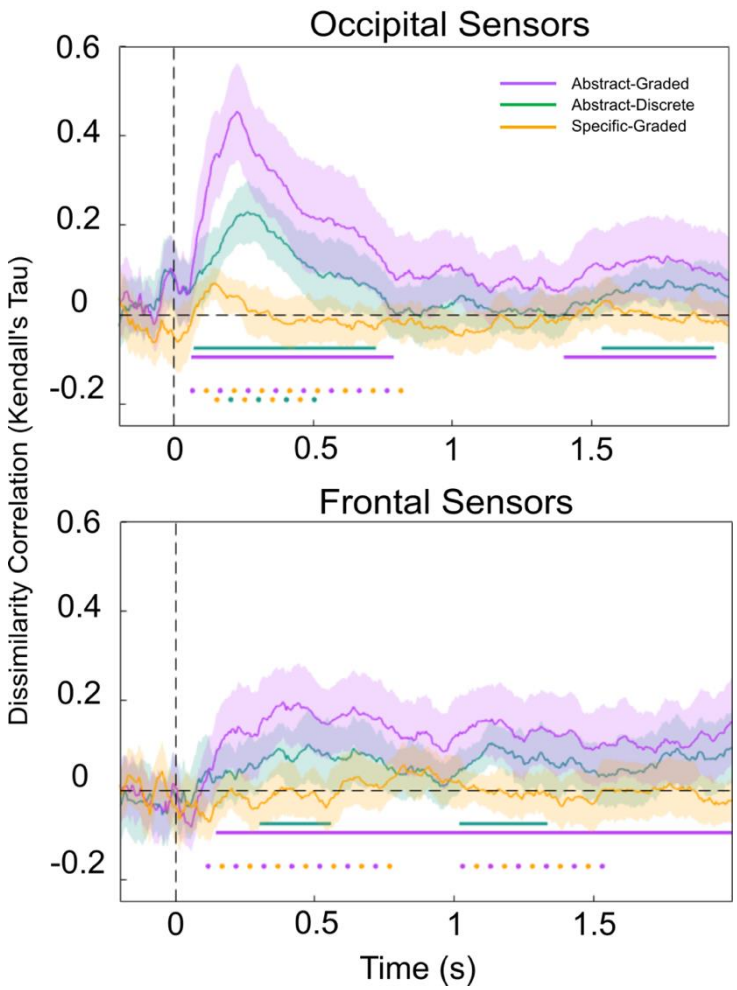
The existence of stimulus-independent representations of perceptual vividness in visual cortical areas was unexpected, since these areas have been shown to distinguish stimulus features rather than subjective vividness in previous studies (Kamitani & Tong, 2005; Kriegeskorte et al., 2008; Mazor et al., 2022). One concern is that neural representations of vividness ratings as revealed by decoding analyses may look similar across stimuli if content-specific information encoded in separate neural populations is treated as belonging to the same population (i.e. within the same voxel). Successful cross-stimulus decoding of vividness ratings could then occur by way of decoding the amplitude of (content-specific) neural responses in these voxels (Fisch et al., 2009; Moutoussis & Zeki, 2002). As a step towards addressing this concern, I show that stimulus-specific decoding remains possible specifically in visual areas on high- (but not low) visibility trials, suggesting that the content-invariant nature of perceptual vividness signals in this region is not due to a lack of power to detect stimulus-specific effects. Moreover, I show anatomical distinctions between content and visibility encoding, again indicating that the unexpected above-chance decoding of visibility in the visual cortex is unlikely to be an artefact of a failure to detect content-specific representations.

Another possibility is that content-invariant signals of perceptual vividness in visual cortex reflect pre-stimulus activations that have been shown to contribute to participants' awareness level in previous studies (Podvalny et al., 2019). Here I could not identify pre-stimulus contributions to visibility codes in the MEG data, supporting a hypothesis that the content-invariant and graded representations I report here are largely stimulus-triggered. As such my results suggest that the content-invariant signals related to awareness level in the current data are partly distinct to those reported by Podvalny *et al*. in temporal profile. In any case, it is worth noting that fluctuations in (pre- or post-stimulus) attention and arousal affecting the intensity of experience (as well as other psychological factors such as emotional state or motivation) may provide domain-general sources of perceptual vividness signals.

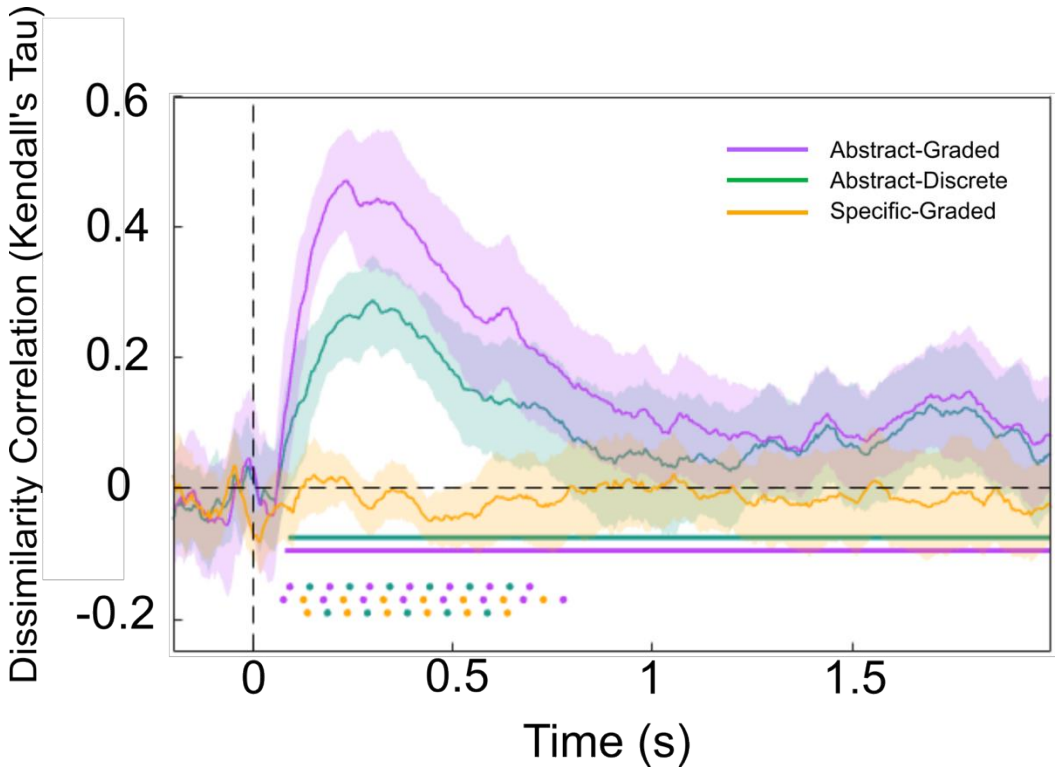
By applying temporal generalisation analysis to the MEG data, I was able to reveal the dynamics of vividness representations over time. This analysis indicated that neural patterns covarying with perceptual vividness are unstable, changing during the course of a trial, consistent with a sequence of different neural populations correlating with awareness level over time (King & Dehaene, 2014). Given that I find that vividness is tracked across a variety of cortical regions, such a rapidly changing temporal profile may reflect dynamic message passing between distinct neural populations, consistent with the reverberation of predictions and prediction errors in hierarchical generative models. Future work to directly test this hypothesis could leverage informational connectivity analyses (e.g. Seeliger et al., 2021) to determine the direction of information flow across interacting brain regions, or use RSA to combine M/EEG and fMRI data collected using the same task and stimuli (Cichy et al., 2014).

In summary, I show that perceptual vividness covaries with content-invariant neural representations that exhibit graded distance effects similar to those observed for analogue magnitude codes in other cognitive domains. These representations are spatially distributed and rapidly evolve over time, consistent with the flow of awareness-related information across the visual, parietal, and frontal cortices. This pattern of results adds to growing evidence for a content-invariant neural component contributing to the strength of conscious experience.

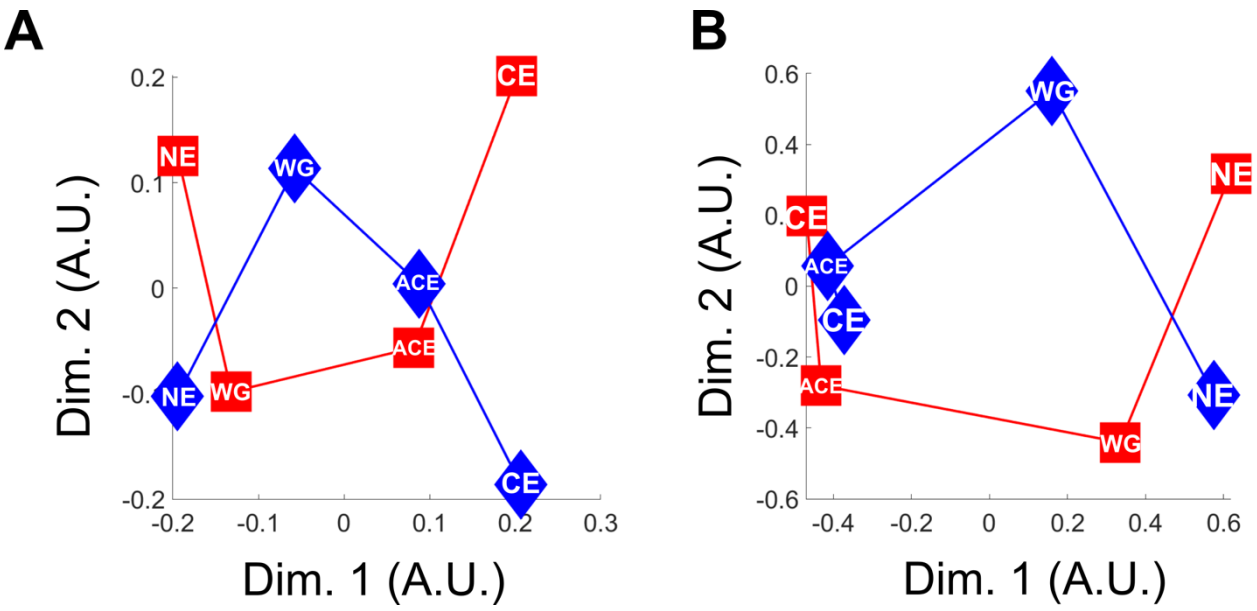
2.5 Supplementary Materials



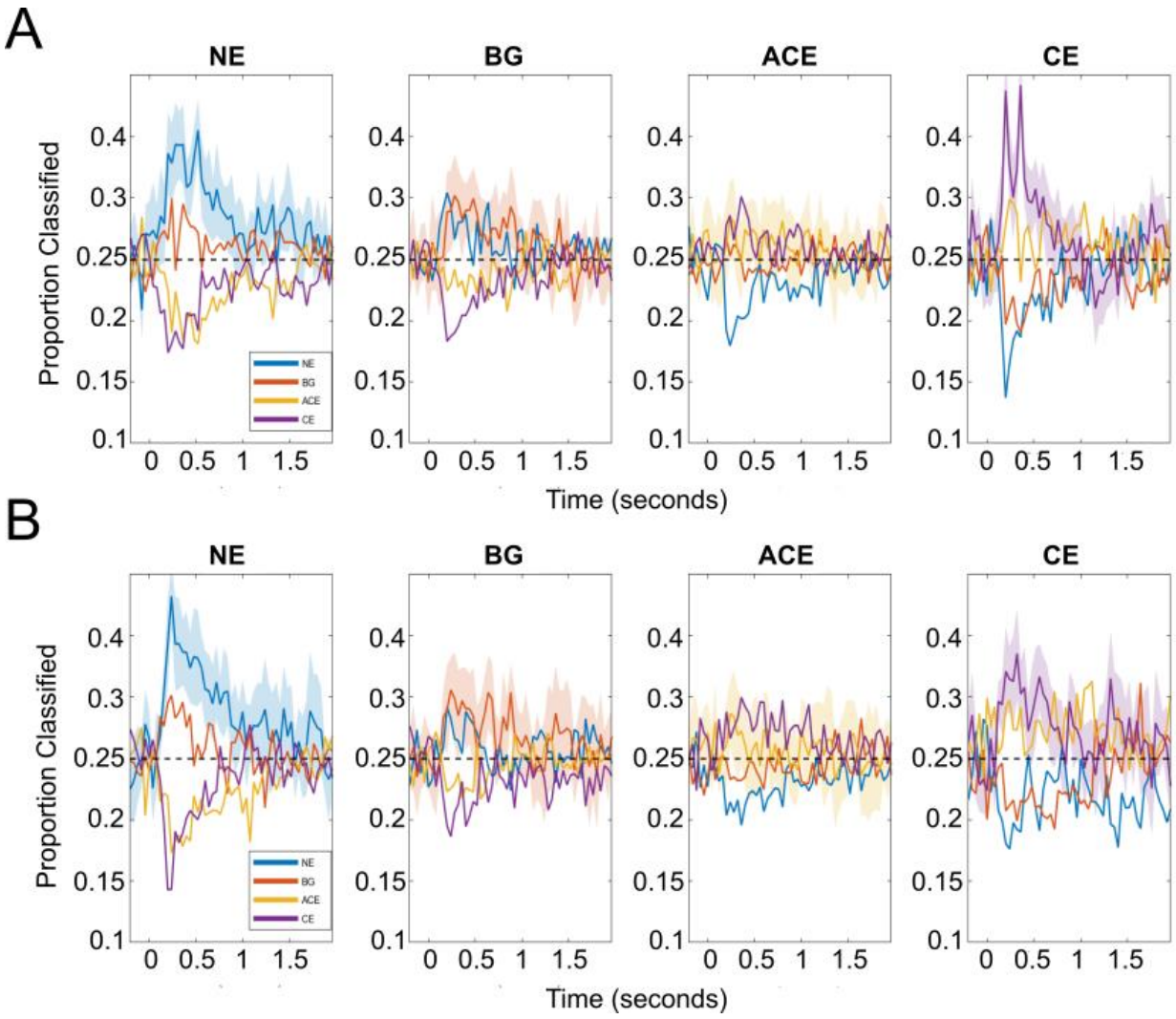
Supplementary Figure 2.1. Awareness Ratings Show Similar Representational Structure Across Occipital and Frontal Sensors. RSA analysis performed over occipital sensors (top) and frontal sensors (bottom) only. Purple, green, and gold lines represent similarity of the Abstract-Graded, Abstract-Discrete, and Specific-Graded models respectively with neural data. Solid horizontal lines represent time points significantly different from 0 for a specific RDM at $p < .05$, corrected for multiple comparisons. The Abstract-Graded model significantly predicted the neural data throughout the majority of the trial (purple line) across frontal sensors, and for a shorter duration when analysis was restricted occipital sensors only. The Abstract-Discrete model was only successful at predicting the neural data across two clusters of time-points post-stimulus when using occipital sensors, but was a significant predictor of neural data for larger portions of the epoch when using frontal sensors. The Specific-Graded model did not significantly predict the neural data at any time point in frontal or occipital sensors. Horizontal dots denote statistically significant paired comparisons between the different models at $p < .05$, corrected for multiple comparisons. Across frontal sensors, the Abstract-Graded model was a significantly better predictor of the neural data than the Specific-Graded model, and likewise across the occipital sensors, both abstract models significantly outperformed the Specific-Graded model. In this split-sensor analysis, the Abstract-Graded model did not significantly outperform the Abstract-Specific model, however there was a noticeable trend in the same direction as in the RSA performed across all sensors, where the difference in performance was significant (**Figure 2.3B**).



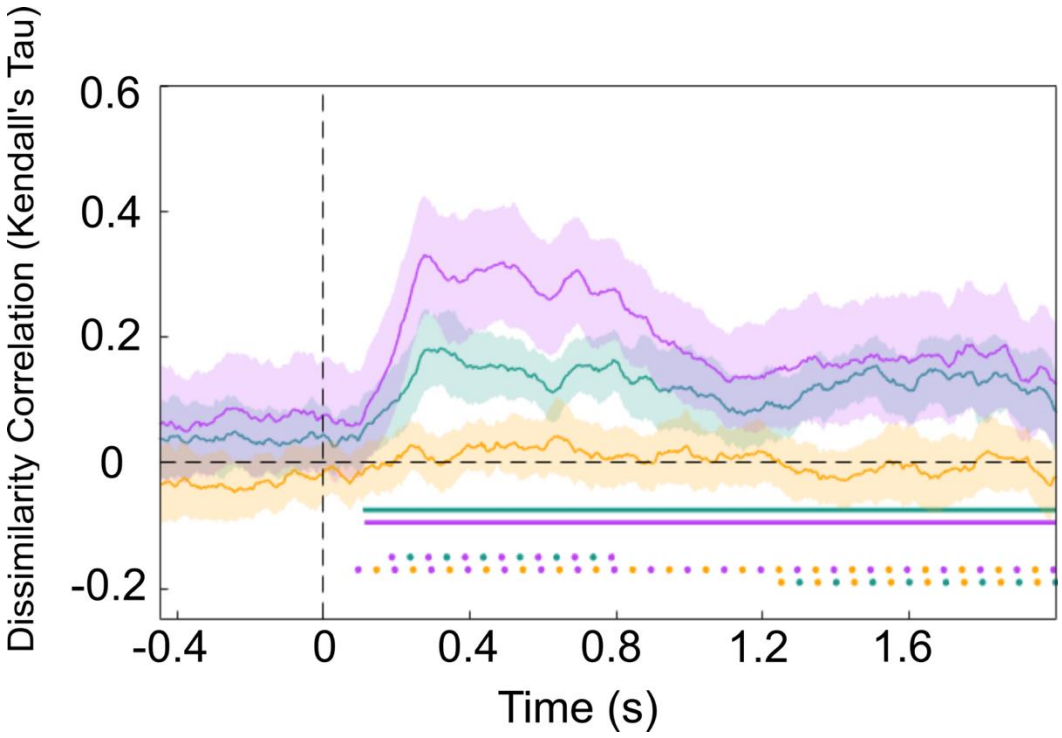
Supplementary Figure 2.2 RSA on MEG data with stimulus contrast level regressed out. When stimulus contrast level was regressed out of the MEG data, an RSA still produced comparable results to the original analysis. The Abstract-Graded model still predicted the neural data better than either alternative model. Solid horizontal lines represent time points significantly different from 0 for a specific RDM at $p < .05$, corrected for multiple comparisons. Horizontal dots denote statistically significant paired comparisons between the different models at $p < .05$, corrected for multiple comparisons.



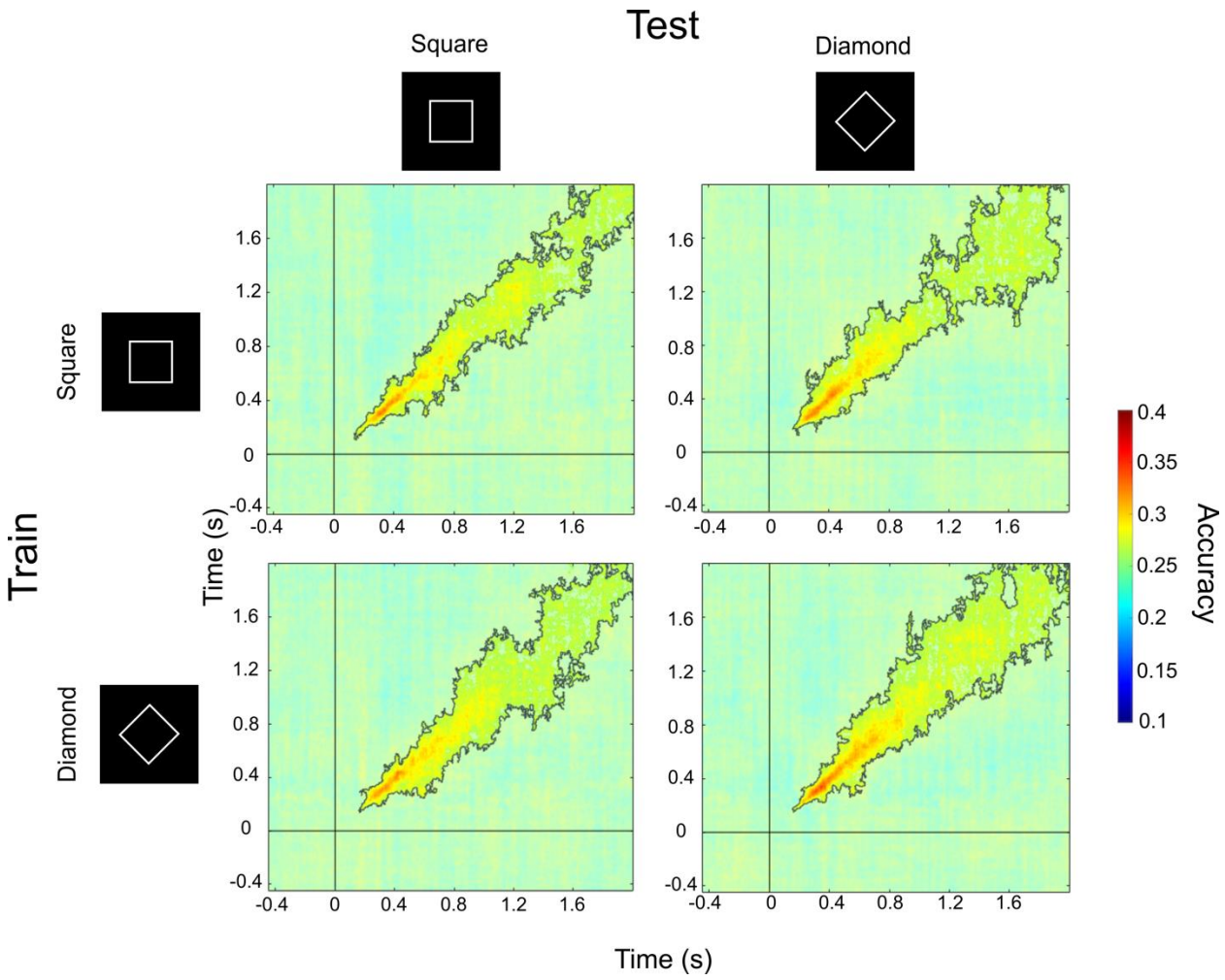
Supplementary Figure 2.3 Multidimensional scaling of neural activity covarying with awareness reports. A: Prior to stimulus contrast being regressed from the data, MEG activity covarying with awareness reports exhibit a linear trend from No Experience to Clear Experience along the first dimension, which tracks perceptual vividness. B: Following the removal of the linear component of stimulus contrast, the dimension which tracks perceptual vividness becomes compressed at the higher end, with ratings of "clear experience" (CE) and "almost clear experience" (ACE) becoming less distinct. Red squares illustrate ratings for squares, and blue diamonds illustrate ratings for diamonds. Data are averaged over the 100ms – 1000ms post-stimulus time window.



Supplementary Figure 2.4 Cross-decoding shows the graded nature of the PAS scale. For each cross-condition decoder (A: Train on Squares; B: Train on Diamonds), the four sub-plots illustrate the proportion of PAS ratings the classifiers decoded trials as. The subplots correspond to trials where the true PAS rating reported by subjects were (from left to right) 'No Experience', 'Weak Glimpse', 'Almost Clear Experience', and 'Clear Experience'. Each coloured line represents the proportion of trials classified as each PAS rating across time. For example, in trials where participants reported No Experience, the majority of trials were classified correctly (blue line), with the classifier most often misclassifying these 'No Experience' trials as 'Weak Glimpse' ratings (orange line), and rarely misclassifying them as 'Almost Clear Experience' or 'Clear Experience' trials (gold and purple lines, respectively). Shaded areas represent 95% confidence intervals.



Supplementary Figure 2.5 RSA analysis for MEG data without baseline correction. None of the model RDMs significantly predicted pre-stimulus activity in non-baseline-corrected data. Instead, the predominant neural signature was stimulus triggered, as in the main analysis (**Figure 2.3B**), with the Abstract-Graded model being the best predictor of neural representations of phenomenal magnitude.



Supplementary Figure 2.6 Temporal Generalisation matrices for MEG decoding analyses on data without baseline-correction. For each row, statistical comparisons between the two columns showed no significant differences in decoding accuracy between within and cross-condition decoding. Pre-stimulus decoding of awareness ratings was not possible, even when data had not been baseline-corrected.

ROI	Cluster Size (Number of Voxels)	MNI coordinates of Central Voxel (X, Y, Z)		
Visual	52	-21.9	-68	-15.7
Visual	51	-19.7	-82	-20.1
Visual	33	-35.9	-84.2	-20.1
Visual	19	-41.8	-54	-23.8
Visual	5	26	-85.7	-20.1
Frontal	60	-4.2	30.1	32.2
Frontal	51	-4.2	38.2	18.2
Frontal	44	8.3	26.7	35.9
Frontal	10	-21.9	32.3	38.1

Supplementary Table 2.1 Clusters within both the visual and frontal regions of interest. Used for fMRI ROI decoding analyses. Clusters smaller than 5 voxels not shown.

Decoding Type	Atlas Label	Cluster Size (voxels)	MNI Coordinates of maximum	Maximum accuracy
Cross (train animate)	Postcentral Gyrus	23,838	-46, -36, 54	0.593
Cross (train animate)	Temporal Fusiform Cortex	474	-31.5, -36.1, -25.8	0.567
Cross (train animate)	Paracingulate Gyrus	192	-8, 44, -6	0.563
Cross (train animate)	Frontal Orbital Cortex	168	-30, 34, -10	0.570
Cross (train animate)	Cingulate Gyrus, posterior division	109	4, -44, 10	0.563
Cross (train animate)	Occipital Pole	89	-22, -96, 2	0.562

Cross (train animate)	Frontal Pole	75	32, 36, -10	0.561
Cross (train animate)	Temporal Occipital Fusiform Cortex	59	20, -52, -18	0.566
Cross (train animate)	Lingual Gyrus	56	10, -38, -6	0.555
Cross (train animate)	Left Cerebral White Matter	56	-22, -38, 8	0.566
Cross (train inanimate)	Supramarginal Gyrus, anterior division	23,564	-54, -36, 30	0.593
Cross (train inanimate)	Subcallosal Cortex	512	5.85, 10.6, -6.63	0.575
Cross (train inanimate)	Left Caudate	176	-10, 2, 16	0.575
Cross (train inanimate)	Middle Temporal Gyrus, posterior division	166	-52, -42, -2	0.585
Cross (train inanimate)	Frontal Orbital Cortex	152	-32, 36, -12	0.570
Cross (train inanimate)	Left Hippocampus	106	-20, -16, -14	0.567
Cross (train inanimate)	Central Opercular Cortex	100	38, 0, 18	0.571
Cross (train inanimate)	Frontal Pole	78	22, 42, -12	0.565
Cross (train inanimate)	Insular Cortex	71	-38, -20, -4	0.564
Cross (train inanimate)	Lateral Occipital Cortex, superior division	62	32, -88, 14	0.555

Within (animate)	Cingulate Gyrus, anterior division	49,130	-2, 36, 20	0.616
Within (inanimate)	Postcentral Gyrus	32,710	62, -8, 18	0.601
Within (inanimate)	Right Caudate	79	18, 10, 16	0.565

Supplementary Table 2.2 fMRI Searchlight Decoding Results. Clusters with above chance decoding of perceptual visibility for both cross-condition and within-condition decoding. Clusters are significant at $p < .05$, corrected for multiple comparisons with an FDR of 0.01. Region names are found for the peak co-ordinate using the Harvard-Oxford Cortical and Subcortical Structural Atlas.

Decoding Type	**Atlas Label**	**Cluster Size (voxels)**	**MNI Coordinates of maximum**	**Maximum accuracy**
Content	Inferior Lateral Occipital Cortex	2768	-48, -70, -8	0.572
Content	Cerebellum	2150	0, -48, -10	0.558
Content	Inferior Frontal Gyrus	1837	-52, 32, 6	0.556
Content	Superior Lateral Occipital Cortex	919	-40, - 58, 54	0.557
Content	Superior Lateral Occipital Cortex	907	-8, -64, 60	0.554
Content	Angular Gyrus	735	44, -52, 20	0.557
Content	Superior Parietal Lobule	626	20, -44, 62	0.548
Content	Frontal Orbital Cortex	375	38, 24, -2	0.557

Content	Paracingulate Cyrus	371	-6, 26, 34	0.547
Content	Inferior Frontal Gyrus	356	46, 20, 20	0.563
Content	Superior Frontal Gyrus	209	16, 24, 58	0.546
Content	Frontal Pole	201	-24, 42, 44	0.546
Content	Superior Temporal Gyrus	192	-54, -2, -10	0.549
Content	Postcentral Gyrus	189	-62, -22, 38	0.547
Content	Posterior Cingulate Gyrus	178	-8, -26, 44	0.547
Content	Left Cerebral White Matter	133	6, -24, 16	0.544
Content	Cerebellum	98	34, -64, -40	0.547
Content	Superior Lateral Occipital Cortex	76	20, -66, 46	0.553
Content	Anterior Cingulate Gyrus	74	6, -4, 40	0.551
Content	Parietal Operculum Cortex	56	44, -32, 20	0.546
Content	Precentral Gyrus	52	44, -16, 58	0.542
Mean Cross-Condition Visibility	Precentral Gyrus	16,638	-50, 8 32	0.584

Mean Cross-Condition Visibility	Postcentral Gyrus	2863	-50, -30, 48	0.58
Mean Cross-Condition Visibility	Right Accumbens	278	10, 14, -6	0.561
Mean Cross-Condition Visibility	Left Cerebral White Matter	167	-16, 0, 16	0.568
Mean Cross-Condition Visibility	Frontal Medial Cortex	166	-4, 48, -10	0.555
Mean Cross-Condition Visibility	Middle Temporal Gyrus	150	-52, -42, -2	0.556
Mean Cross-Condition Visibility	Central Opercular Cortex	138	-36, -10, 18	0.563
Mean Cross-Condition Visibility	Frontal Orbital Cortex	133	32, 28, 2	0.556
Mean Cross-Condition Visibility	Occipital Fusiform Gyrus	116	14, -84, -22	0.557
Mean Cross-Condition Visibility	Cerebellum	103	-28, -42, -34	0.559
Mean Cross-Condition Visibility	Frontal Orbital Cortex	95	-36, 36, -10	0.563
Mean Cross-Condition Visibility	Temporal Occipital Fusiform Cortex	91	-40, -62, -24	0.554
Mean Cross-Condition Visibility	Inferior Lateral Occipital Cortex	58	42, -80, 4	0.553
Mean Cross-Condition Visibility	Frontal Pole	58	22, 38, -12	0.557
Mean Cross-Condition Visibility	Occipital Pole	56	12, -90, -2	0.557

Mean Cross-Condition Visibility	Superior Precentral Gyrus	53	26, -12, 52	0.552
Visibility and Content Overlap	Cerebellum	1587	-6, -82, -32	N/A
Visibility and Content Overlap	Frontal Orbital Cortex	1521	-28, 16, -24	N/A
Visibility and Content Overlap	Superior Lateral Occipital Cortex	782	-32, -74, 20	N/A
Visibility and Content Overlap	Precuneous Cortex	474	-14, -58, 18	N/A
Visibility and Content Overlap	Inferior Lateral Occipital Cortex	412	46, -68, -10	N/A
Visibility and Content Overlap	Supramarginal Gyrus	383	64, -28, 34	N/A
Visibility and Content Overlap	Frontal Pole	339	40, 36, 12	N/A
Visibility and Content Overlap	Cingulate Gyrus	315	0, 36, 16	N/A
Visibility and Content Overlap	Frontal Orbital Cortex	213	36, 22, -14	N/A
Visibility and Content Overlap	Middle Frontal Gyrus	181	-24, 36, 32	N/A
Visibility and Content Overlap	Posterior Cingulate Gyrus	156	-4, -30, 36	N/A
Visibility and Content Overlap	Middle Frontal Gyrus	147	24, 28, 34	N/A
Visibility and Content Overlap	Right Cerebral Cortex	145	20, 2, -14	N/A

Visibility and Content Overlap	Cerebellum	115	26, -58, -26	N/A
Visibility and Content Overlap	Cerebellum	111	-24, -36, -38	N/A
Visibility and Content Overlap	Superior Parietal Lobule	107	34, -50, 36	N/A
Visibility and Content Overlap	Postcentral Gyrus	101	-60, -20, 24	N/A
Visibility and Content Overlap	Postcentral Gyrus	74	-36, -22, 44	N/A
Visibility and Content Overlap	Lingual Gyrus	74	-14, -56, -2	N/A
Visibility and Content Overlap	Cerebellum	63	20, -72, -32	N/A
Visibility and Content Overlap	Cerebellum	52	34, -56, -48	N/A

Supplementary Table 2.3 fMRI Searchlight Content Decoding in High Visibility Trials, Average Cross-Condition Visibility Decoding, and Overlapping Cluster Information. Decoding type = Content: Clusters with above chance decoding of perceptual content (animate vs. inanimate) in high visibility trials. Decoding type = Mean Cross-Condition Visibility: Clusters with above chance decoding of perceptual visibility for cross-condition decoding when accuracy from both decoding directions (training on animate and training on inanimate) was averaged together. To aid the identification of individual clusters in this map, clustering was performed at an increased accuracy threshold of 0.54. Decoding Type = Awareness and Content Overlap: Clusters where decoding of content and cross-decoding of visibility were both successful (i.e. the intersection of the Content and Mean Cross-Condition Visibility clusters). Clusters are significant at $p < .05$, corrected for multiple comparisons with an FDR of 0.01. Region names are found for the peak co-ordinate using the Harvard-Oxford Cortical and Subcortical Structural Atlas and the MNI Structural Atlas.

3. Creating something out of nothing: Symbolic and non-symbolic representations of numerical zero in the human brain

3.1 Introduction

Sparse higher-order theories of consciousness (e.g., Fleming, 2020; Lau, 2019) make the claim that low-dimensional neural codes tracking the reliability of perceptual states may underlie awareness judgements. Chapter 2 provided initial evidence for such a neural code, which abstracts over the content of perception and encodes only the vividness of perceptual experiences. According to these theories, and particularly the HOSS model (Fleming, 2020), such codes should extend to capturing neural activity related to the absence of experience, which is described in HOSS as a higher-order mechanism inferring an absence of perceptual input. Most importantly with respect to absence, the HOSS model describes how the brain should actively represent the absence of experience, rather than simply not representing the presence of sensory stimulation (Fleming, 2020). One way of exploring this aspect of the HOSS model is to examine active representations of absence in other cognitive domains with a view to informing theories of perceptual absences. This chapter follows this method. It offers the first

characterisation of the neural representation of numerical absence (i.e., the number zero) in the human brain and, in addition, begins to explore ideas that suggest perceptual experiences of absence may provide the scaffolding through which more complex absence-related concepts are developed (Nieder, 2016).

The number zero plays a central role in science, mathematics, and human culture (Kaplan, 1999; Nieder, 2016) and its symbolic representation is considered a unique property of abstract human thought (Bialystok & Codd, 2000; Nieder, 2016). The psychological basis of zero is unusual: while natural numbers correspond to the observable number of countable items within a set (e.g., one bird; three clouds), an empty set does not contain any countable elements. To conceptualise zero, one must instead abstract away from the (absence of) sensory evidence to construct a representation of numerical absence: creating 'something' out of 'nothing' (Butterworth, 1999; Nieder, 2016; Wellman & Miller, 1986). Given these differences, it remains an open question as to how zero is represented in relation to other numbers.

In contrast to zero, the neural representation of natural numbers is better understood. Distinct neural populations are selective for specific numerosities, exhibiting overlapping tuning curves with neighbouring populations tuned to adjacent numerosities (Kutter et al., 2018; Piazza et al., 2004). This architecture underpins a so-called distance effect (Dehaene et al., 1998), where numbers close together in numerical space have similar neural representations. For instance, neural responses to numbers one and two are more similar than neural responses to one and ten (Borghesani et al., 2019; Luyckx et al., 2019; Piazza et al., 2004). Importantly, a component of this neural code is thought to be invariant to numerical format (Damarla et al., 2016; Eger et al., 2003, 2009; Piazza et al., 2007; Teichmann et al., 2018) such that, for example, neural representations of 'six' are shared across symbolic and non-symbolic formats (e.g., both the Arabic numeral '6' and six dots; although see (Cohen Kadosh et al., 2007)). In humans, these format-invariant representations of numerical magnitude have been localised to the parietal cortex (Damarla et al., 2016; Eger et al., 2009; Piazza et al., 2007), with topographic maps

underpinning numerosity perception found more broadly across association cortex (Harvey et al., 2013; Harvey & Dumoulin, 2017).

Although behavioural evidence suggests that zero occupies a place at the beginning of this mental number line (Dehaene et al., 1993; Pinhas & Tzelgov, 2012; Zagury et al., 2022), zero is also associated with unique behavioural and developmental profiles compared to natural numbers. For instance, the reading times of human adults are increased for zero compared to non-zero numbers (Brysbaert, 1995) and zero concepts emerge later in children than those for natural numbers (Krajcsi et al., 2021; Merritt & Brannon, 2013; Wellman & Miller, 1986). Distinct behavioural characteristics associated with zero are not surprising given the heightened degree of abstraction required to conceptualise numerical absence. In turn, it is plausible that neural representations of zero are distinct to the scheme that has been discovered for natural numbers (Schubert et al., 2020). Initial research in non-human animals has indicated that numerical zero shares some neural properties with natural numerosities, such as overlapping tuning curves and associated distance effects, along with invariance to particular stimulus properties (Kirschhock et al., 2021; Okuyama et al., 2015; Ramirez-Cardenas et al., 2016). Moreover, behavioural evidence for zero representations has been reported in a number of animals, including macaque monkeys (Merritt et al., 2009), African grey parrots (Pepperberg & Gordon, 2005), and honeybees (Howard et al., 2018). However, it remains unknown whether the symbolic, human conceptualisation of numerical zero, which emerges later than natural numbers in human culture (Ifrah, 1985; Kaplan, 1999), engenders representations of zero that are both distinct from other numbers and which studies in non-human animals may have failed to reveal.

I tackled this question by employing two qualitatively different numerical tasks in humans while leveraging methodological advances to reveal the representational content of neural responses to numerical stimuli in magnetoencephalography (MEG) data (Kriegeskorte & Diedrichsen, 2019; Luyckx et al., 2019). The choice of tasks was guided by previous work examining both non-symbolic (Ramirez-Cardenas et al., 2016) and symbolic (Luyckx et al., 2019) numerosity representations. Importantly, the use of two distinct tasks required

participants to adopt distinct mathematical attitudes towards zero, ensuring that any commonalities between symbolic and non-symbolic neural representations of zero were not confounded by task-related processing. I assay both neural representations of non-symbolic numerosities (dot patterns), including zero (empty sets), and symbolic numerals, including symbolic zero. Numerosities ranged from zero to five, allowing me to examine the fine-grained representations of numbers close to zero. My results reveal that neural representations of zero are situated along a graded neural number line shared with other natural numbers. Notably, symbolic representations of zero generalised to predict non-symbolic empty sets. I go on to localise abstract representations of numerical zero to posterior association cortex, extending the purview of parietal cortex in human numerical cognition to encompass representations of zero (Harvey & Dumoulin, 2017; Piazza et al., 2007).

3.2 Materials and Methods

3.2.1 Study Participant Details

Twenty-nine participants (M_{age}: 29.27 years, SD_{age}: 10.69) took part in the MEG experiment at the Wellcome Centre for Human Neuroimaging, University College London. Five participants either failed to follow task instructions (chance performance on one or more tasks) or did not complete the experiment and were therefore excluded from analysis. All analysis was performed on the remaining sample of 24 participants. Informed consent was given before the experiment and ethical approval was granted by the Research Ethics Committee of University College London (#1825/005).

3.2.2 Stimuli

Numerical dot stimuli were created using custom MATLAB 2021b (Mathworks) scripts and consisted of different numbers of dots (from zero to five) on grey backgrounds (**Figure 3.1E**). There were two sets of dot stimuli, a standard set and a control set. In the

standard set, dot size was pseudorandomly specified, and each dot was pseudorandomly located around the centre of the screen. In the control set, low level visual properties of the stimuli (total dot area, density, luminance) were constant across numerosities. Total dot area was controlled for by systematically reducing the size of the dots as the number of dots increased, such that the total number of pixels included in a stimulus were constant across numerosities. To control for density, dots were located within an invisible circle, with the radius of the circle determined by the number of dots. Larger numerosities had larger circles, thereby ensuring that patterns with more dots were not systematically denser. Finally, in both stimulus sets, 50% of dots were black and 50% were white, such that the contrast of the dot patterns did not increase with numerosity (the increase in contrast generated by an increasing number of black dots was cancelled out by the increased number of white dots). **Supplementary Figure 3.1** reports the correlation between these non-numerical features and numerosity in the controlled stimulus set, highlighting how the stimulus-generating procedure successfully controlled for the association between low-level visual properties and numerosity. Empty set stimuli contained only a grey background in both stimulus sets.

To help prevent participants relying on low level visual cues in identifying empty set stimuli, the background luminance was varied within and across stimulus sets and the background square size was randomly varied across all stimuli. The two stimulus sets allowed me to test whether numerical representations generalised across controlled and uncontrolled stimulus sets, which if successful indicates that numerosity representations are not merely picking up on low-level stimulus features correlated with number. Indeed, a control analysis confirmed that numerical information was extracted from the stimuli independently from physical features (see Representational Similarity Analysis; **Supplementary Figure 3.2**). As such, both standard and control stimulus sets were included in the analysis of the non-symbolic data.

3.2.3 Experimental Procedure

The tasks were presented to subjects using MATLAB (Mathworks) and the Psychophysics Toolbox (Brainard, 1997; Kleiner et al., 2007). Participants practiced the tasks on a computer before the MEG session. In the MEG scanner, the tasks were performed in alternating miniblocks with 35 symbolic trials and 54 non-symbolic trials per MEG recording block. The order of the tasks would swap on each block, and the starting order was counterbalanced across participants. There were 9 MEG blocks in total, resulting in 315 symbolic numeral trials and 486 non-symbolic dot trials across the whole experiment. Participants responded using two buttons on a button box and their right thumb.

3.2.3.1 Non-symbolic Task

Participants performed a match to sample task on dot stimuli (Kirschhock et al., 2021; Ramirez-Cardenas et al., 2016). On each trial, participants saw a sample image containing between zero and five dots for 250ms followed by a fixation cross for 800ms. A test image, also containing between zero and five dots, was then presented for 250ms, followed before another 800ms fixation period (**Figure 3.1A**). Within a trial, a single stimulus set was used for both the sample and test image. Participants reported whether the number of dots in the test stimulus matched that of the sample stimulus, or not. The response was followed by feedback in the form of a coloured rectangle surrounding the response options, with green and red used to indicate correct and incorrect answers, respectively. Response options were positioned randomly on each trial to eliminate any correlation between the decision and motor response. Intertrial intervals were also sampled randomly from a uniform distribution between 500-1000ms.

3.2.3.2 Symbolic Task

I adapted the symbolic numeral averaging task introduced by (Luyckx et al., 2019) to include the number zero. In one trial, ten numerals ranging from zero to five were

presented in a random order (**Figure 3.1B**). Five of the numerals were blue and five were orange. Each numeral was displayed for 250ms with an interstimulus interval of 100ms. The numerals were randomly selected on each trial to obey the constraint that the mean of the blue numerals could not equal the mean of the orange numerals. The response required at the end of each trial was counterbalanced across subjects, with half of the subjects reporting which set of numerals (orange or blue) had the highest average, and the other half reporting the set with the lowest average. Participants had 2000ms to respond, after which they were given feedback in the form of a green (correct) or red (incorrect) rectangle surrounding the response options. Again, to disentangle participants' decisions from motor responses, response options were positioned randomly on each trial. Intertrial intervals were randomly sampled from a uniform distribution between 500-1000ms.

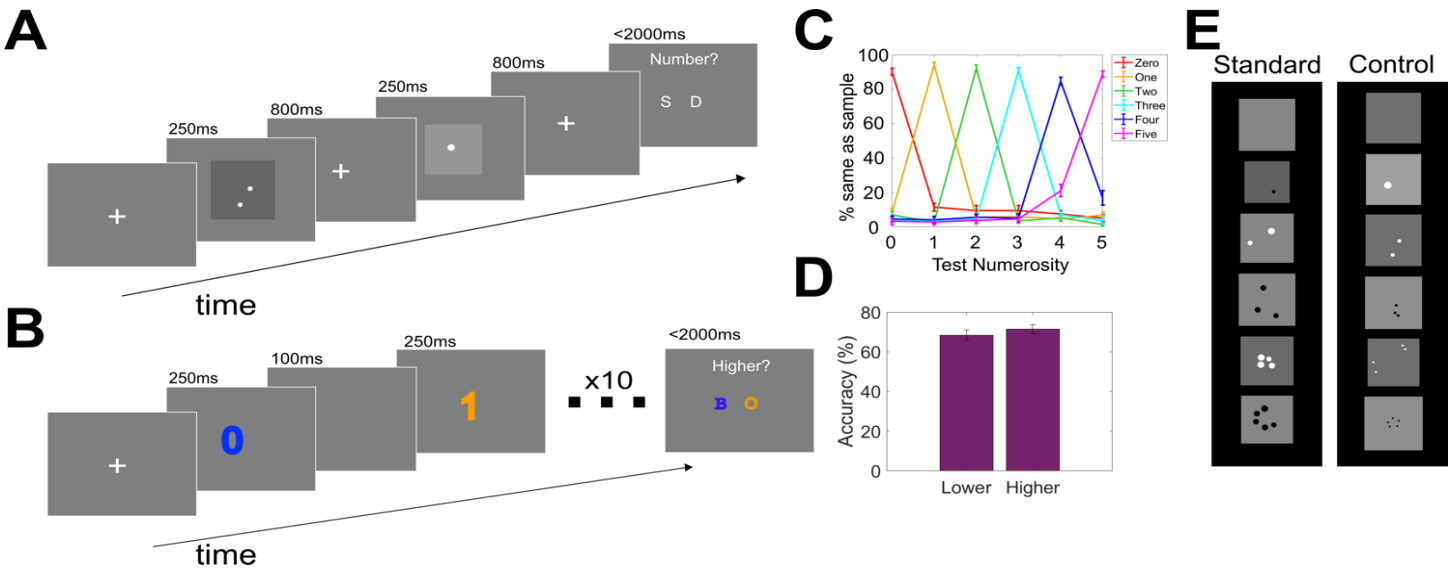


Figure 3.1. Experimental Procedure. A. Trial structure for the non-symbolic match to sample task. Participants observed a sample dot pattern followed by a test dot pattern before reporting whether the two patterns had the same or different numbers of dots. B. Trial structure for the symbolic averaging task. Participants observed a sequence of blue and orange numerals before reporting which set of numerals had the higher or lower average. C. Behavioural tuning curves in the non-symbolic task. Each curve reflects the percentage of trials that participants judged the test numerosity to be the same as the sample numerosity. Each colour represents trials with specific sample numerosities. The peak of each curve illustrates correct performance when the sample and test numerosities matched. Data points either side of the peak represent non-match trials. Error bars indicate SEM. D. Accuracy in the symbolic task split across participants who judged which set of numbers was higher, and those who judged which was lower. E. Stimulus sets for non-symbolic task. Dot size was pseudorandomised in the standard set, while low level properties of the dots including total dot area, density, and luminance were held constant in the control set. Across both sets, frame size of the dot patterns was randomly varied, to limit reliance on visual cues when identifying empty sets.

3.2.4 MEG Preprocessing

MEG data were analysed using FieldTrip (Oostenveld et al., 2011). MEG was recorded continuously at 600Hz using a 273-channel axial gradiometer system (CTF Omega, VSM MedTech) while participants sat upright inside the scanner. To remove line noise, the raw MEG data were preprocessed with a Discrete Fourier Transform and bandstop filter at 50Hz and its harmonics. The numeral task was segmented into epochs of -500ms to 4000ms relative to trial onset. For the dot task, the segments were from -200ms to 2500ms. Baseline correction was performed where, for each trial, activity in the pre-trial window was averaged and subtracted from the entire epoch per channel. The data were downsampled to 300Hz to conserve processing time and improve signal to noise ratio. During artefact rejection, trials with high kurtosis were visually inspected and removed if they were judged to contain excessive artefacts. To assist in removing eye-movement artefacts, an independent components analysis was carried out on the MEG data, and the components with the highest correlation with eye-tracking data were discarded after visual inspection. Components showing topographic and temporal signatures typically associated with cardiac artefacts were also removed by eye. This procedure was performed separately for the numeral and dot task. Finally, a second stage of epoching was performed to generate trials of individual numerosities. In the numeral task, trials were segmented into -100ms to 800ms epochs around each numeral onset. Trials were then baseline corrected again using the pre-stimulus window. In the dot task, trials were segmented into two different -200ms to 800ms epochs with respect to the onsets of the sample and test stimuli. All analyses used the sample images only. Finally, all analyses focusing on shared representations across notational formats were performed on the shared timepoints of -100ms to 800ms relative to stimulus presentation.

3.2.5 Representational Similarity Analysis

Representational Similarity Analysis (RSA) allows us to test specific hypothesis about how neural representations are structured (Kriegeskorte & Kievit, 2013). Here, I tested for the existence of a distance effect across numerosities. To do this, I defined a model

representational dissimilarity matrix (RDM) that describes the dissimilarity of two numerosities as a function of their numerical distance. To compare this model dissimilarity matrix with the neural data I first created a neural dissimilarity matrix that represents the similarity in neural patterns associated with each numerosity. To do this, I first ran a linear regression on the MEG data with dummy coded predictors for each of the six numerosities (trial numerosity coded with a 1, alternative numerosities coded with a 0). This produced a coefficient weight for each numerosity at each time point and sensor. These weights were then combined into a vector, representing the multivariate neural response for each numerosity, averaged over trials. To create the neural RDM, I computed the Pearson distance between each pair of condition weights over sensors, resulting in a 6x6 neural RDM reflecting the pairwise similarity of neural patterns associated with each numerosity. These neural RDMs were smoothed over time via convolution with a 60ms uniform kernel. To compare the neural and model RDMs, at every time point I correlated the lower triangle of each matrix (excluding the diagonal) using Kendall's Tau rank correlation (Nili et al., 2014).

Cross-task RSA was performed in the same manner, except here there were 12 predictors in the linear regression (0-5 symbolic, 0-5 non-symbolic). This resulted in a 12x12 neural RDM, of which I used the quadrant representing the cross-task pairwise similarities between numerosities when comparing with the model RDM. The whole quadrant including the diagonal was used in this analysis. This is because here the diagonal does not contain redundant information, but rather the similarity of the same numerosity across two different formats, and cells in the upper triangle represent different pairwise similarities to those in the lower triangle.

Finally, to test whether numerical information was decodable from non-symbolic stimuli over and above the physical features of the stimuli, I ran a cross-stimulus set RSA in the same manner as above, except now I tested exclusively within the non-symbolic task. As such, the 12 predictors were: 0-5 from the standard set and 0-5 from the control set. This RSA established whether representations of numerical magnitude generalised across

stimulus set, and therefore went beyond information that could be derived solely from physical features of the stimuli.

3.2.6 Decoding Analyses

To examine the representational structure of the number zero more specifically across symbolic and non-symbolic formats, I employed different decoding techniques using both multiclass and binary decoders. First, to reveal the temporal profile of numerosity representations, I trained a multiclass Linear Discriminant Analysis (LDA) decoder to decode numerosities zero to five. This was performed in a temporal generalisation procedure, whereby the classifier was trained on each time point and tested on all other time points (King & Dehaene, 2014). This process results in a train time x test time decoding accuracy matrix, which illustrates how stable representations of numerosity are over time.

I performed both within-format and cross-format decoding procedures. Within-format decoding involved training and testing a classifier to identify numerosities on trials from one format (e.g. numerals or dots). In cross-format decoding, I trained the classifier on one format and tested it on the other (e.g., training on symbolic trials and testing on non-symbolic trials, and vice versa). For the within-format approach, I implemented a 5-fold cross-validation strategy. Prior to decoding, five trials per numerosity were averaged and the resulting average trials was balanced per numerosity. It is worth noting that cross-validation is not required in cross-format decoding because the test data is never seen by the classifier during training, and thus there is no risk of overfitting. Cross-format decoding allows us to empirically assess whether the neural patterns associated with numerals share a common neural code across formats.

To complement the RSA analyses and isolate the representational structure underpinning numerical zero specifically, I extracted the confusion matrices from the decoders. Confusion matrices indicate how often different stimulus classes (i.e., numerosities) are confused for one another, and this information can be used to infer the organisation of

neural representations. For example, a decoder that confuses zero with the number one more than the number two displays evidence for a numerical distance effect. The data used to train the decoders from which these confusion matrices were extracted was time-averaged over the timepoints where the initial multiclass decoder could decode numerosity significantly above chance (non-symbolic: 70ms – 800ms, symbolic 56.7ms – 800ms). I also computed confusion matrices across time.

To examine whether representations of zero could reliably be dissociated from numerosities presented in the alternative format, I created a decoding procedure using a binary LDA classifier to decode zero vs. non-zero numerosities. Within this training regime, the number of trials per non-zero numerosities was kept equal, and the number of zero trials vs. non-zero numerosity trials was also balanced. The resulting 'zero' decoder was uniquely trained to identify neural representations of numerical zero in symbolic or non-symbolic format and was tested on the other format to identify format-invariant representations of zero.

Finally, to reveal whether abstract representations of numerical zero exist on a graded number line, or whether they are categorically distinct from other numbers, I ran a new cross-format decoding analysis using binary classifiers. Here, I trained the decoders to discriminate zero vs. all non-zero numerosities (one to five) separately, and then tested these binary decoders on the corresponding numerosities in the opposite format. This resulted in five different classifiers per format. Specifically, I trained five different decoders to dissociate: symbolic zero vs symbolic one, symbolic zero vs symbolic two, symbolic zero vs symbolic three, symbolic zero vs symbolic four, and symbolic zero vs symbolic five. I then tested these decoders on empty sets vs one dot, empty sets vs two dots, empty sets vs three dots, empty sets vs four dots, and empty sets vs five dots, respectively. This was also done in the reverse direction: training on non-symbolic trials and testing on symbolic numerals. I used the area under the receiver operating characteristic (AUROC) as a metric for discriminability between each pair of classes. In line with the hypothesis that format-invariant representations of zero exist on a graded, abstract neural number line, I expected the discriminability to improve as the numerical

distance from zero increased. To statistically test whether this was the case, I performed one-tailed, paired comparisons between the discriminability of successive numbers with zero (e.g., by comparing 0-2 vs. 0-1, 0-3 vs. 0-2, etc.).

For all decoding analyses, I utilized multiclass or binary LDA decoders in conjunction with the MVPA-light toolbox (Treder, 2020) integrated with FieldTrip. To improve the robustness of the classifier, I applied L1-regularization to the covariance matrix, and the shrinkage parameter was automatically determined using the Ledoit-Wolf formula within each training fold (Ledoit & Wolf, 2004).

3.2.7 Source Reconstruction

Both FieldTrip's template single shell head model and its standard volumetric grid (8mm resolution) were warped to participants' individual fiducial points, generating a subject-specific forward model aligned in MNI space. Source reconstruction was performed using a linearly constrained minimum variance (lcmv) beamformer(Van Veen et al., 1997) which applies spatial filters to the MEG data to generate source-level time courses. To reduce the impact of noise on the source estimates, I used a regularisation parameter of lambda = 5%. For each task, spatial filters were calculated by combining the leadfield matrix with the data covariance matrix across all numerosities and the timepoints coinciding with the stable cluster of significantly above-chance decoding in the zero vs. non-zero cross-task classifier (100 – 450ms). These spatial filters were then applied to zero trials and non-zero trials separately, generating reconstructed maps of source activity for these two trial types. I contrasted the broadband source power of zero > non-zero trials in a mass-univariate procedure across subjects for each task separately with an alpha parameter of $p < .05$, corrected for multiple comparisons. For binary LDA classifiers, this is equivalent to localising the classifier weights (Haufe et al., 2014), and therefore gives an indication of which brain regions drove the decoding results. I computed the conjunction of these two contrasts, revealing the voxels where zero stimuli were dissociable from other numbers in both symbolic and non-symbolic formats.

Multidimensional scaling of source space activity was performed using the same beamforming parameters to calculate spatial filters over combined non-symbolic and symbolic trials. Using these filters, virtual channels were created for each source location within the map defined by the conjunction analysis. The virtual channels were then used to create a cross-task representational dissimilarity matrix in the same manner as described for the cross-task RSA sensor-level analysis. This was then submitted to MATLAB's cmdscale function for multidimensional scaling.

3.2.8 Statistical Inference

Across sensor and source level analyses, cluster-based permutation testing was used to statistically test hypotheses and correct for multiple comparisons (Maris & Oostenveld, 2007). For all analyses (decoding, RSA, and source-level contrasts), 1000 permutations were used with cluster-forming alpha parameter of .05 and a significance threshold of .05. It is important to emphasize that this cluster-based permutation testing approach does not provide information about when neural representations emerge. This limitation arises because the statistical inference process does not focus on individual time points; instead, it relies on cluster-level statistics that encompass multiple time points (Sassenhagen & Draschkow, 2019).

3.3 Results

Twenty-nine human participants (24 after exclusions; see Methods for details) took part in a magnetoencephalography (MEG) experiment involving two numerical tasks. The first was a non-symbolic match-to-sample task (**Figure 3.1A**) where participants observed two sequentially presented dot patterns that ranged in number from zero dots (empty set) to five dots (Ramirez-Cardenas et al., 2016). Participants were asked to report whether the patterns contained the same or different number of dots. I employed two sets of dot patterns: a standard set which randomised the size of dots within each pattern, and a control set which kept total dot area, density, and luminance constant across numerosities

(**Figure 3.1E**). To ensure participants could not rely on low level visual cues in identifying empty set stimuli, the background luminance was varied within and across stimulus sets, the background square size was randomly varied across all stimuli, and 50% of dots were of opposite contrast (white rather than black). The second task was a symbolic averaging task (Luyckx et al., 2019) (**Figure 3.1B**). Here, participants observed a rapid serial presentation of 10 symbolic numerals from zero to five (0, 1, 2, 3, 4, 5), divided into orange and blue sets (5 numbers in each). Participants were asked to report the set of numbers with the higher or lower average. Decision type (higher or lower) was counterbalanced across participants. Employing different tasks per each notational format with different task requirements and decision types also ensured neural patterns induced by the perception of zero are unlikely to be driven by specific task features or calculation requirements. All analyses were exploratory and were not pre-registered prior to data collection.

In the non-symbolic match-to-sample task, participants accurately determined whether dot patterns had the same or different numbers of dots ($Mean_{accuracy}$: 0.92, *SE*: 0.16). Plotting behavioural tuning curves revealed near-ceiling performance across all numerosities (**Figure 3.1C**), with the exception of five-dot patterns which were more often confused with four-dot patterns than three-dot patterns ($t(23) = 4.97$, $p < .001$) – consistent with numerosity tuning curves becoming wider as number increases (Dehaene, Dehaene-Lambertz, et al., 1998). In the symbolic task, participants could reliably perform the task regardless of whether they were reporting the higher ($Mean_{accuracy}$: 0.71, *SE*: 0.23) or lower ($Mean_{accuracy}$: 0.68, *SE*: 0.27) average (**Figure 3.1D**), and there was no difference between performance across decision types ($t(22) = -0.88$, $p = 0.39$). As expected, performance was significantly higher in the non-symbolic match-to-sample task compared to the symbolic averaging task ($t(23) = 8.82$, $p < .001$).

3.3.1 Identifying Neural Representations of Number

I next asked whether neural patterns recorded by MEG were sensitive to numerosity, by timelocking the data to the presentation of the dot pattern/symbolic numeral stimuli.

Multiclass decoders were trained to classify different numerosities (zero to five) in both the non-symbolic and symbolic tasks. The frequency with which the decoders confused numerosities for one another is illustrated in **Figure 3.2A**. Here, individual panels represent trials where a particular numerosity was presented to the classifier, and the coloured lines indicate the proportion of those trials where the classifier predicted each one of the possible classes (zero to five) over the trial epoch. For example, the 'NS-one' panel shows that when one dot is presented in the non-symbolic task, the classifier predominantly and correctly labels this stimulus as numerosity one (yellow curve), with the next most likely error being a misclassification as the number two (green curve). Across all numbers and both formats, the classifiers successfully predicted the numerosity participants were viewing from their neural data, including zero numerosities (time-points where classifier significantly exceed chance level: non-symbolic: 70ms – 800ms; symbolic: 56.7ms – 800ms).

I next leveraged temporal generalisation analysis to ask whether numerosity representations were stable over time (King & Dehaene, 2014). When training and testing on all combinations of time points, stable time-windows where numerical information could be decoded above chance level were identified in both tasks from shortly after stimulus presentation up until the end of the analysed time window **Figure 3.2B**. This analysis was also used to generate **Figure 3.2A**, such that time-points at which classifiers exceed chance level are identical (non-symbolic: 70ms – 800ms; symbolic: 56.7ms – 800ms). These time windows in which stable numerosity representations were identified were used to create time-averaged data for use in subsequent population tuning curve (**Figure 3.2D**) and multidimensional scaling (**Figure 3.2E**) analyses.

3.3.2 A Neural Number Line from Zero to Five

A fundamental feature of neural codes for natural numbers is a distance effect, whereby numbers closer together in numerical space are closer together in representational space (Dehaene et al., 1998; Nieder & Dehaene, 2009). Here I asked whether numerical zero exhibits similar distance effects with other numbers, consistent with it sharing a neural

number line with countable numerosities. A Representational Similarity Matrix (RDM) describing a distance effect from zero to five successfully predicted neural data across both non-symbolic and symbolic numerical formats (**Figure 3.2C**). In the non-symbolic task, an RDM generalising numerical information across the two non-symbolic stimulus sets significantly predicted neural responses throughout the trial, indicating that neural correlates of number were independent of the physical properties of the dot stimuli (**Supplementary Figure 3.2**). Multidimensional scaling of neural representations of numerosity in turn illustrates a distance effect (**Figure 3.2E**), with the numbers zero to five occupying positions along a single, ordered dimension, while a second dimension loosely distinguished intermediate numerosities (one to four) from the extremes (zero and five).

A stronger test of a distance effect in neural data is furnished by examining the confusability between neighbouring numerosities using population tuning curves (**Figure 3.2D**). These plots are time-averaged versions of the classifier confusion matrices in **Figure 3.2A**, i.e., the proportion of trials where the classifier predicted a particular numerosity as a function of the true numerosity within the time window in which numerical information could be reliably decoded (**Figure 3.2B**). For example, the red curve in **Figure 3.2D** indicates that the proportion of trials predicted as being zero peaks when the numerosity seen by the decoder was also zero, is next highest when the numerosity seen by the decoder was one, and so on.

In the non-symbolic task (**Figure 3.2D**, left), the classifier confuses zero with one ($Mean_{proportion\ predicted}$ = 0.218) more often than it confuses zero with two ($Mean_{proportion\ predicted}$ = 0.138) (t(23) = 6.23, p < .001). Similarly, it confuses one with two ($Mean_{proportion\ predicted}$ = 0.206) more often than with three ($Mean_{proportion\ predicted}$ = 0.155) (t(23) = 4.76, p < .001). This pattern of results is indicative of a gradedness in the representation of numerical magnitude across non-symbolic numerosities. In contrast, in the symbolic task (**Figure 3.2D**, right), the multiclass classifier does not confuse zero with one ($Mean_{proportion\ predicted} = 0.159$) significantly more than it confuses zero with two ($Mean_{proportion\ predicted} = 0.163$) (t(23) = -0.61, p = 0 .54), nor does it confuse one

with two ($Mean_{proportion\ predicted} = 0.153$) significantly more than it confuses one with three ($Mean_{proportion\ predicted} = 0.143$) (t(23) = 1.67, p = 0 .11). This difference in distance effects between non-symbolic and symbolic formats was statistically significant for both zero (t(23) = 5.45, $p < .001$) and one (t(23) = 3.48, p = .002), and is suggestive of more gradedness in the representation of non-symbolic than symbolic numerosities, consistent with previous work describing narrower tuning curves for symbolic numerals (Eger et al., 2009; Kutter et al., 2018). I note that a graded RDM still captured a significant portion of variance in the symbolic data (**Figure 3.2C**), due to (graded) confusions between non-neighbouring numerosities (e.g., 5 is predicted more often when the classifier sees a 4 than when it sees a 0).

3.3.3 Representations of Zero are Shared Between Symbols and Empty Sets

Together, the previous analyses establish that neural representations of zero are graded (especially for non-symbolic numerosities) and situated within a number line spanning other countable numerosities from 1 to 5. I next asked whether representations of zero were format- and task-independent – generalising across non-symbolic (empty set) and symbolic (‘0’) stimuli, and across the same/different and averaging tasks. As a first step towards testing for cross-format representations of number, I first computed the Exemplar Discriminability Index (Bang et al., 2020; Luyckx et al., 2019; Nili et al., 2021) as a measure of how similar matching cross-format numerosities were (e.g. ‘1’ and one dot, ‘2’ and two dots, etc.) compared to non-matching numerosities (e.g. ‘1’ and five dots). This EDI analysis indicates a significantly higher degree of similarity for matching cross-format numerosities from ~200ms onwards (**Figure 3.3A**).

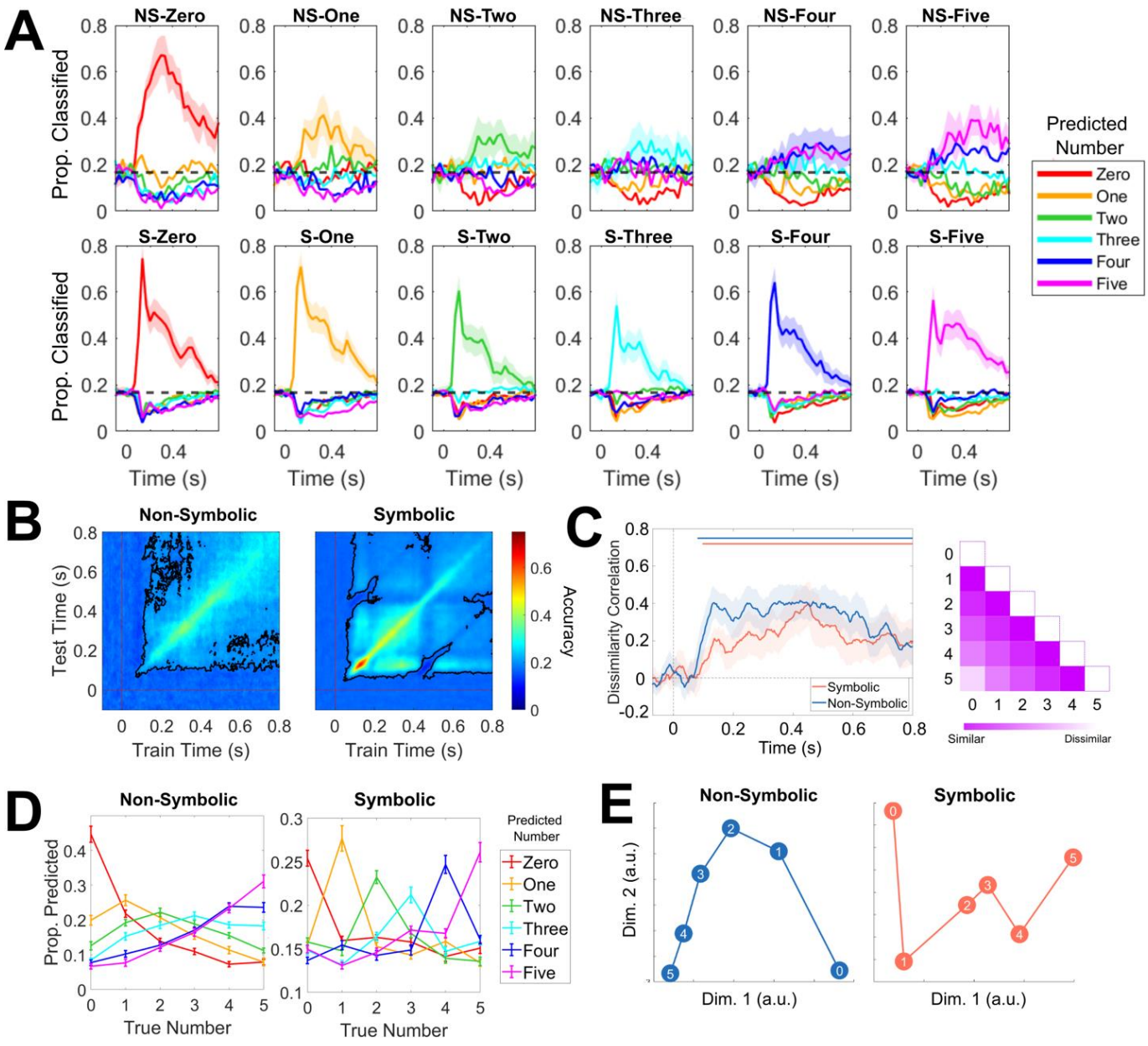


Figure 3.2. A Neural Number Line from Zero to Five. A. Across-time confusion matrices for multiclass decoders classifying non-symbolic (top) and symbolic numerosities (bottom). Individual panels represent trials where particular numerosities were presented to the classifier. Coloured lines indicate the proportion of those trials where the classifier predicted each numerosity. B. Temporal generalisation of multiclass decoders trained to decode numerosities zero to five in the non-symbolic (left) and symbolic (right) task reveals stable numerical representations over time in both tasks emerging shortly after stimulus presentation. Black lines illustrate timepoints where decoding was significantly above chance ($p < .05$, corrected for multiple comparisons). These stable time windows were used in the time-averaged analyses depicted in panels D and E. C: A model representational dissimilarity matrix (RDM) describing a distance effect from zero to five significantly predicted neural data in both non-symbolic and symbolic tasks. The diagonal of the RDM was not included in this analysis, preventing the self-similarity of each number from trivially

explaining the results. Shaded areas indicate 95% confidence intervals. Horizontal lines show clusters of time where dissimilarity correlations were significantly above 0, $p < .05$ corrected for multiple comparisons. D. Population level tuning curves derived from decoder confusion matrices. Each curve represents the proportion of trials the classifier predicted a particular numerosity (indicated by the curve's colour) as a function of the numerosity the decoder actually saw. For example, the red curve illustrates how the prediction of numerosity zero is distributed across different presented numerosities. For non-symbolic numerosities, the classifier confused numbers as a function of their numerical distance, consistent with a graded representation of numerical magnitude. In the symbolic task, representations were more categorical than the non-symbolic task. Error bars represent SEM. E. Multidimensional scaling of numerical representations in both tasks revealed a principal dimension which tracks numerical magnitude and a second dimension distinguishing extreme values from intermediate values.

Next, to specifically test for a cross-format representation of zero, I performed decoding analyses focused on dissociating numerical zero from non-zero numerosities. If a binary classifier trained to distinguish zero from non-zero numerosities in one numerical format is subsequently able to separate zero from non-zero numerosities in another numerical format, this furnishes evidence for an abstract neural representation of numerical absence that is common to both formats.

Decoders trained to distinguish numerical absence within each format separately revealed stable representations of numerical zero from approximately 100ms to 450ms after stimulus presentation, before exhibiting a more dynamic temporal profile until the end of the trial epoch (**Figure 3.3B**, top). Crucially, these decoders could also successfully classify representations of zero in the opposing format to which they had been trained (**Figure 3.3B**, bottom) – both when generalising from empty sets to the decoding of symbolic numerosities, and when generalising from symbolic zero to non-symbolic dot stimuli. This cross-decoding was successful over the initial 350ms period where the within-format decoders identified stable representations of numerical absence, although generalisation was generally stronger when generalising from symbolic zero to empty sets than vice-versa.

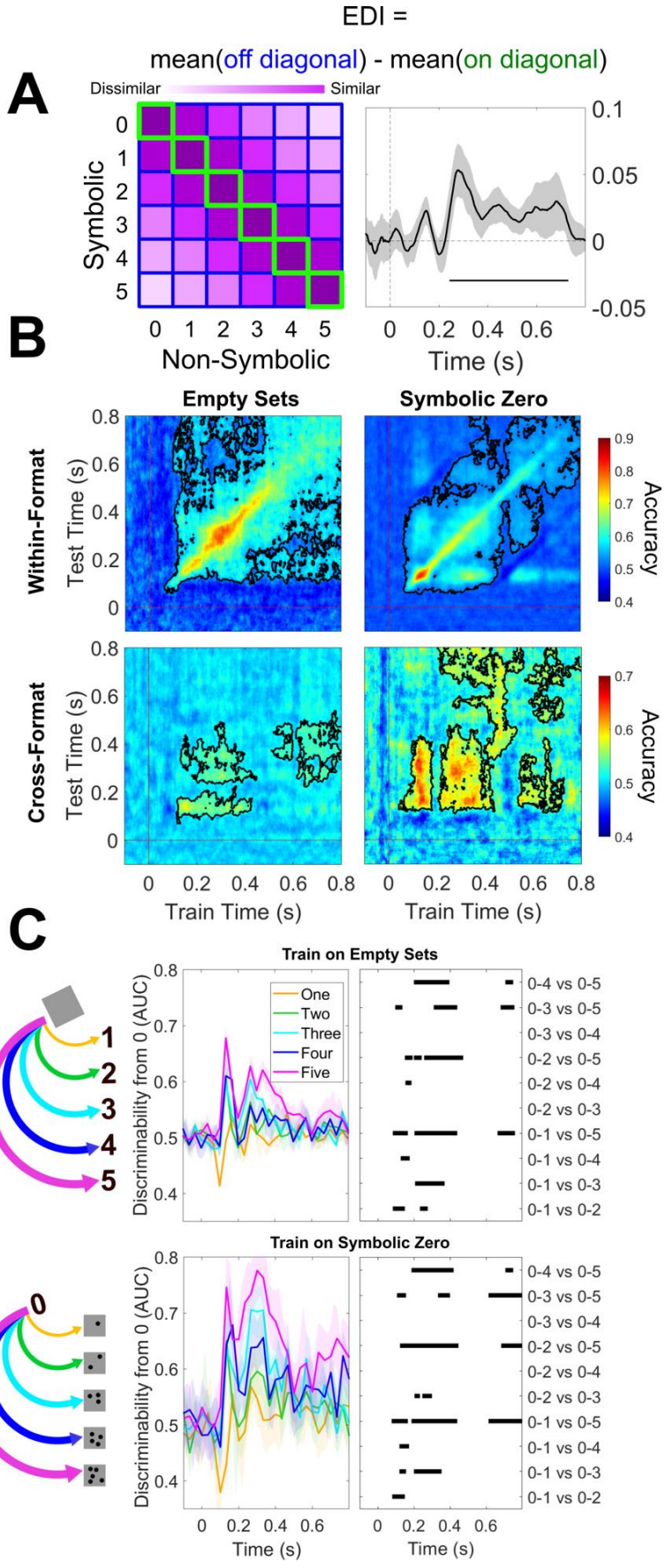


Figure 3.3. Cross-format, Graded Representations of Numerical Absence. A. The Exemplar Discriminability Index was calculated over time as a measure of the similarity numerosities have to their cross-format counterparts (e.g., testing whether symbolic zero was more similar to empty sets than different dot patterns). Significant EDI was found from around ~200ms after stimulus onset. B. Representations of numerical absence generalise over numerical format. Top: A decoder trained to decode zero from natural numbers reveals stable representations of zero up to ~450ms after stimulus presentation for both non-symbolic (left) and symbolic (right) formats, with more dynamic / unstable representations observed towards the end of the epoch. Bottom- left: A decoder trained to decode empty sets also distinguished symbolic zero from non-zero symbolic numerals. Bottom-right: A decoder trained to distinguish symbolic zero from non-zero symbolic numerals also distinguished empty sets from non-symbolic numerosities. Black lines indicate clusters of significantly above chance decoding, p <.05, corrected for multiple comparisons. C. Left: Illustration of the hypothesis that abstract representations of numerical absence are situated on a graded number line that generalises across format, with empty sets represented as more similar to symbolic numeral one than numeral five (top), and symbolic zero as more similar to one dot than five dots (bottom). Centre: Training a classifier to decode non-symbolic empty sets from non-symbolic numerosities and testing it on symbolic numbers in a pairwise manner revealed increasing discriminability as distance from zero increased (top). The same cross-format distance effect is observed when training a classifier on symbolic zero and testing it on non-symbolic numerosities (bottom). Shaded areas represent 95% CIs. Right: clusters of significant differences between different numerosities' discriminability from zero, p<.05, corrected for multiple comparisons. An increase in discriminability for numbers further from zero reveals a cross-format distance effect.

3.3.4 Graded Representations of Zero are Invariant to Numerical Format

The previous analyses were designed to reveal whether representations of numerical zero generalise across formats, but did not provide a test of whether a binary or graded architecture supports zero representations. To probe the representational structure of cross-format representations of zero, I leveraged the numerical distance effect already identified for within-format representations (**Figure 3.2**). To test for such effects, I used a one-vs-one decoding approach to compute the discriminability between zero and each non-zero numerosity in the alternative numerical format (**Figure 3.3C**, middle). This approach allowed me to specifically examine how numerical zero is represented with respect to other numerosities. Tests of cross-format representations across all numerosities from zero to five are presented in **Supplementary Figure 3.3** and **Supplementary Figure 3.4**. Strikingly, neural representations of symbolic zero ('0') were more often confused with one or two dots in the non-symbolic task, than they were with four or five dots (**Figure 3.3C**, middle). Similarly, neural representations induced by non-symbolic zero (empty sets) were more often confused with the symbolic numeral 1 or 2 than they were with symbolic numerals 4 or 5. Pairwise tests comparing the discriminability of different non-zero numerosities from zero revealed clusters of significant differences in discriminability (**Figure 3.3C**, right), with an increased distance from zero increasing discriminability. Together, these cross-format analyses support a hypothesis that an approximate, graded representation of numerical absence is engaged not only by symbolic zero ('0') but also by non-symbolic empty set stimuli.

I also sought to test a more stringent hypothesis that abstract, format-independent neural representations of zero are themselves situated within a cross-format neural number line – thus extending the question of format-independence to now include all numerosities from 0 to 5. A representational dissimilarity matrix situating abstract numerosity representations within a graded number line significantly predicted the neural data (**Supplementary Figure 3.3**, left). Testing for cross-format distance effects between all numerosities using RSA also revealed a qualitative distance effect, although this did not reach statistical significance (**Supplementary Figure 3.3**, right). Finally, multidimensional

scaling of neural representations induced by symbolic and non-symbolic numerosities in a shared space corroborated evidence for a distance effect for zero across tasks (**Supplementary Figure 3.4**).

Finally, to explore how similar cross-format representations of zero were to one another, the zero-specific decoders trained in **Figure 3.3B** were presented with all the cross-format numerosities from zero to five. The decoder's decision evidence was taken as a measure of the discriminability of each numerosity from its cross format zero. I found several timepoints where the zero was significantly less discriminable from its cross-format counterpart than any other numerosity (**Supplementary Figure 3.5**), suggesting that representations of symbolic zero are most closely related to representations of non-symbolic empty sets than other non-symbolic numerosities, and vice versa.

3.3.5 Format-Invariant Representations of Numerical Zero are Localised to Posterior Association Cortex

Finally, I sought to localise representations of format-invariant numerical zero in the brain. To do this, I reconstructed and compared source-level neural activity for zero and non-zero numerosities in both the non-symbolic and symbolic tasks. By performing mass-univariate contrasts of broadband source power (zero > non-zero numerosities) in both the non-symbolic (**Figure 3.4A**, top; peak voxels (*xyz*): left hemisphere = -36, -24, 56; right = 60, -64, -24) and symbolic (**Figure 3.4A,** bottom; peak voxels (*xyz*): left hemisphere = -28, -56, 32; right = 28, -72, 8) tasks and computing the conjunction between these two contrasts (**Figure 3.4B**; peak voxels within conjunction (*xyz*): non-symbolic task: left hemisphere = -20, -64, 32; right = 60, -64, -24; symbolic task: left hemisphere = -28, -56, 32; right = 20, -48, -64), I was able to show that neural activity induced by numerical absence is distributed across the posterior association cortex (**Figure 3.4B**). To explore whether the zero representations localised to this region exhibited a graded structure (as found in sensor-level analyses, **Figure 3.3C**), I performed multidimensional scaling on source level activity patterns for each numerosity and format. Neural responses to zero within this conjunction map were again situated within a number

line populated by non-zero numbers, with numerical magnitude increasing along a single dimension that was similar for both symbolic and non-symbolic formats (**Figure 3.4B**, bottom-right).

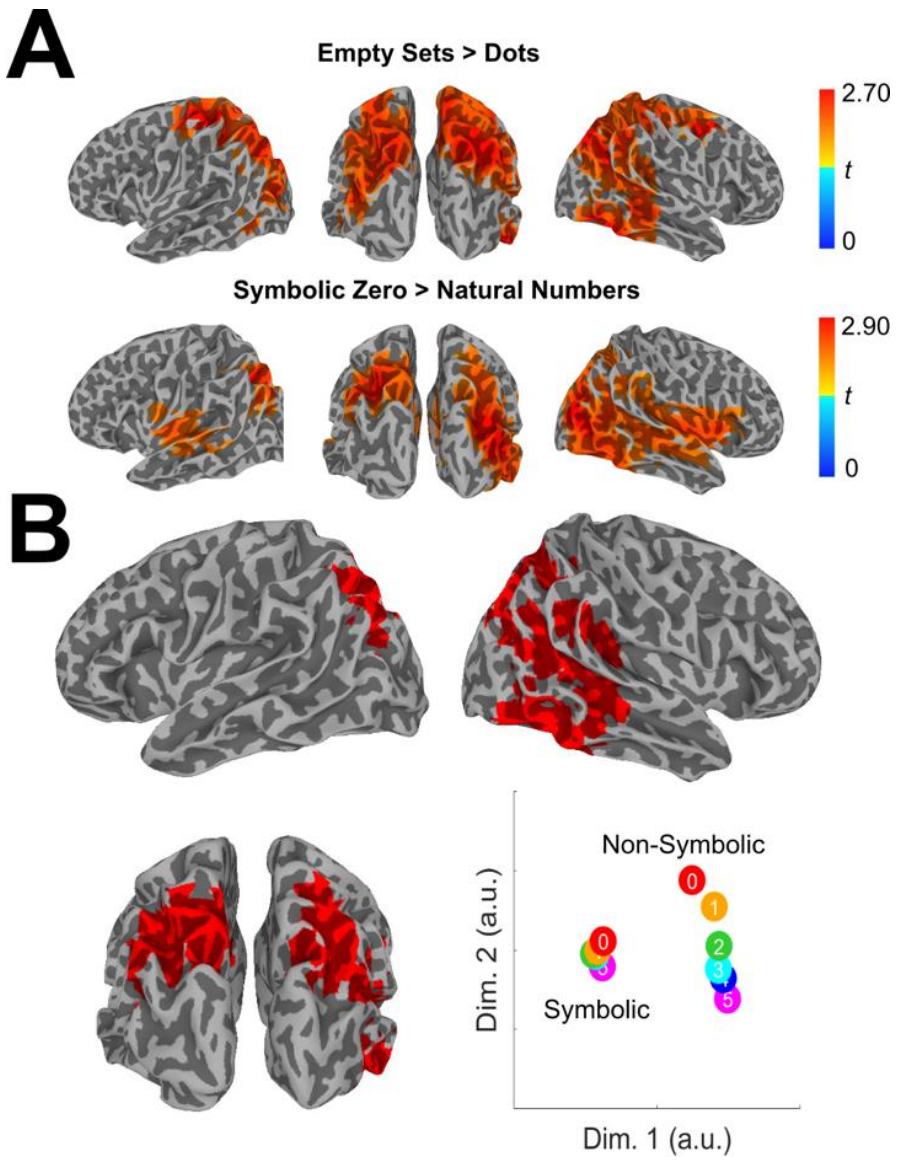


Figure 3.4. Neural activity induced by numerical zero localised to posterior association cortex. A: Mass univariate contrasts of source power revealed regions more active following presentations of zero vs. non-zero numerosities in non-symbolic (top) and symbolic (bottom) tasks. Colour represents t-value and only clusters significant at $p < .05$ are presented, corrected for multiple comparisons. B: A conjunction of zero > non-zero contrasts in both numerical formats yielded a map identifying broad regions of the posterior association cortex as representing numerical absence across numerical formats. Multidimensional scaling of each numerosity's neural pattern within these regions revealed a graded representational structure of numerical magnitude along a single dimension that was similar for both formats.

3.4 Discussion

The number zero is associated with unique psychological properties compared to natural numbers. Here, I characterise the neural representation of numerical zero in the human brain. I describe how numerical zero occupies a slot at the lower end of neural number lines for both symbolic and non-symbolic numerical formats. I go on to show that a component of this representation is both task- and format-independent, such that empty sets – the absence of dots – generalised to predict the neural profiles and distance effects observed for symbolic zero. These abstract, format-invariant representations of zero were situated at the lower end of a neural code for number that was localised across the posterior association cortex.

The finding that representations of numerical absence have a format-invariant component extends previous work documenting neural representations of numerosity that generalise across countable non-symbolic elements and their symbolic counterparts (Eger et al., 2009; Libertus et al., 2007; Piazza et al., 2007). Here I show how neural representations of non-symbolic empty sets, which do not contain any countable items, also share variance with symbolic zero across qualitatively different tasks with distinct behavioural profiles. These abstract representations of zero were localised to regions of the posterior association cortex that have previously been associated with numerical processing (Arsalidou & Taylor, 2011; Eger et al., 2003; Harvey & Dumoulin, 2017; Piazza et al., 2007).

That zero is situated at the lower end of a neural number line in the human brain is consistent with an emerging body of work examining representations of zero in non-human animals (Kirschhock et al., 2021; Okuyama et al., 2015; Ramirez-Cardenas et al., 2016). Across two different studies, single neurons selective for non-symbolic empty sets were found in the parietal and prefrontal cortex of non-human primates (Okuyama et al., 2015; Ramirez-Cardenas et al., 2016). In line with the present results, many of these neurons – but not all – were found to exhibit distance effects with non-zero numbers. When comparing non-symbolic and symbolic instances of zero, I found symbolic

instances were more discrete and less graded than non-symbolic instances, consistent with work describing sharper tuning curves for symbolic number representations (Eger et al., 2009; Kutter et al., 2018). Recent single-cell recordings in the human medial temporal lobe have also identified discrete non-symbolic zero-selective neurons that did not exhibit graded activations in relation to non-zero numerosities (Kutter et al., 2023). Strikingly, however, the majority of my analyses revealed a graded representation of zero that generalised across both symbolic and non-symbolic formats. This is in keeping with 7T fMRI data showing that neural populations at the lower boundary of numerically-tuned topographic maps exhibit a monotonic decrease in response to increasing numerical magnitude (Paul et al., 2022), a finding suggestive of evidence for neural populations tuned to numerosities below one.

I took care to ensure that the neural representations of zero identified in the data were not trivial consequences of zero being classified as the 'lowest' stimulus in the tasks. The concern here is that if the tasks required participants to adopt a particular mathematical attitude towards zero, then decoding of this task-dependent concept would confound any results aimed at identifying task-invariant representations of numerical absence. I consider this explanation of the results as unlikely, however, as, by design, the symbolic and non-symbolic tasks required adopting qualitatively distinct mathematical attitudes towards zero stimuli: the match-to-sample task necessitated deciding whether two dot stimuli were the same or different, whereas the symbolic task required maintenance of condition-specific numerical averages. Because the non-symbolic task did not require participants to order stimuli, any format-invariant representations of zero cannot be explained by a generic requirement to identify lower vs. higher numerosities. Despite these task differences, explicit numerical processing of the stimuli was common to both tasks, and by design, both of these tasks afford the locking of the MEG data to the onset of examples of zero (and other numerical) stimuli that are embedded in a wider task context.

In both cross-decoding and RSA analyses, I observed cross-format representations emerging initially between 100-200ms post-stimulus before diminishing and re-emerging

approximately 300ms after stimulus presentation. The small number of previous studies which have explored the time course of abstract numerical representations in the human brain document generalisation from symbolic formats to non-symbolic formats or alternative magnitude domains from ~300ms onwards (Luyckx et al., 2019; Teichmann et al., 2018), consistent with this later peak in the data. The earlier peak, in contrast, is consistent with other studies documenting decodability of non-numerical conceptual representations (such as animate/inanimate and artificial/natural) as early as ~100ms post-stimulus (Carlson et al., 2013; Cichy et al., 2014). It remains debated whether findings of format-invariant numerical codes are explained by single neurons coding for the same numerosities across formats, or whether they reflect the recruitment of neighbouring format-specific neural populations that are interdigitated within a numerosity map (Cohen Kadosh & Walsh, 2009). Future intracranial recording studies will be required to determine whether single cells in the human brain code for numerical zero in both non-symbolic and symbolic formats. However, the finding of cross-format distance effects is more consistent with a shared neural code, as it is less likely that spatially overlapping but format-specific neural codes would also generalise to exhibit cross-format distance effects with more distant numerosities. Furthermore, the finding of a cross-format code for numerical zero is in keeping with behavioural studies that have found format-invariant distance effects for both symbolic zero and empty sets (Zaks-Ohayon et al., 2022).

Behavioural tasks have previously been used to investigate zero in relation to a mental number line (Dehaene et al., 1993; Fischer & Rottmann, 2005; Merritt & Brannon, 2013; Pinhas & Tzelgov, 2012). For example, the SNARC (spatial-numerical association of response codes) effect extends to numerosity zero, with faster response times when zero responses were given with the left hand (Dehaene et al., 1993). Similarly, the size-congruity effect (Henik & Tzelgov, 1982) has been exploited to suggest that zero occupies a definite position on a mental number line: responses to zero are facilitated when it is physically smaller than an alternative numeral (Pinhas & Tzelgov, 2012). Notably, ‘end effects’ have been established in size congruity paradigms, whereby stimuli perceived to be at the ‘end’ of the number line exhibit facilitated response times. End effects have been found for symbolic zero even when presented amongst negative numbers (Pinhas &

Tzelgov, 2012) and in response to non-symbolic empty sets during numerical comparison tasks (Zagury et al., 2022), suggesting both symbolic and non-symbolic zero may be situated at the beginning of a mental number line. Such findings obtained using elegant behavioural assays are consistent with my findings that numerical zero is incorporated into a graded neural number line amongst other non-zero numbers (at least in some contexts (Zaks-Ohayon et al., 2021, 2022)).

Classical accounts of how symbolic representations of number are mapped to their non-symbolic counterparts do not readily explain how the symbolic concept of zero is generated from non-symbolic empty sets. These models describe a neural architecture in which specific numerosities are mapped onto numerosity-tuned neural populations that form a neural number line, or place-coding system (Dehaene & Changeux, 1993; Piazza et al., 2007). However, according to such models, non-symbolic stimuli only proceed to this place-coding system via a summation procedure, where an increase in the number of non-symbolic elements accumulate a greater degree of activation (DeWind et al., 2019; Park et al., 2016; Verguts & Fias, 2008; Zorzi & Butterworth, 1999). A summation explanation therefore fails to account for how non-symbolic empty sets may be mapped to symbolic zero, since empty sets offer no countable elements to accumulate (Pinhas & Tzelgov, 2012). Computational models of object recognition have, however, documented the emergence of spontaneous representations of non-symbolic zero without any training on numerical stimuli (Nasr et al., 2019), suggesting that the concept of zero can be readily acquired from statistical regularities in visual input without necessarily relying on a summation procedure.

A shared neural representation underpinning both non-symbolic empty sets and symbolic zero is consistent with recent suggestions that representations of numerosity zero may have emerged from more fundamental representations of sensory absence (Nieder, 2016). On this account, low-level perceptual representations indicating an absence of sensory stimulation (Goh et al., 2023; Merten & Nieder, 2012) provide the perceptual grounding for a conceptual representations of numerical zero (Nieder, 2016) – consistent with a broader principle that the human brain co-opts basic sensory and motor functions

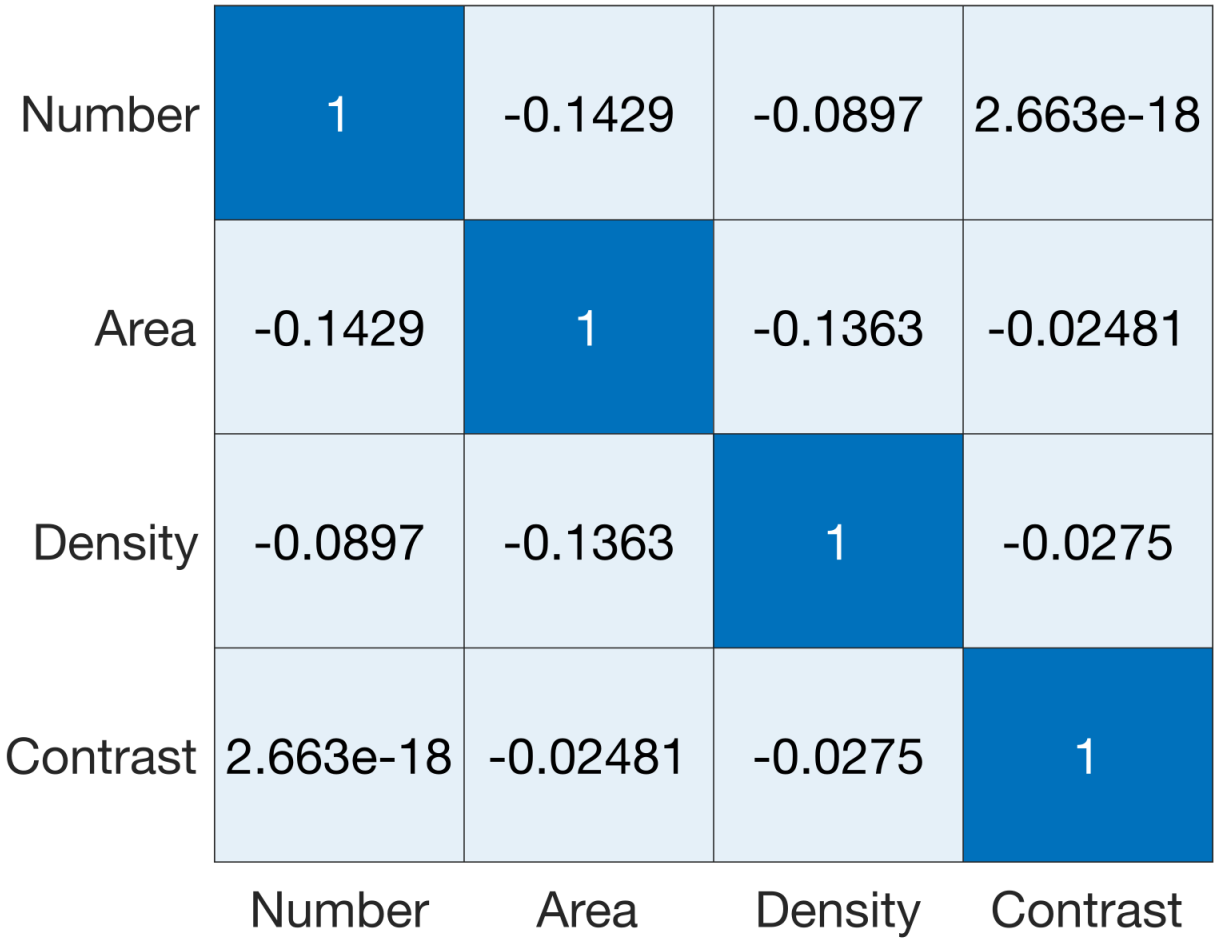
in the service of more complex cognitive abilities (Dehaene & Cohen, 2007). Such a hypothesis aligns with similar behavioural signatures for the processing of absence across perceptual and numerical domains. For instance, reading times are increased for number zero compared to non-zero numbers (Brysbaert, 1995), whilst reaction times for deciding a stimulus is absent are higher than for deciding a stimulus is present (Mazor et al., 2020, 2021). Additionally, judgements about the absence of features mature later in children than judgements about presence (Coldren & Haaf, 2000; Sainsbury, 1971), a developmental pattern mirrored by the late mastery of numerical zero in children (Krajcsi et al., 2021; Merritt & Brannon, 2013; Wellman & Miller, 1986). I note however that the neural responses recorded in this study to empty-set stimuli were still within the context of a numerical task – and, as such, my results are specific to the concept of numerical absence and do not provide a direct test of a link between numerical zero and sensory absence. This study offers a path towards a formal test of this hypothesis in future work - for instance, investigating relationships between numerical absence and non-numerical perceptual absences (Chapter 2; Mazor et al., 2020; Merten & Nieder, 2012).

Non-numerical perceptual absences can be particularly useful in determining the cognitive basis of consciousness, particularly self-awareness (Mazor, 2021). One recent study, for example, found that when participants were required to make detection decisions about stimuli hidden behind differing degrees of occluding barriers, decisions about absence were associated with the modelling of one's own visual system (Mazor et al., 2024). To be specific, when participants decided that a stimulus was absent, they implicitly modelled the likelihood function mapping environmental states to perceptual states, essentially asking the question 'if the stimulus was presented, would my visual system have detected it?' In this way, exploring the unique computations underlying decisions about absence can reveal the extent to which humans maintain an internal model of their own cognition – a fundamental component to self-awareness and consciousness more broadly. Indeed, the unique properties associated with decisions about absence are reflected in the HOSS model (Fleming, 2020), where the asymmetry between decisions about absence vs. presence are used to explain the 'ignition' effects usually referenced by Global Workspace theorists. As such, the HOSS model makes the

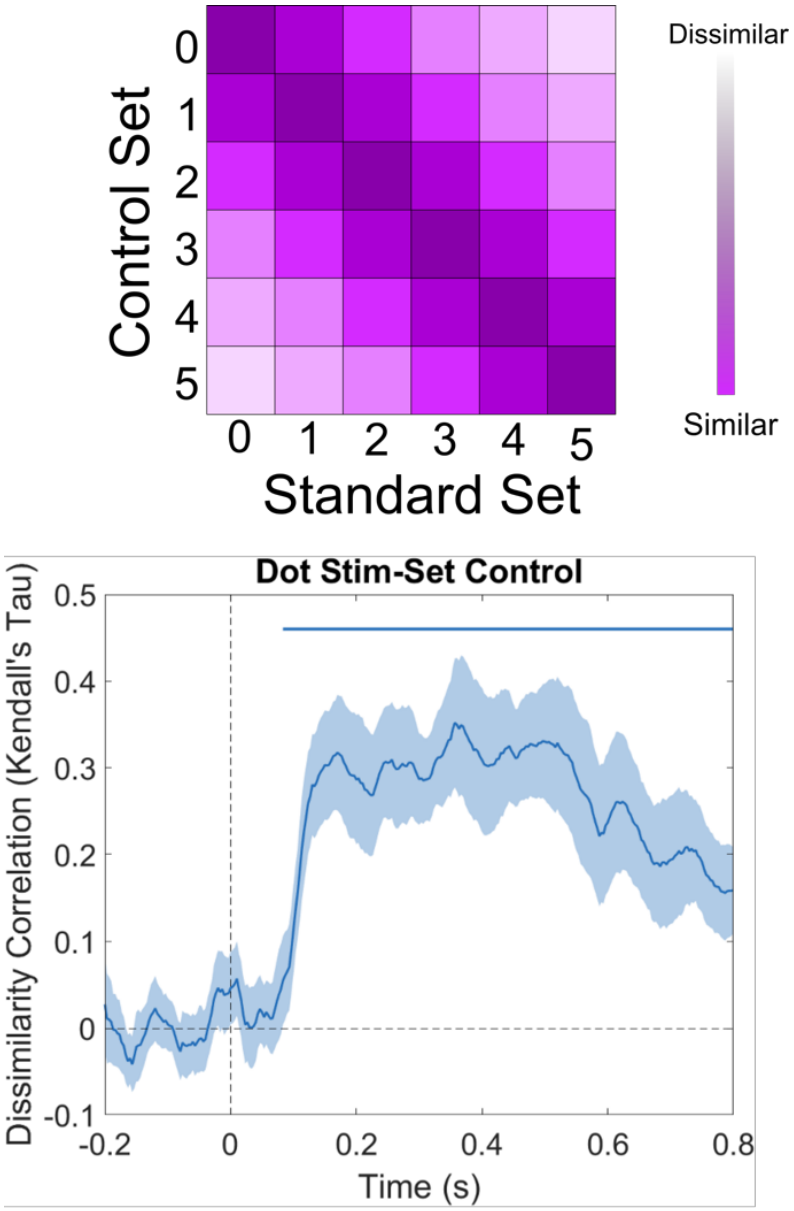
explicit prediction that sensory absences should be actively encoded in neural responses to (an absence of) sensory evidence.

The adoption of the number zero has enabled great advances in science and mathematics (Kaplan, 1999). Here, I show that the human brain represents this unique number by incorporating representations of numerical absence into a broader neural coding scheme that also supports countable numerosities. Representations of numerical zero were found to be format-invariant and graded with respect to non-zero numerosities, and were localised to regions of the posterior association cortex previously implicated in numerical cognition. My results demonstrate that neural number lines include zero, and, more importantly, provide initial evidence that the abstract concept of symbolic zero is linked to representations of non-symbolic empty sets. This study lays the foundations for a deeper understanding of how the human ability to represent the number zero may be grounded in perceptual capacities for detecting an absence of sensory stimulation.

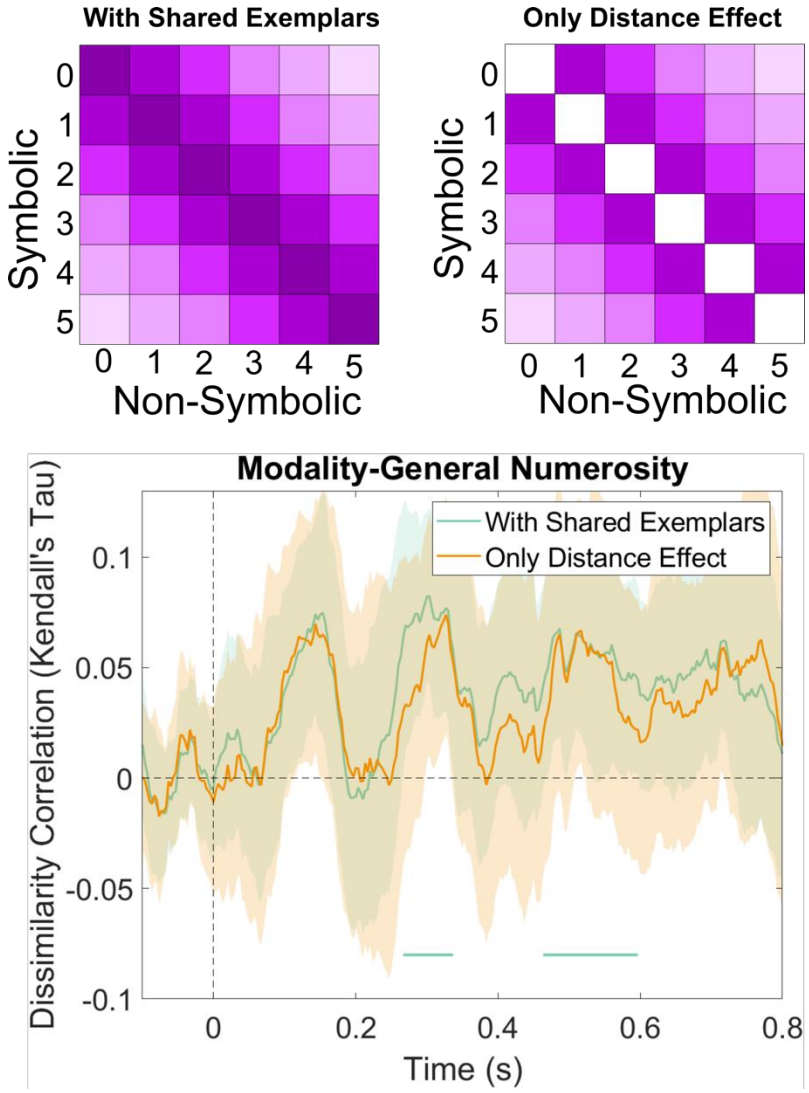
3.5 Supplementary Materials



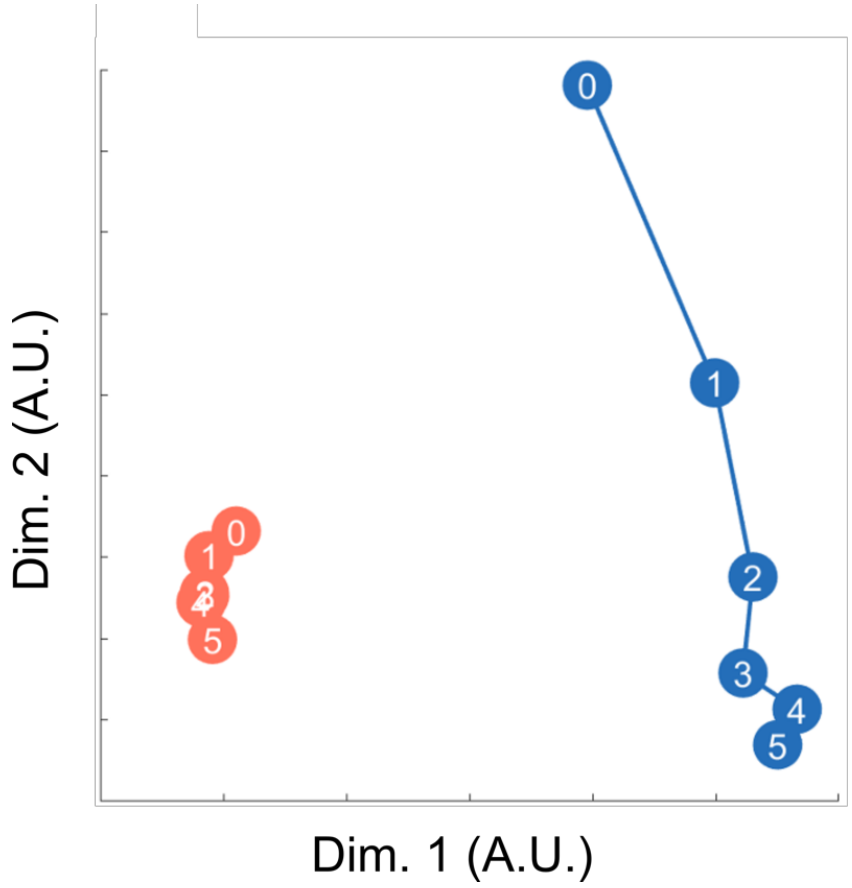
Supplementary Figure 3.1. Non-symbolic control stimuli are successfully controlled for low level visual properties. Total dot area was measured as the number of pixels covered by all dots. Density was computed as the negative median Euclidean distance between dots, such that higher distances between dots results in lower density scores. Contrast was computed using Michelson contrast. Values are Pearson correlation (r) coefficients.



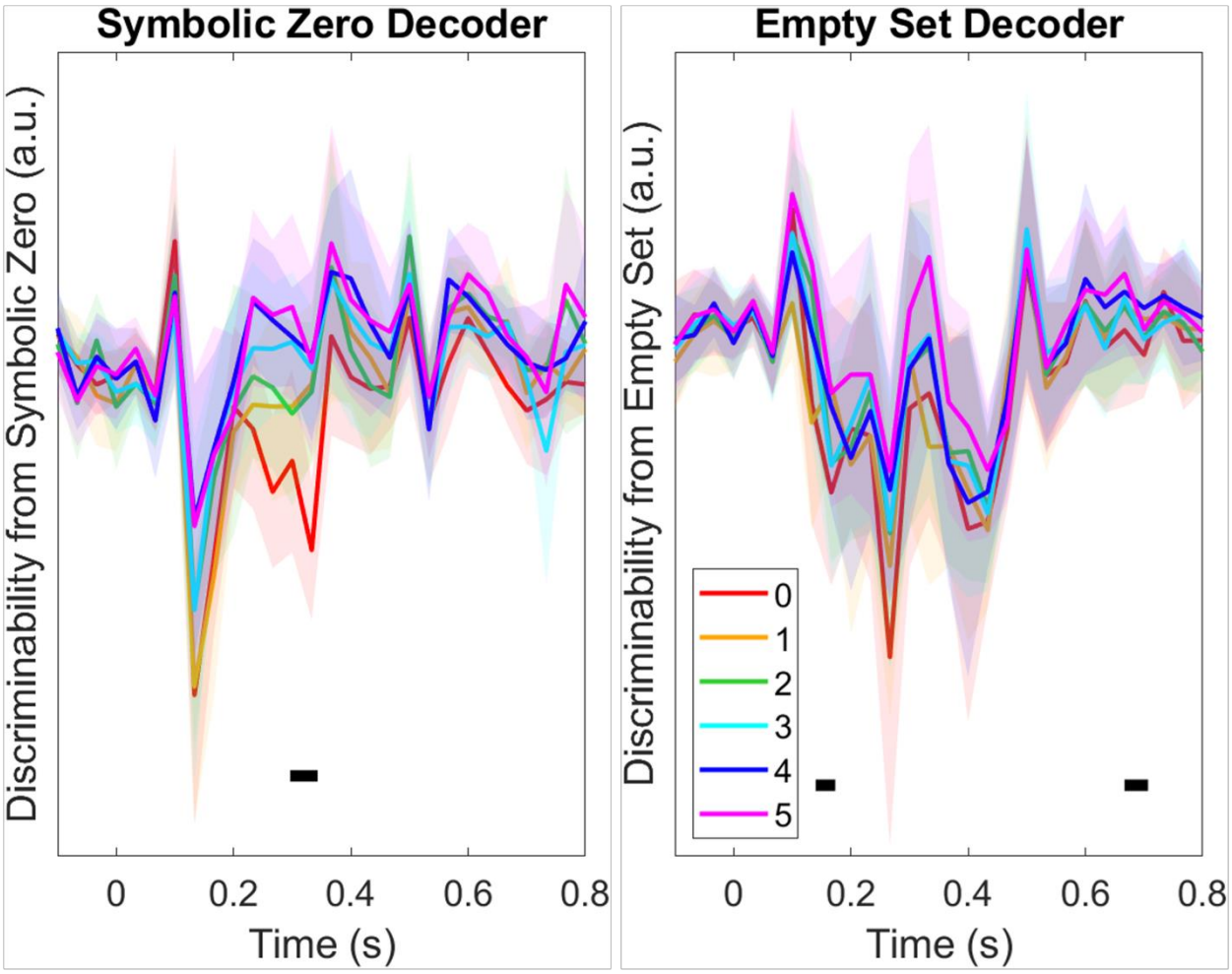
Supplementary Figure 3.2. Cross-stimulus set representations of numerosity in non-symbolic stimuli. A representational dissimilarity matrix (RDM) was constructed that modelled the distance effect between non-symbolic numerosities across stimulus sets (top). This model tests whether numerical information is shared across the standard and control set, independently of their unique physical features. Numerical distance effects could be extracted from the neural data independently of the stimulus set soon after stimulus presentation and for the remainder of the epoch (bottom). Horizontal line represents time points where the correlation of the model RDM with the neural data was significantly above zero with an alpha of $p < .05$, corrected for multiple comparisons.



Supplementary Figure 3.3. Cross-Task RSA reveals a format-invariant neural code for number. An RDM modelling numerical information as shared between numerical format successfully predicted the neural data at two different timepoints. Removing the diagonal from this RDM removes the shared exemplars from the model (empty sets and zero, one dot and symbolic one, etc.) providing a strong test of the hypothesis that abstract numerical information also exhibits a distance effect. This model showed a similar pattern of prediction to the full model with shared exemplars yet failed to reach statistical significance. Broad confidence intervals, represented by the shaded area, suggest this may be an issue of limited statistical power. Horizontal lines indicate clusters where the model RDM correlated with the neural data significantly more than zero, $p < .05$, corrected for multiple comparisons. Shaded areas represent 95% confidence intervals.



Supplementary Figure 3.4. Multidimensional scaling of numbers across format. Performing multidimensional scaling on numerosity representations in a shared space revealed alignment along an axis defining numerical magnitude (dimension two). This illustrates the cross-task distance effect, where empty sets (blue zero) are represented more closely to symbolic one (red one) than symbolic five (red five), and vice versa. Dimension one discriminates between the two tasks.



Supplementary Figure 3.5. Cross-format zeros are most similar to each other. Decoders trained to identify zero stimuli were presented with cross-format numerosities and the decoder's decision evidence was taken as a metric of discriminability of that numerosity from its cross-format zero. The different coloured lines represent the discriminability of a particular number from its cross-format zero. Lower values represent less discriminability (and higher similarity to zero). The lower values of discriminability for zero trials (red lines) indicate that representations of zero are most similar to the cross-format zero compared to non-zero numerosities. Black horizontal bars represent time points where zero was significantly less discriminable from the cross-format zero than one.

4. Dementia as a disorder of consciousness

4.1 Introduction

Alzheimer's disease (AD) is a debilitating cognitive disorder characterised by a slow but steady decline in cognitive functioning, with impairments occurring across multiple cognitive domains (World Health Organization, 2011). Typically, characterisations of AD have centred on the compromised cognitive faculties observed throughout disease progression, such as learning, executive function, planning, and communicating (Sperling et al., 2011). However, as the disease severity increases, patients with AD can lose the capacity to recognise relatives and caregivers, and may also exhibit an unawareness of features of their environment (Clare, 2010; O'Shaughnessy et al., 2021). Awareness-related deficits such as these are the most distressing aspects of AD for both patients and caregivers alike (Rice et al., 2019) and there is an urgent need for the effects of AD on conscious experience to be better understood so that effective and patient-centred care protocols can be developed accordingly (Huntley et al., 2023; Huntley et al., 2021).

Previous research suggests AD patients can suffer from compromised awareness in high-level tasks and judgements (Huntley et al., 2021). For instance, most AD patients suffer from an impaired awareness of their own cognitive deficits and can even be unaware of their diagnosis altogether (Starkstein, 2014). Anosognosia may be driven by a reduced capacity for updating self-relevant knowledge in AD, resulting in the reliance of out-of-date knowledge regarding oneself (Mograbi et al., 2009). Disrupted awareness of the self has also been demonstrated in AD using mirror tasks, where the ability to recognise

oneself becomes increasingly impaired as the disease progresses (Biringer & Anderson, 1992; Grewal, 1994). Given the close theoretical link between metacognition and awareness (Fleming, 2020; Lau, 2019), deficits in metacognitive performance in AD also provide support for disrupted awareness in dementia. Patients with AD, for example, exhibit difficulties in differentiating imagined from external events (Fairfield & Mammarella, 2009), while also showing impaired metacognition in episodic memory tasks, even in mild stages of the disease (Dodson et al., 2011; Mimura & Yano, 2006). Metacognition of memory has even been shown to relate to the degree of tauopathy in cognitively unimpaired adults (Vannini et al., 2019). The above findings provide an initial indication that conscious experience may be systematically altered in AD and motivates calls for an improved understanding of the neurobiological markers of awareness in AD.

Testing whether proposed neural correlates of consciousness (NCCs) are present in patients with AD is a promising method for identifying characteristics of awareness in the disorder (Huntley et al., 2021). The NCCs are defined as the neural processes that are both necessary and sufficient for a conscious experience to occur (Crick and Koch, 1998), and different theories of consciousness propose different NCCs as the determinants of awareness. The Global Neuronal Workspace Theory (GNWT), for example, states that stimuli are consciously perceived when their representations are broadcast across different cognitive domains (e.g., attention, planning, motor systems) in a global workspace that is subserved by frontoparietal regions of the brain (Dehaene and Naccache, 2001; Mashour et al., 2020). According to GNWT, entry into awareness is facilitated by an ‘ignition’ process that underlies the ascendence of sensory representations into these networks (Dehaene and Naccache, 2001; Mashour et al., 2020). Accordingly, frontoparietal activation and late-stage event-related potentials (ERPs) such as the late positivity (LP) are defined as NCCs in a GNWT framework (Dehaene and Changeux, 2011). In a proof-of-concept study, Huntley et al. (2021) used the GNWT framework to examine consciousness in AD, finding that AD patients exhibited a diminished LP response when stimuli were presented above perceptual threshold. This is suggestive of a diminished frontal response to conscious versus unconscious stimuli. However, as a proof-of-concept study involving only four patients with AD, the findings

reported in Huntley et al. (2021) are preliminary and require replication in a larger sample. Here, I conduct a group-level fMRI study (N = 44) using the same task as Huntley et al. (2021) to provide a robust assessment of the neural correlates of awareness in AD under a GNWT framework.

The task used in the present study originates from tests of consciousness in infants (Kouider et al., 2013). Images of faces are presented at varying durations and are preceded and followed by scrambled face masks, rendering the face image unconscious when presented at short durations. This results in trial-by-trial variations in awareness, enabling comparisons of the neural activity associated with visible and subliminal stimuli. To compensate for the cognitive deficits associated with AD, task instructions were minimal and trial-by-trial reports of awareness were not required. Instead, I relied on previous findings using the same procedure and stimuli which indicate that the threshold for conscious perception of faces is approximately 50 ms in adult humans (Gelskov & Kouider, 2010). To mitigate the risk of incorrect trial labelling, trial durations were selected with a large margin either side of this perceptual threshold. As such, I defined subliminal and visible stimuli as those that were presented for 33 ms and 200 ms, respectively.

To anticipate my findings, I provide evidence to suggest that neural activity associated with conscious perception is degraded in mild-moderate AD patients compared to controls. Using multivariate decoding analyses, I find that neural correlates of visibility-related information within visual and frontoparietal regions are reduced in mild-moderate AD patients. Furthermore, I characterise the neural responses of four individual patients with severe AD, finding qualitative signatures of frontal ignition in the only patient who displayed explicit behavioural signs of awareness. This potentially indicates that ignition-like responses to visual stimuli may be a potential biomarker for awareness in AD, although caution is needed when interpreting this result. These findings represent a novel contribution of the neuroscience of consciousness in understanding awareness in AD and justify calls to further examine AD as a disorder of consciousness.

4.2 Materials and Methods

4.2.1 Participants

Forty-four participants took part in the fMRI experiment at the Wellcome Centre for Human Neuroimaging, University College London. The full sample of 44 participants consisted of 26 healthy controls ($Mean_{age}$ = 75.88, SD_{age} = 5.49), 14 mild-to-moderate AD patients ($Mean_{age}$ = 75.92, SD_{age} = 6.01), and 4 patients with severe AD ($Mean_{age}$ = 84.75, SD_{age} = 8.14). Three clinical rating tools were used to classify patients according to the severity of AD symptoms: Clinical Dementia Ratings (CDR; Morris, 1997), The Global Deterioration Scale (GDS; Reisberg et al., 1982), and the standardized Mini-Mental State Examination (sMMSE, Molloy et al., 1991). Participant scores on each of these scales are presented in **Supplementary Table 4.1**. As participants with severe AD lacked capacity to consent, following the legal framework of the Mental Capacity Act (2005), personal consultees were identified and provided a declaration that the person would have wished to take part in the study (HRA, 2017). The study was approved by the Wales 6 NHS ethics committee (18/WA/0012).

4.2.2 Experimental Procedure

The experimental procedure is illustrated in **Figure 4.1**. Trials began with a fixation cross of 1000ms. On each trial, the critical stimulus was either a face or a mask and was presented for either 33ms (subliminal condition) or 200ms (visible condition). The critical stimuli were flanked by a forward mask (300ms) and a backward mask (33ms), followed by a final mask (1500ms). Jitter between trials was randomly drawn from a uniform distribution between 0 and 1 seconds. Stimuli were presented in a random order throughout the experiment, with 60 trials per scanning run across 3 runs, totalling 180 trials overall, with 45 trials per condition (visible-face, subliminal-face, visible-mask, subliminal-mask). Participants were instructed to passively attend to the stimuli and the

experimenter ensured their eyes stayed open throughout the experiment via an eye-tracking monitor.

4.2.3 fMRI Scanning Parameters

fMRI scans were performed using a 3 Tesla Siemens Prisma MRI scanner with a 32-channel head coil. I acquired structural images using an MPRAGE sequence (1x1x1 mm voxels, 176 slices), followed by a double-echo FLASH (gradient echo) sequence with TE1 = 10 ms and TE2 = 12.46 ms (64 slices, slice thickness = 2 mm, gap = 1 mm, in plane FoV = 192 x 192 mm^2, resolution = 3 x 3 mm^2) that was later used for field inhomogeneity correction. Functional scans were acquired using a 2D EPI sequence (3x3x3 mm voxels, TR = 3.36 s, TE = 30 ms, 48 slices, matrix size = 64 x 72, Z-shim = 0).

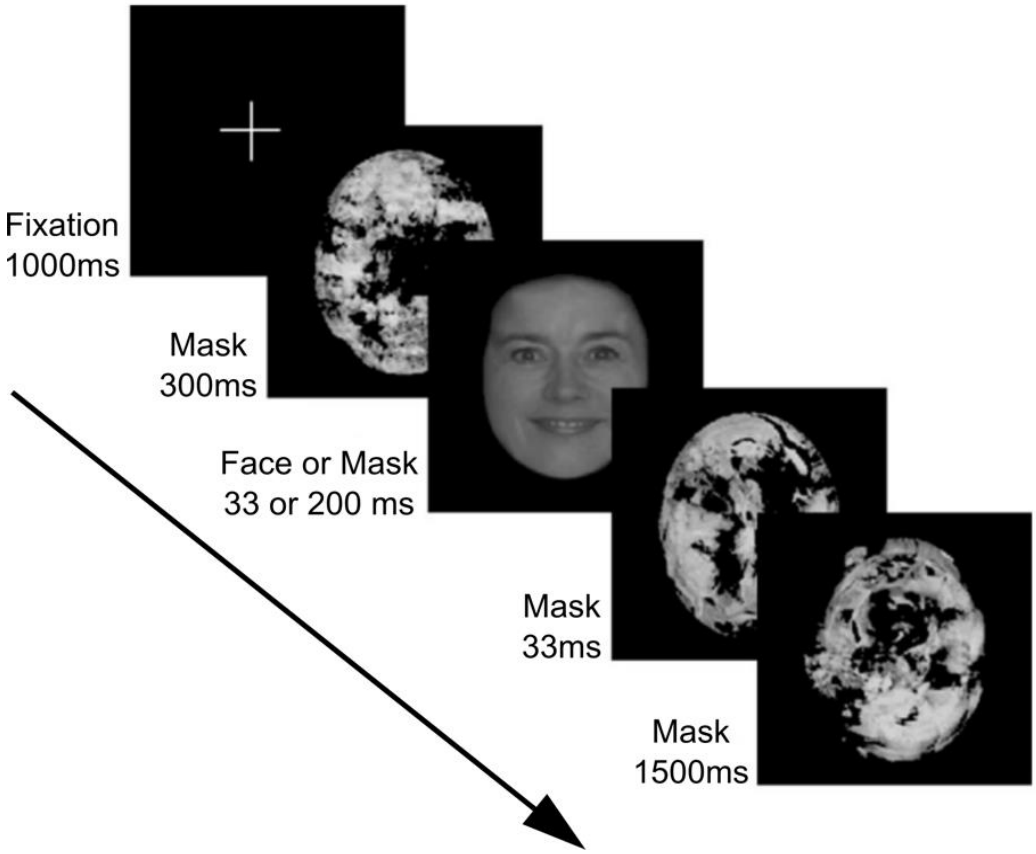


Figure 4.1. Experimental Design. Target stimuli (faces or masks) were presented for 33 (subliminal) or 200 (visible) ms. They were flanked by forward and backwards masked to further manipulate their visibility.

4.2.4 fMRI Data Preprocessing

fMRI preprocessing followed a standard procedure reported in Morales et al. (2018) and Mazor et al. (2020). All fMRI preprocessing was performed using SPM12 (Statistical Parametric Mapping; www.fil.ion.ucl.ac.uk/spm). The first five volumes of each run were discarded to allow for T1 stabilization. Functional images were realigned and unwarped using local field maps (Andersson et al., 2001) and then slice-time corrected (Sladky et al., 2011). Each participant's structural image was segmented into gray matter, white matter, CSF, bone, soft tissue, and air/background images using a nonlinear deformation field to map it onto template tissue probability maps (Ashburner & Friston, 2005). This mapping was applied to both structural and functional images to create normalized images to Montreal Neurological Institute (MNI) space. Normalized images were spatially smoothed using a Gaussian kernel (8 mm FWHM). I set a within-run 1 mm rotation and 3 mm affine motion cut-off criterion, which led to the removal of one block for a patient with severe AD. Preprocessing and construction of first- and second-level models used standardized pipelines and scripts available at https://github.com/metacoglab/MetaLabCore/.

4.2.5 General Linear Model (GLM)

The GLM consisted of four regressors of interest per block, one for each stimulus condition. As such, for each block, there was a regressor for the visible-face, subliminal-face, visible-mask, and subliminal-mask trials. Trials were modeled by specifying stick functions aligned with the onset of the critical stimulus. Motion correction parameters were included as covariates of no interest for each run, alongside a unique constant term per run. Regressors were convolved with the canonical haemodynamic response function. Low-frequency drifts were excluded with a 1/128 Hz high-pass filter.

4.2.6 Univariate Contrasts

All univariate analyses were conducted using (Statistical Parametric Mapping; www.fil.ion.ucl.ac.uk/spm). For control and AD groups, whole-brain single-subject contrast images were computed to examine neural markers of visual awareness (visible-faces > subliminal-faces). Contrast images were then submitted to second-level random-effects analyses. One-sample *t* tests against zero were used to visualise within-group contrasts and two-sample independent *t* tests were used to visualise differences between the control and mild-moderate AD group. For severe AD patients, owing to the small number of patients, only single-subject contrast images were computed. Because of the ethical implications associated with null findings of neural correlates of awareness in patient populations (particularly those that are false negatives), I visualised univariate contrasts at a liberal voxelwise threshold of $p < .05$ uncorrected, with an arbitrary cluster-forming threshold of 250 for the purposes of visualisation. As such, results from univariate contrasts in the present study provide only qualitative illustrations of neural responses to conscious and unconscious stimuli, while my multivariate decoding analyses (below) provide valid whole-brain statistical inference.

4.2.7 Multivariate Decoding

To demonstrate statistical differences in neural correlates of consciousness between patients and controls, I performed whole-brain searchlight decoding analyses. Decoding analyses are sensitive tools used to identify multivariate neural patterns associated with specific cognitive processes and have been used widely to characterise the neural basis of consciousness (Chapter 2; Mazor et al., 2022; Andersen et al., 2016; Taschereau-Dumouchel et al., 2020). Importantly, they can provide greater sensitivity in neuroimaging analyses as they consider the covariance in voxel activity, rather than simply estimating the mean activation over all voxels. As such, I preferred them to univariate analyses to statistically test for differences in neural correlates of consciousness in AD patients. To perform the decoding analysis, I computed beta estimates per trial (Mumford et al., 2012) and used the estimates for visible-face and masked-face trials to train a binary LDA

decoder. Decoders were trained to distinguish beta patterns associated with visible-face trials from patterns associated with subliminal-face trials in a searchlight procedure. Searchlights had a radius of 4 voxels, resulting in 257 voxels per searchlight. The searchlights moved throughout the brain, with each voxel included as the centre voxel in a searchlight. Above chance (50%) accuracy of such a decoder indicates that neural activity covarying with an individual's awareness of a particular stimulus is present within the searchlight, and as such this procedure can be used to examine which neural regions are associated with visual awareness.

The decoder was trained with a 5-fold cross-validation scheme and a shrinkage parameter of 0.1. Trials were balanced such that there was an equal number of visible and invisible trials within each fold. The accuracy of each searchlight's decoder was averaged across folds, and this value was stored at the centre of the searchlight, producing a whole brain map of decoding accuracy. All decoding analyses were performed using custom MATLAB (2021b) scripts.

4.2.8 Statistical Inference on Decoding Accuracies

Distributions of accuracy values from the classification of fMRI data are often non-Gaussian and asymmetric around the chance level. This means that parametric statistical comparisons, such as *t* tests against chance decoding (50%), are unable to provide valid tests of whether group-level accuracy values are significant (Stelzer et al., 2013). Therefore, to determine where decoders had performed significantly above chance, I compared mean performance across all participants with a null distribution created by first permuting the class labels 25 times prior to decoding per participant and then using bootstrapping to form a group-level null distribution of 10,000 bootstrapping samples (Stelzer et al., 2013). To compare decoding performances across AD and control groups, a group-level null distribution was formed by taking the difference between the AD and control groups' mean decoding accuracy values throughout the bootstrapping procedure. The observed empirical accuracy maps were then compared with the null distributions to

compute *p* values. The *p* values for all presented decoding maps are significant at $p < .05$, corrected for multiple comparisons with a false discovery rate of 0.05.

4.3 Results

4.3.1 Controls and AD patients are sensitive to face stimuli

To ensure participants were sensitive to the target face stimuli presented in the experiment, I trained a decoder to decode visible face stimuli from scrambled masks. Across both groups, decoding was successful throughout the brain. In healthy controls, decoding was successful across visual, parietal and frontal cortices (**Figure 4.2,** *top*; **Supplementary Table 4.2**). AD patients also showed sensitivity to faces throughout visual, parietal, and frontal regions (**Figure 4.2,** *middle*; **Supplementary Table 4.2**), suggesting that even during passive viewing, AD patients were sensitive to the experimental paradigm and face stimuli. Controls showed greater sensitivity to faces in small regions of the visual and parietal cortex (**Figure 4.2,** *bottom;* **Supplementary Table 4.3**), but no difference was found in frontal regions.

4.3.2 Ignition related activity in AD

To explore whether the neural activity associated with conscious visual perception is diminished in AD, I first produced visualisations of the neural response to conscious stimuli in AD patients and controls. These maps provide a qualitative demonstration of univariate increases in activity for visible versus subliminal face stimuli at a liberal uncorrected threshold, but do not allow for whole-brain corrected statistical inferences to be drawn from them (see Section 4.3.3 for valid statistical inference with awareness-related analyses). As mentioned above, the reason for producing such maps is to offer protection against false negatives when testing for neural markers of awareness in patient groups. To produce the visualisations, I contrasted the activity from visible face trials with subliminal face trials in both groups. In healthy controls, increased activity in the fusiform

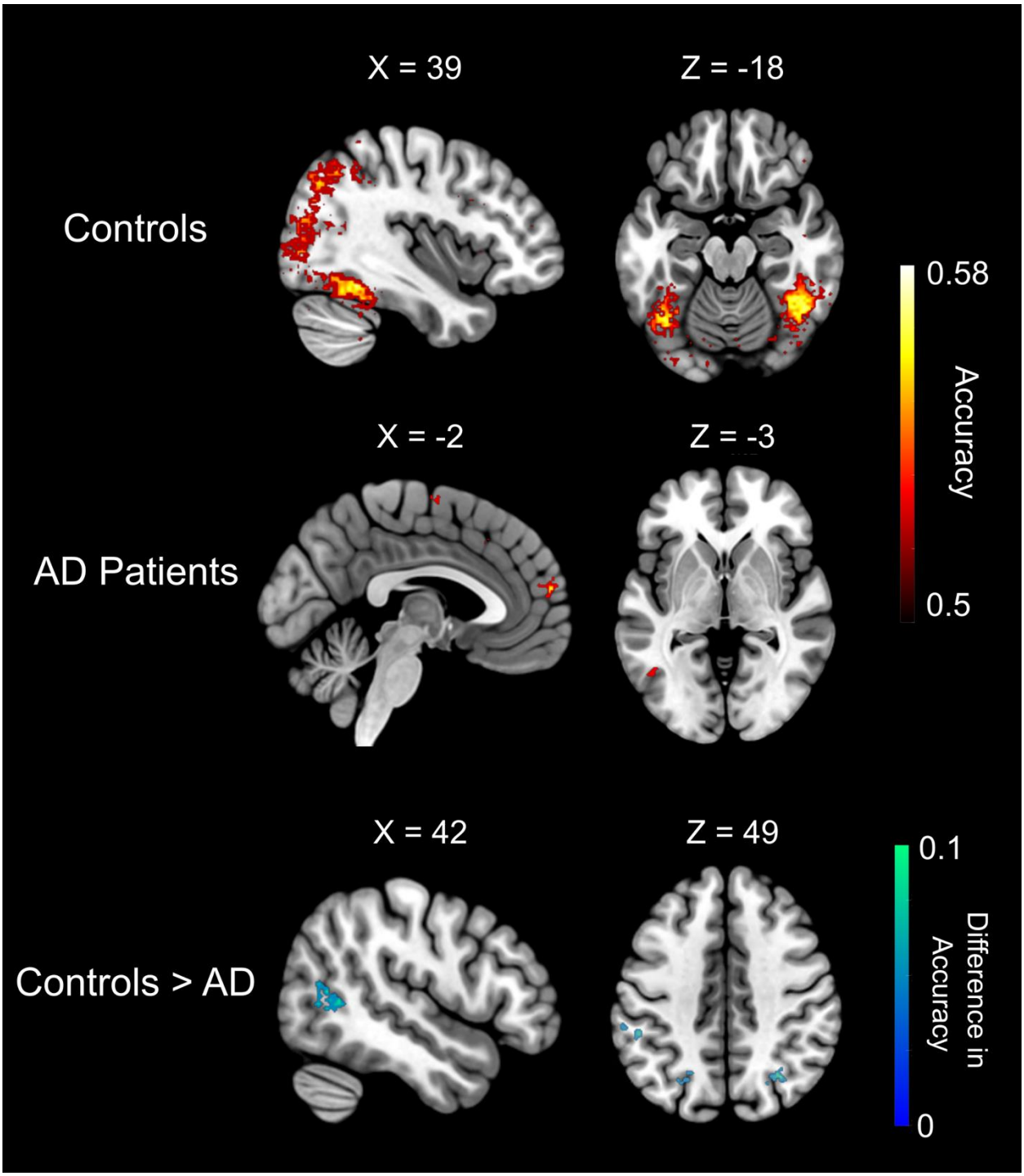


Figure 4.2. Healthy controls and AD patients both show sensitivity to visible faces. Searchlight decoding analysis reveal regions across occipital, fusiform and medial pre-frontal regions where face information could be decoded, indicating that face stimuli were processed by participants. Healthy controls showed significantly better face decoding in regions of the visual and parietal cortex, but no difference was found in frontal cortices. Maps are thresholded at p <.05, corrected for multiple comparisons. Clusters are presented in **Supplementary Table 4.2** and **Supplementary Table 4.3**.

gyrus and dorsolateral prefrontal cortex (dlPFC) was associated with awareness of the face stimuli (**Figure 4.3**, *top*; **Supplementary Table 4.4**). In contrast, a markedly smaller number of voxels across motor and prefrontal regions showed this increase in the AD group, even at a liberal threshold of $p < .05$ uncorrected (**Figure 4.3**, *middle;* **Supplementary Table 4.4**). Most importantly to the endeavour of distinguishing differences in neural markers of awareness in AD, however, is the contrast between awareness-related activity in the control and patient groups. Here, awareness-related activity in the left fusiform gyrus, right anterior insula, and right dlPFC was larger in controls compared to AD patients (**Figure 4.3**, *bottom*; **Supplementary Table 4.4**)

4.3.3 Decoding analyses reveal degraded visibility information in AD

Although qualitative examination of the neural response to conscious vs. unconscious stimuli revealed larger frontal activation in controls than AD patients (**Figure 4.3**), this analysis does not allow for valid statistical inference due to the liberal uncorrected threshold used. To overcome this limitation and to ask whether AD patients show a statistically significant degradation in neural correlates of awareness compared to controls, I performed a searchlight decoding analyses throughout the entire brain to identify multivariate neural patterns that correlated with visual awareness.

In keeping with the univariate visualisations, searchlight decoders trained to classify visible and subliminal trials throughout the brains of healthy controls showed significantly above chance (50%) performance throughout visual, parietal, and frontal cortices (**Figure 4.4**, *top*; **Supplementary Table 4.5**). In AD patients, decoders were also able to isolate representations of visibility across fusiform and prefrontal regions (**Figure 4.4**, *middle*; **Supplementary Table 4.5**), suggesting that with sensitive multivariate techniques, visual and frontal neural patterns associated with visual awareness are still observable in patients with mild to moderate AD. Despite this, decoding of awareness-related activity was significantly better for healthy controls throughout visual, parietal, and prefrontal regions (**Figure 4.4**, *bottom*; **Supplementary Table 4.6**), indicating degraded awareness-related information in the brains of AD patients.

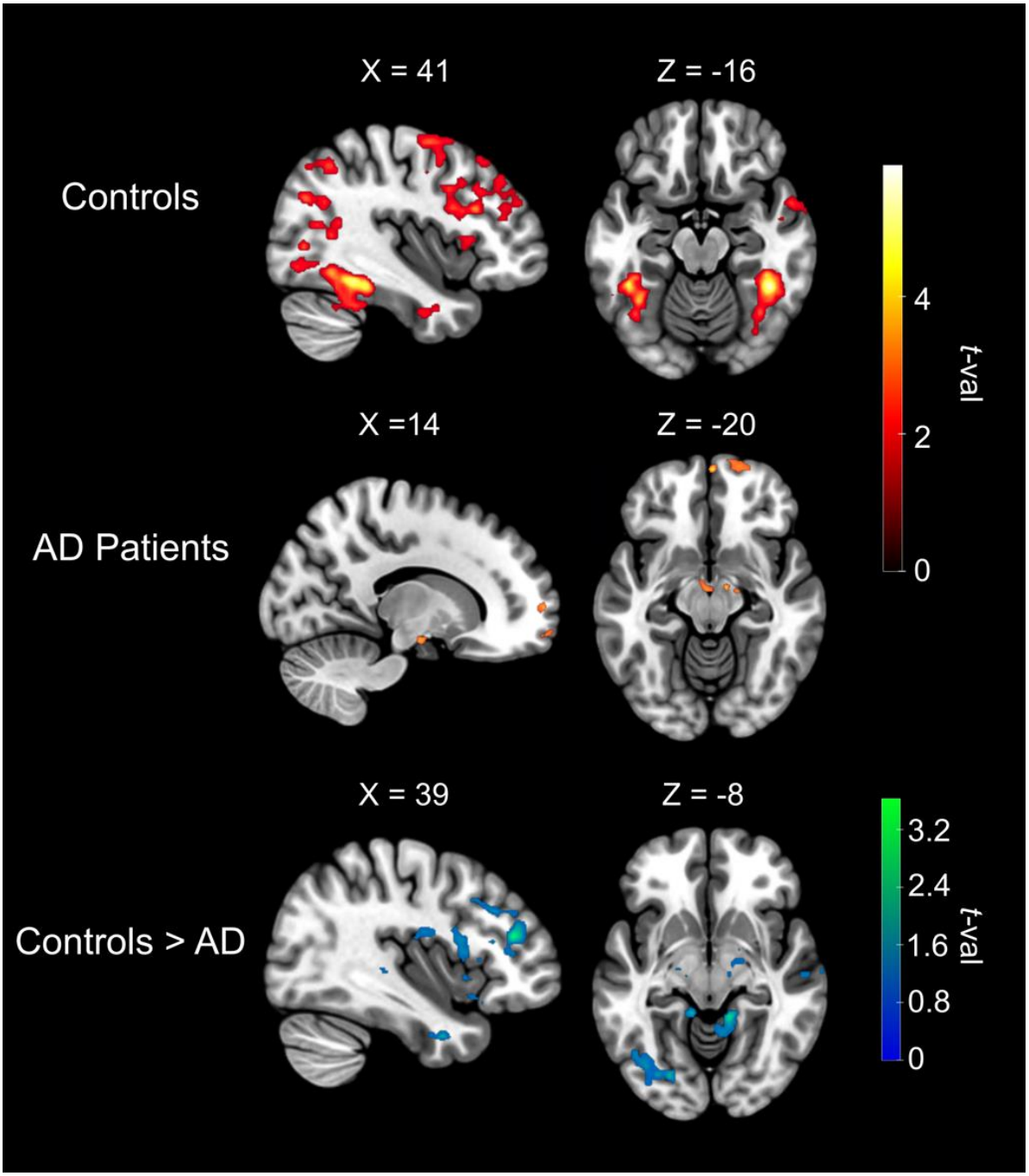


Figure 4.3. Qualitative visualisation of univariate responses to conscious faces in AD patients and controls. Contrasting trials where participants viewed visible faces vs. subliminal faces reveals neural activity associated with visual awareness. In controls, this contrast was associated with activity in the fusiform gyrus and prefrontal cortex (top panel). Only a limited number of prefrontal voxels were identified as increasing for visible vs. subliminal faces in the AD group, even at the liberal threshold of p <.05 uncorrected (middle panel). Notably, when comparing the activity associated with awareness across the two groups, the controls showed greater frontal activation (bottom panel). Maps are presented at p <.05 uncorrected. Clusters are reported in **Supplementary Table 4.4**.

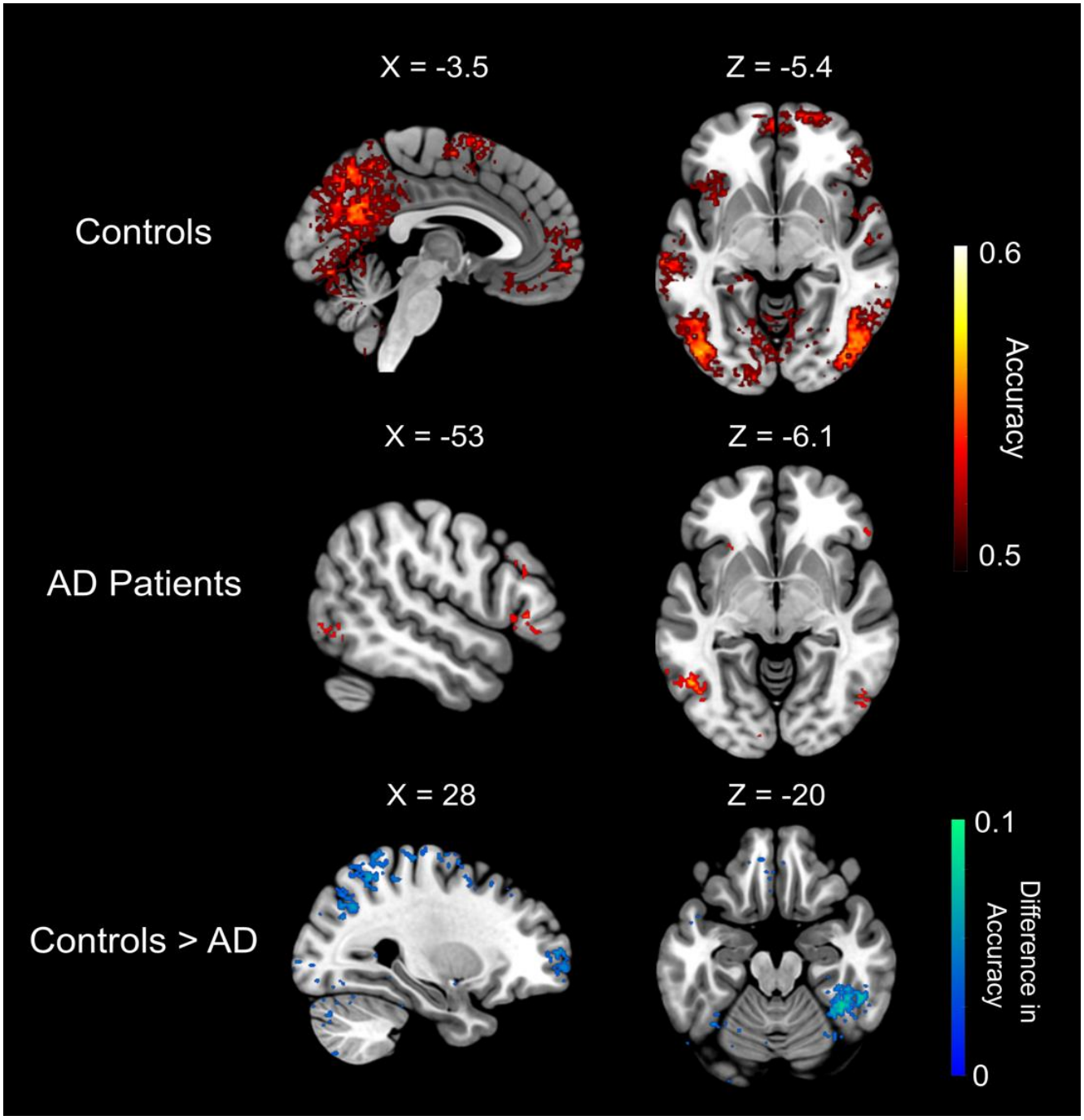


Figure 4.4. Decoding analyses reveal existing, albeit diminished, NCCs in AD. Searchlight decoding analyses where decoders were trained to classify visible and subliminal face trials in searchlights throughout the brain. In healthy controls, the decoder could classify visible from subliminal trials significantly above chance throughout visual, parietal, and frontal regions (top). In AD patients, decoders were also successful in identifying neural patterns associated with awareness in visual and prefrontal regions (middle). In keeping with univariate analyses, there was a significant increase in decoding accuracy for controls vs. AD patients, indicative of diminished neural markers of awareness in AD. Maps are thresholded at $p < .05$, corrected for multiple comparisons. Clusters are reported in **Supplementary Table 4.5** and **Supplementary Table 4.6**.

4.3.4 Global Ignition as a Biomarker for Awareness in Severe AD

Although I have demonstrated differences in the neural correlates of visual awareness in mild-to-moderate AD, a question remains with regards to how the neural responses associated with awareness are affected in latter, more severe, stages of AD. To explore this, I again produced visualisations illustrating increases in univariate neural activity for conscious vs. unconscious stimuli in 4 patients with severe AD and compared these visualisations with the patients' scores on clinical scales of cognitive functioning and disease severity. To do this, I first contrasted visible face trials with subliminal face trials at the single subject level for each patient. In all but one patient, no neural activity related to conscious perception of faces was identifiable. For visualisation purposes, these contrasts are illustrated at different weak thresholds (**Figure 4.5**; *subjects 2-4*). One participant with severe AD, however, still exhibited a classic 'ignition' like neural profile associated with a contrast of visible and subliminal faces, extending across bilateral fusiform, parietal, and medial and lateral pre-frontal cortices (**Figure 4.5**; *subject 1*; **Supplementary Table 4.7**).

Strikingly, patient #1 was the only patient in the severe group to be able to speak in intelligible sentences. Moreover, they could maintain some eye contact and offer appropriate responses to staff. Importantly too, their general cognitive screening (resulted in a 'moderate-severe' rating (sMMSE = 12/30), and their dementia-specific clinical ratings were both 'moderate-severe' as well (GDS = 6; CDR = 14). This contrasts with patients #2-4, who were non-verbal and unable to follow command or engage with any cognitive assessments (sMMSE = 0/30). These patients showed limited eye contact, and while they did offer occasional moments of emotional expression, these seemed unrelated to external circumstances. Family and carers of patients #2-4 reported limited or no evidence of the patients recognising those closest to them.

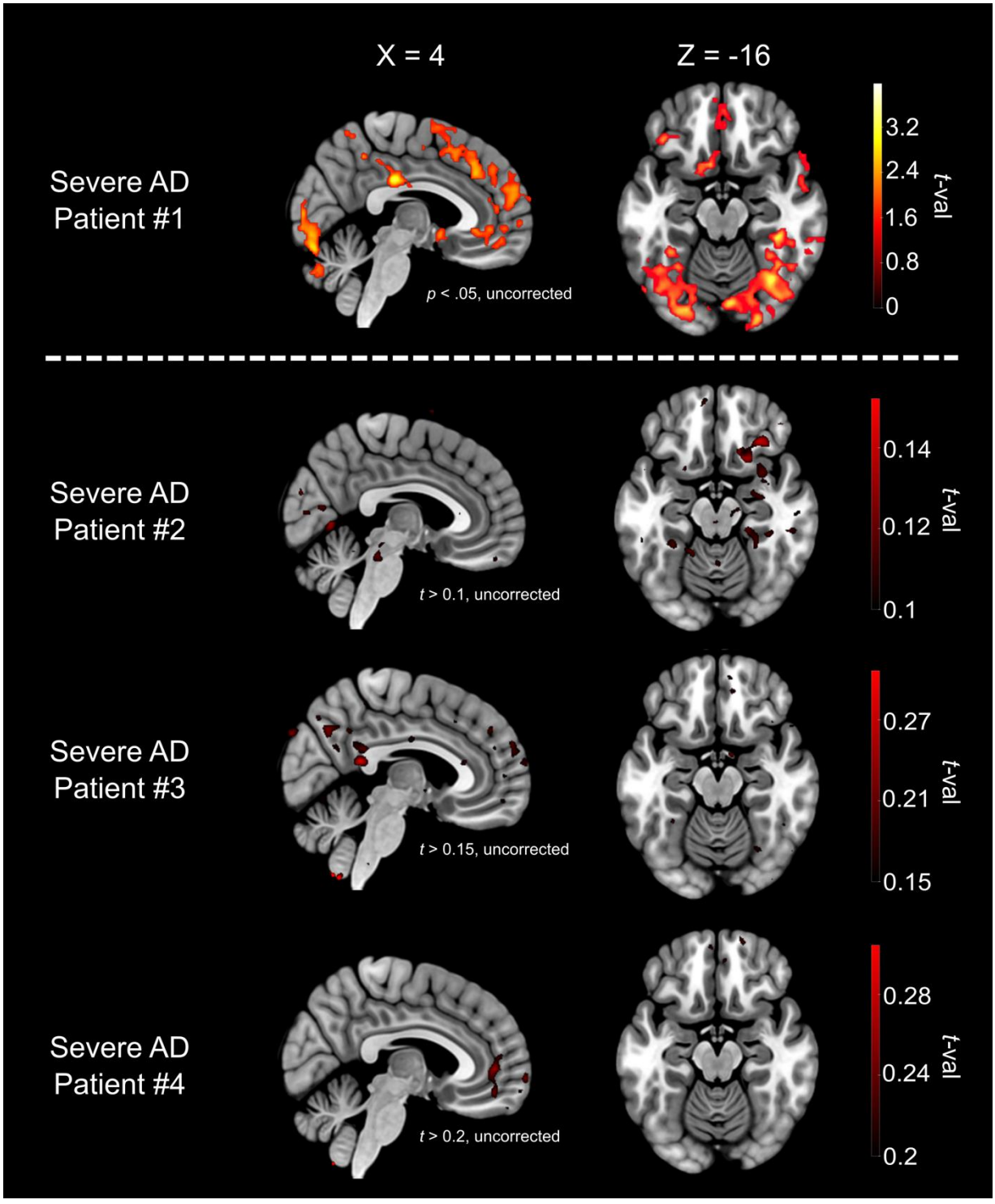


Figure 4.5. Severe AD leads to abolition of responses to visible faces. For single subjects with severe AD, contrasting trials where patients viewed visible faces vs. subliminal faces illustrates the abolition of neural markers of awareness in most patients (#2 – #4). For visualisation purposes, maps for these patients are presented at arbitrary low thresholds. One patient with severe AD (#1) showed awareness-related activity across posterior and frontal regions. Notably, patient #1 was the only patient with severe symptoms to maintain some communicative and cognitive abilities (see main text). All maps are qualitative visualisations of neural activity, produced using a lenient uncorrected threshold of $p < .05$ uncorrected to protect against false negatives. Clusters are reported in **Supplementary Table 4.7**.

To assess whether an absence of single subject clusters in these single-subject qualitative maps was unique to subjects with severe AD, I extracted functional regions of interests (ROIs) from the equivalent group-level contrast in healthy controls. Two ROIs were created. First, an ROI that comprised of the right fusiform cluster (**Supplementary Table 4.4**). Second, by combining the two largest frontal clusters (ACC and dlPFC; **Supplementary Table 4.4**), I created a functional ROI which spanned lateral and medial portions of the PFC. Both these ROIs exhibited sensitivity to the visibility of face stimuli in healthy controls. Extracting the peak t-value in both ROIs for each single subject contrast of visible > subliminal faces allowed me to compare the single subject responses to visible versus subliminal faces across all groups (**Figure 4.6**). As illustrated in **Figure 4.5**, all but one severe AD patient registered t-values near zero in the right fusiform (**Figure 4.6**, top-left) and PFC (**Figure 4.6**, top-right), suggesting no increase in activity for visible versus subliminal faces in these patients. However, across both ROIs, the Control and Mild AD groups also contained subjects that failed to exhibit a single-subject increase in activity for visible versus subliminal faces. As such, despite exclusively finding ignition-like activity in the only severe AD patient to exhibit cognitive and communicative abilities, the null findings in the other severe AD patients cannot conclusively be attributed to an absence of such abilities, as null results also occur in a proportion of healthy individuals (**Figure 4.6**, top). Additionally, within the same ROIs, the t-values computed from a visible face versus scrambled mask contrast were clustered around 0 for all three patients with severe AD who showed no response to the visibility of faces (**Figure 4.6**; bottom). This is an indication that an absence of neural markers of awareness in these patients may have been due to a global failure to respond to face information at all, rather than a specific dissolution of awareness-related mechanisms.

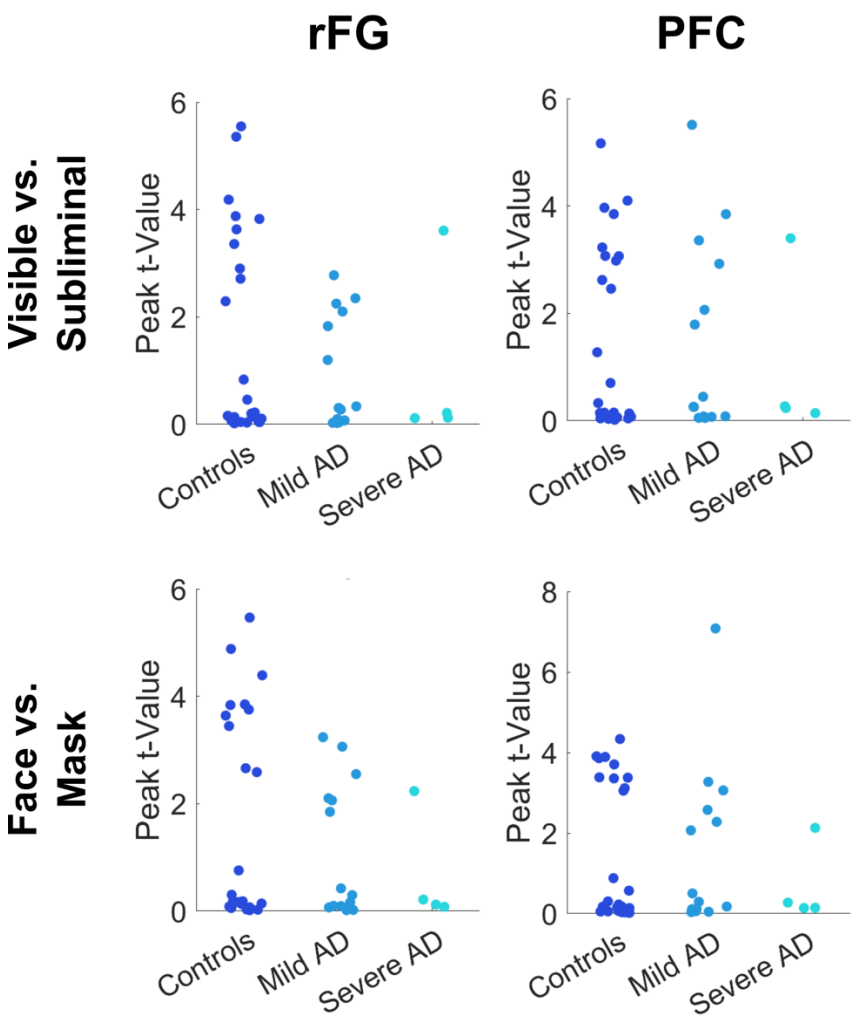


Figure 4.6. Single subject contrasts of visible vs. subliminal faces in functional ROIs. *Top row:* Peak t-values from functional ROIs in fusiform and prefrontal regions reveal subjects within each control and patient group who display no increase in activity for visible vs. subliminal faces at the single subject level. As such, a failure to reach statistical significance in such an analysis cannot conclusively be attributed to a patient's AD diagnosis. *Bottom row:* Peak t-values from the same ROIs for the visible face vs. scrambled mask contrast. Only one severe AD patient showed sensitivity to faces, suggesting that null findings of ignition profiles may be a result of a global insensitivity to the paradigm, rather than being tied to patients' awareness.

4.4 Discussion

People with AD suffer from debilitating cognitive deficits and exhibit an impaired ability to communicate with family members and caregivers. For caregivers and patients, degradations in awareness are the most distressing aspect of the disorder (Rice et al., 2019). In the present study, I provide the first systematic exploration of NCCs in AD in an initial attempt to characterise the subjective experience of AD patients. In a masking paradigm, I found that, while patients with mild-moderate AD still show sensitivity to faces, they exhibit degraded information pertaining to whether they consciously perceived a face or not. This reduction in sensitivity to the visibility of stimuli was present across visual and

frontoparietal regions of the brain. Visualisations of univariate activity also illustrated larger frontal responses to conscious versus unconscious information in controls compared to AD patients. Finally, analyses of patients with severe AD revealed qualitative single-subject NCCs that corroborated clinical scores of disease severity and cognitive functioning. Although this final analysis suggests that 'ignition' processes may be a useful biomarker for awareness in AD in future, a control analysis showing an absence of ignition processes in individual healthy controls suggests this interpretation should be taken with caution.

The term 'disorder of consciousness' is usually applied to conditions where arousal (i.e., the *state* of consciousness) is negatively affected, such as comas and vegetative states (Giacino et al., 2014). However, this negates the impact that alterations in the *content* of consciousness play in cognitive disorders such as AD (Huntley et al., 2021). To improve wellbeing of AD patients, it is not sufficient to ask to what extent they are awake. Instead, we must examine what the content of their awareness appears to be – how do they subjectively experience the world? According to GNWT, whether stimuli are detected depends on propagation of perceptual information from sensory cortices to frontoparietal regions (Dehaene and Changeux, 2011; Mashour et al., 2020). In line with this, reportable experiences have been shown to rely on sustained neural responses in the PFC of monkeys (van Vugt et al., 2018). Moreover, failures to detect stimuli are seemingly driven by a loss of information transmission from both low- and high-level visual regions to frontoparietal areas (van Vugt et al., 2018). Taking this into consideration, the finding of reduced awareness-related information in frontal regions of AD patients may reflect a decrease in the amount of sensory information reaching awareness in individuals with AD.

Neurodegeneration in AD results in the progressive disconnection of cortical networks often associated with awareness and selfhood, such as the default mode and frontoparietal networks (Buckner et al., 2005; Hallam et al., 2020; Menon, 2011; Weiler et al., 2016). These networks are comprised of regions commonly associated with the content of consciousness, such as lateral and medial frontal cortex, anterior cingulate

cortex, and insula (Chapter 2; Bor & Seth, 2012; Dijkstra & Fleming, 2023; Lau & Rosenthal, 2011). Metacognitive deficits in AD are also associated with similar regions, further indicating disruptions to brain areas responsible for self-awareness and insight in AD (Hallam et al., 2020). Dysfunction of large-scale cortical networks in AD may have a destabilising effect on the representation of perceptual content in the global workspace, resulting in weaker frontal responses to visible stimuli. Interestingly, multivariate decoding analyses highlighted significantly improved decoding of stimulus visibility in visual regions for controls compared to patients. This suggests that sensory representations were of greater precision in controls. One mechanism that could explain the improved awareness decoding in both visual and frontal regions is that in AD, sensory representations may lack the stability or precision required to undergo ignition into the global workspace. This would also explain why decoding of perceptual content (i.e., face vs. mask decoding) was more successful in the visual cortex of controls compared to patients. It should be noted, however, that other models of AD progression have been proposed, particularly in relation to a global deterioration of myelination (Bartzokis, 2004). In this case, the effect of AD on consciousness may have less to do with neural circuits that typically support perceptual experience, and more to do with a breakdown of neural functioning more generally. Still, the core claim of this chapter, that AD should be considered a disorder of consciousness, is not impacted by such a position – it remains agnostic as to whether AD is 'only' a disorder of consciousness, or whether it is a global cognitive disorder. Indeed, it is far more likely, given the widespread destruction of neural regions, that the latter be true. What is important is that the conscious experiences of AD patients begin to be considered from a neuroscientific perspective. In any case, to isolate the processes responsible for degraded neural correlates of perceptual experience in AD, further neuroimaging studies involving direct measures of functional and structural connectivity during near-threshold tasks are needed.

Given the difference in ignition-like signatures between AD patients and controls, it may be appealing to classify frontal activity in perceptual threshold tasks as a neurobiological marker of awareness in AD. Indeed, when visualising the neural activity of individual patients with severe AD, the presence of frontal responses to visible stimuli corroborated

clinical scores, with ignition-like activity only present in the one patient that continued to exhibit communicative abilities. However, null findings were also present within the healthy control group, indicating that frontal activity could not reliably distinguish conscious and unconscious processing at the single subject level. To mitigate discomfort for patients partaking in the experiment, experiment duration was kept short and trial numbers were limited, constraining the power of single-subject analyses to detect effects in both patients and controls (Mei et al., 2022) and likely contributing to these null findings. Additionally, there is a concern that null findings in severe AD patients may reflect an inability to attend to the stimuli, rather than an absence of awareness. This was supported by a further control analysis, which showed how the patients for whom no marker of awareness was found also exhibited no response to faces at all, indicating a global failure to respond to the experimental paradigm. For the mild-moderate AD patients, however, the sensitivity to faces and visibility in decoding analyses suggests that patients in this group were visually attending and responsive to the stimuli throughout. Overall, I provide novel evidence in support of the notion that awareness may be degraded in AD at the group level. However, if neural activity in perceptual tasks is to provide a neurobiological marker of awareness in single AD patients, the challenge of recording larger datasets from attentive individual patients must first be overcome.

Behavioural tests of perceptual awareness in AD have thus far been inconclusive. The tools used to explore sensory consciousness in healthy populations (e.g. trial by trial awareness reports, subjective ratings, psychophysical paradigms) are not viable in studies with AD patients who lack basic cognitive functioning. As such, drawing conclusions from such studies can be difficult. For instance, some studies that have argued for maintained perceptual awareness in dementia have monitored facial expressions and physiological measures in response to unpleasant and pleasant stimuli (Asplund et al., 1991; Beach et al., 2021). In these cases, facial expressions, heart rate, and respiratory rate all remain sensitive to the pleasantness of stimuli, ostensibly indicating intact sensory awareness (Asplund et al., 1991; Beach et al., 2021; O'Shaughnessy, 2019). However, different metrics of facial expression have shown limited agreement (Asplund et al., 1995) and physiological markers do not necessarily

track subjective emotional experiences (Taschereau-Dumouchel et al., 2018, 2020). The paucity of viable behavioural measures of perceptual awareness in AD motivates studies like ours, which seek to identify objective neural characteristics of AD that can indicate differences in awareness. However, the approach of using candidate NCCs as a test bed for awareness in AD must also be used with caution. Identifying the neural basis of consciousness is still an ongoing research programme and no single NCC has been identified. Indeed, it is likely that different NCCs pertain to different aspects of awareness (Seth and Bayne, 2022). As such, reduced decoding of visibility in the frontal cortex of AD patients should not be used to classify patients as categorically unaware. GNWT is, after all, a cognitive theory of consciousness (Baars, 1988; Seth and Bayne, 2022), so it may be that reduced ignition in AD reflects an impaired ability to report or evaluate perceptual experiences, rather than an absence or degradation of experience itself. It would be ethically egregious to deny awareness to a patient who still maintained a subjective consciousness, so any reverse inference based on hypothesised NCCs should be caveated with the acknowledgement that results may reflect only one small component of patients' perceptual experiences (Bayne et al., 2016; Pérez et al., 2024). Thus, although I provide evidence in favour of a dysfunction of awareness in AD, future studies will be needed to fully characterise the different dimensions of awareness impacted by the disorder.

In summary, I provide the first neurobiological test of awareness in patients with mild-moderate and severe AD. I find that mild-moderate AD patients exhibit reduced decoding of stimulus visibility across visual and frontoparietal cortices compared to controls. Additionally, single subject visualisations of four patients with severe AD showed an ignition like effect in the only patient to maintain communicative abilities, although control analyses suggested it would be premature to use such effects as biomarkers of awareness in AD. When interpreted in a GNWT framework, my results provide evidence for a degraded or compromised perceptual awareness in patients with AD and motivates further investigation into the dimensions of consciousness impacted by the disorder.

4.5 Supplementary Material

	Control (n= 26)	Mild (n = 14)	Severe (n=4)
AGE (SD)	75.88 (5.49)	75.92 (6.01)	84.75 (8.14)
MALE/FEMALE	10/16	7/7	1/3
CDR (Median, IQR)	0 (0)	4.25 (4.75)*	18 (1)
GDS (Median, IQR)	1 (0)	4 (1.5)*	7 (0.25)
sMMSE	29.62 (0.59)**	22.42 (6.71)*	4.25 (5.68)

*n = 12 as 2 participants refused/would not participate with cognitive assessment.
**n = 21 as added sMMSE to controls later in the course of the study.

Supplementary Table 4.1 Demographic information. Including three different measures of cognitive abilities and dementia symptoms (CDR, GDS, sMMSE). As the CDR and GDS are ordinal scales, the interquartile range (IQR) is reported instead of the SD.

Contrast	Region	Peak Accuracy	No. Voxels	x	y	z
Controls	Fusiform Gyrus	0.59	4430	40	-50	-16
Controls	Lateral Occipital Cortex	0.58	2834	-40	-78	16
Controls	Inferior Frontal Gyrus	0.55	214	-44	8	26
Controls	Posterior Cingulate	0.55	202	-6	-72	16
AD	Supplementary Motor Area	0.59	82	8	-6	72
AD	Inferior Parietal	0.57	52	52	-34	50

AD	Medial Prefrontal	0.58	42	-4	58	14
AD	Lateral Occipital Cortex	0.57	23	-48	-60	-2

Supplementary Table 4.2. fMRI clusters associated with the Visible Face vs. Scrambled Mask decoding. XYZ coordinates are MNI coordinates of peak voxel.

Region	Peak Difference in Accuracy	No. Voxels	x	y	z
Intraparietal Sulcus	0.09	92	-26	-60	54
Posterior Superior Temporal	0.10	75	50	-60	10
Lateral Occipital Cortex	0.09	59	-42	-86	18
Intraparietal Sulcus	0.09	42	32	-64	42

Supplementary Table 4.3 fMRI clusters where controls show greater Visible Face vs. Scrambled Mask decoding than AD patients. XYZ coordinates are MNI coordinates of peak voxel.

Contrast	Region	Peak *t*	No. Voxels	x	y	z
Controls	Right dlPFC	4.72	2440	50	26	24
Controls	Anterior Cingulate	4.08	973	12	18	-18
Controls	Right Fusiform Gyrus	6.26	856	42	-44	-16
Controls	Left Fusiform Gyrus	5.02	778	-44	-42	-18
Controls	Superior Temporal Sulcus	4.05	630	44	-44	14

Controls	Inferior Frontal Gyrus	4.70	521	-38	10	28
Controls	Right Intraparietal Sulcus	3.89	411	36	-60	42
Controls	Anterior Temporal Lobe	3.89	397	46	12	-28
Controls	Left Intraparietal Sulcus	3.97	303	-36	-60	54
Controls	Corpus Callosum	3.32	261	-10	-28	26
AD	Right vmPFC	3.83	593	2	62	-12
AD	Premotor Cortex	3.01	354	-16	4	64
AD	Precentral Sulcus	3.53	326	40	0	36
AD	White Matter	4.26	317	36	-46	16
AD	White Matter	4.02	311	-22	-42	22
AD	Brainstem	5.94	304	-6	-14	-24
AD	Left vmPFC	3.22	275	-22	56	6
Controls > AD	Right dlPFC	3.68	1044	40	38	18
Controls > AD	Intraparietal Sulcus	3.84	800	-40	-40	40
Controls > AD	Anterior Temporal Lobe	4.08	736	62	6	-24
Controls > AD	Parahippocampal Cortex	3.53	486	12	-42	-10
Controls > AD	Superior Temporal Sulcus	3.38	485	52	-40	10

Controls > AD	Cerebellum	3.35	421	-42	-52	-46
Controls > AD	Premotor Cortex	2.81	366	52	-2	34
Controls > AD	Brainstem	3.44	350	8	-30	-16
Controls > AD	Anterior Cingulate	2.72	328	10	16	34
Controls > AD	Inferior Temporal	3.06	320	-54	-48	-18
Controls > AD	Fusiform Gyrus	3.50	318	-34	-36	-24
Controls > AD	Putamen	2.97	318	24	-10	0
Controls > AD	mPFC	3.32	303	26	40	16
Controls > AD	Cingulate Cortex	2.84	290	4	-14	38
Controls > AD	Thalamus	3.41	267	-14	-16	-4
Controls > AD	Posterior Cingulate	3.14	253	8	-36	40

Supplementary Table 4.4 fMRI Clusters associated with the Visible Face > Subliminal Face group-level contrasts and the subsequent between-group comparison. XYZ coordinates are MNI coordinates of peak voxel.

Group	Region	Peak Accuracy	No. Voxels	x	y	z
Controls	Right Fusiform Face Area	0.60	33,692	46	-54	-20
Controls	dmPFC	0.56	2337	18	46	36
Controls	dmPFC	0.56	481	-8	54	36

Controls	Superior Temporal Sulcus	0.55	329	52	-16	-10
Controls	vmPFC	0.55	240	0	26	-18
Controls	Anterior Insula	0.55	223	34	20	-12
Controls	Anterior Temporal Lobe	0.55	208	50	-8	-32
Controls	dlPFC	0.55	178	30	10	54
Controls	Anterior Temporal Lobe	0.55	177	-24	0	-30
AD	Posterior Temporal Sulcus	0.58	473	-58	-64	6
AD	Left Fusiform Gyrus	0.57	172	-38	-44	-18
AD	Occipital Cortex	0.58	152	-12	-84	-20
AD	Inferior Frontal Gyrus	0.58	84	56	22	0
AD	Amygdala	0.56	58	16	0	-24
AD	V5/MT	0.57	58	56	-66	2
AD	Right Fusiform Face Area	0.57	54	40	-54	-26

Supplementary Table 4.5 fMRI Clusters associated with significantly above-chance decoding in searchlight analyses in the control and AD groups separately. Decoders were trained to classify visible face trials from subliminal face trials and therefore reveal brain regions associated with visual awareness of faces. XYZ coordinates are MNI coordinates of peak voxel.

Region	Peak Difference in Accuracy	No. Voxels	x	y	z
Posterior Temporal Sulcus	0.10	531	56	-66	10
Right Fusiform Face Area	0.12	404	40	-48	-16
Intraparietal Sulcus	0.09	340	-28	-66	34
Parahippocampal Cortex	0.09	339	14	-66	18
Intraparietal Sulcus	0.09	334	46	-54	56
Motor Cortex	0.10	326	36	-16	70
Intraparietal Sulcus	0.09	266	42	-36	46
mPFC	0.09	140	24	68	-2
Precuneus	0.11	103	6	-46	50

Supplementary Table 4.6 fMRI Clusters associated with significantly higher decoding accuracies in the control group compared to the AD group. XYZ coordinates are MNI coordinates of peak voxel.

Region	Peak *t*	No. Voxels	x	y	z
Intraparietal Sulcus	4.11	13757	-26	-44	46
Anterior Insula	3.51	774	-26	32	4
Right dlPFC	3.40	1598	52	18	26
Paracingulate Gyrus	3.28	4511	14	8	44

Superior Temporal Sulcus	3.06	287	46	-18	-10

Supplementary Table 4.7 fMRI Clusters associated with the Visible Face > Subliminal Face single-subject contrast for patient #1, the only patient with severe AD who showed a statistically significant response to visible faces versus hidden faces. XYZ coordinates are MNI coordinates of peak voxel.

5. Towards a naturalistic neuroscience of social cognition

5.1 Introduction

Thus far, this thesis has approached questions of perceptual experience by studying individual subjects in isolated and artificial settings. Whilst this approach has succeeded in characterising a vast array of neural phenomena, it is limited with respect to understanding the nature of perceptual experiences in unconstrained and naturalistic settings, such as social interactions. This is particularly relevant to the neuroscience of consciousness when one considers that certain theories of consciousness, such as AST, make explicit claims regarding the role of social cognition in generating perceptual experience (Chapter 1.5; Graziano, 2013; Carruthers, 2009). Newly developed optically-pumped MEG systems (OP-MEG; Boto et al., 2018) offer a promising means towards assessing the interaction between social cognition and perceptual experience, since they enable participants to move freely whilst their neural activity is recorded. Before such studies are possible, however, the use of OPMs in complex, multi-person paradigms needs to be validated. This chapter describes the development of a novel naturalistic perspective-taking task, adapted from an artificial paradigm used in conventional MEG experiments, with the purpose of assessing whether wearable OPM systems can reliably detect known electrophysiological correlates of social cognition.

Theory of Mind - defined as the ability to infer the mental states of other people (Frith & Frith, 2007; Tomasello et al., 2005) – is a fundamental feature of human cognition. Although the brain basis of Theory of Mind is well-studied (Frith & Frith, 2007; Siegal & Varley, 2002), like studies of perceptual experience, most studies examining social cognition have been performed in tightly controlled settings that do not reflect real-world social interactions (Fan et al., 2021; Stangl et al., 2023). Here, I harnessed a fundamental component of Theory of Mind, perspective taking (Kessler & Rutherford, 2010; Seymour et al., 2018; Wang et al., 2016), to explore one aspect of social cognition in an ecological, real-world task while assaying the underlying neural dynamics using wearable OPMs (Boto et al., 2018).

Studying social cognition with artificial proxies of real-world settings can prevent findings from generalising to truly social situations. There is ample evidence to indicate that processing ecologically valid social stimuli may recruit, at least in part, distinct neural mechanisms compared to the processing of 'social' 2D images (Stangl et al., 2023; Fan et al., 2021). For instance, the classic face-sensitive N170 ERP does not respond to realistic 3D faces presented in virtual reality (Sagehorn et al., 2023), challenging the existing consensus that the N170 reflects the neural correlate of face processing (Rossion, 2014). Additionally, gaze-behaviour seemingly operates differently across natural and artificial tasks (Hayward et al., 2017) and is supported by distinct neural correlates across different settings as well (Pönkänen et al., 2011). Even if neural mechanisms are unchanged in ecological settings, more naturalistic tasks often evoke stronger behavioural and neural responses. For instance, real life objects attract more attention (Gomez et al., 2018) and are remembered better (Snow et al., 2014) than artificial counterparts. Moreover, live interactions viewed over video elicit greater neural response in social cognition brain regions than watching a pre-recorded video of the same interaction (Redcay et al., 2010). However, even live interactions over video do not evoke the same neural mechanisms as face-to-face interactions (Fan et al., 2021; Hirsch et al., 2017).

Newly developed OP-MEG offers the potential to record neural activity with a high signal-to-noise ratio from mobile participants (Boto et al., 2018; Seymour et al., 2021), lending itself to the study of social cognition in naturalistic settings. OP-MEG sensors operate at room temperature, which obviates the need for large cooling systems used in conventional MEG (Boto et al., 2018), meaning that sensors can be fitted close to the scalp of participants, improving signal-to-noise, and enabling participants to move freely during experiments (Boto et al., 2018; Seymour et al., 2021). To date, however, OP-MEG systems have largely been used in proof-of-concept studies exploring basic electrophysiological responses in sensorimotor tasks. For instance, OP-MEG has been used to identify beta band activity during finger-tapping (Boto et al., 2018), auditory evoked fields (Borna et al., 2017; Seymour et al., 2021), and visual gamma oscillations (Iivanainen et al., 2020). Despite tantalising progress, no studies have yet validated the use of OP-MEG in social cognition tasks, likely owing to their increased complexity.

To address this, the present study adapted an existing perspective-taking paradigm (Kessler and Rutherford, 2010; Wang et al., 2016; Seymour et al., 2018) to create an analogous “real-world” version that could be examined in OP-MEG alongside a matched computerised task. The task differentiates between perspective-tracking (monitoring what is or isn’t perceived by another individual) and perspective-taking (imagining the world from another’s visuospatial perspective). Perspective-tracking is thought to rely on the relatively simple processing of others in relation to the environment (Kessler & Rutherford, 2010; Michelon & Zacks, 2006), while perspective-taking is understood to be a more complex process, perhaps unique to humans, which requires the capacity to imagine another’s viewpoint (Firth and Frith, 2007; Tomasello et al., 2005). Previous studies using this paradigm have shown that reaction times for perspective-taking decisions are modulated by the angle of rotation required to occupy an avatar’s perspective, whereas rotation has no effect on perspective-tracking decisions (Kessler & Rutherford, 2010; Kessler & Wang, 2012; Surtees et al., 2013; Van Elk & Blanke, 2014). This is explained by the fact that perspective-tracking decisions merely require a geometric computation of the sort required to infer an avatar’s line of sight. In contrast, perspective-taking decisions require the participant to assume the avatar’s perspective, which is increasingly difficult

the further from their own perspective it is. In two MEG studies, theta band activity in the right temporoparietal junction (rTPJ) and medial prefrontal cortex (mPFC) has been shown to correlate with this behavioural effect (Wang et al., 2016; Seymour et al., 2018). This provides a useful paradigm to validate the use of OP-MEG in naturalistic social settings since it allows me to test if OPMs are sensitive to a robust social cognition effect that has been replicated in conventional MEG. Moreover, by developing a naturalistic version of the task, I can examine the present suitability of OPMs for use in ecologically valid social situations.

In a modified perspective taking task, I recorded OP-MEG signals from three participants with the aim of replicating previous findings suggesting theta band activity in social cognition regions contributes to perspective-taking decisions. To validate OPMs for use in real world scenarios, I tested this in both naturalistic and computerised versions of the paradigm. To pre-empt my results, in both naturalistic and computerised versions of the task, I found no evidence for theta band activity underlying decision-making in either perspective-taking or perspective-tracking decisions. There was, however, some evidence for a role of parietal alpha in perspective-taking decisions, although this was only present in the computerised version of the task. Finally, epoching trials with respect to response – rather than stimulus – onset did not reveal any further effects.

5.2 Materials and Methods

5.2.1 Participants

Three female participants aged 43, 25, and 29 took part in the experiment. All participants provided written consent and ethical permission was granted by the University College London Research Ethics Committee.

5.2.2 OPM Data Acquisition

OP-MEG data were acquired in a Magnetically Shielded Room (MSR; Magnetic Shields Ltd) located at University College London. The room has internal dimensions of 438 cm x 338 cm x 218 cm and is constructed from two inner layers of 1 mm mu-metal, a 6 mm copper layer, and then two external layers of 1.5 mm mu-metal.

A combination of dual and tri-axis OPMs (QuSpin Inc., QZFM second and third generation) were used in the study. Participant-specific 3D-printed "scanner-casts" (Boto et al., 2017) were designed using each participant's structural MRI scan (Chalk Studios), see **Figure 5.1A**. The scanner-cast was designed to keep the sensors in slots fixed in relation to the brain during participant movement, and to minimise co-registration errors. As each scannercast was printed by first creating a 3D image in the same coordinate space as the participant's structural MRI brain scan, a sensor's position and orientation could be calculated offline relative to the slot in which it was placed. Sensor position was set as the centre of the cell of the OPM sensor, which was slightly offset from the physical centre. As shown in **Figure 5.1A,** custom plastic clips were used to arrange the OPM sensor ribbon cables for effective cable management. In addition, the larger cables were organised into bundles and fixed to a wearable backpack, to facilitate participant comfort during movement.

As scannercasts were made specifically for each subject, the number of OPM sensors they were designed to hold varied. In two subjects, 54 sensors were used, and in the third subject 56 were fitted. The sensors were arranged to cover the whole head in an even manner.

Before the start of the experiment, the MSR was degaussed to minimise the residual magnetic field in the MSR. Before the start of each experimental run, the OPM sensors were calibrated using a manufacturer-specific procedure. This involves energising coils within an OPM to produce a known field, and the output of the sensor is then measured

and calibrated to this known field. Data were recorded using a 16-bit precision analogue-to-digital converter (National Instruments) with a sample rate of 6000 Hz.

5.2.3 Experimental Procedure

The experiment consisted of both a naturalistic and a computerised session, which were counterbalanced in order across participants. In the naturalistic condition, the participant was joined in the MSR by two confederates, one in blue and one in green, who were unknown to the scanned participant. The confederates were sat at a white, circular table with a 90cm diameter and were positioned exactly opposite one another throughout the entire experiment (**Figure 5.1**, middle). The participant sat in the centre of the MSR facing the table head on. Relative to the participants' perspective, the confederates could occupy four positions around the table (45°, 135°, 225°, 315°), which when collapsed over left and right, created two positions relative to the participant: a small angle of rotation (45° left or right) and a large angle of rotation (135° left or right) (**Figure 5.1**, middle and right). On top of the table sat a cardboard fixture which held four red LED lights at each of the four possible confederate positions and an occlusion screen which blocked the confederates from viewing the light directly opposite their current position. The LED lights were controlled by custom MATLAB (Mathworks) scripts and an Arduino Uno controller.

The participant performed both perspective-taking and perspective-tracking trials in alternating blocks of 32 trials, with 1 practice block and 10 experimental blocks each. Before each block, the participant was instructed by audio cue which task to perform. In the perspective-taking trials, an auditory cue ("blue" or "green") would indicate which confederate's perspective the participant should adopt (**Figure 5.1B**). Following the cue, an LED light on either the left or the right of the cued confederate would turn on. Using a button box, the participant reported whether the light was on the left or right of the cued confederate (LR decision). Confederates remained in their positions in miniblocks of 8 trials, before a small pause indicated by audio cue allowed them to adopt new positions around the table. Within each miniblock, each possible combination of cued confederate

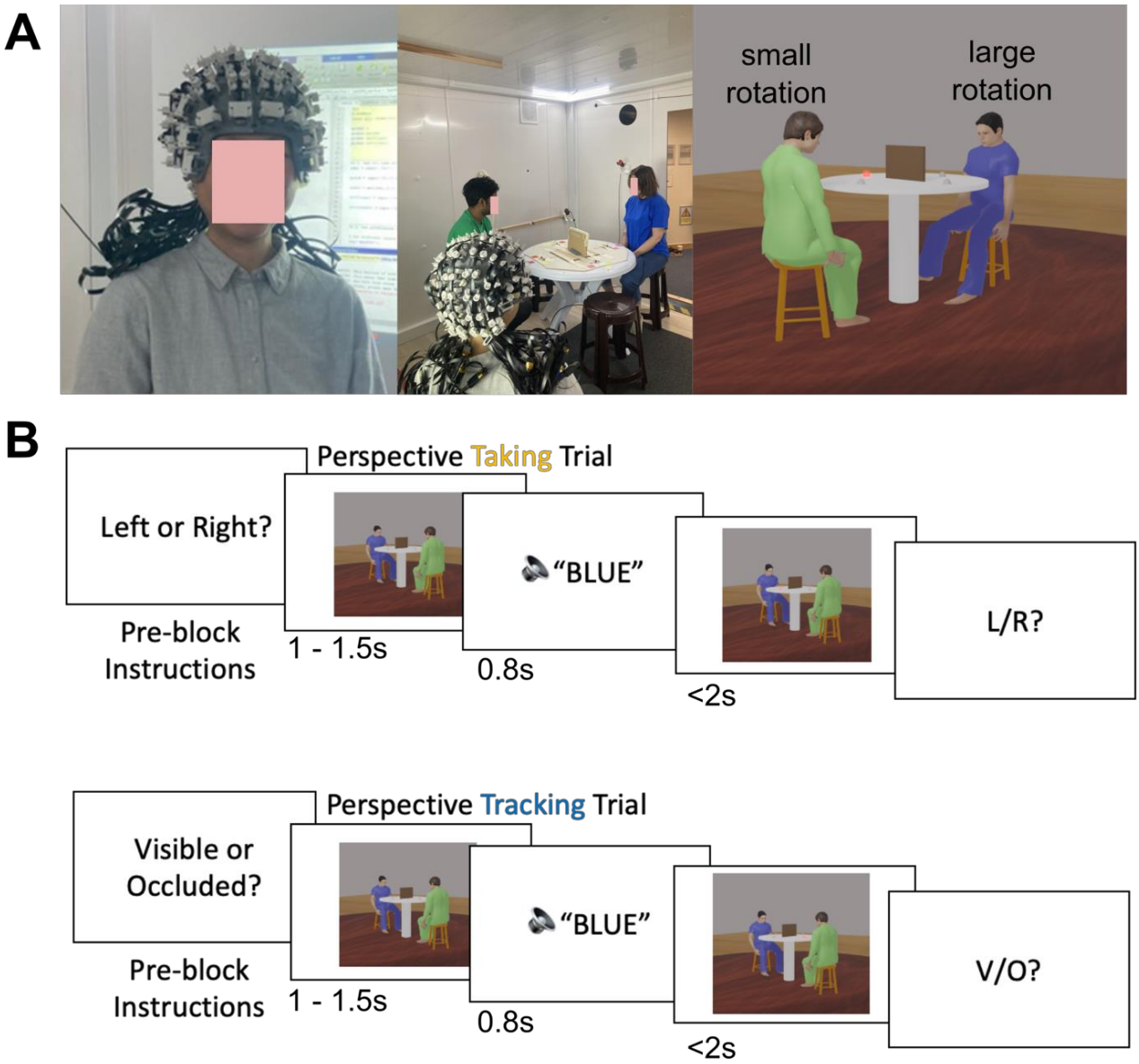


Figure 5.1. Experimental set up and design. A: Left: A participant with their scannercast and OPM sensors. Middle: Experimental Set up with participant and two confederates. Right: Example artificial stimulus from the computerised session. B: Task structure for perspective-taking (top) and perspective-tracking (bottom) tasks.

and light position (i.e., green-right, green-left, blue-right, blue-left) was activated twice in a randomised order. The order of confederate positions was randomised across blocks, and in every block each position was occupied exactly once by both confederates. The perspective-tracking trials followed the same structure, the only difference being the decision the participant had to make. Here, the LED light either in front of or hidden from

the cued confederate was illuminated. The participant was then asked to report whether the light was visible or occluded from the cued participant (VO decision). Again, each possible combination of cued confederate and light position (i.e., green-visible, green-hidden, blue-visible, blue-hidden) was presented twice in each miniblock, with 4 miniblocks per each block.

The computerised session matched the naturalistic session in all aspects except that the participant sat alone in the MSR and viewed confederate avatars on a projector screen. The artificial stimuli consisted of two avatars, one in green and one in blue, sat in the same four positions around a white table relative to the participant's perspective. The table again contained red lights in front of each position (**Figure 5.1A,** right). To match the experience of the naturalistic session as closely as possible, the avatars and table were consistently present on the screen, with the illumination of the light the only activity at trial onset. This prevented any confound caused by the entire stimulus appearing at trial onset. Again, to match the naturalistic session, avatars stayed in position for miniblocks of 8 trials, and the tasks alternated between perspective taking and tracking in blocks of 32 trials.

To ensure the confederates were fixating the light display in the same manner as the artificial avatars, they were instructed to perform a 1-back task on the light position, pressing a button every time a light was illuminated in the same position in a row. Unbeknownst to the confederates, however, due to limitation in the number of recording channels available in the MSR, only data from one confederate was recorded.

Participants had a break from the experiment of at least an hour between the two sessions. The full procedure resulted in 640 trials for both the computerised and naturalistic sessions. Within each session, there were 160 trials per each condition of interest (LR-small, LR-large, VO-small, VO-large).

5.2.4 Paradigm Development

The paradigm described above was developed from the computerised perspective-taking task used in Wang et al. (2016) and Seymour et al. (2018). To modify the task for a naturalistic setting, it was necessary to make several modifications to the original paradigm. This process of paradigm development consisted of iterative changes to the computerised task with subsequent pilot tests to ensure the original behavioural effect of rotation angle on perspective-taking reaction time was still present.

The original task consisted of only a single avatar, with the participant asked to judge whether an illuminated light was either visible/hidden or left/right of the avatar. At the onset of each trial, an image appeared with the avatar sat at a different position at a table that contained the illuminated light. However, in a real-life task, the avatar (i.e., confederate) would not 'appear' in different locations at the onset of each trial, but would be visible before, during, and between trials. This visual information could contribute to pre-trial computations regarding the positions of the lights with respect to the confederate. For example, the participant may have the thought "if I see this particular light turn on, I will press 'left' as quickly as possible." To prevent any pre-trial computations that may have aided participants' decisions, I included a second avatar in the design, sat opposite the original avatar. This way, the participant would not know which avatar was relevant before trial onset and would thus be unable to compute task-relevant information prior to the trial starting and the relevant avatar being cued. The next alteration was to keep the avatars' positions fixed for miniblocks of 8 trials, rather than changing on each trial. This was to limit the amount of movement the confederates would have to make in the naturalistic task. Finally, I modified the intertrial screens to keep the avatars and table visible to the participant, such that at trial onset the only change was the illumination of the light. This was chosen to reflect the naturalistic setting, where participants would be able to see the confederates and table in between trials. Following these changes, participants still exhibited extended reaction times for perspective-taking decisions when the avatar's perspective was far from their own and, as such, was used throughout the OP-MEG recordings.

5.2.5 Preprocessing

Data were initially downsampled to 350 Hz for computational efficiency. The power spectra for the individual channels were plotted to identify channels with excessive levels of noise and these were removed from all analyses. For the first subject, 6 channels were removed from the computational session and 5 from the naturalistic session. For the second subject 5 and 15 channels were removed. For the final subject, 1 and 2 channels were removed. Data were then high-pass filtered at 2 Hz and low-pass filtered below 45 Hz. To remove any artefacts from line noise, a bandstop filter around 50 Hz was also applied. To reduce interference from the background magnetic field, homogenous field correction (HFC) was then applied to the data (Tierney et al., 2021). Independent components analysis was used to identify and remove components related to heartbeats and eye movements. Finally, incorrect trials were removed and the data were epoched and time-locked at -1.7 s to 1.5 s around the light onset.

5.2.6 Source-Level Analysis

A participant's T1-weighted structural MRI scan was used to create a forward model based on a single-shell description of the inner surface of the skull (Nolte, 2003). Using SPM12, a nonlinear spatial normalisation procedure was used to construct a volumetric grid (8 mm resolution) registered to the MNI brain. Source analysis was conducted using a linearly constrained minimum variance (LCMV) beamformer in the time domain (Van Veen et al., 1997) and a Dynamical Imaging of Coherent Sources (DICS) beamformer in the frequency domain (Gross et al., 2001), both of which apply a spatial filter to the MEG data at each point of the 8 mm grid. A lambda regularisation parameter of 1% was used. Beamformer weights were calculated by combining lead-field information with a sensor-level covariance matrix, computed over the entire trial epoch. Source reconstruction maps were baseline corrected against the baseline window (prior to the auditory cue; -1.7 s – -0.9 s) prior to visualisation or comparison with other maps.

5.2.7 Virtual Channel Analysis

I iterated over each parcellation in the Destrieux Cortical Atlas (Destrieux et al., 2010), obtaining a time-course of the data within each parcel. At each atlas location I performed a principal components analysis on the concatenated filters of each grid-point within the parcel, multiplied by the sensor-level covariance matrix, and extracted the first component (Seymour et al., 2021). The pre-processed, sensor-level data were multiplied by this spatial filter to obtain a "virtual channel" at each parcel location. This procedure generated virtual channel information at 64 parcels throughout the cortex which could be submitted to Time-Frequency analyses.

5.2.8 Time-Frequency Analyses

Time-frequency representations (TFRs) were calculated using a single Hanning taper between frequencies of 2–30 Hz in steps of 1 Hz. The entire 3.2 s epoch was used, with a sliding window of 500 ms. TFRs were baseline corrected against the baseline window (prior to the auditory cue; -1.7 s – -0.9 s) prior to visualisation or comparison with other TFRs.

5.3 Results

5.3.1 Reaction Times Indicate Increased Difficulty for Larger Angles During Perspective-Taking

To first ensure the experimental manipulation was successful and that the behavioural effects reported in Seymour et al. (2018) had been replicated, I plotted the reaction times for LR and VO decisions as a function of the angle of rotation (**Figure 5.2A**). In line with both Seymour et al. (2018) and my pilot testing, an interaction between decision type and angle size was observed. Specifically, decisions following large angles of rotation took longer compared to small angles for LR decisions but not for VO decisions. This was true

for both the naturalistic (**Figure 5.2A** left) and computerised tasks (**Figure 5.2A**, right). When analysing each subject's reaction times separately, for subjects 1 and 3 the interaction between angle and decision type was significant in both computerised and naturalistic tasks (subject 1: naturalistic, $F(1, 629) = 41.48$, $p < .001$; computerised, $F(1,625) = 25.07$, $p < .001$; subject 3: naturalistic, $F(1, 611) = 9.68$, $p = .002$; computerised, $F(1,595) = 7.83$, $p = .005$). Subject two, on the other hand, showed no interaction on reaction times in either task (naturalistic, $F(1, 587) = 0$, $p = .97$; computerised, $F(1,596) = 0.08$, $p = .78$).

5.3.2 Observable Motor Response in OPMs

Before I examined the neural correlates of the social decision process, I performed source reconstructions of the motor response ERF to validate both my source reconstruction procedures and the OPM's ability to effectively localise neural signals in a naturalistic task. Reconstructing the broadband power occurring 100ms after participants reported their decision with a right-hand button press revealed a lateralised cluster of activity in the left motor cortex (**Figure 5.2B**). This was replicated across both naturalistic and computerised tasks, illustrating that the OPMs were sensitive to localised neural activity and, importantly, that my reconstruction procedures could identify the loci of such activity.

5.3.3 No Evidence for Theta Oscillations in Perspective-Taking

Seymour et al. (2018) reported an increase in theta power across temporoparietal and prefrontal regions associated with perspective-taking (i.e., Left-Right) decisions. To replicate this result in OPMs, I examined the source-level Time-Frequency Representations (TFRs) associated with both LR and VO decisions in both the naturalistic and computerised tasks. First, a TFR computed over all virtual channels (see Methods) and all trials revealed a spectral profile in line with an involvement of theta-band activity in perspective taking and tracking decisions (**Figure 5.3A**). From approximately 250ms following the light onset, theta band activity increased from baseline, potentially indicating a role for low-frequency oscillations during social-visual decision-making processes.

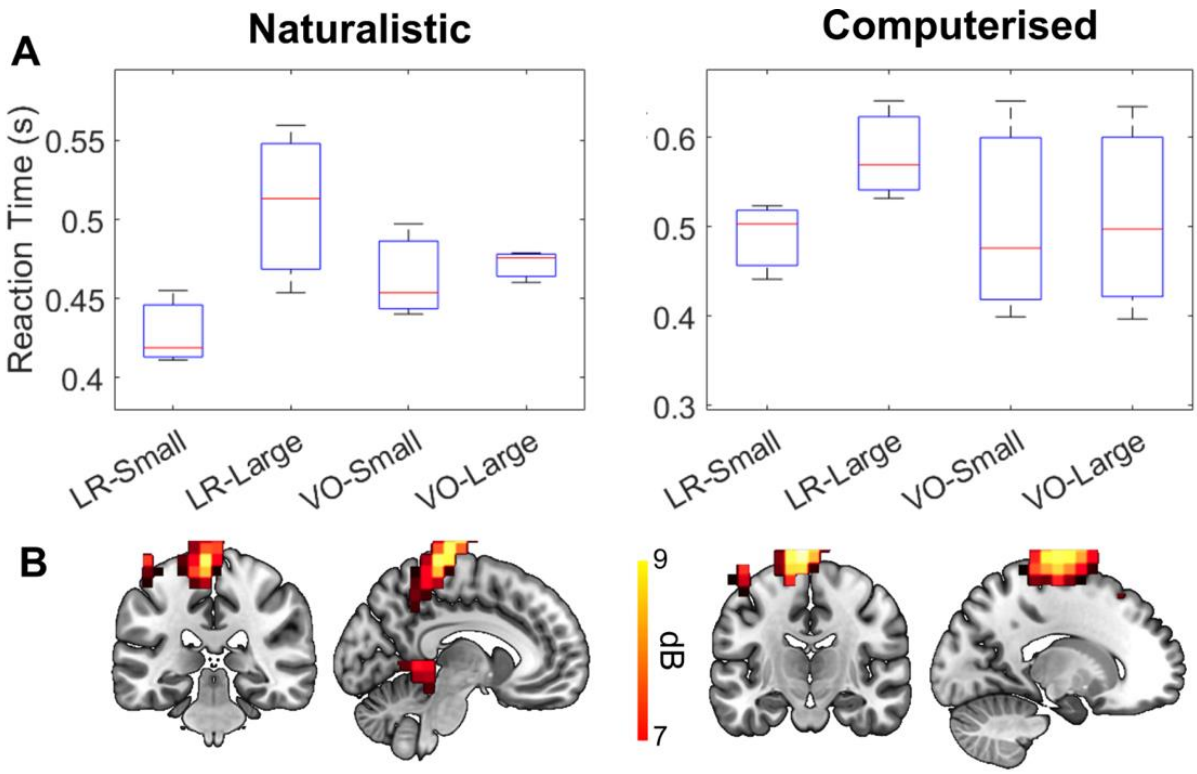


Figure 5.2. Reaction times are greater for LR decisions with large angles of rotation. A. In both naturalistic and computerised tasks, the reaction time for LR decisions was selectively increased in trials with large angles of rotation. The size of the angle did not impact reaction time during VO decisions. This interaction replicates the behavioural effect found in Seymour et al. (2018). B. Average source reconstructions of participants' motor responses indicate successful localisation to the motor cortex in both tasks. Source maps arbitrarily thresholded for visualisation purposes.

To determine whether this theta band activity originated from regions typically associated with social cognition, I localised the increase in theta band power revealed in the source level TFR (3.3 - 6.3 Hz; 220 – 570 ms; **Figure 5.3A**). Surprisingly, in both tasks the low-frequency oscillations following the light onset were localised to left-lateralised motor regions (**Figure 5.3B**). This suggests that the theta power increase illustrated in **Figure 5.3A** is simply a result of motor preparation effects before participants reported their decision. However, it is also possible that, given the rotational demands of the task, task-specific effects may have originated in motor regions. To ascertain whether there were any effects specific to the social-visual decision-making nature of the task in these

regions, I created a virtual channel at the peak voxel of the motor cortex clusters from each task (**Figure 5.3B**) and used the virtual channels to compute and contrast TFRs from trials of different conditions. No features of the spectral profile were systematically higher for LR decisions versus VO decisions (**Figure 5.3C**), suggesting that the theta band activity observed when averaging over all conditions (**Figure 5.3A**) was likely not a consequence of the decision-making task, but instead a confound of response preparation. Likewise, when comparing LR-large with LR-small trials, there was no evidence for an increase in theta power for larger angles of rotation (**Figure 5.3D**).

Rather than testing a virtual channel formed via analyses collapsing over all trials, an alternative method is to compare condition specific TFRs within each virtual channel separately. However, across the brain, there were no ROIs that demonstrated an increase in theta band power for either LR vs. VO or LR-large vs. LR-small decisions. Taken together, I was thus unsuccessful in replicating reports of low-frequency oscillations driving perspective-taking decisions (Wang et al., 2016; Seymour et al., 2018).

5.3.4 Evidence for Visual and Parietal Alpha Desynchronisation in Computational Perspective-Taking

Alongside an increase in theta power following light onset, source level TFRs also revealed a cluster of alpha desynchronisation following light onset in the computerised task (**Figure 5.3A**; **Figure 5.4A**). Across all trials and virtual channels, alpha power was reduced from approximately 100ms to 650ms following light onset in the computerised task only (**Figure 5.3A**; **Figure 5.4A**). Following a similar procedure as for the theta band analyses, source reconstruction of the alpha desynchronisation (9 – 12.5 Hz; 70ms – 660ms; **Figure 5.4A**) yielded a cluster across the parietal cortex (**Figure 5.4B**). However, as in the previous theta band analyses, a virtual channel computed at the peak of this cluster did not return any condition-specific effects (i.e., LR vs VO or LR-large vs. LR-small) in the alpha band.

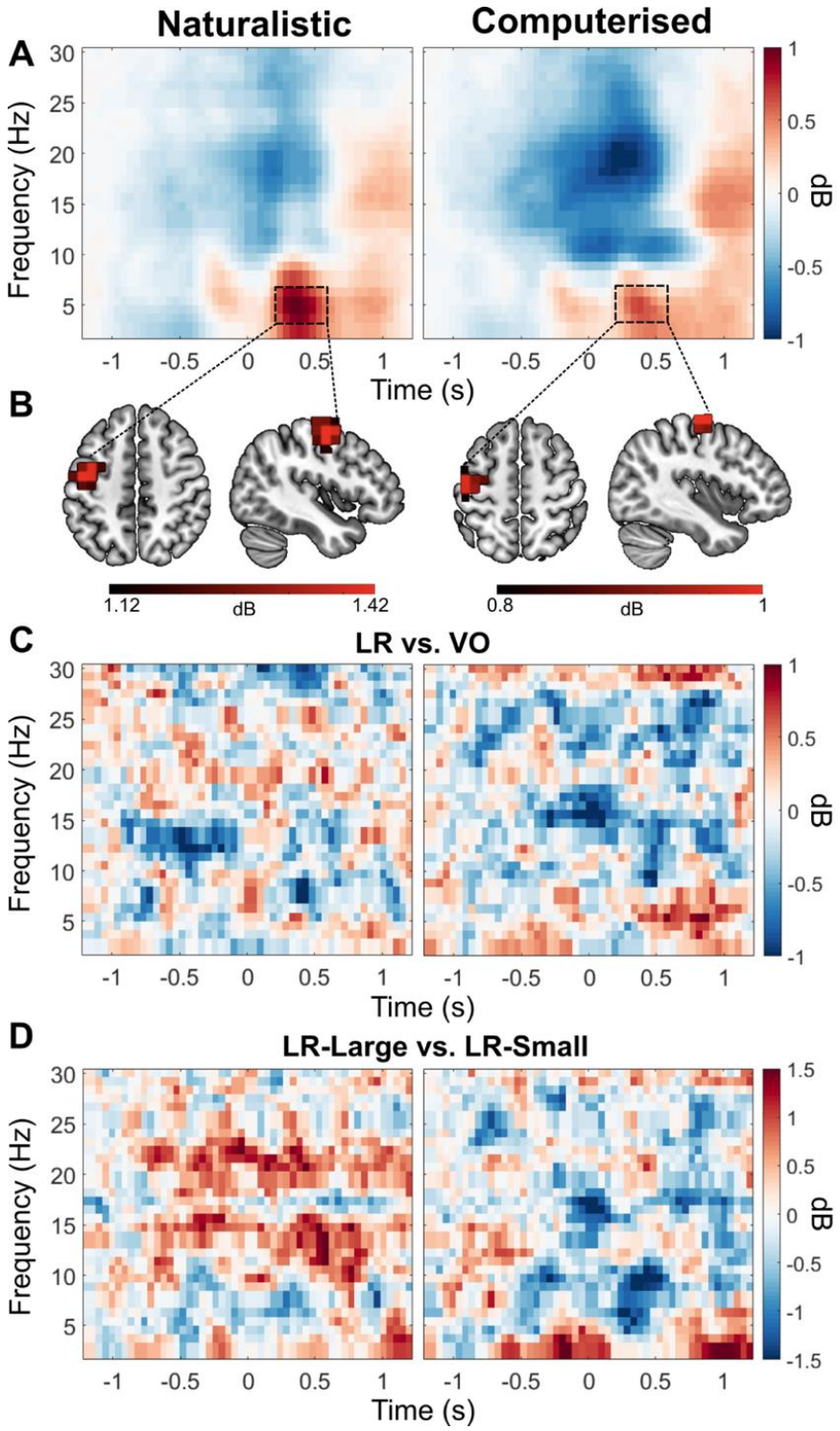


Figure 5.3. No evidence for Theta Oscillations in Perspective-Taking. A. Source-level TFRs computed over all virtual channels and trials reveal an increase in theta power following light onset. B. Theta power was localised to the left-lateralised motor cortex. C. Computing virtual channels at the peak voxel in 2B and contrasting the TFRs from LR and VO trials revealed no oscillatory effects associated with perspective-taking vs. perspective-tracking. D. Comparing LR trials with large angles of rotation to those with small rotations similarly yielded no oscillatory effects in the same virtual channels.

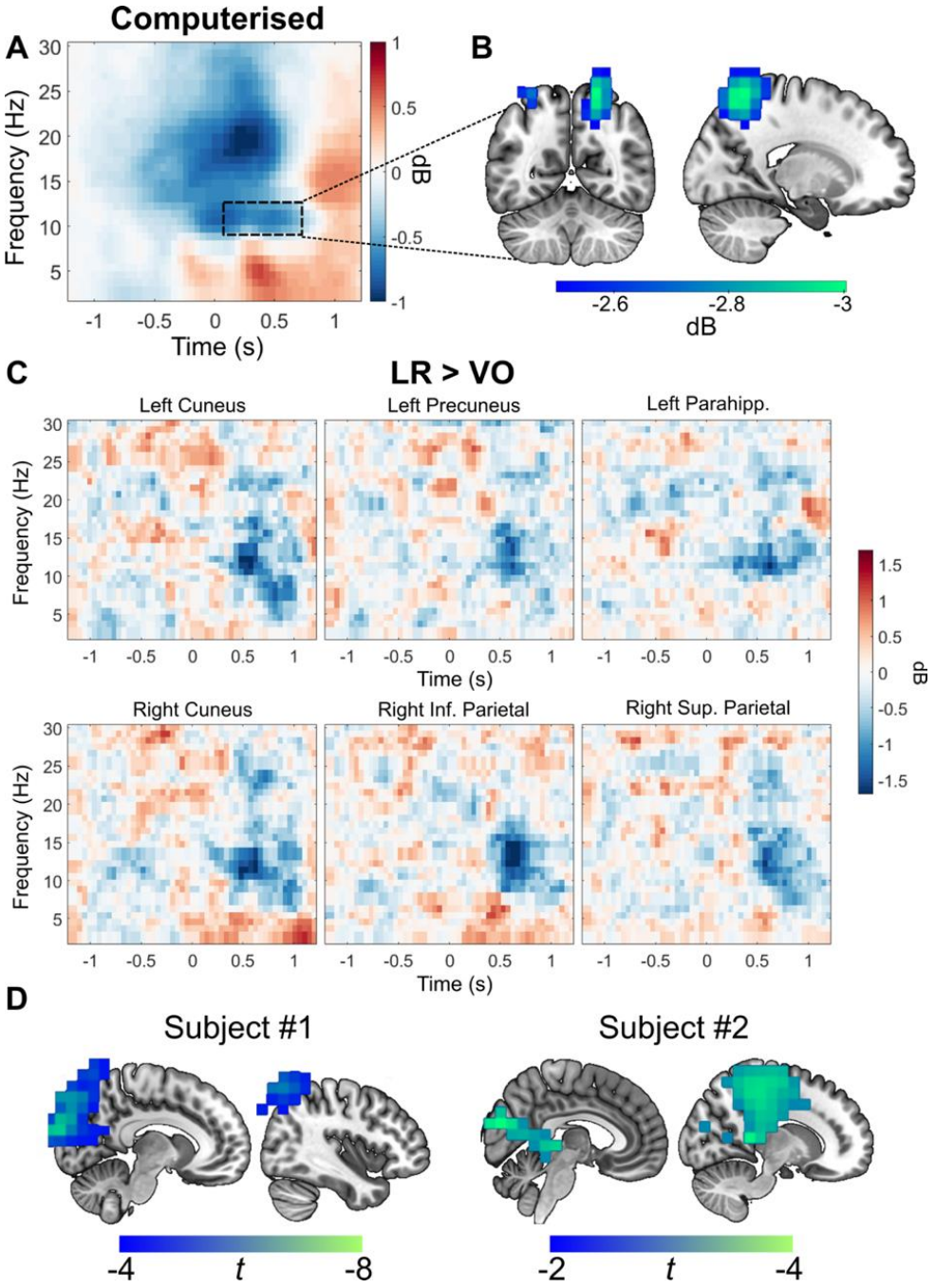


Figure 5.4. Evidence for Visual and Parietal Alpha Desynchronisation in Computational Perspective-Taking. A. In the computerised task, collapsing over all trials and virtual channels revealed a cluster of alpha desynchronisation following light onset. B. This cluster was localised to the parietal cortex. Map thresholded at an arbitrary value for visualisation purposes. C. Contrasting LR decisions with VO decisions at each virtual channel separately revealed a number of channels around parietal and visual cortices where alpha desynchronisation was larger for LR trials. D. A one-tailed cluster-based permutation test contrasting the alpha power associated with LR and VO trials revealed regions of visual and parietal cortex in two out of three participants where alpha desynchronisation was significantly larger for LR decisions ($p < .05$, corrected for multiple comparisons).

When performing condition-specific comparisons at each virtual electrode separately, a consistent pattern of greater alpha desynchronisation following light onset in LR trials compared to VO trials was found in regions surrounding the cuneus and lateral parietal lobe (**Figure 5.4C**), potentially indicating a role for alpha band activity in the occipital and parietal cortex during perspective-taking in the computerised task. To test whether the increased alpha desynchronisation in LR trials was statistically significant, I localised the frequency band and time window that captured this alpha desynchronisation across occipital-parietal virtual channels (10 – 15 Hz; 500 – 800 ms; **Figure 5.4C**) and contrasted maps for LR trials vs. VO trials at the single-subject level. Using a lenient one-tailed cluster-based permutation test (Maris and Oostenveld, 2007), in two out of three subjects, regions across occipital and parietal cortex exhibited significantly increased alpha desynchronisation during LR trials compared to VO trials (**Figure 5.4D**), providing limited evidence for a role of alpha desynchronisation in the computerised perspective-taking task. No significant differences were found in the naturalistic task.

5.3.5 Time-locking to Motor Response Does Not Reveal Condition-Specific Neural Effects

It is possible that the inability to find robust neural responses associated with social perspective taking is due to the previous analyses being locked to the stimulus (i.e., light) onset. However, time-locking the analyses to participants' responses may provide a more sensitive test of the neural bases of such processes. To explore this possibility, I contrasted LR and VO decisions, as well as LR-large and LR-small decisions, across all virtual electrodes. In both the naturalistic and computerised task, there were no differences in frequency profile for the LR trials compared to the VO trials (**Figure 5.5A**), replicating the stimulus-locked analyses. Moreover, comparing LR-Large to LR-Small trials revealed no systematic oscillatory effects (**Figure 5.5B**). A weak increase in beta power was observed around 500ms after motor responses to LR-large decisions in the naturalistic task and a weak increase in alpha power following the same decisions in the computerised task. However, source reconstruction failed to identify robust neural loci of these spectral clusters in either task.

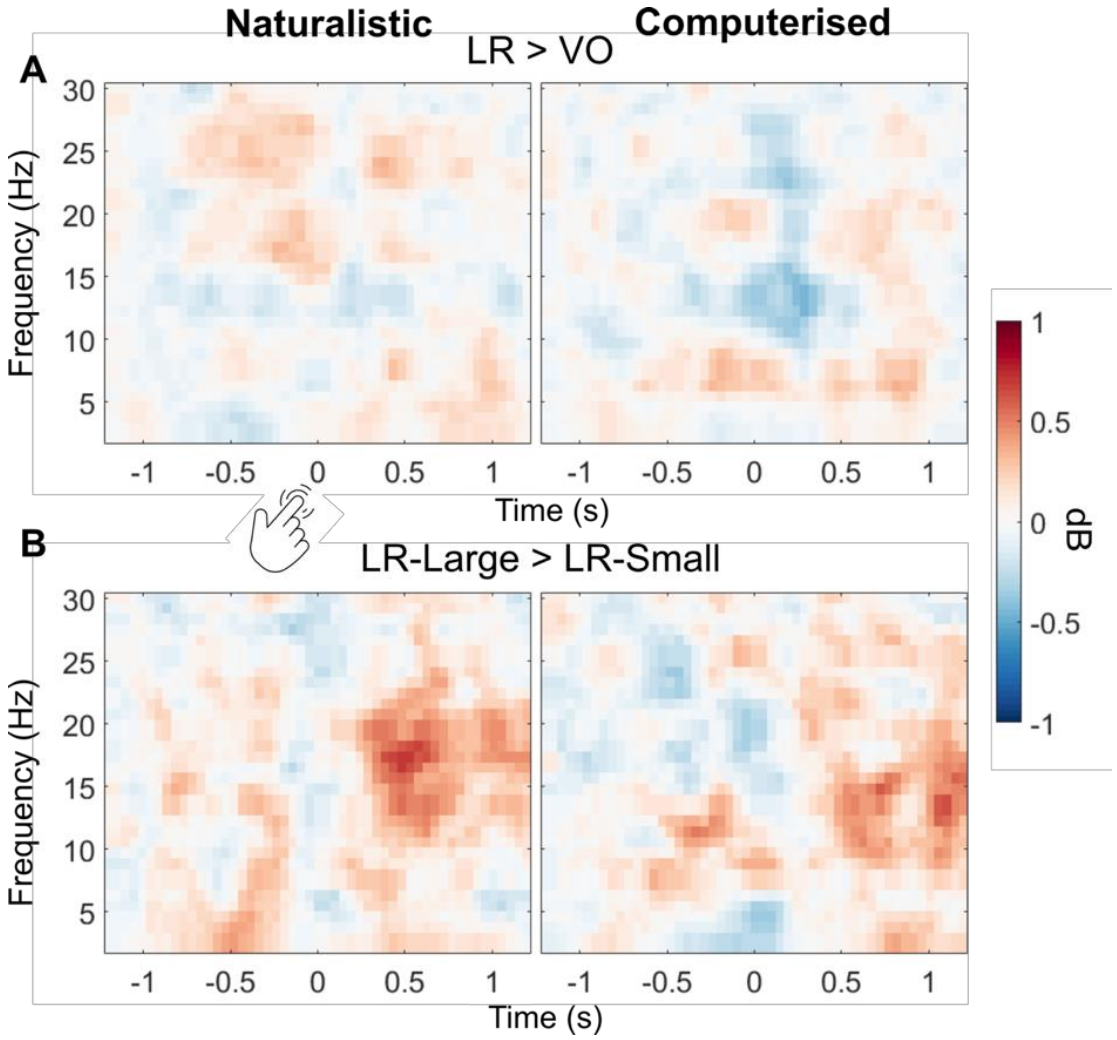


Figure 5.5. Time-locking to motor response does not reveal condition specific effects. A. Contrasting LR and VO trials, time-locked to responses, across all virtual channels failed to reveal any oscillatory difference between decision types. B. Contrasting LR-Large and LR-Small trials showed no systematic oscillatory difference between decision types. A weak increase in beta power accompanied LR-Large trials in the naturalistic task (left) while a small increase in alpha power accompanied LR-Large trials in the computerised task (right).

5.4 Discussion

Theory of Mind is a fundamental component of human cognition (Frith and Frith, 2007; Tomasello et al., 2005), yet most of the neuroscientific research dedicated to elucidating its neural basis has relied on constrained and artificial scenarios that do not reflect the multidimensionality and complexity of naturalistic social settings. To address this, I combined wearable OP-MEG recordings with a novel and ecological perspective-taking task to validate a new generation of mobile neuroimaging devices for use in naturalistic social settings. In a series of exploratory analyses, I failed to replicate previous MEG results identifying theta band activity in social cognition brain regions as neural correlates of perspective taking (Wang et al., 2016; Seymour et al., 2018). There was, however, limited evidence for a role of parietal alpha in perspective-taking, although this was only present in computerised versions of the task.

There is some evidence to suggest oscillations in the alpha band may contribute to perspective-taking processes. In a study using the original version of the present task, alpha band effects similar to those reported here were also observed, with greater alpha desynchronisation observed in perspective-taking trials compared to perspective-tracking trials (Wang et al., 2016). However, the reported alpha effects were also sensitive to the degree of rotation in the perspective-taking trials, with smaller angles of rotation eliciting greater alpha suppression (Wang et al., 2016). In the present study, alpha oscillations were only sensitive to the task, with no difference found between angles of rotation. Furthermore, in Wang et al. (2016), alpha effects were deemed less reliable than effects in the theta band, so further localisation analyses were not performed to better characterise the relationship between alpha oscillations and perspective-taking. There is, however, evidence to suggest that the mu rhythm – alpha oscillations stemming from sensorimotor cortices – is associated with perspective-taking, particularly during action observation (Angelini et al., 2018; Drew et al., 2015; Fu & Franz, 2014). Although mu suppression is typically highest for first-person perspectives (Angelini et al., 2018; Fu and Franz, 2014; Drew et al., 2015, although see Frenkel-Toledo et al., 2013) and the present study required participants to make decisions only in a third person perspective, the

sensitivity of mu oscillations to perspective-taking in general aligns with the fact I found increased alpha suppression when participants were performing perspective-taking, as opposed to perspective-tracking, computations.

It should be emphasised that I only observed perspective-taking related alpha suppression in the computerised version of the task and not in the naturalistic version. Since alpha oscillations are strongly associated with attentional processes (Klimesch, 2012; Klimesch et al., 1998), one potential explanation for this result is that perspective-taking decisions require greater attentional resources than perspective-tracking decisions, but only in computerised tasks. It is possible that perspective-taking is facilitated in real-world settings, since it is a situation more often encountered in day-to-day life. Indeed, reaction times were marginally faster for the naturalistic task, perhaps indicative of reduced attentional load. Studies linking mu rhythms with social cognition have been criticised for lacking control conditions that rule out a confounding role of attention-related alpha oscillations (Hobson & Bishop, 2017), and it is certainly plausible that attention confounds the alpha band result in the present study. However, it represents a strength of the ecological approach to the present study that an easily testable hypothesis regarding the difference between attentional demands in artificial vs. naturalistic settings was generated. Of course, it should be mentioned that effects of alpha oscillations in the computerised task were only present in two out of three participants, and that was only by virtue of lenient and exploratory one-tailed significance tests. As such, the reported effects will need to be replicated in appropriately powered group-level studies before any real conclusion regarding alpha rhythms in naturalistic perspective-taking decisions can be drawn.

Power limitations may also underlie my inability to find differences in the theta band between experimental conditions. In previous work, this task has evoked an increase in theta power between perspective-taking and perspective-tracking decisions, as well as between large and small angles of rotation (Wang et al., 2016; Seymour et al., 2018). Theta oscillations were localised to the rTPJ in both studies. Transcranial magnetic stimulation (TMS) to the rTPJ was found to impact perspective-taking decisions (Wang et

al., 2016) and functional connectivity analyses revealed that theta activity in medial and lateral prefrontal regions acts as inputs to activity in the rTPJ (Seymour et al., 2018), replicating previous findings of mPFC and rTPJ coupling in non-visual perspective-taking tasks (Bögels et al., 2015). An obvious explanation for the present failure to replicate this effect – in both naturalistic and computerised tasks – is because of a lack of statistical power. Previous studies revealed the association between theta oscillations and perspective-taking at the group level, and it may be that only group-level analyses of the current paradigm will suffice to uncover similar effects. However, it may also reflect the current limitations of OPM systems for detecting subtle low frequency effects. OP-MEG is especially sensitive to low-frequency interference, either from motion or features of urban environments, such as moving cars and construction work (Seymour et al., 2021, 2022), which may prevent detection of high-level cognitive effects in low-frequency bands. So far, only one study has identified theta band activity in a cognitive task using OP-MEG (Barry et al., 2019). Here, hippocampal theta activity was shown to underlie imagination behaviour. To show this, imagination was contrasted with a counting task, two vastly different cognitive processes, which may have facilitated detection of theta effects compared to the more high-level and subtle contrast of perspective-taking vs. perspective-tracking computations. Moreover, participants performed this task fixed in place, with movement levels not far beyond conventional MEG. This likely had a beneficial effect on reducing low-frequency movement artefacts but does little to validate OP-MEG for use in a naturalistic cognitive neuroscience. Thus, the challenge of validating OP-MEG for use with high-level social cognition tasks in real world settings is yet to be overcome.

Given the inability to replicate previous results describing electrophysiological correlates of social cognition in OP-MEG, it is a fair question to ask whether other wearable imaging modalities are, at this time, better suited to explore social cognition in naturalistic settings. For instance, wearable electroencephalography (EEG; Gwin et al., 2010; Krugliak & Clarke, 2022; Niso et al., 2023) has supported the investigation of naturalistic social encounters for a number of years and does not require magnetically shielded rooms for proper functioning, meaning it can be used in authentic social situations. Most notably, wearable EEG has been central in hyperscanning studies that examine the role of neural

synchrony (either phase locking or correlations between different individuals' brain activity) in social interactions (Czeszumski et al., 2020). Such studies have been very informative with regards to neural synchrony in real world settings. For instance, in classrooms, increasing levels of brain-to-brain synchrony predict levels of student engagement (Dikker et al., 2017) and students' closeness with teachers (Bevilacqua et al., 2019). Moreover, at museums and festivals, different pairs' empathy, social closeness, and eye contact predicted their level of interbrain coherence (Dikker et al., 2021). At present, the relative ease with which naturalistic data can be collected using wearable EEG may make it more appealing for use in ecological research compared to OP-MEG. However, OP-MEG does offer significant future promise for ecological research. Specifically, artefacts from muscle activation are lower in OP-MEG than EEG by a factor of ten (Boto et al., 2018). Moreover, the improved spatial resolution granted to MEG via insensitivity to volume conduction during source reconstruction is an ever-present advantage compared to EEG techniques (Baillet, 2017). It may be then, that as the validation and development of OP-MEG systems continue, the optimum strategy for ecological research is to opt for more established wearable EEG systems, with the promise of low noise and high spatial resolution in ecological settings being delivered by OP-MEG systems in the near future.

Overall, I attempted to validate novel OP-MEG systems for use in naturalistic social cognition tasks. I failed to replicate conventional MEG results that show an increase in theta band power associated with perspective-taking. I did show limited evidence for a role of alpha desynchronisation in perspective-taking, but this was only present in the computerised version of the task and may be a result of attentional confounds. The present study provides the first test of OP-MEG systems in a truly social task and is an important datapoint in the continued development of this exciting technology, which offers great promise for the investigation of social cognitive processes in naturalistic settings.

6. General Discussion

6.1 An overview

Throughout this thesis, I have examined various hypotheses related to the neural basis of perceptual experience. In Chapter 2, I used findings from magnitude coding studies to generate hypotheses about how perceptual vividness might be encoded in the brain. Using MEG and fMRI data in concert with RSA and decoding analyses, I showed how reports of awareness and perceptual vividness are driven by an ordered and content-invariant neural code that exists throughout visual, parietal, and frontal cortices. In Chapter 3, I demonstrated, for the first time, how zero is represented in the human brain. Using multivariate decoding analyses with MEG data, I revealed how symbolic zero and non-symbolic empty sets exist on a neural number line, with zero situated at the beginning. Moreover, I showed a cross-format effect, whereby representations of empty sets (i.e., blank squares) could generalise to the symbolic format, being most similar to symbolic zero and least similar to symbolic five (and vice versa). I localised this abstract number line to the posterior association cortex, a finding in line with previous work on numerical representations in the human brain. In Chapter 4, I explored the neural correlates of perceptual awareness in patients suffering from Alzheimer's disease using a classic masking paradigm. I illustrated a degradation in the neural correlates of visibility in patients compared to healthy controls, perhaps indicating the contents of consciousness may be altered or diminished in Alzheimer's patients. Furthermore, I explored the potential for using frontoparietal ignition as a neural biomarker for consciousness in noncommunicative patients with severe AD. Finally, in Chapter 5, I conducted a technical development study using wearable OP-MEG in a novel and ecological perspective-taking paradigm. Although this chapter described mostly null

results, it provides a unique data point in the development of wearable neuroimaging technology as the first attempt to explore high-level cognitive effects using OP-MEG in a multi-agent, ecologically valid paradigm.

6.2 Magnitude Codes and Perceptual Experience

Both Chapters 2 and 3 were inspired by work seeking to reveal the architectures supporting different magnitudes in the brain. These chapters provide a rich example of bidirectional influence between different research programmes because not only do they inform the magnitude coding literature and extend its purview – both to the perceptual case of vividness and the numerical case of zero – but they also act as tests of hypotheses derived from specific theories of consciousness, namely Perceptual Reality Monitoring (PRM; Lau, 2019) and the Higher-Order State Space (HOSS; Fleming, 2020).

The PRM theory of consciousness (Lau, 2019) is a higher-order theory describing reality judgements as central to the generation of subjective experience. More specifically, PRM is based on the notion that, at any one time, the brain is generating a vast number of signals. Some of these signals correspond to perceived features of the environment, others underlie internally generated percepts (i.e., imagination), while others simply reflect noise within the system. PRM's central thesis is that there is a higher-order mechanism tracking the extent to which first-order activity represents one of these three options – real, imagined, or noise (Lau, 2019). When this reality monitoring system tags a first-order representation as real, this representation is gifted an 'assertoric force' – an unshakeable belief that this is the current state of the world right now, and in doing so is consciously experienced (Lau, 2019). The HOSS model (Fleming, 2020), on the other hand, is less concerned with determining whether a percept reflects reality. Rather, HOSS describes a computational architecture supporting awareness reports that includes a metacognitive system tracking the reliability of first-order perceptual states (Fleming, 2020). According to HOSS, the high-dimensional mental states supporting perceptual content are monitored by a simple, low-dimensional system that quantifies the precision of the first-order representations, ultimately leading to reports of awareness when high,

or unawareness when low. The critical feature of each of these models with respect to this thesis is their proposal of a 'sparse' higher-order system that does not re-represent perceptual states, but instead monitors their statistical properties in the process of generating reports of awareness.

If sparse HOTs such as PRM and HOSS are correct, the metacognitive systems they describe should not contain information regarding perceptual content, and we should be able to find representations of the statistical properties of perceptual content that are independent of the content itself. Indeed, this is exactly what I found in Chapter 2, where I showed how reports of perceptual visibility are, at least in part, subtended by a content-invariant neural code. This provides one of the first empirical tests of this relatively new generation of HOT theories of consciousness, which are opposed to more classical HOTs that describe higher-order representations as recapitulating the perceptual content in higher-order regions (Brown, 2015; Rosenthal, 2005). To be clear, if higher-order representations re-represented perceptual information in this 'rich' manner, the cross-decoding analysis of perceptual visibility in Chapter 2 would not have been successful, since the features the decoder would use to learn what denotes a highly visible trial for one stimulus would not be present in trials featuring the alternative stimulus.

The finding of content-invariant codes for vividness also leads to more ambitious hypotheses about whether perceptual vividness may rely on an extreme form of content-invariance: domain-general magnitude coding (Walsh, 2003; Summerfield et al., 2020). If sparse HOTs predict a low-dimensional, content-invariant, magnitude code governing awareness, then the prediction that this may be encoded in a system suspected to govern other magnitudes (Walsh, 2003; Summerfield et al., 2020) naturally emerges. Results from Chapter 3 are in line with the proposal that there may indeed be shared representation between visibility and numerical magnitude. Specifically, the fact that empty set stimuli, which – when taken out of context – are simply blank squares, share representational currency with symbolic representations of the number zero, fits with hypotheses suggesting more fundamental representations of sensory absences may have given rise, evolutionarily speaking, to the numerical concept of zero (Nieder, 2016).

Of course, further experiments will be required to test this hypothesis directly by including a non-numerical detection task to evoke representations of sensory absence that are not contaminated by a numerical task framing. Moreover, it remains to be seen whether a shared magnitude representation extends beyond sensory and numerical absence – the shared representation between vividness and other magnitudes need not necessarily stop at zero.

One component of magnitude coding schemes that wasn't examined in this thesis was their logarithmic coding. Magnitudes are typically modelled as symmetrical Gaussian functions when set on a logarithmic scale. This is because neural tuning is less precise as magnitudes increase, causing asymmetric tuning curves unless plotted logarithmically (Dehaene et al., 1998; Tsouli et al., 2021). This is a form of the Weber-Fechner law, a general perceptual rule stating that as magnitudes increase, a larger difference between them is required to maintain a constant discrimination performance (Droit-Volet et al., 2008; Harvey et al., 2020; Merchant et al., 2008; Tudusciuc & Nieder, 2007). Despite being a fundamental component of magnitude codes and perceptual coding more generally, examining whether higher-order representations of vividness or awareness are situated on a logarithmic code was outside the scope of this thesis, most notably because the central aim of this thesis was to test hypotheses generated from different theories of consciousness. Assessing whether reports of vividness were underpinned by content-invariant and graded neural codes were explicit tests of sparse HOTs' hypothesis that low dimensional higher-order systems track properties of perceptual content independently of the content itself (Fleming, 2020; Lau, 2019). These theories, however, make no firm commitment with respect to the linear or logarithmic coding of such phenomenal magnitudes and, as such, this was not explored here. Given the Weber-Fechner law's ubiquity in perceptual domains, however, it would be a natural next step to examine whether neural representations of vividness also demonstrate logarithmic coding properties. If found to be the case, this would further motivate tests of shared representational architectures between awareness reports and other magnitudes.

6.3 Consciousness and Social Cognition

Findings supporting the existence of a content-invariant neural code for awareness also pertain to the hypothesised link between social cognition and consciousness (Section 1.5; Graziano, 2013), and in Chapter 5 I tested the suitability of novel, wearable OP-MEG systems for examining this relationship in future studies. Although AST may be the most prominent exponent of an association between social cognition and consciousness, the link between mindreading and metacognition has been developed previously. For instance, Carruthers (2009) has argued that having direct introspective access to our own propositional attitudes is unlikely, and instead we must turn our Theory of Mind skills inwards to make inferences about our own beliefs and judgements. This idea gains empirical support from a meta-analysis revealing that the brain regions governing metacognitive judgements also largely overlap with those that enable judgements about other minds (Vaccaro & Fleming, 2018), as well as the various examples of altercentric perception introduced in Section 1.5.

How might awareness-specific magnitude codes interface with a shared system for mentalising and self-awareness? It may be that the kinds of codes identified in Chapter 2 are also fundamental in forming inferences about other people's awareness. In the HOSS model, for example, the higher-order awareness layer tracks the precision of an individual's own perceptual states (Fleming, 2020). But it could also perform inference on the observable perceptual states of another individual, for example through monitoring their reaction time, attentional gaze, emotional state, etc. This would result in the low-dimensional neural code for awareness also encoding the inferred awareness of other people. Examples of altercentric perception could be explained by such a phenomenon. Specifically, the facilitation of detection judgements when sharing the perspective of another agent (e.g., Seow & Fleming, 2019) could result from both self and other inputs to the awareness layer driving inference of high detectability. Likewise, cases of altercentric interference – when inconsistent perceptual experiences between oneself and a partner disrupts perceptual judgments (e.g., Nielsen et al., 2015) – may arise as simultaneous but conflicting inferences are made by the same higher-order system.

These ideas remain to be assessed, and Chapter 5 aimed to validate OP-MEG systems to test such questions in ecologically valid scenarios – utilising real human-to-human interactions rather than avatar proxies or other artificial stimuli. Unfortunately, the null results reported in Chapter 5 most likely indicate that the subtle interference or facilitation effects elicited by shared perception are unlikely to be captured by OP-MEG systems at present. To pursue such hypotheses, then, two options seem to be available. The first is to attempt an ecological experiment of this sort using the slightly more established mobile EEG set ups typically used in hyperscanning studies (Czezumski et al., 2020; Krugliak and Clarke, 2022; Stangl et al., 2023). Alternatively, as a scientifically less risky approach, one could resort to less ecologically valid techniques, perhaps conventional MEG or fMRI. Given the lack of neural research on this topic, it may be more sensible to opt for the latter. In such a case, it would be exciting to test, perhaps, whether one could cross-decode reports of one's own awareness with judgements about a partner's – or avatar's – awareness. Additionally, one could examine the interaction between social cognition regions (e.g., rTPJ) and visual cortex during altercentric perception tasks: does the decoding of perceived gratings in early visual cortex improve when a partner shares my perspective, and is this effect driven by a coupling between rTPJ and visual regions? These ideas offer exciting avenues for future research, and importantly this thesis has provided the initial groundwork towards pursing them: first, by characterising the neural basis of reports of one's own awareness, and subsequently by assessing the most promising imaging modality for extending these findings to judgements about others.

6.4 Theories and Disorders of Consciousness

In Chapter 4, I described preliminary evidence in favour of labelling Alzheimer's disease (AD) as a disorder of consciousness. In contrary to the data in this chapter, previous calls to consider the impact of AD on awareness have largely focused on higher-level features of consciousness such as self-awareness (Huntley et al., 2021), owing largely to the high rates of anosognosia co-occurring with the disease (Starkstein et al., 2014). However, Chapter 4 introduced neuroimaging evidence to suggest that AD may impact perceptual features of awareness, at least according to a GWT framework.

The suggestion that AD patients may experience a degraded perceptual awareness of their world is perhaps surprising, not least because previous behavioural reports of sensory awareness in AD have largely proved positive (O'Shaughnessy et al., 2021), and, as mentioned, self-awareness is typically taken to be the major casualty in AD (Huntley et al., 2021; O'Shaughnessy, 2021). Do my results generate evidence for a degradation in perceptual awareness or are they representative of the major cognitive difficulties encountered in AD? This is, of course, dependent on the theory of consciousness used to interpret them. GWT entails some degree of identification of consciousness with cognition since it defines sensory consciousness as the entry of sensory representations into a global, cognitive, workspace (Baars, 1988), and, as such, consciousness and cognition are hard to disentangle here. This poses a challenge, most notably in determining to what extent awareness relies on, or is distinct from, cognition, and how we can interpret biomarkers of awareness accordingly. Exploring awareness in AD may, however, also provide a novel route towards solving this challenge. AD provides a unique testbed for studies that can compare the severity of cognitive aberrations with awareness measures to see if there is a linear relationship between the two or whether the association is more complex – where awareness is not necessarily degraded in line with cognition but shows some degree of independence. This could provide similar evidence to previous experiments highlighting a dissociation of task performance and subjective experience (Lau & Passingham, 2006; Weiskrantz, 1995) but has the potential to provide stronger evidence outside the bounds of matched performance on a narrowly defined task. Specifically, it has the potential to test for markers of awareness when cognitive functioning is disrupted on a global scale. This holds promise for shaping different theories of consciousness, particularly with respect to their reliance on cognition. Of course, the challenge with this approach comes with garnering experiential reports from patients with minimal cognitive functioning. However, the advent of no-report paradigms (Tsuchiya et al., 2015) could play a role here, at least in demarcating when patients become aware of a stimulus or not.

There is further scope for bidirectional influence between theories of consciousness and cognitive disorders, particularly in relation to findings discussed in this thesis. For

instance, if there are shared neural representations across numerical and sensory domains, as posited in Chapter 3, then it may be worth exploring whether individuals with dyscalculia, a developmental impairment of numerical cognition associated with intraparietal regions (Molko et al., 2003), exhibit altered signatures of perceptual experience. Difficulties in visual perception are known to co-occur with dyscalculia (Cheng et al., 2018; Szucs et al., 2013), but psychophysical detection tasks evoking reports of awareness are yet to be conducted in these individuals. Likewise, theories of consciousness that support interactions between awareness and social cognition (Graziano, 2013; Fleming, 2020) could explore how differences in mentalising abilities impact self-awareness. There is already empirical work in support of this relationship, where autistic individuals exhibit lower metacognitive abilities than the general population (Nicholson et al., 2021; van der Plas et al., 2021), for example, but whether or not autistic individuals exhibit significant differences in perceptual experience remains to be tested. One intriguing example of cognitive disorders informing theories of consciousness comes from tests of the meta-problem of consciousness encountered in Section 1.2: exploring why humans perceive consciousness to be ethereal and ineffable (Chalmers, 2018). One account holds that such dualistic tendencies arise because of the folk-psychological understandings that other individuals can act according to invisible beliefs and desires (Berent et al., 2022). Critically, autistic individuals, who hold a compromised Theory of Mind, are less dualistic in their beliefs than the general population, evincing a role for social cognition in the meta-problem of consciousness (Berent, 2022). Theories of consciousness are important in generating hypotheses with regard to the perceptual experiences of individuals with cognitive disorders, but, in this way, they may subsequently benefit from the use of such disorders as testbeds for refining their own theoretical commitments.

6.5 Dogmatic Approaches to Consciousness

Throughout this thesis, I have used different theories of consciousness to motivate experiments and interpret their results. Higher-order theories motivated the studies described in Chapters 2 and 3, while GWT underpinned the interpretation of results in

Chapter 4. Finally, AST influenced the desire to validate ecological approaches to the neuroscience of consciousness in Chapter 5. However, despite revealing several features of the neural basis of perceptual experience, I have not made a commitment to any of the above theories. Indeed, I believe the approach used throughout this thesis, to employ different theories to generate and test different hypotheses without staunchly advocating for any theory in particular, represents one of its strengths.

There is a danger that an individual's commitment to a single theory can impede progress over time. For instance, while HOTs, GWT, and AST are all theories of consciousness, it can be argued that, when being precise, they do not necessarily share the same explanatory target (Seth and Bayne, 2022). More practically, dogmatic commitment to individual theories has resulted in a certain stasis in consciousness research, where theories are coevolving without impacting on one another (Yaron et al., 2021) and progress on the problem of consciousness is being hindered. This results from the fact that post-hoc confirmatory tests are used significantly more than a priori disconfirmatory tests when probing favoured theories of consciousness and because different theories are subject to entirely different methodological procedures when being tested (Yaron et al., 2022).

The solution to this problem is twofold. First, tests of consciousness theories should be better constrained such that predictions neatly distinguish between opposing theories, with results used to amend and develop theory, rather than used as post-hoc support for existing iterations. This thesis followed this rule, particularly in Chapter 2, where competing predictions from sparse vs. rich HOTs of consciousness were tested, with the data being more in line with sparse theories. Indeed, this result, in tandem with the more exploratory analyses of Chapter 3, allows this thesis to fall down on the side of sparse HO theories of consciousness that predict the lean neural representation of graded degrees of awareness, from a complete absence of sensory stimulation to a clear and vivid experience. The second, non-mutually exclusive solution, is to remain light on commitment to any one theory and, critically, to appreciate that each theory can help

guide us towards a better understanding of consciousness in some partial way. This is the approach I have taken throughout the latter half of the thesis.

6.6 Conclusion

Let us return to the question posed in the introduction: how does the inert, electrical activity of the brain generate the rich perceptual experiences that comprise your life as a conscious being? Recall one solution to this problem: solving the 'easy' problems (Chalmers, 1995). By examining and testing these 'easy' problems (the neural systems and processes contributing to our perception of the world), and with a particular focus on the subjective experience of that perception, it has been proposed that the hard problem of consciousness might dissipate (Dennett, 1991; Varela, 1996; Seth, 2016; Frankish, 2019). Across four chapters of this thesis, I have taken this approach, revealing and exploring different neural architectures underpinning perceptual experience.

In Chapter 2, I used MEG and fMRI data to reveal a content-invariant component to the neural basis of vividness and awareness judgements. In Chapter 3, I showed how numerical absence, i.e., the number zero, is situated on a neural number line in the posterior association cortex and illustrated the invariance of this number line to numerical format, laying the groundwork for studies exploring a shared representation of absence across sensory and conceptual domains. In Chapter 4, I used fMRI to show how patients with Alzheimer's disease exhibit degraded neural correlates of consciousness, perhaps indicative of altered perceptual awareness. Finally, I developed and tested a novel perspective-taking paradigm in OP-MEG systems to validate the new, wearable imaging modality in a naturalistic social cognition task, such that we may soon have access to ecologically valid tests of social theories of consciousness.

Perceptual experiences are fundamental to each of us as humans. Without them, there would be no reason to do anything: life would not be worth living (Cleeremans & Tallon-Baudry, 2022). The findings presented throughout this thesis provide a multifaceted

exploration of perceptual experience and its neural basis, drawing inspiration from a variety of psychological domains, and ultimately providing several important steps towards a full understanding of how the brain generates our conscious experience of the world.

Bibliography

Albers, A. M., Kok, P., Toni, I., Dijkerman, H. C., & de Lange, F. P. (2013). Shared Representations for Working Memory and Mental Imagery in Early Visual Cortex. *Current Biology*, *23*(15), 1427–1431. https://doi.org/10.1016/J.CUB.2013.05.065

Albrecht, D. G., & Hamilton, D. B. (1982). Striate cortex of monkey and cat: Contrast response function. *Journal of Neurophysiology*, *48*(1), 217–237. https://doi.org/10.1152/jn.1982.48.1.217

Andersen, L. M., Pedersen, M. N., Sandberg, K., & Overgaard, M. (2016). Occipital MEG Activity in the Early Time Range (<300ms) Predicts Graded Changes in Perceptual Consciousness. *Cerebral Cortex*, *26*(6), 2677–2688. https://doi.org/10.1093/cercor/bhv108

Andersson, J. L. R., Hutton, C., Ashburner, J., Turner, R., & Friston, K. (2001). Modeling Geometric Deformations in EPI Time Series. *NeuroImage*, *13*(5), 903–919. https://doi.org/10.1006/nimg.2001.0746

Angelini, M., Fabbri-Destro, M., Lopomo, N. F., Gobbo, M., Rizzolatti, G., & Avanzini, P. (2018). Perspective-dependent reactivity of sensorimotor mu rhythm in alpha and beta ranges during action observation: An EEG study. *Scientific Reports*, *8*(1), 12429. https://doi.org/10.1038/s41598-018-30912-w

Arnold, D. H. (2011). Why is Binocular Rivalry Uncommon? Discrepant Monocular Images in the Real World. *Frontiers in Human Neuroscience*, *5*. https://doi.org/10.3389/fnhum.2011.00116

Arsalidou, M., & Taylor, M. J. (2011). Is 2+2=4? Meta-analyses of brain areas needed for numbers and calculations. *NeuroImage*, *54*(3), 2382–2393. https://doi.org/10.1016/j.neuroimage.2010.10.009

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

Arzi, A., Rozenkrantz, L., Gorodisky, L., Rozenkrantz, D., Holtzman, Y., Ravia, A., Bekinschtein, T. A., Galperin, T., Krimchansky, B.-Z., Cohen, G., Oksamitni, A., Aidinoff, E., Sacher, Y., & Sobel, N. (2020). Olfactory sniffing signals consciousness in unresponsive patients with brain injuries. *Nature*, *September 2019*. https://doi.org/10.1038/s41586-020-2245-5

Ashburner, J., & Friston, K. J. (2005). Unified segmentation. *NeuroImage*, *26*(3), 839–851. https://doi.org/10.1016/j.neuroimage.2005.02.018

Asplund, K., Jansson, L., & Norberg, A. (1995). Facial Expressions of Patients With Dementia: A Comparison of Two Methods of Interpretation. *International Psychogeriatrics*, *7*(4), 527–534. https://doi.org/10.1017/S1041610295002262

Asplund, K., Norberg, A., & Adolfsson, R. (1991). The Sucking Behaviour of Two Patients in the Final Stage of Dementia of the Alzheimer Type. *Scandinavian*

Journal of Caring Sciences, *5*(3), 141–148. https://doi.org/10.1111/j.1471-6712.1991.tb00099.x
Baars, B. J. (1993). *A cognitive theory of consciousness* (pp. xxiii, 424). Cambridge University Press.
Baillet, S. (2017). Magnetoencephalography for brain electrophysiology and imaging. *Nature Neuroscience*, *20*(3), 327–339. https://doi.org/10.1038/nn.4504
Bang, D., Ershadmanesh, S., Nili, H., & Fleming, S. M. (2020). Private–public mappings in human prefrontal cortex. *eLife*, *9*, e56477. https://doi.org/10.7554/eLife.56477
Barnett, B., Andersen, L. M., Fleming, S. M., & Dijkstra, N. (2024). Identifying content-invariant neural signatures of perceptual vividness. *PNAS Nexus*, *3*(2), pgae061. https://doi.org/10.1093/pnasnexus/pgae061
Barry, D. N., Tierney, T. M., Holmes, N., Boto, E., Roberts, G., Leggett, J., Bowtell, R., Brookes, M. J., Barnes, G. R., & Maguire, E. A. (2019). Imaging the human hippocampus with optically-pumped magnetoencephalography. *Neuroimage*, *203*, 116192. https://doi.org/10.1016/j.neuroimage.2019.116192
Bartlett, J. R., & Doty, R. W. (1974). Response of Units in Striate Cortex of Squirrel Monkeys to Visual and Electrical Stimuli. *Journal of Neurophysiology*, *37*(4), 621–641.
Bartzokis, G. (2004). Age-related myelin breakdown: A developmental model of cognitive decline and Alzheimer's disease. *Neurobiology of Aging*, *25*(1), 5–18. https://doi.org/10.1016/j.neurobiolaging.2003.03.001
Bayne, T., Hohwy, J., & Owen, A. M. (2016). Are There Levels of Consciousness? *Trends in Cognitive Sciences*, *20*(6), 405–413. https://doi.org/10.1016/j.tics.2016.03.009
Beach, P. A., Humbel, A., Dietrich, M. S., Bruehl, S., Cowan, R. L., Moss, K. O., & Monroe, T. B. (2021). A Cross-Sectional Study of Pain Sensitivity and Unpleasantness in People with Vascular Dementia. *Pain Medicine: The Official Journal of the American Academy of Pain Medicine*, *23*(7), 1231–1238. https://doi.org/10.1093/pm/pnab327
Bekinschtein, T. A., Dehaene, S., Rohaut, B., Tadel, F., Cohen, L., & Naccache, L. (2009). Neural signature of the conscious processing of auditory regularities. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(5), 1672–1677. https://doi.org/10.1073/pnas.0809667106
Benwell, C. S. Y., Tagliabue, C. F., Veniero, D., Cecere, R., Savazzi, S., & Thut, G. (2017). Prestimulus EEG Power Predicts Conscious Awareness But Not Objective Visual Performance. *eNeuro*, *4*(6). https://doi.org/10.1523/ENEURO.0182-17.2017
Berent, I., Theodore, R. M., & Valencia, E. (2022). Autism attenuates the perception of the mind-body divide. *Proceedings of the National Academy of Sciences*, *119*(49), e2211628119. https://doi.org/10.1073/pnas.2211628119
Berkovitch, L., Del Cul, A., Maheu, M., & Dehaene, S. (2018). Impaired conscious access and abnormal attentional amplification in schizophrenia. *NeuroImage. Clinical*, *18*, 835–848. https://doi.org/10.1016/j.nicl.2018.03.010
Bernardi, S., Benna, M. K., Rigotti, M., Munuera, J., Fusi, S., & Salzman, C. D. (2020). The Geometry of Abstraction in the Hippocampus and Prefrontal Cortex. *Cell*. https://doi.org/10.1016/j.cell.2020.09.031

Bevilacqua, D., Davidesco, I., Wan, L., Chaloner, K., Rowland, J., Ding, M., Poeppel, D., & Dikker, S. (2019). Brain-to-Brain Synchrony and Learning Outcomes Vary by Student-Teacher Dynamics: Evidence from a Real-world Classroom Electroencephalography Study. *Journal of Cognitive Neuroscience*, *31*(3), 401–411. https://doi.org/10.1162/jocn_a_01274

Bialystok, E., & Codd, J. (2000). Representing quantity beyond whole numbers: Some, none, and part. *Canadian Journal of Experimental Psychology*, *54*(2), 117–128. https://doi.org/10.1037/h0087334

Biringer, F., & Anderson, J. R. (1992). Self-recognition in Alzheimer's disease: A mirror and video study. *Journal of Gerontology*, *47*(6), P385-388. https://doi.org/10.1093/geronj/47.6.p385

Block, N. (2019). What Is Wrong with the No-Report Paradigm and How to Fix It. *Trends in Cognitive Sciences*, *23*(12), 1003–1013. https://doi.org/10.1016/j.tics.2019.10.001

Böckler, A., & Zwickel, J. (2013). Influences of spontaneous perspective taking on spatial and identity processing of faces. *Social Cognitive and Affective Neuroscience*, *8*(7), 735–740. https://doi.org/10.1093/scan/nss061

Boehler, C. N., Schoenfeld, M. A., Heinze, H.-J., & Hopf, J.-M. (2008). Rapid recurrent processing gates awareness in primary visual cortex. *Proceedings of the National Academy of Sciences*, *105*(25), 8742–8747. https://doi.org/10.1073/pnas.0801999105

Bögels, S., Barr, D. J., Garrod, S., & Kessler, K. (2015). Conversational Interaction in the Scanner: Mentalizing during Language Processing as Revealed by MEG. *Cerebral Cortex*, *25*(9), 3219–3234. https://doi.org/10.1093/cercor/bhu116

Bor, D., & Seth, A. K. (2012). Consciousness and the Prefrontal Parietal Network: Insights from Attention, Working Memory, and Chunking. *Frontiers in Psychology*, *3*. https://doi.org/10.3389/fpsyg.2012.00063

Borghesani, V., de Hevia, M. D., Viarouge, A., Pinheiro-Chagas, P., Eger, E., & Piazza, M. (2019). Processing number and length in the parietal cortex: Sharing resources, not a common code. *Cortex*, *114*, 17–27. https://doi.org/10.1016/j.cortex.2018.07.017

Borna, A., Carter, T. R., Goldberg, J. D., Colombo, A. P., Jau, Y.-Y., Berry, C., McKay, J., Stephen, J., Weisend, M., & Schwindt, P. D. D. (2017). A 20-channel magnetoencephalography system based on optically pumped magnetometers. *Physics in Medicine & Biology*, *62*(23), 8909. https://doi.org/10.1088/1361-6560/aa93d1

Boto, E., Holmes, N., Leggett, J., Roberts, G., Shah, V., Meyer, S. S., Muñoz, L. D., Mullinger, K. J., Tierney, T. M., Bestmann, S., Barnes, G. R., Bowtell, R., & Brookes, M. J. (2018). Moving magnetoencephalography towards real-world applications with a wearable system. *Nature*, *555*(7698), Article 7698. https://doi.org/10.1038/nature26147

Boto, E., Meyer, S. S., Shah, V., Alem, O., Knappe, S., Kruger, P., Fromhold, T. M., Lim, M., Glover, P. M., Morris, P. G., Bowtell, R., Barnes, G. R., & Brookes, M. J. (2017). A new generation of magnetoencephalography: Room temperature measurements using optically-pumped magnetometers. *NeuroImage*, *149*, 404–414. https://doi.org/10.1016/j.neuroimage.2017.01.034

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 433–436.
Breese, B. B. (1909). Binocular rivalry. *Psychological Review*, *16*(6), 410–415. https://doi.org/10.1037/h0075805
Brown, R. (2015). The HOROR Theory of Phenomenal Consciousness. *Philosophical Studies*, *172*(7), 1783–1794. https://doi.org/10.1007/S11098-014-0388-7
Brown, R., Lau, H., & LeDoux, J. E. (2019). Understanding the Higher-Order Approach to Consciousness. *Trends in Cognitive Sciences*, *23*(9), 754–768. https://doi.org/10.1016/J.TICS.2019.06.009
Brysbaert, M. (1995). Arabic Number Reading: On the Nature of the Numerical Scale and the Origin of Phonological Recoding. *Journal of Experimental Psychology: General*, *124*(4), 434–452. https://doi.org/10.1037/0096-3445.124.4.434
Buckner, R. L., Snyder, A. Z., Shannon, B. J., LaRossa, G., Sachs, R., Fotenos, A. F., Sheline, Y. I., Klunk, W. E., Mathis, C. A., Morris, J. C., & Mintun, M. A. (2005). Molecular, Structural, and Functional Characterization of Alzheimer's Disease: Evidence for a Relationship between Default Activity, Amyloid, and Memory. *Journal of Neuroscience*, *25*(34), 7709–7717. https://doi.org/10.1523/JNEUROSCI.2177-05.2005
Butterworth, B. (1999). *The mathematical brain* (1. publ). Macmillan.
Carlson, T., Tovar, D. A., Alink, A., & Kriegeskorte, N. (2013). Representational dynamics of object vision: The first 1000 ms. *Journal of Vision*, *13*(10), 1–19. https://doi.org/10.1167/13.10.1
Carrasco, M. (2018). How visual spatial attention alters perception. *Cognitive Processing*, *19*(Suppl 1), 77–88. https://doi.org/10.1007/s10339-018-0883-4
Carrasco, M., Ling, S., & Read, S. (2004). Attention alters appearance. *Nature Neuroscience*, *7*(3), Article 3. https://doi.org/10.1038/nn1194
Carrasco, M., Penpeci-Talgar, C., & Eckstein, M. (2000). Spatial covert attention increases contrast sensitivity across the CSF: Support for signal enhancement. *Vision Research*, *40*(10), 1203–1215. https://doi.org/10.1016/S0042-6989(00)00024-9
Carruthers, P. (2009). How we know our own minds: The relationship between mindreading and metacognition. *Behavioral and Brain Sciences*, *32*(2), 121–182. https://doi.org/10.1017/S0140525X09000545
Casali, A. G., Gosseries, O., Rosanova, M., Boly, M., Sarasso, S., Casali, K. R., Casarotto, S., Bruno, M.-A., Laureys, S., Tononi, G., & Massimini, M. (2013). A theoretically based index of consciousness independent of sensory processing and behavior. *Science Translational Medicine*, *5*(198), 198ra105. https://doi.org/10.1126/scitranslmed.3006294
Casarotto, S., Comanducci, A., Rosanova, M., Sarasso, S., Fecchio, M., Napolitani, M., Pigorini, A., G Casali, A., Trimarchi, P. D., Boly, M., Gosseries, O., Bodart, O., Curto, F., Landi, C., Mariotti, M., Devalle, G., Laureys, S., Tononi, G., & Massimini, M. (2016). Stratification of unresponsive patients by an independently validated index of brain complexity. *Annals of Neurology*, *80*(5), 718–729. https://doi.org/10.1002/ana.24779
Chalmers, D. (1995). Facing Up to the Problem of Consciousness. *Journal of Consciousness Studies*, *2*(3), 200–219.
Chalmers, D. (2018). *The Meta-Problem of Consciousness*.

Charles, L., King, J. R., & Dehaene, S. (2014). Decoding the dynamics of action, intention, and error detection for conscious and subliminal stimuli. *Journal of Neuroscience*, *34*(4), 1158–1170. https://doi.org/10.1523/JNEUROSCI.2465-13.2014

Cheng, D., Xiao, Q., Chen, Q., Cui, J., & Zhou, X. (2018). Dyslexia and dyscalculia are characterized by common visual perception deficits. *Developmental Neuropsychology*, *43*(6), 497–507. https://doi.org/10.1080/87565641.2018.1481068

Chib, V. S., Rangel, A., Shimojo, S., & O'Doherty, J. P. (2009). Evidence for a Common Representation of Decision Values for Dissimilar Goods in Human Ventromedial Prefrontal Cortex. *Journal of Neuroscience*, *29*(39), 12315–12320. https://doi.org/10.1523/JNEUROSCI.2575-09.2009

Cichy, R. M., Pantazis, D., & Oliva, A. (2014). Resolving human object recognition in space and time. *Nature Neuroscience*, *17*(3), 455–462. https://doi.org/10.1038/nn.3635

Clare, L. (2010). Awareness in people with severe dementia: Review and integration. *Aging & Mental Health*, *14*(1), 20–32. https://doi.org/10.1080/13607860903421029

Cleeremans, A., Achoui, D., Beauny, A., Keuninckx, L., Martin, J. R., Muñoz-Moldes, S., Vuillaume, L., & de Heering, A. (2020). Learning to Be Conscious. *Trends in Cognitive Sciences*, *24*(2), 112–123. https://doi.org/10.1016/j.tics.2019.11.011

Cleeremans, A., & Tallon-Baudry, C. (2022). Consciousness matters: Phenomenal experience has functional value. *Neuroscience of Consciousness*, *2022*(1), niac007. https://doi.org/10.1093/nc/niac007

Cohen Kadosh, R., Cohen Kadosh, K., Kaas, A., Henik, A., & Goebel, R. (2007). Notation-Dependent and -Independent Representations of Numbers in the Parietal Lobes. *Neuron*, *53*(2), 307–314. https://doi.org/10.1016/j.neuron.2006.12.025

Cohen Kadosh, R., & Walsh, V. (2009). Numerical representation in the parietal lobes: Abstract or not abstract? *Behavioral and Brain Sciences*, *32*(3–4), 313–328. https://doi.org/10.1017/S0140525X09990938

Cohen, M. A., Botch, T. L., & Robertson, C. E. (2020). The limits of color awareness during active, real-world vision. *Proceedings of the National Academy of Sciences*, *117*(24), 13821–13827. https://doi.org/10.1073/pnas.1922294117

Coldren, J. T., & Haaf, R. A. (2000). Asymmetries in infants' attention to the presence or absence of features. *Journal of Genetic Psychology*, *161*(4), 420–434. https://doi.org/10.1080/00221320009596722

Cortese, A., Amano, K., Koizumi, A., Kawato, M., & Lau, H. (2016). Multivoxel neurofeedback selectively modulates confidence without changing perceptual performance. *Nature Communications*, *7*(1), 13669. https://doi.org/10.1038/ncomms13669

Czeszumski, A., Eustergerling, S., Lang, A., Menrath, D., Gerstenberger, M., Schuberth, S., Schreiber, F., Rendon, Z. Z., & König, P. (2020). Hyperscanning: A Valid Method to Study Neural Inter-brain Underpinnings of Social Interaction. *Frontiers in Human Neuroscience*, *14*. https://doi.org/10.3389/fnhum.2020.00039

Damarla, S. R., Cherkassky, V. L., & Just, M. A. (2016). Modality-independent representations of small quantities based on brain activation patterns. *Human Brain Mapping*, *37*(4), 1296–1307. https://doi.org/10.1002/hbm.23102
Dehaene, S., Bossini, S., & Giraux, P. (1993). The mental representation of parity and number magnitude. *Journal of Experimental Psychology: General*, *122*(3), 371–396. https://doi.org/10.1037/0096-3445.122.3.371
Dehaene, S., & Changeux, J. P. (1993). Development of elementary numerical abilities: A neuronal model. *Journal of Cognitive Neuroscience*, *5*(4), 390–407. https://doi.org/10.1162/jocn.1993.5.4.390
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., & Cohen, L. (2007). Cultural Recycling of Cortical Maps. *Neuron*, *56*(2), 384–398. https://doi.org/10.1016/j.neuron.2007.10.004
Dehaene, S., Dehaene-Lambertz, G., & Cohen, L. (1998). Abstract representations of numbers in the animal and human brain. *Trends in Neurosciences*, *21*(8), 355–361. https://doi.org/10.1016/S0166-2236(98)01263-6
Dehaene, S., Kerszberg, M., & Changeux, J. P. (1998). A neuronal model of a global workspace in effortful cognitive tasks. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(24), 14529–14534. https://doi.org/10.1073/pnas.95.24.14529
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
Dehaene, S., Naccache, L., Le Clec'H, G., Koechlin, E., Mueller, M., Dehaene-Lambertz, G., van de Moortele, P.-F., & Le Bihan, D. (1998). Imaging unconscious semantic priming. *Nature*, *395*(6702), Article 6702. https://doi.org/10.1038/26967
Dennett, D. C. (1988). Quining Qualia. In A. J. Marcel & E. Bisiach (Eds.), *Consciousness in Contemporary Science*. Oxford University Press.
Dennett, D. C. (1991). *Consciousness Explained*. Penguin Books.
Destrieux, C., Fischl, B., Dale, A., & Halgren, E. (2010). Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *NeuroImage*, *53*(1), 1–15. https://doi.org/10.1016/j.neuroimage.2010.06.010
DeWind, N. K., Park, J., Woldorff, M. G., & Brannon, E. M. (2019). Numerical encoding in early visual cortex. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *114*, 76–89. https://doi.org/10.1016/j.cortex.2018.03.027
Dijkstra, N., Ambrogioni, L., & Gerven, M. A. J. van. (2020). Neural dynamics of perceptual inference and its reversal during imagery. *eLife*, 781294. https://doi.org/10.1101/781294
Dijkstra, N., & Fleming, S. M. (2023). Subjective signal strength distinguishes reality from imagination. *Nature Communications*, *14*(1), 1627. https://doi.org/10.1038/s41467-023-37322-1
Dijkstra, N., Gaal, S. van, Geerligs, L., Bosch, S. E., & Gerven, M. A. J. van. (2021). No Evidence for Neural Overlap between Unconsciously Processed and Imagined

Stimuli. *eNeuro*, *8*(5), ENEURO.0228-21.2021. https://doi.org/10.1523/ENEURO.0228-21.2021
Dijkstra, N., Mostert, P., de Lange, F. P., Bosch, S., & van Gerven, M. A. J. (2018). Differential temporal dynamics during visual imagery and perception. *eLife*, *7*, 1–16. https://doi.org/10.7554/eLife.33904
Dikker, S., Michalareas, G., Oostrik, M., Serafimaki, A., Kahraman, H. M., Struiksma, M. E., & Poeppel, D. (2021). Crowdsourcing neuroscience: Inter-brain coupling during face-to-face interactions outside the laboratory. *NeuroImage*, *227*, 117436. https://doi.org/10.1016/j.neuroimage.2020.117436
Dikker, S., Wan, L., Davidesco, I., Kaggen, L., Oostrik, M., McClintock, J., Rowland, J., Michalareas, G., Van Bavel, J. J., Ding, M., & Poeppel, D. (2017). Brain-to-Brain Synchrony Tracks Real-World Dynamic Group Interactions in the Classroom. *Current Biology*, *27*(9), 1375–1380. https://doi.org/10.1016/j.cub.2017.04.002
Dodson, C. S., Spaniol, M., O'Connor, M. K., Deason, R. G., Ally, B. A., & Budson, A. E. (2011). Alzheimer's disease and memory-monitoring impairment: Alzheimer's patients show a monitoring deficit that is greater than their accuracy deficit. *Neuropsychologia*, *49*(9), 2609–2618. https://doi.org/10.1016/j.neuropsychologia.2011.05.008
Drew, A. R., Quandt, L. C., & Marshall, P. J. (2015). Visual influences on sensorimotor EEG responses during observation of hand actions. *Brain Research*, *1597*, 119–128. https://doi.org/10.1016/j.brainres.2014.11.048
Droit-Volet, S., Clément, A., & Fayol, M. (2008). Time, Number and Length: Similarities and Differences in Discrimination in Adults and Children. *Quarterly Journal of Experimental Psychology*, *61*(12), 1827–1846. https://doi.org/10.1080/17470210701743643
Dubner, R., & Zeki, S. M. (1971). Response properties and receptive fields of cells in an anatomically defined region of the superior temporal sulcus in the monkey. *Brain Research*, *35*(2), 528–532. https://doi.org/10.1016/0006-8993(71)90494-x
Duysens, J., Schaafsma, S. J., & Orban, G. A. (1996). Cortical Off Response Tuning for Stimulus Duration. *Vision Research*, *36*(20), 3243–3251. https://doi.org/10.1016/0042-6989(96)00040-5
Eger, E., Michel, V., Thirion, B., Amadon, A., Dehaene, S., & Kleinschmidt, A. (2009). Deciphering Cortical Number Coding from Human Brain Activity Patterns. *Current Biology*, *19*(19), 1608–1615. https://doi.org/10.1016/j.cub.2009.08.047
Eger, E., Sterzer, P., Russ, M. O., Giraud, A.-L., & Kleinschmidt, A. (2003). A supramodal number representation in human intraparietal cortex. *Neuron*, *37*(4), 719–725. https://doi.org/10.1016/s0896-6273(03)00036-9
Eiselt, A. K., & Nieder, A. (2016). Single-cell coding of sensory, spatial and numerical magnitudes in primate prefrontal, premotor and cingulate motor cortices. *Experimental Brain Research*, *234*(1), 241–254. https://doi.org/10.1007/s00221-015-4449-8
Fairfield, B., & Mammarella, N. (2009). The role of cognitive operations in reality monitoring: A study with healthy older adults and Alzheimer's-type dementia. *The Journal of General Psychology*, *136*(1), 21–39. https://doi.org/10.3200/GENP.136.1.21-40

Fan, S., Monte, O. D., & Chang, S. W. C. (2021). Levels of naturalism in social neuroscience research. *iScience*, *24*(7). https://doi.org/10.1016/j.isci.2021.102702
Fisch, L., Privman, E., Ramot, M., Harel, M., Nir, Y., Kipervasser, S., Andelman, F., Neufeld, M. Y., Kramer, U., Fried, I., & Malach, R. (2009). Neural 'Ignition': Enhanced Activation Linked to Perceptual Awareness in Human Ventral Stream Visual Cortex. *Neuron*, *64*(4), 562–574. https://doi.org/10.1016/j.neuron.2009.11.001
Fischer, M., & Rottmann, J. (2005). Do negative numbers have a place on the mental number line. *Psychology Science*, *47*, 22–32.
Fleming, S. M. (2020). Awareness as inference in a higher-order state space. *Neuroscience of Consciousness*, *2020*(1). https://doi.org/10.1093/nc/niz020
Frankish, K. (2019). The Meta-Problem is the Problem of Consciousness. *Journal of Consciousness Studies*, *26*(9–10), 83–94.
Frenkel-Toledo, S., Bentin, S., Perry, A., Liebermann, D. G., & Soroker, N. (2013). Dynamics of the EEG power in the frequency and spatial domains during observation and execution of manual movements. *Brain Research*, *1509*, 43–57. https://doi.org/10.1016/j.brainres.2013.03.004
Friston, K. (2009). The free-energy principle: A rough guide to the brain? *Trends in Cognitive Sciences*, *13*(7), 293–301. https://doi.org/10.1016/j.tics.2009.04.005
Frith, C. D., & Frith, U. (2007). Social cognition in humans. *Current Biology: CB*, *17*(16), R724-732. https://doi.org/10.1016/j.cub.2007.05.068
Fu, Y., & Franz, E. A. (2014). Viewer perspective in the mirroring of actions. *Experimental Brain Research*, *232*(11), 3665–3674. https://doi.org/10.1007/s00221-014-4042-6
Gardumi, A., Ivanov, D., Hausfeld, L., Valente, G., Formisano, E., & Uludağ, K. (2016). The effect of spatial resolution on decoding accuracy in fMRI multivariate pattern analysis. *NeuroImage*, *132*, 32–42. https://doi.org/10.1016/J.NEUROIMAGE.2016.02.033
Gelskov, S. V., & Kouider, S. (2010). *Psychophysical thresholds of face visibility during infancy*.
Giacino, J. T., Fins, J. J., Laureys, S., & Schiff, N. D. (2014). Disorders of consciousness after acquired brain injury: The state of the science. *Nature Reviews. Neurology*, *10*(2), 99–114. https://doi.org/10.1038/nrneurol.2013.279
Giacino, J. T., Katz, D. I., Schiff, N. D., Whyte, J., Ashman, E. J., Ashwal, S., Barbano, R., Hammond, F. M., Laureys, S., Ling, G. S. F., Nakase-Richardson, R., Seel, R. T., Yablon, S., Getchius, T. S. D., Gronseth, G. S., & Armstrong, M. J. (2018). Practice guideline update recommendations summary: Disorders of consciousness: Report of the Guideline Development, Dissemination, and Implementation Subcommittee of the American Academy of Neurology; the American Congress of Rehabilitation Medicine; and the National Institute on Disability, Independent Living, and Rehabilitation Research. *Neurology*, *91*(10), 450–460. https://doi.org/10.1212/WNL.0000000000005926
Goff, P. (2019). *Galileo's Error: Foundations for a New Science of Consciousness*. Pantheon Books.

Goh, R. Z., Phillips, I. B., & Firestone, C. (2023). The perception of silence. *Proceedings of the National Academy of Sciences*, *120*(29), e2301463120. https://doi.org/10.1073/pnas.2301463120
Gomez, M. A., Skiba, R. M., & Snow, J. C. (2018). Graspable Objects Grab Attention More Than Images Do. *Psychological Science*, *29*(2), 206–218. https://doi.org/10.1177/0956797617730599
Graziano, M. S. A. (2013). *Consciousness and the Social Brain*. Oxford University Press.
Graziano, M. S. A. (2019). *Attributing Awareness to Others*. 21.
Grewal, R. P. (1994). Self-recognition in dementia of the Alzheimer type. *Perceptual and Motor Skills*, *79*(2), 1009–1010. https://doi.org/10.2466/pms.1994.79.2.1009
Gross, J., Kujala, J., Hämäläinen, M., Timmermann, L., Schnitzler, A., & Salmelin, R. (2001). Dynamic imaging of coherent sources: Studying neural interactions in the human brain. *Proceedings of the National Academy of Sciences*, *98*(2), 694–699. https://doi.org/10.1073/pnas.98.2.694
Gwin, J. T., Gramann, K., Makeig, S., & Ferris, D. P. (2010). Removal of movement artifact from high-density EEG recorded during walking and running. *Journal of Neurophysiology*, *103*(6), 3526–3534. https://doi.org/10.1152/jn.00105.2010
Hallam, B., Chan, J., Gonzalez Costafreda, S., Bhome, R., & Huntley, J. (2020). What are the neural correlates of meta-cognition and anosognosia in Alzheimer's disease? A systematic review. *Neurobiology of Aging*, *94*, 250–264. https://doi.org/10.1016/j.neurobiolaging.2020.06.011
Harvey, B. M., & Dumoulin, S. O. (2017). A network of topographic numerosity maps in human association cortex. *Nature Human Behaviour*, *1*(2). https://doi.org/10.1038/s41562-016-0036
Harvey, B. M., Dumoulin, S. O., Fracasso, A., & Paul, J. M. (2020). A Network of Topographic Maps in Human Association Cortex Hierarchically Transforms Visual Timing-Selective Responses. *Current Biology*, *30*(8), 1424-1434.e6. https://doi.org/10.1016/j.cub.2020.01.090
Harvey, B. M., Fracasso, A., Petridou, N., & Dumoulin, S. O. (2015). Topographic representations of object size and relationships with numerosity reveal generalized quantity processing in human parietal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(44), 13525–13530. https://doi.org/10.1073/pnas.1515414112
Harvey, B. M., Klein, B. P., Petridou, N., & Dumoulin, S. O. (2013). Topographic Representation of Numerosity in the Human Parietal Cortex. *Science*, *341*(6150), 1123–1126. https://doi.org/10.1126/science.1240405
Hatamimajoumerd, E., Ratan Murty, N. A., Pitts, M., & Cohen, M. A. (2022). Decoding perceptual awareness across the brain with a no-report fMRI masking paradigm. *Current Biology*. https://doi.org/10.1016/j.cub.2022.07.068
Haufe, S., Meinecke, F., Görgen, K., Dähne, S., Haynes, J. D., Blankertz, B., & Bießmann, F. (2014). On the interpretation of weight vectors of linear models in multivariate neuroimaging. *NeuroImage*, *87*, 96–110. https://doi.org/10.1016/j.neuroimage.2013.10.067
Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral

temporal cortex. *Science (New York, N.Y.)*, *293*(5539), 2425–2430. https://doi.org/10.1126/science.1063736
Haynes, J.-D. (2015). A Primer on Pattern-Based Approaches to fMRI: Principles, Pitfalls, and Perspectives. *Neuron*, *87*(2), 257–270. https://doi.org/10.1016/j.neuron.2015.05.025
Haynes, J.-D., & Rees, G. (2006). Decoding mental states from brain activity in humans. *Nature Reviews. Neuroscience*, *7*(7), 523–534. https://doi.org/10.1038/nrn1931
Hayward, D. A., Voorhies, W., Morris, J. L., Capozzi, F., & Ristic, J. (2017). Staring Reality in the Face: A Comparison of Social Attention Across Laboratory and Real World Measures Suggests Little Common Ground. *Canadian Journal of Experimental Psychology*, *71*(3), 212–225. https://doi.org/10.1037/cep0000117
He, J., Hashikawa, T., Ojima, H., & Kinouchi, Y. (1997). Temporal Integration and Duration Tuning in the Dorsal Zone of Cat Auditory Cortex. *Journal of Neuroscience*, *17*(7), 2615–2625. https://doi.org/10.1523/JNEUROSCI.17-07-02615.1997
Hendriks, M. H. A., Daniels, N., Pegado, F., & Op de Beeck, H. P. (2017). The Effect of Spatial Smoothing on Representational Similarity in a Simple Motor Paradigm. *Frontiers in Neurology*, *0*(MAY), 222. https://doi.org/10.3389/FNEUR.2017.00222
Henik, A., & Tzelgov, J. (1982). Is three greater than five: The relation between physical and semantic size in comparison tasks. *Memory & Cognition*, *10*(4), 389–395. https://doi.org/10.3758/BF03202431
Hirsch, J., Zhang, X., Noah, J. A., & Ono, Y. (2017). Frontal temporal and parietal systems synchronize within and across brains during live eye-to-eye contact. *NeuroImage*, *157*, 314–330. https://doi.org/10.1016/j.neuroimage.2017.06.018
Hobson, H. M., & Bishop, D. V. M. (2017). The interpretation of mu suppression as an index of mirror neuron activity: Past, present and future. *Royal Society Open Science*, *4*(3), 160662. https://doi.org/10.1098/rsos.160662
Hofstetter, S., Cai, Y., Harvey, B. M., & Dumoulin, S. O. (2021). Topographic maps representing haptic numerosity reveals distinct sensory representations in supramodal networks. *Nature Communications*, *12*(1), 221. https://doi.org/10.1038/s41467-020-20567-5
Hohwy, J., & Seth, A. (2020). Predictive processing as a systematic basis for identifying the neural correlates of consciousness. *Philosophy and the Mind Sciences*, *1*(II), Article II. https://doi.org/10.33735/phimisci.2020.II.64
Howard, J. D., Gottfried, J. A., Tobler, P. N., & Kahnt, T. (2015). Identity-specific coding of future rewards in the human orbitofrontal cortex. *Proceedings of the National Academy of Sciences*, *112*(16), 5195–5200. https://doi.org/10.1073/pnas.1503550112
Howard, S. R., Avarguès-Weber, A., Garcia, J. E., Greentree, A. D., & Dyer, A. G. (2018). Numerical ordering of zero in honey bees. *Science*, *360*(6393), 1124–1126. https://doi.org/10.1126/SCIENCE.AAR4975
Hu, Y., Yin, C., Zhang, J., & Wang, Y. (2018). Partial least square aided beamforming algorithm in magnetoencephalography source imaging. *Frontiers in Neuroscience*, 616.

Hubel, D. H., & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *The Journal of Physiology*, *195*(1), 215–243. https://doi.org/10.1113/jphysiol.1968.sp008455
Hume, D. (2000). *1739. A treatise of human nature, ed. DF Norton and MJ Norton*. Oxford: Oxford University Press.
Huntley, J., Bor, D., Deng, F., Mancuso, M., Mediano, P. A. M., Naci, L., Owen, A. M., Rocchi, L., Sternin, A., & Howard, R. (2023). Assessing awareness in severe Alzheimer's disease. *Frontiers in Human Neuroscience*, *16*. https://www.frontiersin.org/articles/10.3389/fnhum.2022.1035195
Huntley, J. D., Fleming, S. M., Mograbi, D. C., Bor, D., Naci, L., Owen, A. M., & Howard, R. (2021). Understanding Alzheimer's disease as a disorder of consciousness. *Alzheimer's & Dementia: Translational Research & Clinical Interventions*, *7*(1), e12203. https://doi.org/10.1002/trc2.12203
Ifrah, G. (1985). From one to zero: A universal history of numbers. *(No Title)*.
Iivanainen, J., Zetter, R., & Parkkonen, L. (2020). Potential of on-scalp MEG: Robust detection of human visual gamma-band responses. *Human Brain Mapping*, *41*(1), 150–161. https://doi.org/10.1002/hbm.24795
Itti, L., & Baldi, P. (2009a). Bayesian surprise attracts human attention. *Vision Research*, *49*(10), 1295–1306. https://doi.org/10.1016/j.visres.2008.09.007
Itti, L., & Baldi, P. (2009b). Bayesian surprise attracts human attention. *Vision Research*, *49*(10), 1295–1306. https://doi.org/10.1016/j.visres.2008.09.007
Kamitani, Y., & Tong, F. (2005). Decoding the visual and subjective contents of the human brain. *Nature Neuroscience*, *8*(5), 679–685. https://doi.org/10.1038/nn1444
Kampis, D., & Southgate, V. (2020). Altercentric Cognition: How Others Influence Our Cognitive Processing. *Trends in Cognitive Sciences*, *24*(11), 945–959. https://doi.org/10.1016/j.tics.2020.09.003
Kaplan, R. (1999). *The Nothing that Is: A Natural History of Zero*. Oxford University Press.
Kelly, Y. T., Webb, T. W., Meier, J. D., Arcaro, M. J., & Graziano, M. S. A. (2014). Attributing awareness to oneself and to others. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(13), 5012–5017. https://doi.org/10.1073/pnas.1401201111
Kessler, K., & Rutherford, H. (2010). The Two Forms of Visuo-Spatial Perspective Taking are Differently Embodied and Subserve Different Spatial Prepositions. *Frontiers in Psychology*, *1*. https://www.frontiersin.org/articles/10.3389/fpsyg.2010.00213
Kessler, K., & Wang, H. (2012). Spatial perspective taking is an embodied process, but not for everyone in the same way: Differences predicted by sex and social skills score. *Spatial Cognition and Computation*, *12*(2–3), 133–158. https://doi.org/10.1080/13875868.2011.634533
Kim, I., Hong, S. W., Shevell, S. K., & Shim, W. M. (2020). Neural representations of perceptual color experience in the human ventral visual pathway. *Proceedings of the National Academy of Sciences*, *117*(23), 13145–13150. https://doi.org/10.1073/pnas.1911041117

King, J. R., Faugeras, F., Gramfort, A., Schurger, A., El Karoui, I., Sitt, J. D., Rohaut, B., Wacongne, C., Labyt, E., Bekinschtein, T., Cohen, L., Naccache, L., & Dehaene, S. (2013). Single-trial decoding of auditory novelty responses facilitates the detection of residual consciousness. *NeuroImage*, *83*, 726–738. https://doi.org/10.1016/j.neuroimage.2013.07.013

King, J.-R., & Dehaene, S. (2014). Characterizing the dynamics of mental representations: The temporal generalization method. *Trends in Cognitive Sciences*, *18*(4), 203–210. https://doi.org/10.1016/J.TICS.2014.01.002

Kirschhock, M. E., Ditz, H. M., & Nieder, A. (2021). Behavioral and Neuronal Representation of Numerosity Zero in the Crow. *The Journal of Neuroscience*, *41*(22), 4889–4896. https://doi.org/10.1523/jneurosci.0090-21.2021

Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007). What's new in psychtoolbox-3. *Perception*, *36*(14), 1–16.

Klein-Flügge, M. C., Barron, H. C., Brodersen, K. H., Dolan, R. J., & Behrens, T. E. J. (2013). Segregated Encoding of Reward–Identity and Stimulus–Reward Associations in Human Orbitofrontal Cortex. *Journal of Neuroscience*, *33*(7), 3202–3211. https://doi.org/10.1523/JNEUROSCI.2532-12.2013

Klimesch, W. (2012). Alpha-band oscillations, attention, and controlled access to stored information. *Trends in Cognitive Sciences*, *16*(12), 606–617. https://doi.org/10.1016/j.tics.2012.10.007

Klimesch, W., Doppelmayr, M., Russegger, H., Pachinger, T., & Schwaiger, J. (1998). Induced alpha band power changes in the human EEG and attention. *Neuroscience Letters*, *244*(2), 73–76. https://doi.org/10.1016/s0304-3940(98)00122-0

Knotts, J. D., Michel, M., & Odegaard, B. (2020). Defending subjective inflation: An inference to the best explanation. *Neuroscience of Consciousness*, *2020*(1), 25. https://doi.org/10.1093/nc/niaa025

Korisky, U., Hirschhorn, R., & Mudrik, L. (2019). "Real-life" continuous flash suppression (CFS)-CFS with real-world objects using augmented reality goggles. *Behavior Research Methods*, *51*(6), 2827–2839. https://doi.org/10.3758/s13428-018-1162-0

Korisky, U., & Mudrik, L. (2021). Dimensions of Perception: 3D Real-Life Objects Are More Readily Detected Than Their 2D Images. *Psychological Science*, *32*(10), 1636–1648. https://doi.org/10.1177/09567976211010718

Kouider, S., Stahlhut, C., Gelskov, S. V., Barbosa, L. S., Dutat, M., de Gardelle, V., Christophe, A., Dehaene, S., & Dehaene-Lambertz, G. (2013). A Neural Marker of Perceptual Consciousness in Infants. *Science*, *340*(6130), 376–380. https://doi.org/10.1126/science.1232509

Kovalenko, L. Y., Chaumon, M., & Busch, N. A. (2012). A pool of pairs of related objects (POPORO) for investigating visual semantic integration: Behavioral and electrophysiological validation. *Brain Topography*, *25*(3), 272–284. https://doi.org/10.1007/S10548-011-0216-8

Kragh Nielsen, M., Slade, L., Levy, J. P., & Holmes, A. (2015). Inclined to see it your way: Do altercentric intrusion effects in visual perspective taking reflect an intrinsically social process? *Quarterly Journal of Experimental Psychology (2006)*, *68*(10), 1931–1951. https://doi.org/10.1080/17470218.2015.1023206

Krajcsi, A., Kojouharova, P., & Lengyel, G. (2021). Development of Preschoolers' Understanding of Zero. *Frontiers in Psychology*, *0*, 3169. https://doi.org/10.3389/FPSYG.2021.583734
Krantz, H. (1972). A Theory of Magnitude Estimation and Cross-Modality. *Journal of Mathematical Psychology*, *9*(2), 168–199.
Kriegeskorte, N., & Diedrichsen, J. (2019). Peeling the Onion of Brain Representations. *Annual Review of Neuroscience*, *42*(1), 407–432. https://doi.org/10.1146/annurev-neuro-080317-061906
Kriegeskorte, N., & Kievit, R. A. (2013). Representational geometry: Integrating cognition, computation, and the brain. *Trends in Cognitive Sciences*, *17*(8), 401–412. https://doi.org/10.1016/J.TICS.2013.06.007
Kriegeskorte, N., Mur, M., Ruff, D. A., Kiani, R., Bodurka, J., Esteky, H., Tanaka, K., & Bandettini, P. A. (2008a). Matching Categorical Object Representations in Inferior Temporal Cortex of Man and Monkey. *Neuron*, *60*(6), 1126–1141. https://doi.org/10.1016/j.neuron.2008.10.043
Kriegeskorte, N., Mur, M., Ruff, D. A., Kiani, R., Bodurka, J., Esteky, H., Tanaka, K., & Bandettini, P. A. (2008b). Matching Categorical Object Representations in Inferior Temporal Cortex of Man and Monkey. *Neuron*, *60*(6), 1126–1141. https://doi.org/10.1016/j.neuron.2008.10.043
Krugliak, A., & Clarke, A. (2022). Towards real-world neuroscience using mobile EEG and augmented reality. *Scientific Reports*, *12*(1), 2291. https://doi.org/10.1038/s41598-022-06296-3
Kutter, E. F., Bostroem, J., Elger, C. E., Mormann, F., & Nieder, A. (2018a). Single Neurons in the Human Brain Encode Numbers. *Neuron*, *100*(3), 753-761.e4. https://doi.org/10.1016/j.neuron.2018.08.036
Kutter, E. F., Bostroem, J., Elger, C. E., Mormann, F., & Nieder, A. (2018b). Single Neurons in the Human Brain Encode Numbers. *Neuron*, *100*(3), 753-761.e4. https://doi.org/10.1016/j.neuron.2018.08.036
Kutter, E. F., Dehnen, G., Borger, V., Surges, R., Mormann, F., & Nieder, A. (2023). Distinct neuronal representation of small and large numbers in the human medial temporal lobe. *Nature Human Behaviour*, 1–10. https://doi.org/10.1038/s41562-023-01709-3
Lamme, V. A. F. (2006). Towards a true neural stance on consciousness. *Trends in Cognitive Sciences*, *10*(11), 494–501. https://doi.org/10.1016/J.TICS.2006.09.001
Lamme, V. A. F., & Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends in Neurosciences*, *23*(11), 571–579. https://doi.org/10.1016/S0166-2236(00)01657-X
Lau, H. (2022a). *In Consciousness we Trust: The Cognitive Neuroscience of Subjective Experience*. Oxford University Press.
Lau, H. (2022b). *In consciousness we trust: The cognitive neuroscience of subjective experience*. Oxford University Press.
Lau, H. C. (2019). *Consciousness, Metacognition & Perceptual Reality Monitoring*.
Lau, H. C., & Passingham, R. E. (2006). Relative blindsight in normal observers and the neural correlate of visual consciousness. *Proceedings of the National Academy of Sciences*, *103*(49), 18763–18768. https://doi.org/10.1073/pnas.0607716103

Lau, H., Michel, M., LeDoux, J. E., & Fleming, S. M. (2022). The mnemonic basis of subjective experience. *Nature Reviews Psychology*. https://doi.org/10.1038/s44159-022-00068-6

Lau, H., & Rosenthal, D. (2011). Empirical support for higher-order theories of conscious awareness. *Trends in Cognitive Sciences*, *15*(8), 365–373. https://doi.org/10.1016/j.tics.2011.05.009

Ledoit, O., & Wolf, M. (2004a). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, *88*(2), 365–411. https://doi.org/10.1016/S0047-259X(03)00096-4

Ledoit, O., & Wolf, M. (2004b). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, *88*(2), 365–411. https://doi.org/10.1016/S0047-259X(03)00096-4

LeDoux, J. E., Michel, M., & Lau, H. (2020). A little history goes a long way toward understanding why we study consciousness the way we do today. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(13), 6976–6984. https://doi.org/10.1073/pnas.1921623117

Lee, H., & Chen, J. (2022). Predicting memory from the network structure of naturalistic events. *Nature Communications*, *13*(1), 4235. https://doi.org/10.1038/s41467-022-31965-2

Leopold, D. A., & Logothetis, N. K. (1996). Activity changes in early visual cortex reflect monkeys' percepts during binocular rivalry. *Nature*, *379*(6565), 549–553. https://doi.org/10.1038/379549a0

Levin, D. T., & Simons, D. J. (1997). Failure to detect changes to attended objects in motion pictures. *Psychonomic Bulletin & Review*, *4*(4), 501–506. https://doi.org/10.3758/BF03214339

Levinson, M., Podvalny, E., Baete, S. H., & He, B. J. (2021). Cortical and subcortical signatures of conscious object recognition. *Nature Communications*, *12*(1). https://doi.org/10.1038/s41467-021-23266-x

Libertus, M. E., Woldorff, M. G., & Brannon, E. M. (2007). Electrophysiological evidence for notation independence in numerical processing. *Behavioral and Brain Functions*, *3*. https://doi.org/10.1186/1744-9081-3-1

Luyckx, F., Nili, H., Spitzer, B., & Summerfield, C. (2019). Neural structure mapping in human probabilistic reward learning. *eLife*, *8*. https://doi.org/10.7554/eLife.42816

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177–190. https://doi.org/10.1016/j.jneumeth.2007.03.024

Marks, L. E., Stevens, J. C., Bartoshuk, L. M., Gent, J. F., Rifkin, B., & Stone, V. K. (1988). Magnitude-matching: The measurement of taste and smell. *Chemical Senses*, *13*(1), 63–87. https://doi.org/10.1093/chemse/13.1.63

Marks, L. E., Szczesiul, R., & Ohlott, P. (1986). On the Cross-Modal Perception of Intensity. *Journal of Experimental Psychology: Human Perception and Performance*, *12*(4), 517–534.

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

Mazor, M. (2021). *Inference about Absence as a Window into the Mental Self-Model*. OSF. https://doi.org/10.31234/osf.io/zgf6s
Mazor, M., Friston, K. J., & Fleming, S. M. (2020). Distinct neural contributions to metacognition for detecting, but not discriminating visual stimuli. *eLife*, *9*, e53900. https://doi.org/10.7554/eLife.53900
Mazor, M., Moran, R., & Fleming, S. M. (2021). Metacognitive asymmetries in visual perception. *Neuroscience of Consciousness*, *2021*(1), 1–15. https://doi.org/10.1093/nc/niab025
Mazor, M., Moran, R., & Press, C. (2024). *The role of beliefs about perception in perceptual inference*. OSF. https://doi.org/10.31234/osf.io/gx58j
McNamee, D., Rangel, A., & O'Doherty, J. P. (2013). Category-dependent and category-independent goal-value codes in human ventromedial prefrontal cortex. *Nature Neuroscience*, *16*(4), 479–485. https://doi.org/10.1038/nn.3337
Mei, N., Santana, R., & Soto, D. (2022). Informative neural representations of unseen contents during higher-order processing in human brains and deep artificial networks. *Nature Human Behaviour*, *6*(May). https://doi.org/10.1038/s41562-021-01274-7
Menon, V. (2011). Large-scale brain networks and psychopathology: A unifying triple network model. *Trends in Cognitive Sciences*, *15*(10), 483–506. https://doi.org/10.1016/j.tics.2011.08.003
Merchant, H., Zarco, W., & Prado, L. (2008). Do We Have a Common Mechanism for Measuring Time in the Hundreds of Millisecond Range? Evidence From Multiple-Interval Timing Tasks. *Journal of Neurophysiology*, *99*(2), 939–949. https://doi.org/10.1152/jn.01225.2007
Merritt, D. J., & Brannon, E. M. (2013). Nothing to it: Precursors to a zero concept in preschoolers. *Behavioural Processes*, *93*, 91–97. https://doi.org/10.1016/j.beproc.2012.11.001
Merritt, D. J., Rugani, R., & Brannon, E. M. (2009). Empty sets as part of the numerical continuum: Conceptual precursors to the zero concept in rhesus monkeys. *Journal of Experimental Psychology. General*, *138*(2), 258–269. https://doi.org/10.1037/a0015231
Merten, K., & Nieder, A. (2012). Active encoding of decisions about stimulus absence in primate prefrontal cortex neurons. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(16), 6289–6294. https://doi.org/10.1073/pnas.1121084109
Michelon, P., & Zacks, J. M. (2006). Two kinds of visual perspective taking. *Perception & Psychophysics*, *68*(2), 327–337. https://doi.org/10.3758/BF03193680
Mimura, M., & Yano, M. (2006). Memory impairment and awareness of memory deficits in early-stage Alzheimer's disease. *Reviews in the Neurosciences*, *17*(1–2), 253–266. https://doi.org/10.1515/revneuro.2006.17.1-2.253
Misaki, M., Luh, W. M., & Bandettini, P. A. (2013). The effect of spatial smoothing on fMRI decoding of columnar-level organization with linear support vector machine. *Journal of Neuroscience Methods*, *212*(2), 355–361. https://doi.org/10.1016/J.JNEUMETH.2012.11.004

Mograbi, D. C., Brown, R. G., & Morris, R. G. (2009). Anosognosia in Alzheimer's disease—The petrified self. *Consciousness and Cognition*, *18*(4), 989–1003. https://doi.org/10.1016/j.concog.2009.07.005

Molko, N., Cachia, A., Rivière, D., Mangin, J. F., Bruandet, M., Le Bihan, D., Cohen, L., & Dehaene, S. (2003). Functional and structural alterations of the intraparietal sulcus in a developmental dyscalculia of genetic origin. *Neuron*, *40*(4), 847–858. https://doi.org/10.1016/s0896-6273(03)00670-6

Molloy, D. W., Alemayehu, E., & Roberts, R. (1991). Reliability of a Standardized Mini-Mental State Examination compared with the traditional Mini-Mental State Examination. *The American Journal of Psychiatry*, *148*(1), 102–105. https://doi.org/10.1176/ajp.148.1.102

Morales, J. (2018). *The strength of the mind: Essays on consciousness and introspection*. Columbia University.

Morales, J. (2021). Introspection Is Signal Detection. *The British Journal for the Philosophy of Science*. https://doi.org/10.1086/715184

Morris, J. C. (1997). Clinical dementia rating: A reliable and valid diagnostic and staging measure for dementia of the Alzheimer type. *International Psychogeriatrics*, *9 Suppl 1*, 173–176; discussion 177-178. https://doi.org/10.1017/s1041610297004870

Moutoussis, K., & Zeki, S. (2002). The relationship between cortical activation and perception investigated with invisible stimuli. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(14), 9527–9532. https://doi.org/10.1073/pnas.142305699

Mudrik, L., Hirschhorn, R., & Korisky, U. (2024). Taking consciousness for real: Increasing the ecological validity of the study of conscious vs. unconscious processes. *Neuron*, *112*(10), 1642–1656. https://doi.org/10.1016/j.neuron.2024.03.031

Nasr, K., Viswanathan, P., & Nieder, A. (2019). Number detectors spontaneously emerge in a deep neural network designed for visual object recognition. *Science Advances*, *5*(5), eaav7903. https://doi.org/10.1126/sciadv.aav7903

Neubert, F. X., Mars, R. B., Sallet, J., & Rushworth, M. F. S. (2015). Connectivity reveals relationship of brain areas for reward-guided learning and decision making in human and monkey frontal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(20), E2695–E2704. https://doi.org/10.1073/pnas.1410767112

Nicholson, T., Williams, D., Lind, S., Grainger, C., & Carruthers, P. (2021). Linking metacognition and mindreading: Evidence from autism and dual-task investigations. *Journal of Experimental Psychology: General*, *150*(2), 206–220. https://doi.org/10.1037/xge0000878

Nieder, A. (2016). Representing Something Out of Nothing: The Dawning of Zero. *Trends in Cognitive Sciences*, *20*(11), 830–842. https://doi.org/10.1016/j.tics.2016.08.008

Nieder, A., & Dehaene, S. (2009). Representation of number in the brain. *Annual Review of Neuroscience*, *32*, 185–208. https://doi.org/10.1146/annurev.neuro.051508.135550

Nieder, A., & Miller, E. K. (2003). Coding of cognitive magnitude: Compressed scaling of numerical information in the primate prefrontal cortex. *Neuron*, *37*(1), 149–157. https://doi.org/10.1016/S0896-6273(02)01144-3

Nieder, A., & Miller, E. K. (2004). A parieto-frontal network for visual numerical information in the monkey. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(19), 7457–7462. https://doi.org/10.1073/pnas.0402239101

Nili, H., Walther, A., Alink, A., & Kriegeskorte, N. (2021). Correction: Inferring exemplar discriminability in brain representations. *PLOS ONE*, *16*(4), e0250474. https://doi.org/10.1371/journal.pone.0250474

Nili, H., Wingfield, C., Walther, A., Su, L., Marslen-Wilson, W., & Kriegeskorte, N. (2014a). A Toolbox for Representational Similarity Analysis. *PLoS Computational Biology*, *10*(4). https://doi.org/10.1371/journal.pcbi.1003553

Nili, H., Wingfield, C., Walther, A., Su, L., Marslen-Wilson, W., & Kriegeskorte, N. (2014b). A Toolbox for Representational Similarity Analysis. *PLoS Computational Biology*, *10*(4). https://doi.org/10.1371/journal.pcbi.1003553

Niso, G., Romero, E., Moreau, J. T., Araujo, A., & Krol, L. R. (2023). Wireless EEG: A survey of systems and studies. *NeuroImage*, *269*, 119774. https://doi.org/10.1016/j.neuroimage.2022.119774

Noel, J.-P., Ishizawa, Y., Patel, S. R., Eskandar, E. N., & Wallace, M. T. (2019). Leveraging Nonhuman Primate Multisensory Neurons and Circuits in Assessing Consciousness Theory. *Journal of Neuroscience*, *39*(38), 7485–7500. https://doi.org/10.1523/JNEUROSCI.0934-19.2019

Odegaard, B., Chang, M. Y., Lau, H., & Cheung, S. H. (2018). Inflation versus filling-in: Why we feel we see more than we actually do in peripheral vision. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *373*(1755). https://doi.org/10.1098/rstb.2017.0345

Okuyama, S., Kuki, T., & Mushiake, H. (2015). Representation of the Numerosity 'zero' in the Parietal Cortex of the Monkey. *Scientific Reports*, *5*(1), Article 1. https://doi.org/10.1038/srep10059

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J. M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, *2011*. https://doi.org/10.1155/2011/156869

O'Shaughnessy, N. J., Chan, J. E., Bhome, R., Gallagher, P., Zhang, H., Clare, L., Sampson, E. L., Stone, P., & Huntley, J. (2021). Awareness in severe Alzheimer's disease: A systematic review. *Aging & Mental Health*, *25*(4), 602–612. https://doi.org/10.1080/13607863.2020.1711859

O'Shea, R. P. (2011). Binocular Rivalry Stimuli are Common but Rivalry is not. *Frontiers in Human Neuroscience*, *5*. https://doi.org/10.3389/fnhum.2011.00148

Ozana, A., & Ganel, T. (2019). Weber's law in 2D and 3D grasping. *Psychological Research*, *83*(5), 977–988. https://doi.org/10.1007/s00426-017-0913-3

Padoa-Schioppa, C., & Assad, J. A. (2006). Neurons in the orbitofrontal cortex encode economic value. *Nature*, *441*(7090), 223–226. https://doi.org/10.1038/nature04676

Park, J., DeWind, N. K., Woldorff, M. G., & Brannon, E. M. (2016). Rapid and Direct Encoding of Numerosity in the Visual Stream. *Cerebral Cortex*, *26*(2), 748–763. https://doi.org/10.1093/cercor/bhv017

Paul, J. M., van Ackooij, M., ten Cate, T. C., & Harvey, B. M. (2022). Numerosity tuning in human association cortices and local image contrast representations in early visual cortex. *Nature Communications*, *13*(1), Article 1. https://doi.org/10.1038/s41467-022-29030-z

Pearson, J. (2019). The human imagination: The cognitive neuroscience of visual mental imagery. *Nature Reviews Neuroscience*, *20*(10), 624–634. https://doi.org/10.1038/s41583-019-0202-9

Pepperberg, I. M., & Gordon, J. D. (2005). Number Comprehension by a Grey Parrot (Psittacus erithacus), Including a Zero-Like Concept. *Journal of Comparative Psychology*, *119*(2), 197–209. https://doi.org/10.1037/0735-7036.119.2.197

Pereira, M., Megevand, P., Tan, M. X., Chang, W., Wang, S., Rezai, A., Seeck, M., Corniola, M., Momjian, S., Bernasconi, F., Blanke, O., & Faivre, N. (2021). Evidence accumulation relates to perceptual consciousness and monitoring. *Nature Communications*, *12*(1), Article 1. https://doi.org/10.1038/s41467-021-23540-y

Pérez, P., Manasova, D., Hermann, B., Raimondo, F., Rohaut, B., Bekinschtein, T. A., Naccache, L., Arzi, A., & Sitt, J. D. (2024). Content–state dimensions characterize different types of neuronal markers of consciousness. *Neuroscience of Consciousness*, *2024*(1), niae027. https://doi.org/10.1093/nc/niae027

Persaud, N., Davidson, M., Maniscalco, B., Mobbs, D., Passingham, R. E., Cowey, A., & Lau, H. (2011). Awareness-related activity in prefrontal and parietal cortices in blindsight reflects more than superior visual performance. *NeuroImage*, *58*(2), 605–611. https://doi.org/10.1016/j.neuroimage.2011.06.081

Peters, M. A. K., Thesen, T., Ko, Y. D., Maniscalco, B., Carlson, C., Davidson, M., Doyle, W., Kuzniecky, R., Devinsky, O., Halgren, E., & Lau, H. (2017). Perceptual confidence neglects decision-incongruent evidence in the brain. *Nature Human Behaviour*, *1*, 0139. https://doi.org/10.1038/s41562-017-0139

Piazza, M., Izard, V., Pinel, P., Le Bihan, D., & Dehaene, S. (2004a). Tuning curves for approximate numerosity in the human intraparietal sulcus. *Neuron*, *44*(3), 547–555. https://doi.org/10.1016/j.neuron.2004.10.014

Piazza, M., Izard, V., Pinel, P., Le Bihan, D., & Dehaene, S. (2004b). Tuning curves for approximate numerosity in the human intraparietal sulcus. *Neuron*, *44*(3), 547–555. https://doi.org/10.1016/j.neuron.2004.10.014

Piazza, M., Pinel, P., Le Bihan, D., & Dehaene, S. (2007). A Magnitude Code Common to Numerosities and Number Symbols in Human Intraparietal Cortex. *Neuron*, *53*(2), 293–305. https://doi.org/10.1016/j.neuron.2006.11.022

Pinel, P., Piazza, M., Le Bihan, D., & Dehaene, S. (2004). Distributed and overlapping cerebral representations of number, size, and luminance during comparative judgments. *Neuron*, *41*(6), 983–993. https://doi.org/10.1016/S0896-6273(04)00107-2

Pinhas, M., & Tzelgov, J. (2012). Expanding on the mental number line: Zero is perceived as the “smallest”. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(5), 1187–1205. https://doi.org/10.1037/a0027390

Plas, E. van der, Mason, D., Livingston, L. A., Craigie, J., Happe, F., & Fleming, S. (2021). *Computations of confidence are modulated by mentalizing ability*. PsyArXiv. https://doi.org/10.31234/osf.io/c4pzj

Podvalny, E., Flounders, M. W., King, L. E., Holroyd, T., & He, B. J. (2019). A dual role of prestimulus spontaneous neural activity in visual object recognition. *Nature Communications*, *10*(1), 1–13. https://doi.org/10.1038/s41467-019-11877-4

Pönkänen, L. M., Alhoniemi, A., Leppänen, J. M., & Hietanen, J. K. (2011). Does it make a difference if I have an eye contact with you or with your picture? An ERP study. *Social Cognitive and Affective Neuroscience*, *6*(4), 486–494. https://doi.org/10.1093/scan/nsq068

Ramirez-Cardenas, A., Moskaleva, M., & Nieder, A. (2016). Neuronal Representation of Numerosity Zero in the Primate Parieto-Frontal Number Network Article Neuronal Representation of Numerosity Zero in the Primate Parieto-Frontal Number Network. *Current Biology*, *26*, 1285–1294. https://doi.org/10.1016/j.cub.2016.03.052

Rao, R. P. N., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, *2*(1), 79–87. https://doi.org/10.1038/4580

Redcay, E., Dodell-Feder, D., Pearrow, M. J., Mavros, P. L., Kleiner, M., Gabrieli, J. D. E., & Saxe, R. (2010). Live face-to-face interaction during fMRI: A new tool for social cognitive neuroscience. *NeuroImage*, *50*(4), 1639–1647. https://doi.org/10.1016/j.neuroimage.2010.01.052

Reisberg, B., Ferris, S. H., de Leon, M. J., & Crook, T. (1982). The Global Deterioration Scale for assessment of primary degenerative dementia. *The American Journal of Psychiatry*, *139*(9), 1136–1139. https://doi.org/10.1176/ajp.139.9.1136

Rice, H., Howard, R., & Huntley, J. (2019). Professional caregivers' knowledge, beliefs and attitudes about awareness in advanced dementia: A systematic review of qualitative studies. *International Psychogeriatrics*, *31*(11), 1599–1609. https://doi.org/10.1017/S1041610218002272

Ritchie, J. B., Bracci, S., & Op de Beeck, H. (2017). Avoiding illusory effects in representational similarity analysis: What (not) to do with the diagonal. *NeuroImage*, *148*(January), 197–200. https://doi.org/10.1016/j.neuroimage.2016.12.079

Rosenthal, D. M. (2005). *Consciousness and Mind*. Oxford University Press UK.

Rossion, B. (2014). Understanding face perception by means of human electrophysiology. *Trends in Cognitive Sciences*, *18*(6), 310–318. https://doi.org/10.1016/j.tics.2014.02.013

Rounis, E., Maniscalco, B., Rothwell, J. C., Passingham, R. E., & Lau, H. (2010). Theta-burst transcranial magnetic stimulation to the prefrontal cortex impairs metacognitive visual awareness. *Cognitive Neuroscience*, *1*(3), 165–175. https://doi.org/10.1080/17588921003632529

Sadaghiani, S., Hesselmann, G., & Kleinschmidt, A. (2009). Distributed and Antagonistic Contributions of Ongoing Activity Fluctuations to Auditory Stimulus Detection. *Journal of Neuroscience*, *29*(42), 13410–13417. https://doi.org/10.1523/JNEUROSCI.2592-09.2009

Sagehorn, M., Johnsdorf, M., Kisker, J., Sylvester, S., Gruber, T., & Schöne, B. (2023). Real-life relevant face perception is not captured by the N170 but reflected in later potentials: A comparison of 2D and virtual reality stimuli. *Frontiers in Psychology*, *14*. https://doi.org/10.3389/fpsyg.2023.1050892

Sainsbury, R. (1971). The “feature positive effect” and simultaneous discrimination learning. *Journal of Experimental Child Psychology*, *11*(3), 347–356. https://doi.org/10.1016/0022-0965(71)90039-7

Samaha, J., Iemi, L., & Postle, B. R. (2017). Prestimulus alpha-band power biases visual discrimination confidence, but not accuracy. *Consciousness and Cognition*, *54*, 47–55. https://doi.org/10.1016/j.concog.2017.02.005

Sanchez, G., Hartmann, T., Fuscà, M., Demarchi, G., & Weisz, N. (2020). Decoding across sensory modalities reveals common supramodal signatures of conscious perception. *Proceedings of the National Academy of Sciences*, *117*(13), 7437–7446. https://doi.org/10.1073/pnas.1912584117

Sandberg, K., Bahrami, B., Kanai, R., Barnes, G. R., Overgaard, M., & Rees, G. (2013). Early Visual Responses Predict Conscious Face Perception within and between Subjects during Binocular Rivalry. *Journal of Cognitive Neuroscience*, *25*(6), 969–985. https://doi.org/10.1162/jocn_a_00353

Sassenhagen, J., & Draschkow, D. (2019a). Cluster-based permutation tests of MEG/EEG data do not establish significance of effect latency or location. *Psychophysiology*, *56*(6), 1–8. https://doi.org/10.1111/psyp.13335

Sassenhagen, J., & Draschkow, D. (2019b). Cluster-based permutation tests of MEG/EEG data do not establish significance of effect latency or location. *Psychophysiology*, *56*(6), 1–8. https://doi.org/10.1111/psyp.13335

Schnakers, C., Vanhaudenhuyse, A., Giacino, J., Ventura, M., Boly, M., Majerus, S., Moonen, G., & Laureys, S. (2009). Diagnostic accuracy of the vegetative and minimally conscious state: Clinical consensus versus standardized neurobehavioral assessment. *BMC Neurology*, *9*(1), 35. https://doi.org/10.1186/1471-2377-9-35

Schubert, T. M., Rothlein, D., Brothers, T., Coderre, E. L., Ledoux, K., Gordon, B., & McCloskey, M. (2020). Lack of awareness despite complex visual processing: Evidence from event-related potentials in a case of selective metamorphopsia. *Proceedings of the National Academy of Sciences*, *117*(27), 16055–16064. https://doi.org/10.1073/pnas.2000424117

Seeliger, K., Ambrogioni, L., Gucluturk, Y., Van Den Bulk, L. M., Guclu, U., & Van Gerven, M. A. J. (2021). End-to-end neural system identification with neural information flow. *PLoS Computational Biology*, *17*(2), 1–22. https://doi.org/10.1371/JOURNAL.PCBI.1008558

Seow, T., & Fleming, S. M. (2019). Perceptual sensitivity is modulated by what others can see. *Attention, Perception, and Psychophysics*, *81*(6), 1979–1990. https://doi.org/10.3758/s13414-019-01724-5

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. *Nature Communications*, *12*(1), 1–19. https://doi.org/10.1038/s41467-021-21393-z

Seth, A. (2021). *Being you A new science of consciousness.* Penguin Audio.

Seth, A. K., & Bayne, T. (2022). Theories of consciousness. *Nature Reviews Neuroscience*, *23*(7), 439–452. https://doi.org/10.1038/s41583-022-00587-4
Seymour, R. A., Alexander, N., Mellor, S., O'Neill, G. C., Tierney, T. M., Barnes, G. R., & Maguire, E. A. (2021). Using OPMs to measure neural activity in standing, mobile participants. *NeuroImage*, *244*, 118604. https://doi.org/10.1016/j.neuroimage.2021.118604
Seymour, R. A., Alexander, N., Mellor, S., O'Neill, G. C., Tierney, T. M., Barnes, G. R., & Maguire, E. A. (2022). Interference suppression techniques for OPM-based MEG: Opportunities and challenges. *NeuroImage*, *247*, 118834. https://doi.org/10.1016/j.neuroimage.2021.118834
Seymour, R. A., Wang, H., Rippon, G., & Kessler, K. (2018). Oscillatory networks of high-level mental alignment: A perspective-taking MEG study. *NeuroImage*, *177*, 98–107. https://doi.org/10.1016/j.neuroimage.2018.05.016
Shenhav, A., & Greene, J. D. (2010). Moral judgments recruit domain-general valuation mechanisms to integrate representations of probability and magnitude. *Neuron*, *67*(4), 667–677. https://doi.org/10.1016/j.neuron.2010.07.020
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. *Nature Neuroscience*, *20*(6), 872–878. https://doi.org/10.1038/nn.4545
Siegal, M., & Varley, R. (2002). Neural systems involved in 'theory of mind'. *Nature Reviews. Neuroscience*, *3*(6), 463–471. https://doi.org/10.1038/nrn844
Sladky, R., Friston, K. J., Tröstl, J., Cunnington, R., Moser, E., & Windischberger, C. (2011). Slice-timing effects and their correction in functional MRI. *NeuroImage*, *58*(2), 588–594. https://doi.org/10.1016/j.neuroimage.2011.06.078
Snow, J. C., Skiba, R. M., Coleman, T. L., & Berryhill, M. E. (2014). Real-world objects are more memorable than photographs of objects. *Frontiers in Human Neuroscience*, *8*. https://doi.org/10.3389/fnhum.2014.00837
Sperling, R. A., Aisen, P. S., Beckett, L. A., Bennett, D. A., Craft, S., Fagan, A. M., Iwatsubo, T., Jack, C. R., Kaye, J., Montine, T. J., Park, D. C., Reiman, E. M., Rowe, C. C., Siemers, E., Stern, Y., Yaffe, K., Carrillo, M. C., Thies, B., Morrison-Bogorad, M., … Phelps, C. H. (2011). Toward defining the preclinical stages of Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association*, *7*(3), 280–292. https://doi.org/10.1016/j.jalz.2011.03.003
Stangl, M., Maoz, S. L., & Suthana, N. (2023). Mobile cognition: Imaging the human brain in the 'real world'. *Nature Reviews Neuroscience*, 1–16. https://doi.org/10.1038/s41583-023-00692-y
Starkstein, S. E. (2014). Anosognosia in Alzheimer's disease: Diagnosis, frequency, mechanism and clinical correlates. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *61*, 64–73. https://doi.org/10.1016/j.cortex.2014.07.019
Stelzer, J., Chen, Y., & Turner, R. (2013). Statistical inference and multiple testing correction in classification-based multi-voxel pattern analysis (MVPA): Random permutations and cluster size control. *NeuroImage*, *65*, 69–82. https://doi.org/10.1016/j.neuroimage.2012.09.063

Stevens, J. C., & Marks, L. E. (1965). Cross-modality matching of brightness and loudness. *Proceedings of the National Academy of Sciences*, *54*(2), 407–411. https://doi.org/10.1073/pnas.54.2.407

Strawson, G. (2016). Mind and Being: The Primacy of Panpsychism. In G. Brüntrup & L. Jaskolla (Eds.), *Panpsychism: Contemporary Perspectives* (pp. 000–00). Oxford University Press USA. https://philarchive.org/rec/STRMAB

Summerfield, C., Luyckx, F., & Sheahan, H. (2020). Structure learning and the posterior parietal cortex. *Progress in Neurobiology*, *184*, 101717. https://doi.org/10.1016/j.pneurobio.2019.101717

Surtees, A., Apperly, I., & Samson, D. (2013). Similarities and differences in visual and spatial perspective-taking processes. *Cognition*, *129*(2), 426–438. https://doi.org/10.1016/j.cognition.2013.06.008

Suzuki, K., Schwartzman, D. J., Augusto, R., & Seth, A. K. (2019). Sensorimotor contingency modulates breakthrough of virtual 3D objects during a breaking continuous flash suppression paradigm. *Cognition*, *187*, 95–107. https://doi.org/10.1016/j.cognition.2019.03.003

Szucs, D., Devine, A., Soltesz, F., Nobes, A., & Gabriel, F. (2013). Developmental dyscalculia is related to visuo-spatial memory and inhibition impairment. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *49*(10), 2674–2688. https://doi.org/10.1016/j.cortex.2013.06.007

Taschereau-Dumouchel, V., Cortese, A., Chiba, T., Knotts, J. D., Kawato, M., & Lau, H. (2018). Towards an unconscious neural reinforcement intervention for common fears. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(13), 3470–3475. https://doi.org/10.1073/pnas.1721572115

Taschereau-Dumouchel, V., Kawato, M., & Lau, H. (2020). Multivoxel pattern analysis reveals dissociations between subjective fear and its physiological correlates. *Molecular Psychiatry*, *25*(10), 2342–2354. https://doi.org/10.1038/s41380-019-0520-3

Teichmann, L., Grootswagers, T., Carlson, T., & Rich, A. N. (2018). Decoding Digits and Dice with Magnetoencephalography: Evidence for a Shared Representation of Magnitude. *Journal of Cognitive Neuroscience*, *30*(7), 999–1010. https://doi.org/10.1162/jocn_a_01257

Teng, L. (2022). A metacognitive account of phenomenal force. *Mind & Language*, 1–21. https://doi.org/10.1111/mila.12442

Thibaut, A., Schiff, N., Giacino, J., Laureys, S., & Gosseries, O. (2019). Therapeutic interventions in patients with prolonged disorders of consciousness. *The Lancet. Neurology*, *18*(6), 600–614. https://doi.org/10.1016/S1474-4422(19)30031-6

Tierney, T. M., Alexander, N., Mellor, S., Holmes, N., Seymour, R., O'Neill, G. C., Maguire, E. A., & Barnes, G. R. (2021). Modelling optically pumped magnetometer interference in MEG as a spatially homogeneous magnetic field. *NeuroImage*, *244*, 118484. https://doi.org/10.1016/j.neuroimage.2021.118484

Tierney, T. M., Holmes, N., Mellor, S., López, J. D., Roberts, G., Hill, R. M., Boto, E., Leggett, J., Shah, V., Brookes, M. J., Bowtell, R., & Barnes, G. R. (2019). Optically pumped magnetometers: From quantum origins to multi-channel magnetoencephalography. *Neuroimage*, *199*, 598–608. https://doi.org/10.1016/j.neuroimage.2019.05.063

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *The Behavioral and Brain Sciences*, *28*(5), 675–691; discussion 691-735. https://doi.org/10.1017/S0140525X05000129

Tononi, G. (2008). Consciousness as integrated information: A provisional manifesto. *The Biological Bulletin*, *215*(3), 216–242. https://doi.org/10.2307/25470707

Tononi, G., & Edelman, G. M. (1998). Consciousness and complexity. *Science (New York, N.Y.)*, *282*(5395), 1846–1851. https://doi.org/10.1126/science.282.5395.1846

Treder, M. S. (2020). MVPA-Light: A Classification and Regression Toolbox for Multi-Dimensional Data. *Frontiers in Neuroscience*, *0*, 289. https://doi.org/10.3389/FNINS.2020.00289

Tsouli, A., Harvey, B. M., Hofstetter, S., Cai, Y., van der Smagt, M. J., te Pas, S. F., & Dumoulin, S. O. (2021). The Role of Neural Tuning in Quantity Perception. *Trends in Cognitive Sciences*, *xx*(xx), 1–14. https://doi.org/10.1016/j.tics.2021.10.004

Tsuchiya, N., & Koch, C. (2005). Continuous flash suppression reduces negative afterimages. *Nature Neuroscience*, *8*(8), 1096–1101. https://doi.org/10.1038/nn1500

Tsuchiya, N., Wilke, M., Frässle, S., & Lamme, V. A. F. (2015). No-Report Paradigms: Extracting the True Neural Correlates of Consciousness. *Trends in Cognitive Sciences*, *19*(12), 757–770. https://doi.org/10.1016/J.TICS.2015.10.002

Tudusciuc, O., & Nieder, A. (2007). Neuronal population coding of continuous and discrete quantity in the primate posterior parietal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(36), 14513–14518. https://doi.org/10.1073/pnas.0705495104

Vaccaro, A. G., & Fleming, S. M. (2018). Thinking about thinking: A coordinate-based meta-analysis of neuroimaging studies of metacognitive judgements. *Brain and Neuroscience Advances*, *2*, 2398212818810591. https://doi.org/10.1177/2398212818810591

Van Elk, M., & Blanke, O. (2014). Imagined own-body transformations during passive self-motion. *Psychological Research*, *78*(1), 18–27. https://doi.org/10.1007/s00426-013-0486-8

Van Veen, B. D., van Drongelen, W., Yuchtman, M., & Suzuki, A. (1997). Localization of brain electrical activity via linearly constrained minimum variance spatial filtering. *IEEE Transactions on Bio-Medical Engineering*, *44*(9), 867–880. https://doi.org/10.1109/10.623056

Van Vugt, B., Dagnino, B., Vartak, D., Safaai, H., Panzeri, S., Dehaene, S., & Roelfsema, P. R. (2018a). The threshold for conscious report: Signal loss and response bias in visual and frontal cortex. *Science*, *360*(6388), 537–542. https://doi.org/10.1126/science.aar7186

Van Vugt, B., Dagnino, B., Vartak, D., Safaai, H., Panzeri, S., Dehaene, S., & Roelfsema, P. R. (2018b). The threshold for conscious report: Signal loss and response bias in visual and frontal cortex. *Science*, *360*(6388), 537–542. https://doi.org/10.1126/science.aar7186

Vannini, P., d'Oleire Uquillas, F., Jacobs, H. I. L., Sepulcre, J., Gatchel, J., Amariglio, R. E., Hanseeuw, B., Papp, K. V., Hedden, T., Rentz, D. M., Pascual-Leone, A., Johnson, K. A., & Sperling, R. A. (2019). Decreased meta-memory is associated with early tauopathy in cognitively unimpaired older adults. *NeuroImage Clinical*, *24*, 102097. https://doi.org/10.1016/j.nicl.2019.102097

Varela, F. J. (1996). Neurophenomenology: A Methodological Remedy for the Hard Problem. *Journal of Consciousness Studies*, *3*(4), 330–349.

Verguts, T., & Fias, W. (2008). Symbolic and Nonsymbolic Pathways of Number Processing. *Philosophical Psychology*, *21*(4), 539–554. https://doi.org/10.1080/09515080802285545

Vishne, G., Gerber, E. M., Knight, R. T., & Deouell, L. Y. (2022). *Representation of sustained visual experience by time-invariant distributed neural patterns* (p. 2022.08.02.502469). bioRxiv. https://doi.org/10.1101/2022.08.02.502469

Walsh, V. (2003). A theory of magnitude: Common cortical metrics of time, space and quantity. *Trends in Cognitive Sciences*, *7*(11), 483–488. https://doi.org/10.1016/j.tics.2003.09.002

Wang, H., Callaghan, E., Gooding-Williams, G., McAllister, C., & Kessler, K. (2016). Rhythm makes the world go round: An MEG-TMS study on the role of right TPJ theta oscillations in embodied perspective taking. *Cortex*, *75*, 68–81. https://doi.org/10.1016/j.cortex.2015.11.011

Wang, L., Mruczek, R. E. B., Arcaro, M. J., & Kastner, S. (2015). Probabilistic maps of visual topography in human cortex. *Cerebral Cortex*, *25*(10), 3911–3931. https://doi.org/10.1093/cercor/bhu277

Ward, E., Ganis, G., & Bach, P. (2019). Spontaneous Vicarious Perception of the Content of Another's Visual Perspective. *Current Biology*, *29*(5), 874-880.e4. https://doi.org/10.1016/j.cub.2019.01.046

Webb, T. W., & Graziano, M. S. A. (2015). The attention schema theory: A mechanistic account of subjective awareness. *Frontiers in Psychology*, *6*(APR), 1–11. https://doi.org/10.3389/fpsyg.2015.00500

Weiler, M., Northoff, G., Damasceno, B. P., & Balthazar, M. L. F. (2016). Self, cortical midline structures and the resting state: Implications for Alzheimer's disease. *Neuroscience & Biobehavioral Reviews*, *68*, 245–255. https://doi.org/10.1016/j.neubiorev.2016.05.028

Weiskrantz, L. (1995). Blindsight—Not an Island Unto Itself. *Current Directions in Psychological Science*, *4*(5), 146–151. https://doi.org/10.1111/j.1467-8721.1995.tb00264.x

Wellman, H. M., & Miller, K. F. (1986). Thinking about nothing: Development of concepts of zero. *British Journal of Developmental Psychology*, *4*(1), 31–42. https://doi.org/10.1111/j.2044-835X.1986.tb00995.x

Whyte, C. J., & Smith, R. (2021). The predictive global neuronal workspace: A formal active inference model of visual consciousness. *Progress in Neurobiology*, *199*, 101918. https://doi.org/10.1016/j.pneurobio.2020.101918

Yallak, E., & Balcı, F. (2021). Metric error monitoring: Another generalized mechanism for magnitude representations? *Cognition*, *210*. https://doi.org/10.1016/j.cognition.2020.104532

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

Zagury, Y., Zaks-Ohayon, R., Tzelgov, J., & Pinhas, M. (2022). Sometimes nothing is simply nothing: Automatic processing of empty sets. *Quarterly Journal of Experimental Psychology*, *75*(10), 1810–1827. https://doi.org/10.1177/17470218211066436

Zaks-Ohayon, R., Pinhas, M., & Tzelgov, J. (2021). On the indicators for perceiving empty sets as zero. *Acta Psychologica*, *213*, 103237. https://doi.org/10.1016/j.actpsy.2020.103237

Zaks-Ohayon, R., Pinhas, M., & Tzelgov, J. (2022). Nonsymbolic and symbolic representations of null numerosity. *Psychological Research*, *86*(2), 386–403. https://doi.org/10.1007/s00426-021-01515-4

Zhang, Y., Li, R., Du, J., Huo, S., Hao, J., & Song, W. (2017). Coherence in P300 as a predictor for the recovery from disorders of consciousness. *Neuroscience Letters*, *653*, 332–336. https://doi.org/10.1016/j.neulet.2017.06.013

Zorzi, M., & Butterworth, B. (1999). A Computational Model of Number Comparison. In *Proceedings of the Twenty-first Annual Conference of the Cognitive Science Society*. Psychology Press.