Accessing the mind: relating metacognition to attention, memory, and cognitive function

Julian Rodney Matthews

A thesis submitted for the Degree of Doctor of Philosophy at Monash University in 2018

Philosophy Department
School of Philosophical, Historical and International Studies
Faculty of Arts, Monash University

Copyright notice

© Julian Matthews (2018). Except as provided in the Copyright Act 1968, this thesis may not be reproduced in any form without the written permission of the author.

Under the Copyright Act 1968, this thesis must be used only under the normal conditions of scholarly fair dealing. In particular, no results or conclusions should be extracted from it, nor should it be copied or closely paraphrased in whole or in part without the written consent of the author. Proper written acknowledgement should be made for any assistance obtained from this thesis.

I certify that I have made all reasonable efforts to secure copyright permissions for third-party content included in this thesis and have not knowingly added copyright content to my work without the owner's permission.

Although the methodology continues to improve, the empirical approach to metacognition seems useful already, both in terms of reliable findings about subjective reports and in terms of increasingly better inferences about the kinds of private experiences that occur. Accordingly, the problem of consciousness does seem to beg for a cooperative solution, and perhaps the territory can be shared profitably both by philosophers and psychologists, with research on metacognition producing synergy between the philosophical and psychological approaches.

Thomas O. Nelson
Consciousness and Metacognition, 1996

Table of contents

Abstract

Metacognition, *cognition about cognition*, is one of the foremost topics of interest in the behavioural study of consciousness. This is because the introspective quality of metacognition appears to capture an essential property of conscious awareness—its subjective and reflective nature. Metacognition is also exciting because metacognitive measures appear to quantify the relationship between objective and subjective aspects of decision-making. This has led some to view metacognitive measures as a holy grail of consciousness science—a dependable behavioural tool for distinguishing conscious from unconscious mental states.

Our understanding of metacognition is however still in its infancy. A comprehensive treatment needs to examine how metacognition relates not just to other core processes involved in consciousness but also cognition and perception more broadly. To this end, this thesis explores the relationship between metacognition, consciousness, and key cognitive functions including attention, expectations, memory, and perceptual sensitivity. Metacognitive measures are employed across a variety of psychophysical experiments to extend otherwise mechanistic accounts of human visual perception and behavioural report. Beyond their methodological contribution, the studies in this thesis provide evidence for the increasingly compelling stance that selective attention is doubly dissociable from consciousness. That is, selective attention is neither necessary nor sufficient for consciousness. Thus, by operationalising metacognition in the understanding of consciousness, it is possible to address one of the core questions in consciousness science.

Almost all studies of metacognition focus on healthy humans. However, there is growing interest in interdisciplinary research between cognitive and clinical science. Informed by the predictive processing framework of brain function, the thesis includes a study of metacognition and perception in functional and organic motor disorders. This work revealed novel irregularities in perceptual sensitivity and decision-making in motor disorders, highlighting the counter-intuitive notion that perception is a topic of interest for these groups.

The overall view that emerges from this thesis is that metacognition is a cognitive process that has much to tell us about brain function, perception, as well as the nature and origin of conscious awareness. It is likely that metacognition processes are independent from selective attention and working memory but share close associations with perception. However, the thesis cautions against the view that in metacognition we have found the holy grail of consciousness science. Rather, the role of metacognition in the scientific quest for consciousness will be in framing theories and neurobiological models of introspection and certainty that might then be differentiated from the minimal neural mechanisms sufficient for consciousness.

General declaration

This thesis contains no material which has been accepted for the award of any other degree or diploma at any university or equivalent institution and that, to the best of my knowledge and belief, this thesis contains no material previously published or written by another person, except where due reference is made in the text of the thesis.

Signature: [redacted]

Print name: Julian Matthews

Date: 29 October, 2018

Publications during enrolment

Matthews, J, Schröder, P, Kaunitz, L, van Boxtel, J, Tsuchiya, N. (2018).
Conscious access in the near absence of attention: critical extensions on the dual-task paradigm. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 373 (1755), 20170352. DOI: 10.1098/rstb.2017.0352
https://osf.io/7dqm4/

Matthews, J, Wu, J, Corneille, V, Hohwy, J, van Boxtel, J, Tsuchiya, N. (2018).
Sustained conscious access to incidental memories in RSVP. *Attention, Perception, & Psychophysics*, 1-17. DOI: 10.3758/s13414-018-1600-1
https://osf.io/kb8yr/

Preprint publication

Matthews, J, Nagao, K, Ding, C, Newby, R, Kempster, P, Hohwy, J. (2018).
Impaired perceptual sensitivity with intact attention and metacognition in functional motor disorder. *PsyArXiv*. DOI: 10.31234/osf.io/fz3j2
https://osf.io/sm85w/

Declaration for thesis including published works

I hereby declare that this thesis contains no material which has been accepted for the award of any other degree or diploma at any university or equivalent institution and that, to the best of my knowledge and belief, this thesis contains no material previously published or written by another person, except where due reference is made in the text of the thesis.

This thesis includes 2 original papers published in peer reviewed journals and 1 original paper currently submitted for publication. The core theme of the thesis is the relationship between metacognition and consciousness. The ideas, development and writing up of all the papers in the thesis were the principal responsibility of myself, the student, working within the School of Philosophy under the supervision of Professor Jakob Hohwy, Associate Professor Naotsugu Tsuchiya, and Associate Professor Jeroen van Boxtel.

The inclusion of co-authors reflects the fact that the work came from active collaboration between researchers and acknowledges input into team-based research.

I have renumbered sections of published papers.

[redacted]

Julian Matthews
Student
Date: 29/10/2018

The undersigned hereby certifies that the below declaration correctly reflects the nature and extent of the student's and co-authors' contributions to this work. In instances where I am not the responsible author I have consulted with the responsible author to agree on the respective contributions of the authors.

[redacted]

Jakob Hohwy
Main Supervisor
Date: 29/10/2018

In the case of the chapters that comprise this thesis my contribution to the work involved the following:

Thesis Chapter	Publication Title	Status	Contribution	Co-author contributions	Monash student Y/N
2	Conscious access in the near absence of attention: critical extensions on the dual-task paradigm	Published	50%. Contributed to study design and interpretation, some programming & data collection, all data analysis and wrote first draft.	1) Nao Tsuchiya: 25% 2) Jeroen van Boxtel: 15% 3) Lisandro Kaunitz: 5% 4) Pia Schröder: 5% + some programming	None
3	Sustained conscious access to incidental memories in RSVP	Published	60%. Contributed to study design and interpretation, all programming, all data analysis and wrote first draft.	1) Nao Tsuchiya: 15% 2) Jeroen van Boxtel: 10% 3) Jakob Hohwy: 10% 4) Vanessa Corneille: 2% + data collection 5) Jamin Wu: 3%	Yes (4,5)
4	Impaired perceptual sensitivity with intact attention and metacognition in functional motor disorder	Submitted	65%. Study design, contributed to interpretation, all programming, all data analysis and wrote first draft.	1) Jakob Hohwy: 15% 2) Peter Kempster: 10% 3) Rachel Newby: 5% 4) Catherine Ding: 3% 5) Kanae Nagao: 2% + data collection	None

Acknowledgements

I would like to first acknowledge my three supervisors: Jakob Hohwy, Nao Tsuchiya, and Jeroen van Boxtel. Your guidance and mentorship throughout my candidature have been critical for the production of this document and the papers it contains. Nao, thank you for taking a gamble on me—an unknown Honours grad living two states away and fascinated by consciousness. I will never forget the commitment you have for supporting junior researchers and the unwavering passion you have for this quest for consciousness. I hope to one day pay the opportunity forward as you have done on behalf of Christof. Jeroen, you have been an incredible emotional support as I have traveled this journey. I am very fortunate to have had a supervisor that can relate to me at my level and remind me of the bigger picture. Jakob, thank you for keeping us moving forward. You've assembled an amazing team here at Monash and have lit the spark for a new era of Australian consciousness research. This is just the beginning.

Thank you to my many friends and colleagues at CogPhilLab + tLab, especially Alex, Andrew, Dror and Laura, Iwan, Jenny Windt, Kelsey, Mateusz, Matt and Jane, Nessie, Noam, Steve, Tess, Thomas, and Winnie. I've been amazingly fortunate to have your support over the years and hope our discussions can continue for many years to come. To friends more broadly, thank you for weathering my obsession to this topic and keeping the real-world only a beer or two away. Special mentions go to Clive and Georgia, Edward, Eden, and Thami.

Thank you to my mentors at UQ. In particular, Jenny Burt for your guidance leading up to my return to academic studies. To Jason Tangen, your incredible seminar on consciousness in 2010 lit a spark that would inspire me to travel halfway across the country to change my life. I hope to one day match the enthusiasm you show for your work and inspire people as you inspired me. To your partner, Melanie McKenzie—I will never have the opportunity to thank you for your support during 2012. Your passing in 2015 was a tragedy but the tenacity you showed in completing your PhD despite knowledge of your illness was a constant source of strength.

To family, thank you for your encouragement. This has been a dream for a long time and your voices on the other end of the phone line have made each step easier. Special thanks to Mum and Dad, Alexander, Nicholas and Sujata, Lol-Lol, Nana and Gov, Steve and Henni, Euan and Laney, Alex and Abby, Amaya and Sophia.

Finally, Anna. None of this would have been possible without you. You've supported me through an Honours degree and now a PhD. It's fair to say you're one of the most patient people in history (and not just because of my stubborness!). I'm very excited for us to take the next step in our lives together and to start on that long list of post-PhD celebrations.

***

Preface

The science of consciousness seeks to explain how the thoughts and sensations that make up conscious experience might arise from physical activity in the brain. Whether this avenue of empirical enquiry will answer metaphysical questions that occurred to some of philosophy's earliest practitioners remains open for debate (Chalmers, 1996; Fodor, 1981; Searle, Dennett, & Chalmers, 1997). Despite this potential metaphysical limitation on our inquiry, cognitive science pays great dividends to advances on the *body* side of the mind-body problem—a so-called *reductionist bias* that has only received noticeable scrutiny in recent times (Krakauer, Ghazanfar, Gomez-Marin, MacIver, & Poeppel, 2017). Understanding the body (or at least the brain) is the aim of the neurobiological *Quest for Consciousness* (Koch, 2004) but without a comparative regard for the *mind*-side of the mind-body distinction, this quest will, at best, downplay the extraordinary properties of consciousness, or, at worst, misidentify their origin.

I emphasise a reciprocal approach in this thesis and use advances from science and philosophy to examine one of the human brain's most profound, functional qualities—the capacity to monitor itself. *Thinking about thinking* or *cognition about cognition*, often termed *metacognition*, has risen to prominence in cognitive science in part due to this functional quality but also due to influential neuroscientists and philosophers championing its role in the behavioural study of conscious awareness (Dehaene, Lau, & Kouider, 2017; Graham & Neisser, 2000; Grimaldi, Lau, & Basso, 2015; Lau & Rosenthal, 2011; Nelson, 1996; Terrace & Metcalfe, 2005). But an understanding of metacognition must start by examining the role it plays in cognition and perception more broadly. To this end, this thesis examines how metacognition is placed relative to not only consciousness but core processes that underlie cognitive function and perceptual decision-making, including attention and memory. The thesis identifies key associations and dissociations between these processes and provides insights into the role metacognition might play in a maturing science of consciousness and cognition.

Aim and scope

The aim of this thesis is to identify how metacognition relates to consciousness and the core processes that underlie perception and cognition including attention and memory. I argue that metacognition and conscious awareness are closely related but distinct phenomena. However, I believe nuanced treatment of this distinction can advance basic and applied research on each process and how they relate to cognitive function more broadly. I demonstrate this using *empirical* and *theoretical* enquiry. Visual psychophysics provides the medium for my empirical work. I use experimental paradigms to compare and contrast the cognitive mechanisms involved in metacognitive judgments with those cognitive mechanisms associated with *visual perception* in normal and abnormal human brains. Specifically, I examine the relationship between metacognition and *selective attention* (Chapter 2), metacognition and *working memory* (Chapter 3), as well as metacognition and *perception in functional and neurological motor disorders* (Chapter 4). This focus necessitates that my critique is broadly confined to behavioural considerations and an *operational* treatment of metacognition (i.e., the correspondence between the accuracy of perceptual decisions and confidence in those decisions) but my opening chapter and concluding remarks consider metacognition, consciousness, and cognitive function more broadly.

Chapter 1

Towards an understanding of metacognition, consciousness, and cognitive function

Julian Matthews[1,2]

1. Cognition and Philosophy Lab, Monash University, Clayton, Australia

2. School of Psychological Sciences, Monash University, Clayton, Australia

An early history of metacognition

The term *metacognition* first entered the scientific lexicon in developmental psychology texts from the mid-1970s. There it was used to describe changes in the awareness and knowledge of thinking processes, especially in children during early development (J. Brown, Lewis, & Monk, 1977; Flavell, 1979; Flavell & Wellman, 1977). But metacognition has a conceptual history that can be traced back much further. For instance, William James expressed that metacognitive processes associated with conscious awareness might dissociate from conscious perception:

> A mind which has become conscious of its own cognitive function, plays what we have called 'the psychologist' upon itself. It not only knows the things that appear before it; it *knows that it knows them*... It cannot, however, be regarded as primitive. The consciousness of objects must come first.
> (James, 1890, p. 272)

Earlier still, Aristotle posited that memory requires reflection (Sorabji, 1972). Critically, he noticed imperfections in this process: "the moment of the original experience and the moment of the memory of it are never identical" (Beare, 2010, p. 89). Aristotle suggested this discrepancy arose as a product of imperfect *encoding*, likening the 'implantation' of memories to impressions made using a seal. Such impressions were sustained and highly detailed when made in wax but they did not last when imprinted on running water or the decaying surface of old chamber walls (Beare, 2010). However, Aristotle missed that discrepancies might arise as a product of imperfect *access*. That is, a sustained and detailed wax impression might be misinterpreted when viewed in low light or in the wrong orientation. It is this idea of access that forms the core of the present thesis on metacognition.

The contemporary study of metacognition and consciousness emerged in the 1990s as consciousness was slowly regaining its credibility as a topic of serious scientific regard. Endel Tulving noted the tone in which metacognition was discussed in his foreword to a highly influential text on metacognition and consciousness (Metcalfe &

Shimamura, 1994). Specifically, how researchers used behaviouralistically safe expressions such as 'metamemory', 'memory monitoring' and 'mnemonic behaviour' as if avoiding, in Tulving's view, *the big bad 'C' word*. At this stage, the study of metacognition had split into two broadly independent streams: one that examined developmental psychology, and the other within experimental memory research. This division is still seen today though work on 'metamemory' has largely evolved into experimental-cognitive research on the relationship between metacognition and consciousness more broadly. The larger developmental stream remains and is joined by other subfields including psychopathological research which I will return to below.

A common theme in these earlier studies of metacognition and consciousness was *feeling of knowing* (FOK) judgments, a concept pioneered during the reign of behaviourism (Hart, 1965, 1967a, 1967b). FOK tasks presented people with lists of general-knowledge questions (e.g., "How many bones are in an adult human skeleton?") and asked them to recall the correct answer. If they were unable to recall the answer, participants were instructed to predict the likelihood they would recognise the answer in a forced-choice recognition test. By comparing what people thought they knew with what they actually knew, the accuracy of these FOK judgments gave a crude estimation of participants' *metacognitive performance* though it was not discussed in those terms at the time.

The FOK paradigm was examined in several studies of patients with Korsakoff syndrome and in patients with lesions restricted to the frontal lobes (Janowsky, Shimamura, & Squire, 1989; Shimamura & Squire, 1986a, 1986b, 1988). Both groups demonstrated significant impairments in FOK judgments. Korsakoff patients exhibited additional impairments in memory recall and recognition but patients with purely frontal lobe lesions had otherwise intact memory function. These results were the first empirical evidence that memory and metacognition might be dissociable, and foreshadowed the important role prefrontal cortex would play in future neurocognitive models of metacognition. Drawing from these and other findings, several prominent reviews and books highlighted the importance of metacognition for studying conscious awareness and suggested a rigorous science of metacognition might offer a bridge between psychological and philosophical perspectives on consciousness (Flanagan, 1992; A. Koriat, 1993; Asher Koriat, 2007; Nelson, 1996). These authors raised an important challenge for metacognition research that echoes through to modern times, namely its measurement. Behavioural measures of metacognition are prone to technical and methodological problems that limit their explanatory power with respect to consciousness.

Thomas Nelson (1984) was perhaps the first to contrast measures of metacognition. He compared eight methods for defining the accuracy of FOK judgments and concluded that even the most popular measure of the time had “serious shortcomings” because it could not quantify a subject’s degree of metacognitive insight. Further problems and pitfalls for measuring metacognition have been identified, centreing on inter-individual differences in participants’ decision criterion when judging FOK (John Dunlosky & Nelson, 1994; Masson & Rotello, 2009; Nelson, 1996). Other problems included that FOK judgments are affected by the number of alternatives offered and that participants’ reports are influenced by their expectations regarding the difficulty of the recognition test that will follow (Schwartz & Metcalfe, 1994).

A hugely influential concept that emerged during this time was the *Metacognitive Model* (Nelson, 1990, 1996; Nelson & Narens, 1994). The Metacognitive Model took as its starting point the idea that individuals monitor their own cognitions but that this process was imperfect:

> If the object-level aspect of the individual's cognitions is operationalised in terms of some kind of criterion performance, then we can take the critical step of *assessing the correspondence between what the individual believes is cognitively occurring and the empirical reality of what is actually occurring*. (Nelson, 1996, p. 106)

On the basis of the Metacognitive Model, metacognition could be empirically realised by assessing the correspondence between the *object-level* criterion and a *meta-level* subjective response (e.g., a verbal report, FOK judgment, or confidence rating). Importantly, the Metacognitive Model could be applied to limitless task designs which meant the study of metacognition was free from the FOK paradigm.

Metacognition, knowledge, and Higher Order theories of consciousness

As the study of metacognition embraced consciousness so did the study of consciousness embrace metacognition. One of the central reasons for this was that the Metacognitive Model seemed able to differentiate between conscious and unconscious mental states, provided one accepted certain assumptions about the origin of consciousness. A method that reliably distinguishes between conscious and unconscious mental states is a holy grail of consciousness research (Metcalfe & Schwartz, 2016). Such a method could reveal stimulus conditions associated with either conscious or unconscious perception in healthy participants and settle some of the field's longest lasting debates, namely whether selective attention is necessary for consciousness or whether working memory is necessarily conscious. A related reason is that this method would be a critical tool for identifying the *Neural Correlates of Consciousness* (NCCs)[1]. That is, the minimal set of neuronal mechanisms jointly sufficient for any one specific conscious percept (Crick & Koch, 1990; Koch, 2004; Rees, Kreiman, & Koch, 2002).

An early example of metacognition being used to systematically distinguish between conscious and unconscious mental states can be found in studies of knowledge; notably, explicit and implicit knowledge. The precise definition of these terms is the subject of extensive discussion that is outside the scope of this introduction (Dienes, 2008a; Dienes & Perner, 1996, 1999) but explicit knowledge is widely regarded as

[1] It would be unlikely to reveal NCCs directly because it is widely believed that certain background conditions are necessary for consciousness (Seth, Dienes, Cleeremans, Overgaard, & Pessoa, 2008) and behavioural paradigms involve neural processing associated with conscious perception but also neural processing that is a prerequisite for behavioural report (Tsuchiya, Wilke, Frässle, & Lamme, 2015).

knowledge that can be readily accessed and articulated by the holder of that knowledge. In contrast, implicit knowledge can be viewed as knowledge that is 'inaccessible' to the holder (under a broadly declarative view of access) but nevertheless drives decision-making or behaviour. Specifically, implicit knowledge is knowledge that represents properties in memory without relating them to any particular entity (Dienes & Perner, 1999). The existence of implicit knowledge was fairly uncontroversial throughout the mid-to-late 1990s but the field lacked a method to demonstrate its existence empirically and tie this method to conscious or unconscious mental states.

An important breakthrough for this project was found in philosophy of mind. Specifically, the qualities that make knowledge explicit are, under certain philosophical accounts of consciousness, necessary and sufficient for that knowledge to be regarded as conscious. These views are expressed most clearly under *Higher Order theories of consciousness*[2] that view mental states as conscious because they are the subject (or potential subject) of higher order mental states of various cognitive or perceptual kinds (Richard Brown, 2015; Carruthers, 2000; LeDoux & Brown, 2017; Lycan, 1996; Rosenthal, 1986, 2005). Higher Order theories are contrasted with first-order accounts that view consciousness as representing

[2] Metacognition is relevant for all Higher Order theories but different theorists describe the relation between metacognitive processing and consciousness to a greater or lesser extent. David Rosenthal contrasts the higher order nature of conscious awareness (i.e., *Higher Order Awareness* or HOA) with metacognitive regulation and suggests that they both involve higher order psychological states but have little more in common (Rosenthal, 2000, 2005, 2012, 2018). Metacognitive regulation is characterised by its utility for cognitive function but there is no reason to assume that (conscious) mental states that are the subject of HOA must have utility over and above (unconscious) mental states that are not the subject of HOA. Likewise, metacognition is distinct from consciousness under the *Radical Plasticity Thesis* (RPT) (Cleeremans, 2008, 2011; Timmermans, Schilbach, Pasquali, & Cleeremans, 2012). RPT shares features of Higher Order theories but also *enactivism* (Noë, 2004, 2009). The capacity for metacognition is an instance of a larger class of 'predictive redescription' processes that occur unconsciously and automatically as one learns implicit associations between first-order mental representations. Only a subset of these 'meta-representations' are consciously experienced on the basis of their relevance for the organism (e.g., to motivate action). Therefore, consciousness has utility according to RPT.

properties of the world directly and that view our capacity to reflect on that knowledge (i.e., metacognition) as a consequence of cognitive access (Block, 2011a, 2011b), recurrent processing (Dennett, 2001), or global availability (Baars, 1988; Dehaene, 2014; Dehaene, Kerszberg, & Changeux, 1998).

Returning to the relationship between metacognition and consciousness, meta-level reports (e.g., confidence judgments) were viewed as a means to test for relevant Higher Order thoughts. Consequently, the Metacognitive Model was now an empirical approach for examining consciousness that could be expressly tied to Higher Order theories of consciousness and the broader philosophy of mind (Rosenthal, 2000). Specifically, the correspondence between accuracy and confidence judgments (*metacognitive performance* henceforth) could be used to define two criteria for identifying when a behavioural report (e.g., recognition of an artificial grammar) was made *without* conscious knowledge: 1) the *guessing* criterion, and 2) the *zero-correlation* criterion (Chan, 1992; Cheesman & Merikle, 1984; Dienes, 2004; Dienes, Altmann, Kwan, & Goode, 1995; Dienes & Berry, 1997; Dienes & Perner, 2004).

First, if subjects believe they are *literally guessing* on a cognitive task (according to their meta-level confidence judgments) yet their object-level performance on that task is above-chance then the guessing criterion views the knowledge that underlies their performance as unconscious. Second, if there is *no measurable correspondence* between the accuracy of subjects' decisions and their confidence in those decisions then the zero-correlation criterion views the knowledge that underlies their decision making as unconscious. Conversely, a high degree of correspondence between object and meta-level reports was thought to reflect *conscious knowledge*.

Metacognition, perception, and consciousness

The approach to operationalising metacognition that was established by the Metacognitive Model and that found support in the knowledge literature and Higher Order theories of consciousness continues today. However, metacognition is often conceptualised in an even narrower sense – as metacognitive performance in visual perceptual decision-making tasks (Fleming & Lau, 2014; Kunimoto, Miller, & Pashler, 2001; Norman & Price, 2015) (although see the following review on metacognition in multisensory perception (Deroy, Spence, & Noppeney, 2016)). This approach relates the *first-order* response (i.e., the accuracy of a perceptual decision) to the *second-order* response (i.e., confidence in that perceptual decision). If one accepts the widely held assumption that the sensory information that leads to a perceptual decision is also used (albeit transformed) when rating confidence in that decision (that is, second-order information is constituted by first-order information (Baranski & Petrusic, 1998; Galvin, Podd, Drga, & Whitmore, 2003; Ko & Lau, 2012; Kunimoto et al., 2001; Maniscalco & Lau, 2012)), then a wealth of analytical methods can be employed to measure metacognitive performance but also relate conscious and unconscious knowledge to *perception*.

To this end, novel approaches for measuring metacognition, which employ hierarchical Bayesian estimation (Fleming, 2017), logistic regression (Kristensen, Sandberg, & Bibby, 2018; Rausch & Zehetleitner, 2017), mixed modelling (Matthews, Schröder, Kaunitz, van Boxtel, & Tsuchiya, 2018; Matthews, Wu, et al., 2018), and confidence thresholds (Gallagher, Suddendorf, & Arnold, 2018) have begun to appear. However, the most prominent metacognitive measures today are still those based on signal detection theory (Green & Swets, 1966; Macmillan & Creelman, 2005), an approach for assessing how faithfully decisions separate signal from noise. The conventions for classifying sensitivity in a perceptual decision (the *type 1 task*) are extended to participants' ability to assign confidence in that decision (the *type 2 task*) resulting in probability estimates of metacognitive *sensitivity*, *bias* and *efficiency* that relate type 1 to type 2 performance (Clarke, Birdsall, & Tanner, 1959; Fleming & Lau, 2014; Galvin et al., 2003; Kunimoto et al., 2001; Maniscalco & Lau, 2012; Sherman, Seth, & Barrett, 2018). These measures are popular because they can be applied in many behavioural tasks, account for various sources of response bias (although see (Barrett, Dienes, & Seth, 2013; S. Evans & Azzopardi, 2007)), and allow metacognitive performance to be compared with established measures of objective performance from the signal detection literature.

As with conscious and unconscious knowledge, metacognitive measures become especially interesting for studying perception when the correspondence between accuracy and confidence is either very high or very low. In the first case (i.e., a situation where, on average, accurate perceptual judgments are afforded ratings of high confidence and inaccurate perceptual judgments are those made with low confidence), high metacognition is thought to reflect decision-making that is accompanied by a high degree of insight into the perceptual decision. It is widely regarded that such insight is only possible if the decision-maker is *consciously aware* of the perceptual content. That is, high metacognitive performance is widely believed to demonstrate conscious perception under whatever conditions the perception took place.

In the second case (i.e., a situation where accuracy of perceptual judgments rarely corresponds with the magnitude of confidence ratings), low metacognition is thought to reflect those cases where subjects have little or no insight into their perceptual decisions. This is not surprising if decision accuracy is no better than chance. However, this form of metacognitive measure has built its reputation for an alternate situation–the case where subjects exhibit low metacognitive performance but high decision accuracy. Following from the assumption that sensory information is the basis for decision confidence (Galvin et al., 2003; Ko & Lau, 2012) and the logic that underlies the guessing criterion, this situation is widely believed to demonstrate the decision-maker successfully processed the stimulus but was not consciously aware of the perceptual content (although this is critiqued elsewhere (Dienes, 2008a; Fleming & Lau, 2014; M. Peters, Kentridge, Phillips, & Block, 2017; M. Peters, Ro, & Lau, 2016)). That is, low metacognitive performance with high decision accuracy is frequently believed to demonstrate unconscious perception[3].

[3] *Blindsight* is widely regarded as the prototypical example of unconscious perception. Blindsight patients are capable of discriminating visual stimuli presented in their cortically 'blind' field yet report no visual experience of those stimuli. For many years it was considered a purely clinical phenomenon yet was still regarded as one of the most important contributions to philosophy of mind from experimental psychology (K. Martin, 2004). Blindsight is important for consciousness science

Envisioned in this way, metacognitive performance has been used in countless behavioural studies and perceptual tasks to relate cognitive and perceptual processing to conscious awareness or the absence of conscious awareness. This approach has also been applied to map metacognition to the brain and to justify associations between the neural basis of metacognition (especially prefrontal cortex) and the NCCs in humans (Lau & Rosenthal, 2011; Metcalfe & Schwartz, 2016). Controversially, this work has also been extended to the debate regarding machine consciousness. Under this account, machines require some capacity for self-monitoring that is functionally equivalent to human metacognitive processing in order to be conscious in the way the term is generally understood (Dehaene et al., 2017).

The neuroscience of metacognition

The foundations for a cognitive neuroscience of metacognition were outlined in a special issue of *Consciousness & Cognition* (Nelson & Rey, 2000). There, metacognition was broadly differentiated into two core functional aspects with respect to first-order cognitive processes: *monitoring* and *control*. As noted before, foundational studies of metacognition had linked impairments in metacognitive monitoring to lesions in frontal cortex (e.g., Korsakoff patients (Shimamura & Squire, 1988), Alzheimer's disease (Schacter, Moscovitch, Tulving, McLachlan, &

---

because the disorder is commonly associated with lesions to *primary visual cortex* (i.e., V1). This has ramifications for the NCC project because, if blindsight patients are believed, areas towards the 'back' of the cortex may represent the locus for dissociating visual function from visual experience which places less emphasis on areas towards the 'front' of the cortex, an area that certain major theories of consciousness regard as necessary (Boly et al., 2017; Bor & Seth, 2012; Del Cul, Dehaene, Reyes, Bravo, & Slachevsky, 2009; Koch, Massimini, Boly, & Tononi, 2016; Laureys & Schiff, 2012). However, the precise nature of clinical blindsight and its relationship to consciousness remains a topic of heated debate (Cowey, 2010; Overgaard, 2012), not least because the literature on clinical blindsight rests on studies of very few patients, such as the professional research participants G.Y. and D.B. (Weiskrantz, 1986, 1997). It follows that reliable methods for producing unconscious perception in healthy participants would open the study of blindsight to new task designs and even research in nonhuman animals; a paradigm shift for the NCC project.

Freedman, 1986), and traumatic brain injury (McGlynn & Kaszniak, 1991)) but it was noted that metacognitive monitoring did not lend itself easily to study in metacognitive tasks at the time (Shimamura, 2000). On the basis of this, early neuroscientific studies of human metacognition prioritised its control aspect and drew upon the existing neuroscientific literature on executive control processes to tie metacognitive regulation to sites that achieve these functions in frontal regions of the brain. Notably, metacognitive regulation was associated with prefrontal regions involved in selective attention and working memory (Fernandez-Duque, Baird, & Posner, 2000; Shimamura, 2000, 2008) but also regions involved in decision-making and error detection such as dorsal anterior cingulate cortex (Bush et al., 2002; Carter, Botvinick, & Cohen, 1999).

Independent from these examinations of metacognitive control, several prominent neuroscientific studies examined awareness associated with FOK and *tip-of-the-tongue* (TOT) states (i.e., the feeling of confidence that information exists in one's long-term memory and is on the verge of recovery but remains temporarily inaccessible (Roger Brown & McNeill, 1966; Schwartz, 2001; Schwartz & Metcalfe, 2011)). In a semantic recollection task, both TOT and FOK were associated with activation in largely parietal regions (Maril, Simons, Mitchell, Schwartz, & Schacter, 2003; Maril, Wagner, & Schacter, 2001). However, when both report procedures were compared directly, activation of frontal regions including anterior cingulate cortex and dorsolateral prefrontal cortex were associated with TOT but not FOK (Maril, Simons, Weaver, & Schacter, 2005). Similar dissociations were found between FOK and retrospective confidence judgments such that damage to right ventro-medial prefrontal cortex and dorsal anterior cingulate cortex were each associated with impairments in FOK but a preserved capacity for rating confidence (Modirrousta & Fellows, 2008; Schnyer et al., 2004). Conversely, intact FOK with impairments in confidence rating were identified in patients with lesions in lateral frontal cortex (Pannu, Kaszniak, & Rapcsak, 2005). Collectively, these studies formed early evidence that states of awareness associated with metacognition are unlikely to be instantiated in a single neurocognitive mechanism despite qualitative similarities between, for instance, FOK and TOT.

As the methods for operationalising metacognition converged on metacognitive performance and the signal detection framework, neuroscientific studies increasingly targeted this aspect. An advantage of this approach when compared to those using FOK or TOT was that precise psychophysical task designs could control or closely monitor participants' decision accuracy while independently investigating the effects of different task conditions on metacognition and its neural substrate. Results from within and between-subjects studies identified overwhelming evidence that metacognitive sensitivity was associated with prefrontal cortex; dorsolateral, rostrolateral, and ventromedial regions in particular (Del Cul, Dehaene, Reyes, Bravo, & Slachevsky, 2009; De Martino, Fleming, Garrett, & Dolan, 2013; Fleming, Huijgen, & Dolan, 2012; Fleming, Weil, Nagy, Dolan, & Rees, 2010; Hilgenstock, Weiss, & Witte, 2014; Lau & Passingham, 2006; Yokoyama et al., 2010). These results strongly imply that prefrontal cortex has a mediating role in the accuracy of retrospective confidence judgments (see (Fleming & Dolan, 2012) for review). This finding was used as empirical support for Higher Order theories of consciousness (Lau & Rosenthal, 2011) although Rosenthal has recently expressed concerns about the extent to which confidence judgments might reflect higher order awareness (Rosenthal, 2018).

A separate but related line of neuroscientific inquiry examined how humans and other species optimise their decision-making behaviour by modeling their internal uncertainty (Knill & Pouget, 2004). A key finding from this was humans use their knowledge of uncertainty to bias perceptual and motor decision-making in a purportedly Bayes optimal fashion ((Daw, O'Doherty, Dayan, Seymour, & Dolan, 2006; Deneve, 2012; Whiteley & Sahani, 2008) although see (Jarvstad, Hahn, Warren, & Rushton, 2014; Rahnev & Denison, 2018)). Importantly for the relationship between metacognition and consciousness, this process of optimisation could be achieved implicitly. That is, Bayesian inference optimises decision-making behaviour employing information we are not consciously aware we know or are learning (Whiteley & Sahani, 2008).

Recent studies have extended the above framework in an effort to dissociate those neural processes involved in computing certainty (i.e., evidence reliability) from those involved in confidence judgments (Aitchison, Bang, Bahrami, & Latham, 2015; Bang & Fleming, 2018; Fleming & Daw, 2017). This is an important step in identifying the neural basis of metacognition but, with respect to consciousness, it is surprising that this project has not directly addressed the question of implicit learning raised in the decision-making literature. Identifying how explicit and implicit knowledge factor into these computations, the role they play in decision optimisation, and to what extent explicit knowledge might be necessary to form confidence judgments or other metacognitive reports[4] are vital considerations for future research on the neural basis of metacognition and broader consideration of the possible functions of conscious awareness. The role TOT states play in curiosity (Bloom, Friedman, Xu, Vuorre, & Metcalfe, 2018; Metcalfe, Schwartz, & Bloom, 2017; Schwartz & Cleary, 2016) is one topic that might probe deeper qualities of metacognitive thought. The link to more mechanistic conceptions of metacognition might exist in research that shows future value-based decisions are influenced by explicit representations of confidence that underlie real-world behaviours such as gambling (Folke, Jacobsen, Fleming, & De Martino, 2016; Otto, Fleming, & Glimcher, 2016).

---

[4] Another topic of interest in contemporary metacognition research is relevant to this discussion—whether metacognition is supported by domain-specific or domain-general processes. This research concerns whether metacognition relies on separate mechanisms that are engaged when performance is evaluated in different domains or whether a single overarching resource is applied to the task at hand. The consensus from a recent meta-analysis was that both mechanisms co-exist in the human brain and that some tasks are more likely to rely on shared metacognitive processes than other tasks (Rouault, McWilliams, Allen, & Fleming, 2018). However, the overwhelming majority of studies that compare metacognition between domains operationalise metacognition using signal detection theoretic methods. There is reason to believe metacognition is a more complex process than these methods assume (Fleming & Daw, 2017; Grimaldi, Lau, & Basso, 2015; Moran, Teodorescu, & Usher, 2015). Future examinations of domain-general versus domain-specific metacognition might build on these more complex models of metacognitive processing and consider standardised tasks between domains (Ruby, Giles, & Lau, 2017).

Two cognitive processes related to consciousness: attention and short-term memory

The history of metacognition research demonstrates that a holistic understanding of metacognition requires that it is placed relative to consciousness but also broader cognitive functions such as attention and memory. This principle applies to an even greater extent for consciousness research since consciousness famously eludes an all-encompassing, functional definition (Chalmers, 1995, 1996; Searle et al., 1997). To this end, in the next section I illuminate two cognitive processes that a frequently associated, and sometimes equated, with consciousness—attention and short-term memory.

Attention

Debates about the relationship between attention and consciousness can be traced back to at least the beginning of psychology as a scientific discipline (Wundt, 1874). Putting aside the incontrovertible mysteries of consciousness, attention is a multifaceted cognitive function with enough complexity to motivate its own line of philosophical inquiry (Watzl, 2011a, 2011b). Endogenous (top-down, motivationally-driven) and exogenous (bottom-up, saliency-driven) attention, a distinction documented since the 19th century (James, 1890), distinguishes how attention is studied by the sciences. While *bottom-up attention* is typically linked to arousal or alertness, it is *top-down attention* that is of primary interest when we consider consciousness. Top-down attention is the volitional aspect of attention we typically employ to 'select' from a multitude of competing sensory inputs. This functional quality is responsible for its common alternative name, selective attention, but it has also been dubbed an 'analyser' (van Boxtel, Tsuchiya, & Koch, 2010a).

A core feature of debates regarding top-down attention and consciousness concerns whether top-down attention is necessary for consciousness (Jennings, 2015). The question of necessity is important because major theories of consciousness disagree on the answer. Notably, the *integrated information theory* (IIT) (Tononi, 2008; Tononi, Boly, Massimini, & Koch, 2016) which considers consciousness an independent phenomenon from attention and *global neuronal workspace theory* (GNWT) which considers top-down attentional amplification a prerequisite for conscious awareness (Dehaene, 2014; Dehaene et al., 1998). More broadly, the necessity of top-down attention for consciousness has important ramifications for whether attentional networks in the 'front' of the brain should be considered part of the NCCs (Boly et al., 2017; Koch, Massimini, Boly, & Tononi, 2016; Tsuchiya, Wilke, Frässle, & Lamme, 2015).

Certain branches of the attention and consciousness debate express doubt that the two processes can be disentangled with some going as far as to question whether they even represent distinct phenomena (Cohen & Dennett, 2011; De Brigard & Prinz, 2010; Prinz, 2012). And though an equally representative group agrees that the functions and neuronal mechanisms of attention and consciousness might be separable (Baars, 2005; Hohwy, 2012; Koch, 2004) the precise nature of their interaction generates ground for further disagreement. Crucially, if consciousness can be fully dissociated from top-down attention, the breakdown of the necessity claim would demand revision of the core feature of GNWT and other theories that regard top-down attention as necessary for conscious experience (Jennings, 2015; Prinz, 2012).

Support for GNWT and other theories that rely on the necessity claim is found in behavioural studies that demonstrate conscious report failing in the absence of attention. These include but are not limited to studies of *inattentional blindness* (Mack & Rock, 1998), the *attentional blink* (Chun & Potter, 1995; Raymond, Shapiro, & Arnell, 1992), and *change blindness* (Rensink, O'Regan, & Clark, 1997; Simons & Levin, 1997; Tse, 2004). The necessity claim also draws support from neurological evidence. For instance, *visual neglect* (the invisibility of stimuli in one hemifield) which is characteristic of damage to attentional complexes in the cortex and subcortical regions of the brain (Swan, 2001; Vallar, 2001, 2007; Werth, Von Cramon, & Zihl, 1986).

But does a failure to report imply a lack of conscious experience or some breakdown in the cognitive processes underlying report itself? This question is at the core of arguments against the necessity claim; that the richness of *phenomenal consciousness* 'overflows' our capacity to *cognitively access* experience (Block, 2007b; Fazekas & Overgaard, 2018; Mole, 2008). Those that challenge the necessity claim typically draw upon perception of visual *gist* to exemplify this distinction (Koch & Tsuchiya, 2007a; Tsuchiya & Koch, 2008; van Boxtel et al., 2010a). Phenomenologically, we have an intuitive sense that the gist of our visual world is accessible to consciousness at almost all times. Empirical support for this intuition is found in studies that demonstrate inattentional blindness and attentional blinks rarely occur if the gist of an image is impacted (Einhäuser, Koch, & Makeig, 2007; K. K. Evans & Treisman, 2005; Mack & Rock, 1998; Sampanes, Tseng, & Bridgeman, 2008). Additionally, participants can perceive the gist of an image at stimulus durations as short as 30ms (Fei-Fei, Iyer, Koch, & Perona, 2007) or even possibly a single 13-16ms visual frame (Pavlopoulou & Yu, 2010; Potter, Wyble, Hagmann, & McCourt, 2014). It seems unlikely that top-down attention could play a critical role in conscious perception at these speeds. Moreover, top-down attention conceivably functions to enhance local sensory features, such as specific contents of a visual scene. This function serves limited utility when the fundamental property of gist is to summarise sensory contents in their entirety.

One paradigm that has surfaced to examine the relationship between attention and visual consciousness is the *dual-task*. Using an identical visual display, participants are instructed to respond either to an attention-demanding *central task* or a challenging *peripheral task*. Critically, in the *dual-task condition*, they are instructed to respond to both tasks together while fixing the focus of their selective attention on the central task (Braun & Julesz, 1998; Sperling & Dosher, 1986). By comparing performance in each *single-task condition* with the dual-task condition, it is possible to quantify to what extent the attention-demanding central task affects discrimination of peripheral stimuli.

If the engagement of attention in one task can be said to reduce available resources for another, then it follows that dual-task conditions should impact performance for most if not all categories of stimuli presented in the visual periphery (Jennings, 2015). However, a growing list of stimulus types show very little or even no drop in discrimination performance under dual-task conditions. This is fairly unremarkable in the case of low-level perceptual features, such as colour or size (Braun & Julesz, 1998; Lee, Koch, & Braun, 1999) but has also been demonstrated with complex stimulus features including face-gender, face identity, and even the presence of animals or vehicles in natural scenes (García-Gutiérrez, Aguado, Romero-Ferreiro, & Pérez-Moreno, 2017; Li, VanRullen, Koch, & Perona, 2002; Reddy, Reddy, & Koch, 2006; Reddy, Wilken, & Koch, 2004). Intriguingly, several categories of simple visual stimuli cannot be discriminated under dual-task conditions, including rotated letters or the orientation of coloured discs (Li et al., 2002; Pastukhov, Fischer, & Braun, 2009; Reddy et al., 2004). Clearly, whether peripheral discrimination is possible under dual-task conditions does not derive from stimulus complexity alone (Tsuchiya & Koch, 2008; VanRullen, Reddy, & Koch, 2004).

While at first glance these results present a sobering picture for the necessity claim, several noteworthy caveats exist before interpretation is entirely clear. Dual-task experiments typically employ onerous training over many hours before participants achieve proficiency. Such requirements may lead to the engagement of separate brain regions, rendering the experience of trained participants quite different to that of naïve subjects (Joseph, Chun, & Nakayama, 1997). Furthermore, dual-task studies typically require that participants make a response to the central task followed by the peripheral task. This order of reports does not detract from stimuli that have been successfully discriminated in past dual-task studies but it does bring into question whether those stimulus categories that we fail to discriminate only fail because of report demands.

However, the key qualification with past dual-task studies is whether peripheral discriminations are being performed consciously at all. In addition to the phenomenon of *unconscious perception* and *blindsight* discussed above, a body of evidence has established that certain tasks (such as word-stem completion) may be possible for stimuli that are presented subconsciously (Debner & Jacoby, 1994; Merikle & Joordens, 1997). This raises the possibility that successful discrimination under dual-task conditions may be a product of subconscious processing. Thus, performance alone is insufficient to determine whether conscious experience is possible in the absence of attention and a more direct measure of participants' internal state is required (Jennings, 2015).

One category of measure that meets this criteria is metacognition. High metacognitive performance (i.e., a strong correlation between the accuracy of participants perceptual decisions and confidence they hold in those decisions) is a reflection of conscious knowledge (Dienes, 2008a) and strongly implies that the participant is consciously aware of the perceptual content on which the decision is based (Fleming & Lau, 2014; Lau & Rosenthal, 2011; Nelson, 1996; M. Peters et al., 2016). Therefore, metacognition is an ideal behavioural tool for examining whether peripheral stimuli presented under dual-task conditions might be consciously perceived.

On the basis of this, we examined how metacognition relates to consciousness and selective attention (Chapter 2). Specifically, we employed several measures of metacognitive performance in an extended version of the dual-task paradigm (Matthews, Schröder, et al., 2018). Our extended dual-task paradigm used psychometric thresholding to vastly reduce the need for training and we used a partial-report procedure to account for report demands. In addition to this, a final experiment employed 'blended' stimuli to address the inherent saliency of faces (Cerf, Frady, & Koch, 2009); an ongoing limitation of dual-task studies that use face stimuli. Participants exhibited high metacognitive sensitivity when discriminating face-gender even under dual-task conditions. We concluded that complex stimuli features such as the gender of faces are accessible to consciousness in the near absence of attention.

Short-term memory

The relationship between memory and consciousness is another central issue in cognitive science. Discussion centres on *short-term memory* and how consciousness relates to *working memory* in particular (Baars & Franklin, 2003; Baddeley, 1986, 2003; Gross, 2018; Persuh, LaRock, & Berger, 2018; Sergent, 2018). Short-term memory is the term used to describe our overall capacity to retain information over a short delay when this information is no longer present in the external world. If this information is explicitly selected and retained for a future purpose, it is called working memory. For example, we store the location of our lunch in working memory when we temporarily leave it to collect cutlery or a drink.

As with much of cognitive science, working memory research is dominated by studies that examine the visual domain (although see (Cowan, 1998; Quak, London, & Talsma, 2015; Salmela, Moisala, & Alho, 2014)). To this end, *visual working memory* (VWM) has been defined as that subset of working memory used in the active maintenance of visual representations for the service of cognitive tasks (Luck & Vogel, 2013). VWM has been associated with critical executive functions including selective attention (J. R. Anderson, Matessa, & Lebiere, 1997; Awh & Jonides, 2001), motor skill acquisition (Maxwell, Masters, & Eves, 2003), and object tracking during visual search (Drew, Horowitz, Wolfe, & Vogel, 2011; Drew & Vogel, 2008; Luria & Vogel, 2011).

Despite these broad functional qualities, VWM has a remarkably limited capacity. Most agree that 3 or 4 items can be stored in working memory at any one time (Cowan, 2010; Luck & Vogel, 1997). However, there are conflicting accounts regarding the effect of stimulus complexity on these limits (Alvarez & Cavanagh, 2004; Awh, Barton, & Vogel, 2007; Eng, Chen, & Jiang, 2005; Fougnie & Alvarez, 2011) and how these limits might be affected by encoding rate (Luck & Vogel, 2013; Vogel, Woodman, & Luck, 2006).

With respect to consciousness, the functional qualities of working memory have seen it equated with the eponymous *workspace* of GNWT, where thinking and cognition are believed to take place (Baars, 2005; Baars & Franklin, 2003). The primary argument is that capacities of the workspace that are essentially conscious, such as mental rehearsal and visual imagery, are elements intimately linked with working memory function. Conversely, cognitive mechanisms that support but are distinct from working memory are unconscious according to GNWT because they are not necessary for global availability (Baars & Franklin, 2003). Further support for a the link between GNWT and working memory is provided by Alan Baddeley, the architect of the leading model of working memory. He associated one of the core components of working memory, the *episodic buffer*, with conscious memory retrieval using the global workspace theoretic framework (Baddeley, 2000, 2003). Recently, GNWT was tied to other features of working memory (e.g., *latent* working memory–information that is not actively maintained) to support the claim that working memory is separate from but conceivably underlies conscious access when the contents of memory are amplified by selective attention (Sergent, 2018).

VWM has been linked to perceptual consciousness. For many years VWM was believed to operate on and maintain only information that had been consciously perceived (Baddeley, 1986; Carruthers, 2015; Prinz, 2012). Recently, this view has been questioned on the basis of evidence for the existence of unconscious VWM (Bergström & Eriksson, 2014; Soto, Mäntylä, & Silvanto, 2011) (although see (Persuh et al., 2018)). While it is necessary that VWM is conscious according to GNWT (although see (Baars, 1997)), other leading theories of consciousness (e.g., IIT and Higher Order theories) do not require that the contents of VWM are consciously accessible. Independent from any one theory of consciousness, an argument is made that working memory is distinct from conscious awareness because working memory can operate on unconscious representations and can be engaged without awareness (Soto & Silvanto, 2014). However, studies that attempt to distinguish between conscious and unconscious memory stores encounter the same technical and methodological challenges that are associated with broader studies of perception (M. Peters et al., 2017, 2016). Metacognition is an ideal tool for meeting many of these challenges and on that basis we employed metacognitive measures to illuminate how VWM relates to consciousness and cognition.

In Chapter 3, we examined how metacognition relates to consciousness, selective attention, and VWM. Specifically, we contrasted VWM (here, *explicit memory*) with *incidental memory*, our capacity to remember items viewed outside the focus of our primary attentional goal (i.e., memory that is not actively maintained) (Beck, Peterson, Boot, Vomela, & Kramer, 2006; Castelhano & Henderson, 2005; Dickinson & Zelinsky, 2007; Kaunitz, Rowe, & Tsuchiya, 2016; Williams, Henderson, & Zacks, 2005). Incidental memory has almost exclusively been studied in the visual search domain (although see (Williams, 2010)) which means past studies cannot rule out that memory was supported by the cognitive processing that underlies self-directed visual search. Also, past studies of incidental memory have separated when stimuli were encoded into memory and when they were tested which means these studies cannot rule out that memory capacity or access diminishes over time.

To address these and other limitations, we contrasted explicit and incidental memory using the *Rapid Serial Visual Presentation* paradigm (Potter, 1976; Spence & Witkowski, 2013). We examined objective performance, confidence, and metacognitive sensitivity and found that participants have high capacity and metacognitively accessible incidental memory for faces in both upright and inverted orientation. Moreover, incidental memory was broadly equivalent to explicit memory which suggests that selective attention may be less important for memory encoding and/or access than it seems. In novel category analysis we found that semantic features of faces did not influence conscious access to memories. We concluded that incidental memory has a large capacity, is consciously accessible, and arises naturally as a consequence of perception (Matthews, Wu, et al., 2018).

Metacognition at the borderland of psychiatry and cognitive science

The contemporary scientific study of metacognition spans at least six subfields of cognitive science: social and comparative psychology, experimental and developmental psychology, perceptual neuroscience, and psychopathology (Proust, 2013). My final chapter lies at the intersection of these last two subfields: perceptual neuroscience and psychopathology. In the following section I briefly introduce metacognition in the clinical domain and describe the contribution that cognitive science (specifically, the *predictive processing* framework of brain function) is making for studying perception in an unexpected branch of neurology—motor disorders.

For some time, metacognition has been seen as a topic of interest in psychiatry, particularly schizophrenia, anxiety, and depression (Lysaker et al., 2005, 2008; Wells, 2011). Cognitive-behavioural therapies that employ metacognitive techniques have recently been championed in the treatment of psychosis, compulsivity, and other aberrant behaviours (Hauser, Allen, NSPN Consortium, Rees, & Dolan, 2017; Lysaker, Hamm, Hasson-Ohayon, Pattison, & Leonhardt, 2018; Moritz & Woodward, 2007; Moritz, Woodward, & Balzan, 2016). Disappointingly, the psychiatric subfield of metacognitive research and the perceptual subfield that examines metacognition and consciousness remain largely separate. This is, in no small part, due to the methods by which each subfield operationalise metacognition which differ markedly and limit the extent to which breakthroughs in either subfield might cross-pollinate. However, this state of affairs may change given the excitement that surrounds the application of cognitive science in other areas of psychiatry. A notable example of which is predictive processing, hailed by some as a candidate unifying theory of brain function (Clark, 2013, 2016; Friston, 2010; Hohwy, 2013).

An exhaustive account of predictive processing lies outside the scope of this introduction but, in a nutshell, predictive processing casts the brain as a *hypothesis-testing system* (Hohwy, 2013). That is, a system that updates its expectations—*predictive beliefs*, or *prior probabilities*—in response to new evidence received by the senses. Perception and cognition are then processes for inferring the causes of sensory information received from the external world. In the brain, Bayesian inference is approximated by a continuous process of prediction error minimisation between anticipated and recorded sensory input, across all levels of the cortical hierarchy. Prediction error minimisation can be achieved through refinement of predictions—*perceptual inference*—or by changing sensation through action to make it fit with existing expectations—*active inference*.

Attentional processes play a key role in determining the balance between perceptual and active inference. Each prediction error signal is afforded a certain precision weighting. Those with higher precision have greater modulatory access to prior probabilities encoded at higher levels. That is, they can drive associative learning at a higher rate. A salient environmental signal will attract more attentional resources (it will receive greater precision up-weighting) and thus have a greater capacity to modify predictive beliefs. Top-down attention is thus cast in Bayesian terms as the brain's *optimisation of expected precisions*. When functioning effectively, attentional processes filter sensory input so that the most reliable and relevant data has the greatest capacity to refine predictive beliefs. Predictive beliefs are then the brain's store of prior probabilities—its *expectations*.

The predictive processing framework has been employed in descriptions of numerous features of high-level cognition and perception including theory of mind (Koster-Hale & Saxe, 2013), cognitive control (Pezzulo, 2012), synesthesia (Seth, 2014), binocular rivalry (Hohwy, Roepstorff, & Friston, 2008), and the placebo effect (Büchel, Geuter, Sprenger, & Eippert, 2014). Whether it is appropriate to employ the framework so widely is itself a subject of critique (Hutto, 2018; Menary, 2015). It is surprising then that a comprehensive account of metacognition under the predictive processing framework is yet to surface. A challenge for this project is taking a framework of cognitive function that emphasises top-down processes and uniting it with prevailing, signal-detection theoretic models of metacognition that assume second-order information is constituted by first-order information (i.e., bottom-up signals carry content) (Barrett et al., 2013). One proposal speculates that expectations reshape the probability distributions of decision thresholds (Scott, Dienes, Barrett, Bor, & Seth, 2014; Sherman, Seth, Barrett, & Kanai, 2015). Metacognition might then be regarded as internal assessment of the precision of this process although determining what information is available at the personal level (e.g., subjective confidence) remains unclear.

Predictive processing offers persuasive descriptions of the symptoms that characterise psychosis and autism (Friston, Stephan, Montague, & Dolan, 2014; Horga, Schatz, Abi-Dargham, & Peterson, 2014; Palmer, Lawson, & Hohwy, 2017; Powers, Mathys, & Corlett, 2017; van Boxtel & Lu, 2013; Van de Cruys et al., 2014). It also offers accounts of less defined disorders of mind, a prime candidate being *functional motor disorders*[5] (FMDs). FMDs are characterised by the motor abnormalities seen in *organic movement disorders* but lacking their distinct psychiatric or physical basis. The challenge this presents for diagnosis has made it necessary to draw on cognitive neuroscience for novel interpretations. Predictive coding hypothesizes that individuals with a sufficient conjunction of predisposing factors (such as cognitive biases or traumatic events) will develop flaws in the brain's processing of sensory input which may elaborate into motor symptoms. In the case of FMD, it is argued that misallocation of attention and abnormal predictive beliefs lead to movements that are conducted without a sense of agency (Edwards, Adams, Brown, Pareés, & Friston, 2012; Newby, Alty, & Kempster, 2016).

The predictive processing view applies to broader functional neurological and sensory disorders (Edwards et al., 2012; Stenner & Haggard, 2016). A challenging upshot of this account is the suggestion that flaws in sensory processing that produce motor abnormalities might also extend to flaws in perceptual and cognitive processing. FMD has long been tied to attention such that functional movements demonstrate *distractibility* (i.e., attenuation or extinction when selective attention is directed elsewhere). However, evidence for other perceptual abnormalities is less clear.

[5] The terms *functional* and *organic* are used to this day but are categories that derive from historic accounts of psychosis (Dilsaver, 1992). Organic disorders are those where the cause is a structural lesion or physiologic dysfunction of the brain. Functional disorders are associated with altered behaviour or experience but the cause has not yet been identified. By DSM-5 terminology, most FMD patients are categorised as having Conversion Disorder—characterised by neurological symptoms that typically arise after a stressful experience but cannot be explained by medical evaluation (Hallett, 2016).

Some evidence of cognitive impairment in FMD is found such as 'jumping to conclusions' in probabilistic tasks (Pareés et al., 2012) and consistent errors of commission in go/no-go tasks (Voon et al., 2013). There is also evidence that processing of sensory inputs is altered in FMD. For instance, *sensory attenuation*, the ability to selectively down-weight the sensory consequences of one's own actions, is impaired in FMD (Pareés et al., 2014). FMD might also be explained by appeal to metacognition—viewed here as reduced awareness of how precision is assigned to incoming sensory input. Metacognitive impairment is suggested when FMD patients' actigraphic records are compared with their self-reported tremor. While patients with organic motor disorders tend to overestimate the frequency of their motor symptoms, this mismatch is much greater in FMD (Pareés et al., 2011).

In Chapter 4, we investigated perceptual processing in functional and organic motor disorders; to our knowledge, the most rigorous study of perception in FMD to date (Matthews, Nagao, et al., 2018). Specifically, we identified four domains associated with the predictive processing account of FMD: attention, expectations, sensation, and metacognition. We augmented a visual dual-task paradigm (Matthews, Schröder, et al., 2018; Sherman et al., 2015) to contrast these four domains in a single experimental design. Our behavioural task was used to test FMD patients and healthy controls but also patients with phenotypically-matched organic motor disorders. We found that core executive processes (e.g., attention and metacognition) function normally in FMD compared to healthy controls. However, FMD patients exhibited a significant impairment in their sensitivity to visual perceptual contrast; equivalent to the sensory impairment identified in organic patients. The organic group was further distinguished from FMD and controls by differences in their use of attention and expectations for perception. Metacognitive sensitivity and efficiency was broadly equivalent between the groups. We concluded that perception is a domain of interest in motor disorders and that understanding functional neurological disorders under the predictive processing framework can consolidate and refine existing pathophysiological theories of these conditions.

Linking text between chapter 1 and 2

Chapter 1 established an overarching account of how neurocognitive science has approached the empirical study of metacognition and its relation to consciousness and broader cognitive function. I introduced the challenges posed for a subjective measure of conscious and unconscious mental states, and how metacognition has been operationalised in response to these challenges. I also gave an account of two cognitive processes that are closely related to metacognition and have been intimately associated or even equated with consciousness: selective attention and working memory. Finally, I introduced the psychopathological study of metacognition and how the predictive processing framework is being used to outline a holistic account of cognitive function (including metacognitive processes) in functional motor disorders.

In the following chapters (2 and 3), we employed metacognitive measures to examine how attention and short-term memory relate to visual perceptual consciousness. Consciousness is notorious for eluding a functional definition. To address this we deliberately bracketed our investigation to *conscious access*—the capacity for conscious insight into perceptual experiences (Fazekas & Overgaard, 2018). The philosopher Ned Block makes a famous distinction between this highly functional cognitive process and *phenomenal consciousness* (Block, 1995, 2007b, 2011a). Phenomenal consciousness is equated with the overall character of conscious experience and is sometimes encapsulated using the pithy expression 'what it is like' made famous by Thomas Nagel (Nagel, 1974). The distinction between phenomenal and access consciousness was initially raised to appeal to our subjective intuition that consciousness is phenomenally 'rich' (i.e., perceptual experiences 'overflow' cognitive access (Block, 2007a, 2011a, 2014)) but it has found wide application in consciousness science despite protestation from noteworthy detractors (Richard Brown, 2014; Cohen & Dennett, 2011; Cohen, Dennett, & Kanwisher, 2016; Naccache, 2018).

One of the most pressing questions posed by the phenomenal and access distinction is whether the neural basis of perceptual consciousness can be dissociated from access mechanisms in the brain that make processes such as metacognition possible (Boly et al., 2017; Fazekas & Overgaard, 2018; Odegaard, Knight, & Lau, 2017; Phillips, 2018; Tononi et al., 2016). A critical factor is the extent to which conscious access might operate independently from attention (Pitts, Lutsyshyna, & Hillyard, 2018). On the basis of this and the broader debate regarding the necessity of attention for consciousness we employed metacognitive performance to examine whether conscious access was possible with little or possibly no attentional amplification (Chapter 2).

To achieve this we made critical extensions to the dual-task paradigm—a robust psychophysical tool for examining how the focus of selective attention influences perception and behavioural reports (Matthews, Schröder, et al., 2018). Our changes reduced training requirements by several orders of magnitude, addressed the possibility of working memory decay, and accounted for the possibility that stimulus salience might attract focal attention. We revealed that face gender can be differentiated with high metacognitive sensitivity despite perception occuring in the near absence of attention. In contrast, participants were unable to make a simple orientation discrimination under the same conditions, even if the simple stimulus and the face are blended together and presented at the same time and location. We concluded that conscious access can be achieved with little or possibly no attentional amplification. This implies that for some stimulus features the cognitive mechanisms that support metacognition are unlikely to require that those features are perceived and encoded into memory using selective attention.

Chapter 2

Conscious access in the near absence of attention: critical extensions on the dual-task paradigm

Julian Matthews[1,2], Pia Schroder[2], Lisandro Kaunitz[2],
Jeroen van Boxtel[2,3], Naotsugu Tsuchiya[2,3]

1. Cognition and Philosophy Lab, Monash University, Clayton, Australia
2. School of Psychological Sciences, Monash University, Clayton, Australia
3. Monash Institute of Cognitive and Clinical Neurosciences, Clayton, Australia

Downloaded from http://rstb.royalsocietypublishing.org/ on October 29, 2018

PHILOSOPHICAL TRANSACTIONS B

rstb.royalsocietypublishing.org

Conscious access in the near absence of attention: critical extensions on the dual-task paradigm

Julian Matthews[1,2], Pia Schröder[2], Lisandro Kaunitz[2], Jeroen J. A. van Boxtel[2,3] and Naotsugu Tsuchiya[2,3]

[1]Cognition and Philosophy Lab, Faculty of Arts, [2]School of Psychological Sciences, and [3]Monash Institute of Cognitive and Clinical Neurosciences, Monash University, Clayton, Victoria 3800, Australia

JM, 0000-0002-1895-3403; JJAvB, 0000-0003-2643-0474; NT, 0000-0003-4216-8701

Research

Check for updates

Cite this article: Matthews J, Schröder P, Kaunitz L, van Boxtel JJA, Tsuchiya N. 2018 Conscious access in the near absence of attention: critical extensions on the dual-task paradigm. *Phil. Trans. R. Soc. B* **373**: 20170352.
http://dx.doi.org/10.1098/rstb.2017.0352

Accepted: 21 May 2018

One contribution of 17 to a theme issue 'Perceptual consciousness and cognitive access'.

Subject Areas:
behaviour, cognition

Keywords:
attention, consciousness, dual task, metacognition, faces, confidence ratings

Authors for correspondence:
Julian Matthews
e-mail: julian.r.matthews@gmail.com
Naotsugu Tsuchiya
e-mail: naotsugu.tsuchiya@monash.edu

Electronic supplementary material is available online at https://dx.doi.org/10.6084/m9.figshare.c.4137836.

Whether conscious perception requires attention remains a topic of intense debate. While certain complex stimuli such as faces and animals can be discriminated outside the focus of spatial attention, many simpler stimuli cannot. Because such evidence was obtained in *dual-task paradigms* involving no measure of subjective insight, it remains unclear whether accurate discrimination of unattended complex stimuli is the product of automatic, unconscious processing, as in blindsight, or is accessible to consciousness. Furthermore, these paradigms typically require extensive training over many hours, bringing into question whether this phenomenon can be achieved in naive subjects. We developed a novel dual-task paradigm incorporating confidence ratings to calculate metacognition and adaptive staircase procedures to reduce training. With minimal training, subjects were able to discriminate face-gender in the near absence of top–down attentional amplification, while also displaying above-chance metacognitive accuracy. By contrast, the discrimination of simple coloured discs was significantly impaired and metacognitive accuracy dropped to chance-level, even in a partial-report condition. In a final experiment, we used blended face/disc stimuli and confirmed that face-gender but not colour orientation can be discriminated in the dual task. Our results show direct evidence for metacognitive conscious access in the near absence of attention for complex, but not simple, stimuli.

This article is part of the theme issue 'Perceptual consciousness and cognitive access'.

1. Introduction

The perplexing co-dependency between attention and consciousness has been the subject of philosophical and scientific debate for well over a century [1]. One feature of this debate that has risen to prominence in recent years concerns the necessity of top–down attentional amplification for conscious perception (the *necessity claim*) [2–4]. While several noteworthy theories of consciousness remain divided on this claim, scientific enquiry has made progress in its attempts to independently manipulate top–down attention and visual consciousness using a variety of tasks and illusions[1] [5,17,18].

The relationship between visual consciousness and top–down attentional amplification has been primarily investigated with the *dual-task paradigm* [19–21]. In this paradigm, a subject's attention is spatially drawn to a very demanding central task at the same time as a secondary stimulus is briefly presented in the periphery. If performance on the peripheral task in this diverted attention condition is identical to that without the central task, top–down attention is claimed unnecessary for the peripheral task. This paradigm has been employed to examine the requirement of top–down attention for discriminating many categories of stimuli. For example, a simple, low-level visual distinction such as discriminating the orientation of a coloured shape can be 80% correct

© 2018 The Author(s) Published by the Royal Society. All rights reserved.

THE ROYAL SOCIETY PUBLISHING

Downloaded from http://rstb.royalsocietypublishing.org/ on October 29, 2018

under full attention and yet fall to chance (50% correct) when attention is diverted, demonstrating a necessity of spatial attention [22,23].

Remarkably, however, performance for certain visual discriminations that seem intuitively more complex such as categorizing face-gender or classifying whether natural scenes include animals does not differ between the full and diverted attention conditions [23–25]. These results suggest that for certain complex stimuli, attention may not be necessary for conscious perception, a result taken as empirical evidence for a dissociation between top–down attention and consciousness [5,26].

However, dual-task studies investigating such a dissociation are confronted with various criticisms that question this conclusion. Among these, we address the most critical four in our experiments. First, dual-task experiments typically employ extensive training of thousands of trials. In the case of Reddy *et al*. [23], training took between 6 and 12 h before subjects achieved proficiency. This involved as many as 5760 training trials, more than the experiment itself. Training is known to influence the attentional requirements of the dual task [27–30] and other paradigms including the attentional blink [31,32]. Specifically, the extent to which tasks are influenced by inattention differs between highly trained and naive subjects, which poses critical limitations on the conclusions that can be drawn when such tasks are employed to examine the relationship between attention and consciousness [28–30].

The second challenge concerns the nature of the control experiments employed in the studies listed above [23–25]. In these experiments and under the dual-task condition, certain categories of stimuli such as rotated letters or bisected discs are shown to be impossible to discriminate at above-chance levels [17]. A typical account of this scenario, inattentional blindness [33], suggests that without top–down attentional amplification, these stimuli fail to reach consciousness. An alternative account is that the inability to discriminate such items is not perceptual but results from response interference. That is, without top–down attention, while subjects are responding to the central task, they forget the stimulus because peripheral representations decay very quickly. In other words, the standard dual-task paradigm cannot exclude a possibility of conscious visibility with rapid forgetting of those stimuli that result in chance-level performance under the diverted attention condition [34–37].

The third criticism concerns the difference in attentional draw between different categories of stimuli. Of particular concern for the studies by Reddy *et al*. [23,25], faces are known to strongly attract (bottom–up) attention, possibly due to ecological importance and the presence of dedicated neural resources [38–40]. They capture attention in individuals as young as six weeks old [41], and they impair processing of visual objects that are presented elsewhere at the same time [42,43]. Perhaps this attraction to faces accounts for why this category of stimulus can be discriminated in the dual task while less salient items, such as bisected discs, are missed?

Finally, it remains unclear whether successful dual-task performance for categories such as face-gender is in fact conducted using information accessible to consciousness. Recent literature has accumulated considerable evidence of above-chance behavioural performance that is not accompanied with consciousness, such as blindsight [44–46]. Non-conscious stimulus processing has been observed not only for simple discriminations, but also for complex, high-level stimuli such as upright faces [47–50] (for reviews, see [51,52]). Thus, achieving highly accurate dual-task performance in tasks with complex stimuli such as faces does not guarantee that these discriminations are performed consciously [4]. In fact, in combination with excessive training, there is reason to suspect that at least some aspects of successful discrimination under dual-task conditions may be a product of unconscious processing [49].

In this paper, we addressed these four criticisms by substantially improving the dual-task paradigm in four ways. First, we employed an adaptive staircase procedure [53] to reduce the amount of training typical of dual-task studies [23,24]. Second, our partial-report condition made both the central and peripheral targets task-relevant but required subjects make only one response per trial, removing the concern that peripheral target representations decay beyond reportability while subjects respond to the central task. Third, we accounted for faces attracting bottom–up spatial attention by blending face and disc stimuli through transparency (alpha (α) blending) and examining whether the colour orientation of the disc is reportable when the stimuli are co-located. Fourth, we directly assessed metacognitive insight for unattended peripheral stimuli, a signifier of conscious access, by quantifying trial-by-trial confidence ratings and perceptual awareness judgements as a function of task accuracy [54–57]. With these improvements over previous approaches, we critically assessed the necessity of top–down attention for conscious access of stimuli with high and low complexity.

2. General methods

(a) Subjects

Twenty-four subjects participated in our study, eight for each of our three experiments. Participant numbers were determined from those studies that employed the dual-task paradigm: typically between four and eight [23–25,58]. A power analysis (power of 0.8, assumed correlation of 0.5, one-tailed *t*-test) based on observations from the disc task in Reddy *et al*. [23] revealed that a sample size of 3 would be sufficient to find a difference comparable to or larger than their study. Subjects were recruited from the student and staff bodies of Monash University and were paid for their involvement in the study. All had normal or corrected-to-normal vision and provided informed written consent in accordance with the guidelines of the Monash University Human Research Ethics Committee and the recommendations of the Declaration of Helsinki.

(b) Apparatus

All experiments were performed on a MacBook Pro laptop connected to a 22-inch SMI monitor approximately 60 cm from the subject. Refresh rate of the monitor was fixed at 60 Hz with 1680 × 1050 pixels screen resolution. The experiments were programmed and conducted using the Psychophysics toolbox extension for MATLAB.

(c) Stimuli

(i) Central letter discrimination

The central stimulus for all experiments was a cluster of five uppercase characters presented in white, Helvetica script at

Downloaded from http://rstb.royalsocietypublishing.org/ on October 29, 2018

35 pixels text height (approx. 1° visual angle on our set-up), each rotated at a randomly selected angle. The coordinates of these five letters were fixed, one presented centrally at fixation and the remaining four located directly above, below and on both sides of this point, approximately 3° from fixation. These five letters were either all the same (all 'T' or all 'L') or contained a single differing character (i.e. one 'T' among four 'L's and vice versa). An uppercase letter 'F' individually masked each character for the remainder of the trial following a short, temporal delay that was adjusted to achieve 70% discrimination accuracy across central, single-task blocks (see quick estimate of threshold (QUEST) staircase procedure, §2e). This letter discrimination task has proven effective in maintaining the focus of attention at the fixational point, leaving little spatial attention available at the periphery [22–25].

(ii) Peripheral discrimination

The peripheral stimulus categories consisted of faces (experiment 1), discs (experiment 2) or blended face/discs (experiment 3) (figure 1*a*). On each trial, one such stimulus subtending 2.5° of visual angle was displayed at the periphery. This peripheral stimulus was randomly positioned at one of four locations centred on the corners of an imaginary rectangle 8° × 10° of visual angle in dimensions. After a short temporal delay, a mask replaced the peripheral stimulus (figure 1*a*). This delay was adjusted such that single-task discrimination was held at 70% accuracy (see QUEST staircase procedure, §2e, and figure 1*b*).

(d) Procedure

Dual-task experiments contrast performance in the single-task condition, where a central or a peripheral task is conducted in isolation, against that in the dual-task condition, where both the central and peripheral stimuli are task-relevant. The physical appearance of the experiment should be identical across these conditions, with the only difference being the task relevance of the stimuli (figure 1*c*). Written instructions at the beginning of each block informed subjects which task was required.

(i) Single-task conditions

In the central and peripheral single-task conditions, subjects were presented with one response screen and made a single eight-alternative-forced-choice (8AFC) report per trial (figure 1*b*: response panels enclosed in green and yellow boxes). Once subjects had signalled their readiness using a mouse click, each trial began with the presentation of a fixation cross for 200, 300 or 400 ms with an equal probability for each. This was followed by the central stimulus and, on the next frame, the peripheral stimulus. Following a short temporal delay, central (or peripheral) stimulus onset asynchrony (SOA), the central (or peripheral) stimulus was masked (figure 1*b*). SOAs were controlled such that discrimination accuracy for each stimulus in the single-task condition was held at 70% (see QUEST description below, §2e).

(ii) Dual-task condition with whole report

In the dual-task condition, both central and peripheral stimuli were task-relevant. The presentation of visual items proceeded as above; however, subjects were required to make two responses per trial, first on the central then on the peripheral stimulus (figure 1*b* in red). We termed this dual-task condition 'whole-report' to contrast it with our 'partial-report' procedure in experiment 2.

In order to contrast performance in the dual-task against the single-task conditions, SOAs were not updated during the dual-task condition. Instead, SOAs were fixed at the threshold duration defined by the preceding single-task block for each task type (see QUEST description, §2e). As is typical for this paradigm, we instructed subjects to prioritize performance for the central stimulus in the dual-task condition. We did not give subjects any feedback regarding their performance and did not inform them of the staircase procedure.

(iii) 8AFC response screen

Mask presentation was followed with the display of a response screen comprising eight evenly split segments (figure 1*b*). With a single mouse click, this screen allowed subjects to register their two-alternative-forced-choice (2AFC) discrimination response as well as a four-level subjective rating. Prior to the experiment, and during practice, subjects were verbally instructed to express their confidence from a complete guess (rating 1) to certainty (rating 4) in experiments 1 and 2 or perceptual awareness from complete invisibility (rating 1) to complete visibility (rating 4) in experiment 3. Verbal descriptors for ratings 2 or 3 were not made explicit; however, the experimenter encouraged subjects to fix these criteria across the sessions as best as they could, and use all four levels. The labels 'sure' and 'not sure' in experiments 1 and 2 (or 'easy to see' and 'hard to see' in experiment 3) were displayed at the top and bottom of the screen to remind subjects of the scale of subjective rating.

At the centre of the display, we presented the discrimination options. For the central stimuli, subjects indicated whether the target letters they had seen were all the same (S) or one was different (D) by clicking on one of the confidence segments on the side with the labels either 'S' or 'D' (figure 1*b*). The 'S' option was always on the left. For the peripheral face-gender discrimination (experiments 1 and 3), subjects indicated whether the target was either male (M) or female (F). The 'M' option was always on the left. For the peripheral coloured disc stimuli (experiments 2 and 3), the discrimination options were substituted with images of red–green and green–red discs (the red–green option on the left) and subjects selected a segment to indicate their percept and subjective rating.

(iv) Perceptual awareness scales

As subjective ratings for experiment 3, we employed perceptual awareness scales (PAS) [59,60], which are a more direct measure of conscious perception. For this purpose, we changed the display labels from 'sure' and 'not sure' into 'easy to see' and 'hard to see' and instructed subjects to rate '1' when the stimulus was very hard to see and '4' when the stimulus was very easy to see. We did not explicitly describe the ratings of '2' and '3', but we encouraged subjects to use all four levels when appropriate.

(v) Analysis

In addition to objective and subjective signal detection metrics to measure performance, we employed linear mixed effects (LME) modelling to examine metacognitive accuracy [57]. We also examined trade-off in performance between

Downloaded from http://rstb.royalsocietypublishing.org/ on October 29, 2018

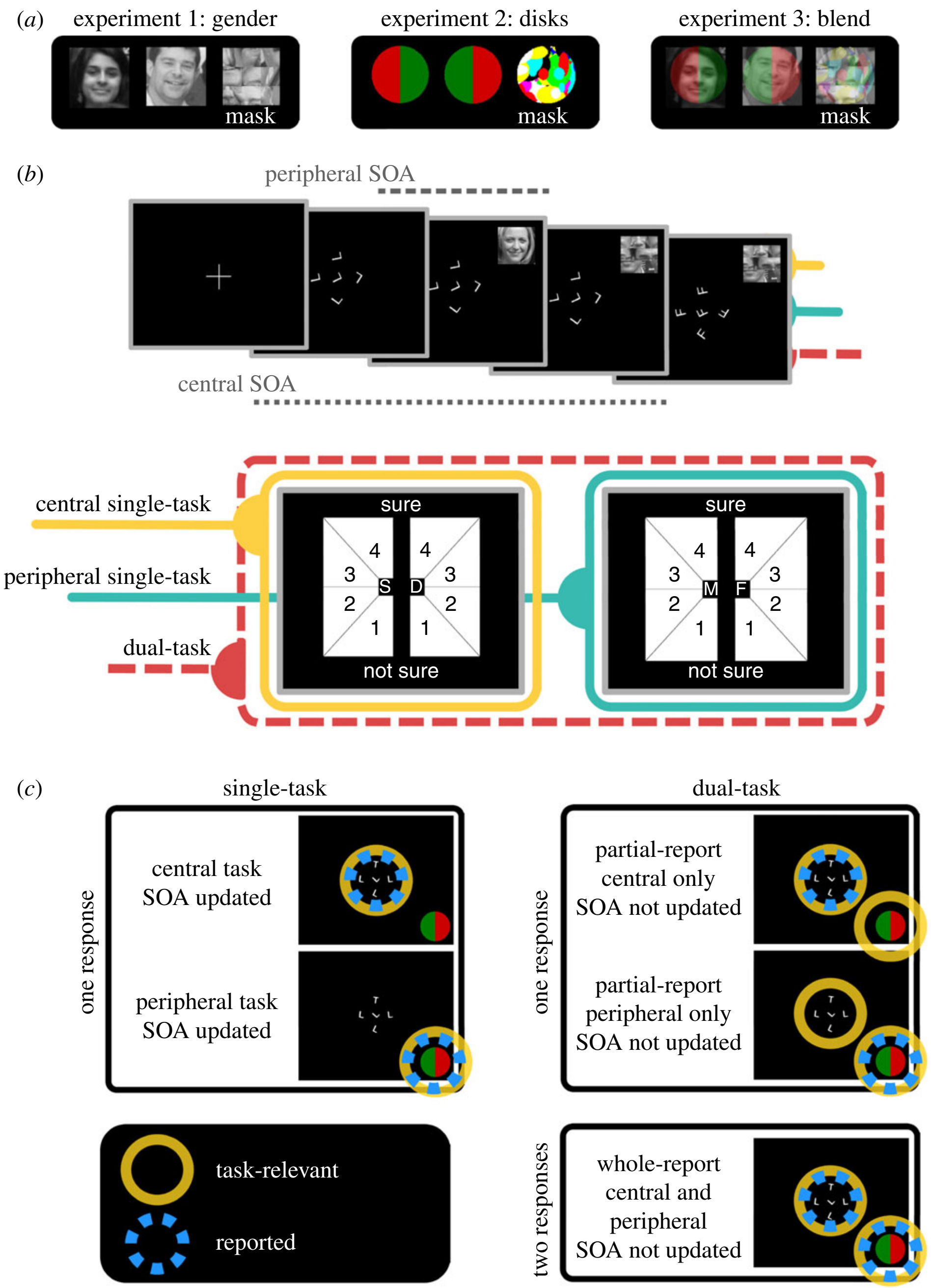

Figure 1. Stimuli and task structure. (*a*) Stimuli used for experiment 1: gender discrimination, experiment 2: bisected disc discrimination, and experiment 3: blended face/disc discrimination. (*b*) Trial sequence. After a variable period of fixation (200–400 ms), five randomly rotated letters (Ls and/or Ts) are presented in the centre. After one frame (16.7 ms), a peripheral target (here, a female face) appeared in the periphery. Following a short delay, or stimulus onset asynchrony (SOA), both central and peripheral targets were masked. In blocks involving the central task (the central single- and dual-task conditions), subjects reported on the letter stimuli at the centre of the display. Subjects made an eight-alternative-forced-choice (8AFC) response with a single mouse click, signalling their discrimination ('S' for same (all Ts or all Ls) and 'D' for different (T among Ls or L among Ts)) and four-level subjective rating (confidence in experiments 1 and 2, perceptual awareness in experiment 3). In blocks involving the peripheral task (the peripheral single- and dual-task conditions), subjects reported on the peripheral stimulus. Again, a single mouse click was used for an 8AFC decision on face-gender (experiments 1 and 3: M, male; F, female) or disc colour-orientation (experiments 2 and 3: a picture of a red–green disc and a green–red disc). (*c*) Summary of partial- and whole-report conditions in experiment 2. During dual-task blocks, task relevance remained consistent but depending on the report condition, either one or two responses were required.

the central and peripheral tasks by defining a measure we call *actual trigonometric altitude* (TA^{actual}). See the electronic supplementary material for details.

(e) QUEST staircase procedure and reduced training

Previous dual-task studies employed large amounts of training to stabilize subjects' performance at the threshold SOAs [19,23–25]. For example, in Li *et al*. [24], training is described as usually taking 'more than 10 h (12 000 trials of all tasks combined)'. Reddy *et al*. [23] trained subjects until they achieved 80% accuracy for the central letter discrimination task with an SOA below 250 ms for an entire 1 h session. This procedure took 'between 6 and 12 h per subject'. A separate experiment by Reddy *et al*. [25] trained subjects for only 2 h, but all had participated in Li *et al*. [24] so had approximately 10 h of prior exposure to the central letter discrimination task.

To generalize conclusions of the dual-task paradigm into an untrained population, we reduced training to a minimal

Downloaded from http://rstb.royalsocietypublishing.org/ on October 29, 2018

level by rapidly and robustly setting SOAs that yielded threshold performance levels equated between subjects. SOAs were adjusted on a trial-by-trial basis during single-task blocks using the (QUEST) adaptive staircase procedure [53]. The initial SOA for central targets was 500 and 250 ms for peripheral stimuli. We set the β parameter for QUEST to be 2 and the standard deviations to be 70% of the respective initial SOA during training. Once training was complete, the standard deviation parameter was reduced to 50 ms for both central and peripheral stimuli. On each trial of the respective single-task block, we updated either the central or peripheral SOA such that discrimination performance was fixed at 70% correct for that condition. To contrast performance in the single-task conditions against the dual-task, we did not update SOAs in the dual-task condition (figure 1*c*). Central SOA in the dual task was drawn from the preceding single-task central block and vice versa for the peripheral SOA.

Training in our experiments only took approximately 20 min, with subjects completing two single-task central blocks followed by two single-task peripheral blocks and 20 practice trials under the dual-task condition. This procedure reduced 12 h of training by more than 97% relative to Reddy [23].

3. Experiment 1: gender discrimination

Our first experiment examined whether gender discrimination in the near absence of attention was associated with conscious access. In addition, we employed a staircase procedure to greatly reduce the amount of training.

(a) Methods

Eight subjects (3 M, 5 F, ages 18–34) took part in experiment 1. The procedure for the experiment was identical to our general method but employed greyscale photographs of human faces as the peripheral stimulus.

A set of 65 male and 65 female faces were selected from natural crowd scenes, details of which are described elsewhere [57,61]. All were facing forward with major features (i.e. eyes, mouth, nose) clearly visible. To generate masking textures, we used 20 male and 20 female faces (out of 130 used for the experiment), each cut into 3×3 squares of equal size and rearranged randomly without rotation or flipping (figure 1*a*). Faces and masks were randomly selected on each trial.

(i) Data collection

Data collection took place over three sessions on three consecutive days. The first session consisted of training (i.e. 2 blocks x 48 trials of the single-central-letter task, 2 blocks x 48 trials of the single-peripheral-gender task and 20 trials of the dual-task), followed by two runs of the experiment. Each run comprised one block each of the single-central, single-peripheral, and the dual-task whole-report condition. The order of these blocks in each run was randomized. SOAs were updated during each single-task block and these updated SOAs were used for the dual-task block that immediately followed. Each block consisted of 48 trials. The second and third sessions skipped training and consisted of three experimental runs. This resulted in eight experimental runs overall, and thus, eight blocks for each condition (single-central, single-peripheral, dual-task) per subject.

(b) Results

(i) Face-gender discrimination in near absence of attention

Even with our minimal training procedure, we largely replicated previous findings [23,25] (figure 2*a*; electronic supplementary material, table S1). Objective performance (type 1 area under receiver operating characteristic curve (AUC)) for the peripheral face-gender and central letter tasks were much higher than chance (0.75 and 0.76, respectively) when they were performed simultaneously in the dual-task condition. Dual-task performance was slightly worse than each respective single-task condition (0.77 and 0.80). These differences were statistically significant ($p = 0.039$, $\eta_p^2 = 0.037$ for peripheral and $p = 0.025$, $\eta_p^2 = 0.044$ for central) according to a two-way within-subject ANOVA (attention condition (single- versus dual-task) and block as factors). Neither the interaction nor main effect of block was significant ($p > 0.25$; see electronic supplementary material, table S1). The TA^{actual} for objective performance was 0.77, indicating almost no trade-off.

(ii) Confidence remains stable in the near absence of attention

As a subjective measure, we asked subjects to rate trial-by-trial confidence on their discrimination in experiment 1 (figures 2*b* and 3*a*; electronic supplementary material, table S2). We conducted LME analysis to examine the relationship between attention and correctness on trial-by-trial confidence ratings (see the electronic supplementary material). (We also performed the analyses including block as a factor, but block was never significant in any analysis across all experiments and measures (all $p > 0.25$), thus we will not report it further.) For the peripheral face-gender task, the main effect of attention did not reach significance ($p > 0.05$ for either correct or incorrect trials), but that for correctness was highly significant (each $p < 0.001$ for the single and dual task). The interaction between attention and correctness was also significant, though the effect size was small ($\chi^2(1) = 4.26$, $p = 0.039$).

(iii) Metacognitive accuracy in the near absence of attention

That confidence ratings broadly correspond with subjects' accuracy implies that subjects had metacognitive insight into their decisions. We examined this relationship quantitatively by computing metacognitive accuracy (measured as type 2 AUC; see the electronic supplementary material) (figure 2*c*; electronic supplementary material, table S3). While metacognitive accuracies for both peripheral faces and central letters were well above chance during the dual task (0.59 and 0.60), they were significantly lower than those in the single task (0.62 and 0.64, two-way within-subject ANOVA: main effect of attention $p < 0.05$, $\eta_p^2 = 0.036$). TA^{actual} was 0.50, reflecting metacognitive accuracy roughly half that of single-task equivalence.

Taken together, these results confirm that despite their attention being consumed by the central letter task, subjects exhibited near intact objective accuracy and confidence ratings on their peripheral face discriminations in the dual task. While some performance decrement was seen, the magnitude was small by Cohen's conventions. Given that subjects were minimally trained in our protocol (20 min), many causes may explain this slight deficit (e.g. motor coordination errors, confusion due to task switching across blocks), and rather it is remarkable that we found highly similar results to the original findings, which required training subjects for

Downloaded from http://rstb.royalsocietypublishing.org/ on October 29, 2018

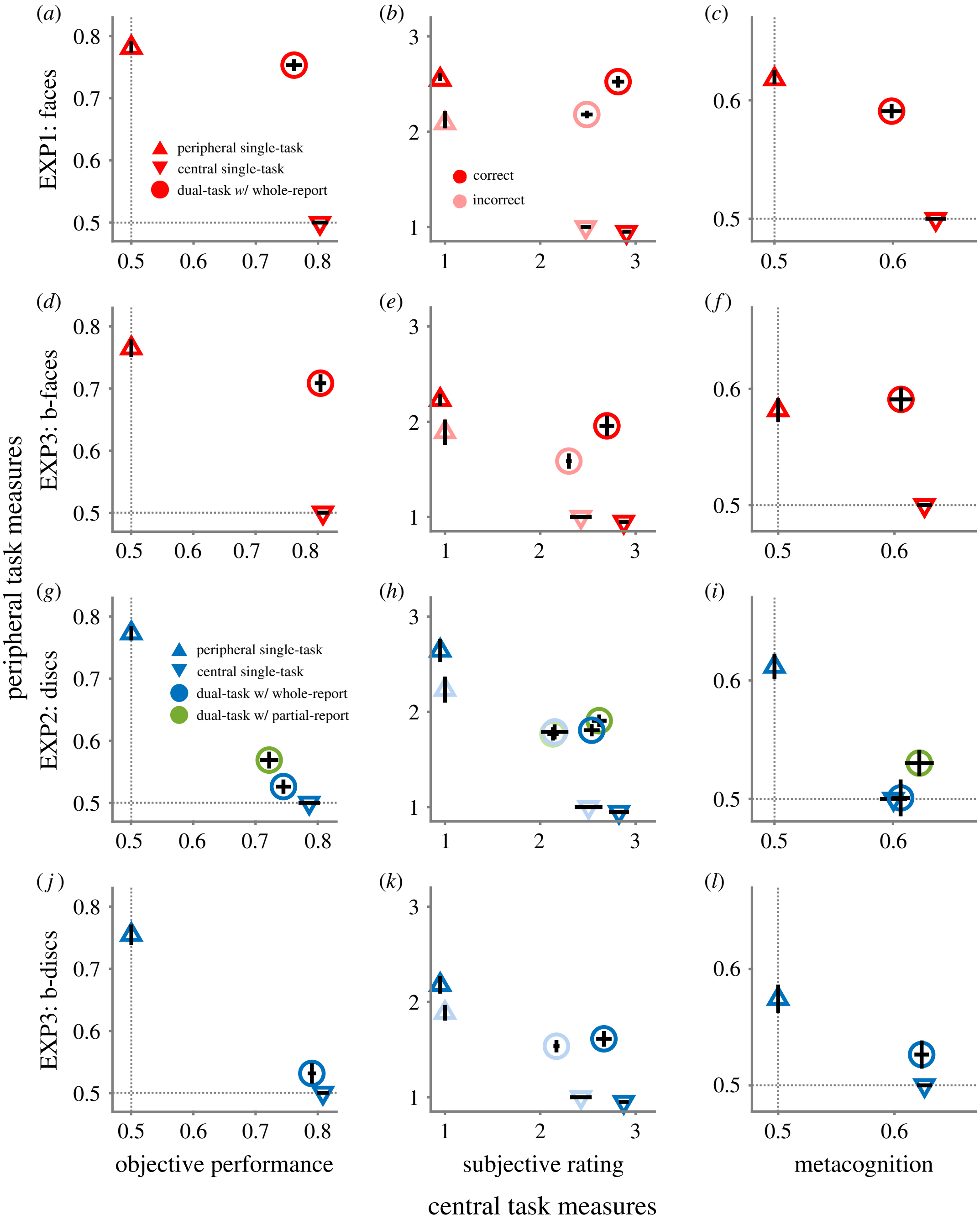

Figure 2. Conscious face-gender, but not disc-colour, discrimination is possible in the near absence of attention despite minimal training. Presented in red, the first (*a*–*c*) and second (*d*–*f*) row reflect discrimination of face-gender in experiments 1 and 3. Presented in blue, rows three (*g*–*i*) and four (*j*–*l*) reflect disc-colour discrimination in experiments 2 and 3. Experiment 3 used a blended face and disc. Each column represents a performance measure; column 1 (*a,d,g,j*) is objective performance (type 1 AUC) and column 3 (*c,f,i,l*) is metacognitive accuracy (type 2 AUC). Column 2 reflects subjective ratings (confidence judgements in experiment 1 (*b*) and 2 (*h*), PAS in experiment 3 (*e* and *k*)). Subjective ratings are plotted separately for correct and incorrect trials (lighter shades reflect mean ratings for incorrect judgements). In all panels, measures along the *x*-axis refer to performance on the central letter task. Measures along the *y*-axis refer to performance on the peripheral task. Single-task performance is plotted as upright (for peripheral tasks) and inverted (for central task) triangles near their respective axes. Dual-task performance is plotted as circles. Additional green markers are included for experiment 2 to reflect results from the partial-report procedure. Error bars signify within-subjects standard error of the mean [62,63].

6–12 h. Further, no statistical analyses on performance measures showed significant effects of training (i.e. the main effect of block or interaction between block and attention), implying that our block-by-block SOA adjustment was successful in achieving stable performance. Our QUEST protocol was robust and stable enough that we saw no effects of block across any measures in experiments 2 and 3 as well (thus they are not reported further).

While subjects retain the capacity to discriminate face-gender in the near absence of attention, their metacognitive insight into these discriminations was affected in the dual-task condition. This might be due to the slight dual-task trade-off in objective accuracy and confidence ratings. We will return to this issue in our general discussion (see §6).

4. Experiment 2: disc discrimination

In experiment 2, we employed the dual-task design for coloured discs, a stimulus category regarded as

Downloaded from http://rstb.royalsocietypublishing.org/ on October 29, 2018

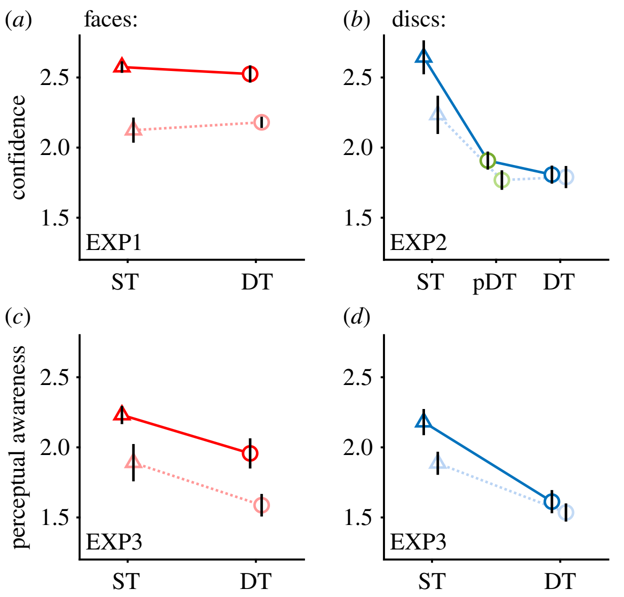

Figure 3. Subjective ratings for peripheral stimuli as a function of attention (ST, single-task; DT, dual-task whole-report; pDT, dual-task partial-report) and correctness. Top row represents confidence ratings for experiments 1 (*a*) and 2 (*b*), respectively. Bottom row represents perceptual awareness scales (PAS) in experiment 3 (*c* for faces, *d* for discs). Lighter shading indicates mean subjective rating for incorrect responses and additional green markers in (*b*) for partial-report. Error bars reflect within-subjects standard error of the mean [62,63].

indiscriminable when selective attention is diverted [22,23,64]. To test if response interference can explain this phenomenon, we used a partial-report procedure (figure 1*c*, see §4a).

(a) Methods

Eight new subjects (four females, ages 19–38) participated in experiment 2. Methods for this experiment were identical to our general methods apart from the inclusion of the partial-report condition and the use of coloured discs as the peripheral stimulus (figure 1*a*). Discs were masked by one of 10 circular, multi-coloured Mondrian patches precomputed before the experiment (figure 1*a*).

(i) Dual-task condition with partial report

In dual-task partial-report blocks, central and peripheral stimuli presentation proceeded as usual, but subjects were required to respond to just the central or the peripheral stimulus on a given trial. Participants did not know in advance to which stimulus they should respond, thus both tasks remained relevant while the report demands were equivalent to the single-task conditions (figure 1*c*).

(ii) Data collection

As in experiment 1, training and testing were conducted over the course of three sessions on three consecutive days. In session 1, training involved two blocks of 30 trials for the single-central letter task and single-peripheral disc task (120 trials in total) and 20 trials of the dual-task under whole-report conditions. No separate training was given under partial-report conditions apart from verbal and written instructions at the beginning of each block.

After training, two runs of the main experiment followed in the first session. Each run contained four blocks of 30 trials in length: one block of single-central-letter task (updated central SOA), one block of single-peripheral-disc task (updated peripheral SOA), one block of whole-report dual-task (fixed SOAs) and one block of partial-report dual-task (fixed SOAs) (figure 1*c*). The order of these four experimental blocks was randomized within each run. Sessions 2 and 3 consisted of three runs, resulting in a total of eight blocks of each condition per subject.

(b) Results

(i) Discs cannot be discriminated in dual-task conditions even with partial-report

In experiment 2, subjects discriminated the orientation of coloured discs in the single task, as well as the dual task under both whole- and partial-report conditions (figure 2*g*; electronic supplementary material, table S4). Replicating previous studies [22,23], objective accuracy (type 1 AUC) for the peripheral disc task was near chance in the dual task in both traditional whole- and our novel partial-report conditions (0.53 and 0.57, respectively) when the central letter task was prioritized (0.74 and 0.72). By contrast, both disc and letter tasks could be performed well above chance in the single-task conditions (0.77 and 0.79, respectively). Critically, we found complete trade-off ($TA^{actual} = -0.04$ and 0.02) for both the whole- and partial-report conditions (electronic supplementary material, table S4). A follow-up *t*-test confirmed no difference between these conditions (paired-sample *t*-test: $t_7 = -0.608$, $p > 0.25$, 95% CI [−0.290, 0.171]).

(ii) Confidence in discriminations for both letter and disc tasks is reduced when attention is diverted

As in experiment 1, we asked subjects to rate confidence in their discrimination judgements (figures 2*h* and 3*b*; electronic supplementary material, table S5). To examine the relationship between attention and correctness on confidence ratings in experiment 2, we conducted LME analysis (see the electronic supplementary material). For peripheral disc discriminations, the full model with interaction term differed substantially from the reduced model ($\chi^2(2) = 68.9$, $p < 0.001$). Subsetting data into correct and incorrect judgements revealed significant main effects of attention ($\chi^2(2) = 769.9$, $p < 0.001$, and $\chi^2(2) = 146.1$, $p < 0.001$, respectively). Subsetting by attention condition, the main effect of correctness reached significance for the single-task ($\chi^2(1) = 111.4$, $p < 0.001$) and dual-task partial-report ($\chi^2(1) = 13.6$, $p < 0.001$) conditions but not dual-task whole-report ($\chi^2(1) = 1.3$, $p > 0.25$). This confirmed that the relationship between correctness and confidence ratings was moderated by attention. Higher confidence corresponded with correct judgements when subjects fully attended to the disc task, but this relationship was largely extinguished when attention was diverted (figure 3*b*).

(iii) Metacognitive accuracy for disc discriminations reduces to chance in the near absence of attention

Metacognitive accuracy for the peripheral disc task is summarized in figure 2*I* and electronic supplementary material, table S6. It fell near chance-level under the dual task, in both the partial (0.53) and whole-report (0.50) conditions, though it was much above chance in the single-task condition (0.61). As we expected from our instruction to prioritize the central task, we found metacognitive accuracy of the central

Downloaded from http://rstb.royalsocietypublishing.org/ on October 29, 2018

task was similar across conditions (approx. 0.61). In terms of trade-off analysis, we observed complete trade-off under the whole-report condition, but not in the partial-report condition (see electronic supplementary material, table S6). We will return to this and related problems regarding trade-off analysis of metacognition in our general discussion (see §6).

5. Experiment 3: blended face/disc discrimination

In our final experiment, we addressed the potential concern that gender discrimination of a face is made possible by its inherent saliency, that is, attraction of bottom–up spatial attention [38,39]. To address this issue, we developed a novel stimulus category by α-blending a face and a disc. If the saliency of faces explains our results, two aspects of the blended object, face-gender and colour orientation, should be equally discriminable under the dual task.

(a) Methods

Eight subjects participated in experiment 3 (three females, ages 21–32). Methods were identical to our general methods except that an α-blended face/disc image was used as the peripheral stimulus (see α parameter staircasing in the electronic supplementary material). The disc aspect of this stimulus was identical (but lowered in contrast through transparency) to that described in experiment 2. In order to create the face aspect, a novel set of 518 faces (half of them female) were selected from the natural crowd scenes cited above [61]. The addition of these extra faces ensured that subjects could not simply learn the face stimuli. In addition, we did not repeat a given face until every face had been presented (resulting in a maximum of four presentations of a given face across the entire experiment). As the masking stimuli, we generated approximately 3500 scrambled face textures and Mondrians, α-blended using the same technique as the blended face/disc stimulus.

(i) Data collection during the main testing runs

Training and testing for experiment 3 were conducted over three sessions on three separate days. In session 1, after the training described above, we tested subjects in two runs of the main experiment. Each run included a 30 trial block of each of our five conditions: single-central-letter, single-peripheral-disc, single-peripheral-face, dual-letter-disc and dual-letter-face tasks. We adjusted SOAs for the central letter task with QUEST in every single-central-letter block. In alternate runs of the single-peripheral blocks, we adjusted the α level (even runs) or SOA (odd runs) for the peripheral stimuli. During the dual-task blocks, we used α levels and SOAs that were updated in the preceding single-task blocks. Sessions 2 and 3 involved three runs of the main experiment each. Thus, excluding training, subjects completed eight blocks for each of the five task conditions.

(b) Results and discussion

(i) Face-aspect, but not colour-aspect, of the blended stimulus can be discriminated in the near absence of attention

When we merged a face and a disc through α-blending in experiment 3, subjects continued to possess discriminability for the face-, but not the disc-, aspect of the blended stimulus (figure 2*d,j*; electronic supplementary material, table S7). When subjects paid close attention to the central letter task, maintaining objective performance (type 1 AUC) at 0.81, objective discrimination performance for the face-gender aspect of the blended stimulus remained above chance (0.71) while disc-colour discrimination fell to chance (0.53). Both stimulus aspects were performed well above chance (approx. 0.75) in the respective single-task conditions. Two-way within-subject ANOVAs confirmed significant main effects of attention (single versus dual, $p < 0.001$), stimulus type (face versus disc, $p < 0.001$) and their interaction ($p < 0.001$) for peripheral task performance. TA^{actual} for face discrimination was 0.76. By contrast, TA^{actual} for disc-colour was 0.08, implying complete trade-off in this feature.

(ii) Perceptual awareness reduced for both features in the near absence of attention

In experiment 3, we employed a PAS (see the electronic supplementary material) to assess subjective experience [59] (figure 2*e,k*; figure 3*c,d*; electronic supplementary material, table S8). The relationship between attention and correctness on PAS ratings in experiment 3 was examined using LME analyses for each stimulus type. For blended-disc discriminations, the full model with interaction term differed substantially from the reduced model ($\chi^2(1) = 38.73$, $p < 0.001$). By contrast, for blended-face discriminations, a reduced model without interaction between attention and correctness did not differ from the full model ($\chi^2(1) = 0.29$, $p > 0.25$). The main effects of attention and correctness were highly significant when LME was applied to subset data (all $p < 0.001$). This confirmed that despite using the same stimuli and presentation parameters, the relationship between correctness and PAS was moderated by attention for disc-colour orientation but remained consistent between attention conditions when discriminating face-gender.

(iii) Metacognitive accuracy

The results of metacognitive accuracy based on PAS in experiment 3 were largely as expected from experiments 1 and 2 with some exceptions (figure 2*f,l*; electronic supplementary material, table S9). Under the dual task, type 2 AUC was high (0.59) for face-gender discrimination but near chance for disc-colour (0.53). In the single-task condition, each was individually higher than chance (approx. 0.58). Subjects maintained similar levels of metacognition for the central letter task (approx. 0.62) across all conditions. A two-way within-subject ANOVA on metacognitive accuracy for peripheral stimuli found a significant main effect of stimulus type ($p = 0.032$), but not for attention or their interaction (each $p > 0.05$). For the trade-off analysis, we identified two outlier subjects (see below). After removal of the outliers, we confirmed that while metacognitive accuracy for the face-aspect remained intact ($TA^{actual} = 0.90$ (against 0, $p = 0.031$; against 1, $p > 0.25$)), the disc-aspect dropped (0.59, $p > 0.05$ against both 1 and 0).

Figure 4 lists the results of objective performance, PAS and metacognitive accuracy for two subjects whose TA^{actual} values were greater than three standard deviations from the mean TA^{actual} results for $n = 32$ datasets pooled across experiments 1–3. Figure 4*b,e* shows that the two subjects did not discriminate correct from incorrect trials under the single-task condition (i.e. overlapping red (correct) and faint red

Downloaded from http://rstb.royalsocietypublishing.org/ on October 29, 2018

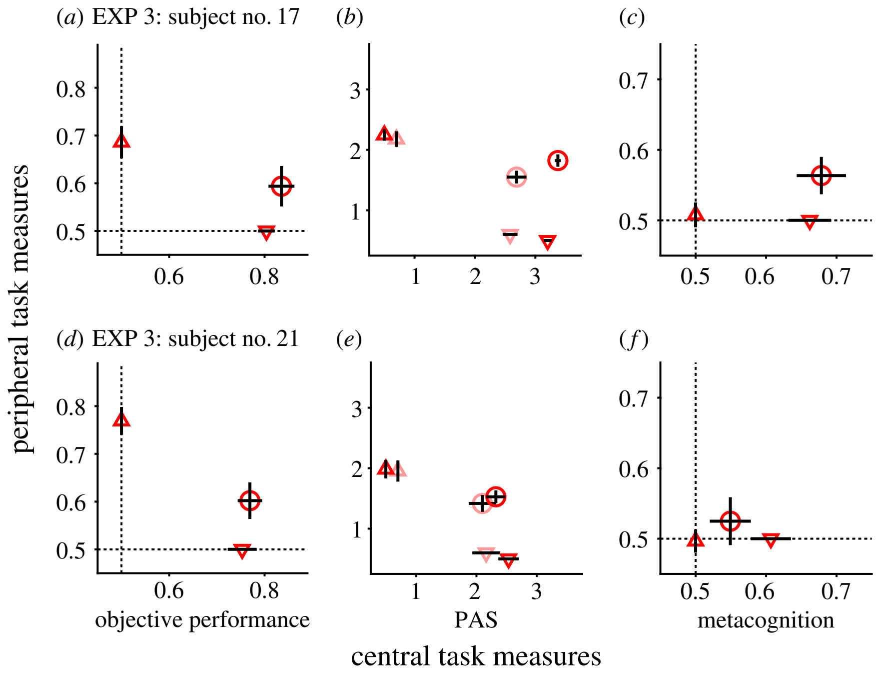

Figure 4. Two outlier subjects removed for trade-off analysis of face-aspect metacognition in experiment 3. (*a–c*) for subject no. 17, (*d–f*) for subject no. 21. (*a*) and (*d*) for objective performance, (*b*) and (*e*) for PAS, (*c*) and (*f*) for metacognition. Note higher metacognitive accuracy for dual-task face discriminations than those in the single-task condition despite lower mean performance and perceptual awareness. Error bars reflect standard error of the mean over blocks. See electronic supplementary material, figure S2*a–c* for all subjects' data in this format. Red, correct; faint red, incorrect. Upright triangles, peripheral single-task; inverted triangles, central single-task. (Online version in colour.)

(incorrect) upright triangles along *y*-axis), which resulted in chance metacognitive accuracy for the single-task condition (upright triangles in figure 4*c*,*f*). Because these subjects were able to discriminate between correct and incorrect trials in the dual-task condition, our method for determining trade-off resulted in TA^{actual} values that were massively positive (8.23 for figure 4*c*) and negative (−7.96 for figure 4*f*). We will return to this issue in our general discussion (see §6).

6. General discussion

In this paper, we addressed four criticisms of the dual-task paradigm: the issues of (i) metacognition, (ii) attraction of bottom–up spatial attention, or saliency, of faces, (iii) excessive training and (iv) response interference. The most important of these was metacognition: whether successful discrimination of faces in the near absence of top–down attentional amplification is achieved with conscious access. We assessed this by computing metacognitive accuracy, the correspondence between subjects' accuracy and subjective reports [54,55,57,61,65,66]. Metacognitive accuracy under the dual task has rarely been investigated (but see [56,67]). Addressing conscious access for stimuli in the dual task remains a critical limitation if this paradigm seeks to address the necessity debate [4].

We assessed the correspondence between subjects' discrimination accuracy and subjective reports of either confidence (experiments 1 and 2) or perceptual awareness (PAS, experiment 3) using type 2 AUC [54,57,61,66,68]. For gender discrimination, objective performance, confidence ratings and metacognitive accuracy did not differ greatly between the single (i.e. attended) and dual (i.e. unattended) task conditions. When the same procedure was applied to the simple disc stimuli, performance collapsed in the near absence of attention. Any evidence of above-chance discrimination in this condition was accounted for by attentional trade-off: subjects sacrificed performance on the central task in order to respond to disc stimuli.

This pattern of results was also found when we used PAS. Subjects' PAS for gender discriminations was broadly unchanged in the dual task and corresponded well with accuracy. This suggests that phenomenology of this perceptual feature (or simply, their appearance) remains largely unchanged when attention is diverted. We note that this result and failure to differentiate PAS for disc-colour in the near absence of attention are inconsistent with a study by Rahnev *et al*. [69]. Using simple, grating stimuli, they found that visibility ratings for orientation judgements were lower when subjects attended.

Our findings support the claim that certain stimulus categories, such as face-gender, remain consciously visible in the near absence of attention, while other features, such as colour orientation, do not. However, accepting this conclusion hinges on our addressing the remaining criticisms of the dual task: (ii) the saliency of face stimuli, (iii) excessive training and (iv) response interference.

It is possible that subjects' ability to distinguish gender in the periphery is a product of the inherent salience of face stimuli; a category known to attract our attention [38–40]. In experiment 3, using an α-blended, face/disc stimulus, we examined whether gender and/or colour orientation could be discriminated in the dual task. Critically, despite these stimuli being co-located and presented using equivalent SOAs in both conditions, face-gender but not colour orientation could be successfully categorized when attention was diverted to the central task. While diverting attention critically impairs subjects' ability to distinguish colour orientation, we conclude that face-gender remains discriminable not because faces attract spatial attention but because this feature remains consciously accessible in the near absence of spatial attention.

Downloaded from http://rstb.royalsocietypublishing.org/ on October 29, 2018

The third issue relates to training in the dual-task paradigm, which in previous studies typically required subjects to complete thousands of trials before the main experiment began [23–25]. Such extensive training may drastically alter neural circuits [70], affect consciousness [49] and impact performance [30–32]. We employed a psychometric staircase procedure in which criterion was reached in fewer than 100 trials, a reduction in excess of 97% when compared with traditional dual-task studies [23–25]. We found that face-gender discrimination was still possible in the near absence of attention, despite limited training (experiments 1 and 3).

Reducing training potentially leaves room for subjects to improve their performance during the main experiment, but we did not observe such improvement. None of our ANOVA analyses found significant main effects of block or interaction between block and other factors in any experiments or measures. Further, in experiment 3, we used more than 500 unique faces and showed each face only five times per subject over 3 days. Given these, face-gender discrimination under dual-task conditions is extremely unlikely to result from perceptual learning during the training phase, but is an inherent capacity of the visual system or perceptual learning through life.

The final issue is response-interference: could making a response on the central task result in subjects forgetting their answer for the peripheral disc task? Supporting this potential explanation, previous studies found some evidence that compared to complex stimuli, simple stimuli can be more effectively masked leaving shorter perceptual availability for subsequent reports [71–73]. Thus, in experiment 2, we examined whether disc discriminations can be rescued if the reporting procedure is simplified. This was achieved using a partial-report paradigm to reduce the load on perceptual memory. Our results clearly indicate that discrimination of colour orientation under the dual task was not rescued even when the partial-report condition minimized the influence of response interference.

(a) Limitations of the study and future directions

It is possible to think that attentional processing differs for different categories of discrimination. However, previous studies of the dual-task paradigm provide evidence that top–down attention is an undifferentiated resource [22,74,75]. For instance, Lee *et al.* [22] employed experiments to compare the concurrent discrimination of form, colour and motion. Interference was indistinguishable for similar (e.g. central letter task versus peripheral letter task) and dissimilar (e.g. central letter task versus peripheral motion task) task combinations, which highlights that different visual discriminations likely exhaust the same attentional capacity.

However, it is not clear if the same logic can be applied to subjective ratings and metacognition (type 2 AUC). In fact, we identified two outlier subjects (figure 4) who clearly violate the assumptions of dual-task studies. These two subjects did not distinguish correct and incorrect trials in terms of PAS in the single task but did so in the dual task. There can be several possible reasons for this behaviour.

One possibility is that the instructions for rating perceptual awareness were unclear. Because we did not include examples of invisible and visible stimuli, it is possible that subjects were not sure when to assign PAS of 1 and 4. Another possibility is that these outlier subjects used PAS 'across' the single- and dual-task conditions, a phenomenon we call *metacognitive saturation*. Inspection of these subjects' type 1 AUC and PAS reveals that their type 1 AUC and PAS for the peripheral faces were higher for the single task than for the dual task. When subjects apply a single criterion for subjective ratings across tasks of dissimilar difficulty, subjective ratings in the simple task can saturate [54,76]. This *metacognitive saturation* prevents type 2 measures from adequately discriminating metacognitive sensitivity in the easy task and, by comparison, inflates the metacognitive accuracy of the difficult task despite subjective ratings being lower on average.

Both of these potential issues may also relate to genuine inter-individual differences in metacognition [55,77]. In future studies, we may be able to reduce apparent individual differences by including stimuli that are clearly visible or invisible, thus setting clear reference stimuli for all participants. We can take advantage of our expedited training procedure to test many more subjects and investigate true individual differences in metacognition.

Further, these improvements permit investigation of the attentional requirements for perceiving a wide variety of stimulus features and categories. In particular, central and peripheral task discriminations could be constructed to involve stimuli that have, or do not have, supposed overlapping neural channels or receptive fields [78–80], using e.g. face stimuli in the central and peripheral tasks. This will allow one to investigate potential interference caused by overlapping neural representations of the central and peripheral stimuli, thereby clarifying whether the differentiation of these features in the near absence of spatial attention might be limited by the representational architecture of the visual system itself.

7. Conclusion

By using subjective ratings and metacognitive accuracy, we showed that certain aspects (face-gender) of peripheral vision but not others (colour-orientation) are consciously accessible in the near absence of top–down attention. This result was achieved despite minimal training and conforms with subjective reports and inferences from the dual-task literature in suggesting that the phenomenological distinction of features such as face-gender might be independent from selective attention [5,17–19,23,24,26]. Using the methods we present here, future studies might explore a range of stimulus types and features to reveal many categories of conscious perception in the near absence of attention. In doing so, we expect the distinction between top–down attention and consciousness might be clarified, permitting a deeper understanding of the functional and neuronal properties of each phenomenon [5,17,26].

Data accessibility. Data and code employed for analysing and conducting the experiments presented in this paper are available in the electronic supplementary material and on GitHub: github.com/julian-matthews.

Competing interests. We declare we have no competing interests.

Funding. L.K. and N.T. were supported by the Australian Research Council Discovery Project (DP130100194, http://www.arc.gov.au/discovery-projects). L.K. was supported by a fellowship from the Japanese Society for the Promotion of Science (JSPS P15048). N.T. was supported by Precursory Research for Embryonic Science and Technology project from the JST (3630, http://www.jst.go.jp/kisoken/presto/en/about/) and Australian Research Council Future Fellowship (FT120100619, http://www.arc.gov.au/future-fellowships).

Downloaded from http://rstb.royalsocietypublishing.org/ on October 29, 2018

Acknowledgements. We thank Alex Robinson and Vanessa Corneille for helping in the data collection and construction of experiment 3.

Endnote

[1]An independent line of research considers the dissociation between visual consciousness and other attentional processes (for review, see [5]). For instance, a series of studies of blindsight patient GY revealed that spatial cues could attract bottom–up attention whether those cues appeared in his spared or blind visual field [6,7]. A clear dissociation between bottom–up attention and visual consciousness has subsequently been demonstrated in healthy subjects using a wide variety of paradigms [8–12]. A similar dissociation has been reported between consciousness and feature-based attention [13–16].

References

1. Wundt W. 1874 *Grundzüge der physiologischen Psychologie*. Leipzig, Germany: Wilhelm Engelmann.
2. Cohen MA, Dennett DC, Kanwisher N. 2016 What is the bandwidth of perceptual experience? *Trends Cogn. Sci.* **20**, 324–335. (doi:10.1016/j.tics.2016.03.006)
3. Dehaene S. 2014 *Consciousness and the brain*. New York, NY: Viking Press.
4. Jennings CD. 2015 Consciousness without attention. *J. Am. Philos. Assoc.* **1**, 276–295. (doi:10.1017/apa.2014.14)
5. van Boxtel JJA, Tsuchiya N, Koch C. 2010 Consciousness and attention: on sufficiency and necessity. *Front. Psychol.* **1**, 217. (doi:10.3389/fpsyg.2010.00217)
6. Kentridge RW, Heywood CA, Weiskrantz L. 1999 Attention without awareness in blindsight. *Proc. R. Soc. B* **266**, 1805–1811. (doi:10.1098/rspb.1999.0850)
7. Kentridge RW, Heywood CA, Weiskrantz L. 2004 Spatial attention speeds discrimination without awareness in blindsight. *Neuropsychologia* **42**, 831–835. (doi:10.1016/j.neuropsychologia.2003.11.001)
8. Perry CJ, Fallah M. 2012 Color improves speed of processing but not perception in a motion illusion. *Front. Psychol.* **3**, 92. (doi:10.3389/fpsyg.2012.00092)
9. Norman LJ, Heywood CA, Kentridge RW. 2015 Exogenous attention to unseen objects? *Conscious. Cogn.* **35**, 319–329. (doi:10.1016/j.concog.2015.02.015)
10. Herreros L, Lambert AJ, Chica AB. 2017 Orienting of attention with and without cue awareness. *Neuropsychologia* **99**, 165–171. (doi:10.1016/j.neuropsychologia.2017.03.011)
11. McCormick PA. 1997 Orienting attention without awareness. *J. Exp. Psychol. Hum. Percept. Perform.* **23**, 168–180. (doi:10.1037/0096-1523.23.1.168)
12. Kentridge RW, Nijboer TCW, Heywood CA. 2008 Attended but unseen: visual attention is not sufficient for visual awareness. *Neuropsychologia* **46**, 864–869. (doi:10.1016/j.neuropsychologia.2007.11.036)
13. Schmidt F, Schmidt T. 2010 Feature-based attention to unconscious shapes and colors. *Atten. Percept. Psychophys.* **72**, 1480–1494. (doi:10.3758/APP.72.6.1480)
14. Tapia E, Breitmeyer BG, Shooner CR. 2010 Role of task-directed attention in nonconscious and conscious response priming by form and color. *J. Exp. Psychol. Hum. Percept. Perform.* **36**, 74–87. (doi:10.1037/a0017166)
15. Melcher D, Papathomas TV, Vidnyánszky Z. 2005 Implicit attentional selection of bound visual features. *Neuron* **46**, 723–729. (doi:10.1016/j.neuron.2005.04.023)
16. Kanai R, Tsuchiya N, Verstraten FAJ. 2006 The scope and limits of top-down attention in unconscious visual processing. *Curr. Biol.* **16**, 2332–2336. (doi:10.1016/j.cub.2006.10.001)
17. Tsuchiya N, Koch C. 2016 The relationship between consciousness and top-down attention. In *The neurology of consciousness* (eds S Laureys, O Gosseries, G Tononi), pp. 71–91. Cambridge, MA: Academic. (doi:10.1016/B978-0-12-800948-2.00005-4)
18. van Boxtel JJA. 2017 Different signal enhancement pathways of attention and consciousness underlie perception in humans. *J. Neurosci.* **37**, 5912–5922. (doi:10.1523/JNEUROSCI.1908-16.2017)
19. Braun J, Julesz B. 1998 Withdrawing attention at little or no cost: detection and discrimination tasks. *Percept. Psychophys.* **60**, 1–23. (doi:10.3758/BF03211915)
20. Braun J, Sagi D. 1990 Vision outside the focus of attention. *Percept. Psychophys.* **48**, 45–58. (doi:10.3758/BF03205010)
21. Sperling G, Dosher BA. 1986 Strategy and optimization in human information processing. In *Handbook of perception and human performance* (eds KR Boff, L Kaufman, JP Thomas), pp. 1–65. New York, NY: Wiley.
22. Lee DK, Koch C, Braun J. 1999 Attentional capacity is undifferentiated: concurrent discrimination of form, color, and motion. *Percept. Psychophys.* **61**, 1241–1255. (doi:10.3758/BF03206177)
23. Reddy L, Wilken P, Koch C. 2004 Face-gender discrimination is possible in the near-absence of attention. *J. Vis.* **4**, 106–117. (doi:10.1167/4.2.4)
24. Li FF, VanRullen R, Koch C, Perona P. 2002 Rapid natural scene categorization in the near absence of attention. *Proc. Natl Acad. Sci. USA* **99**, 9596–9601. (doi:10.1073/pnas.092277599)
25. Reddy L, Reddy L, Koch C. 2006 Face identification in the near-absence of focal attention. *Vision Res.* **46**, 2336–2343. (doi:10.1016/j.visres.2006.01.020)
26. Koch C, Tsuchiya N. 2007 Attention and consciousness: two distinct brain processes. *Trends Cogn. Sci.* **11**, 16–22. (doi:10.1016/j.tics.2006.10.012)
27. Bherer L, Kramer AF, Peterson MS, Colcombe S, Erickson K, Becic E. 2005 Training effects on dual-task performance: are there age-related differences in plasticity of attentional control? *Psychol. Aging* **20**, 695–709. (doi:10.1037/0882-7974.20.4.695)
28. Braun J. 1998 Vision and attention: the role of training. *Nature* **393**, 424–425. (doi:10.1038/30875)
29. Joseph JS, Chun MM, Nakayama K. 1997 Attentional requirements in a 'preattentive' feature search task. *Nature* **387**, 805–807. (doi:10.1038/42940)
30. Sigman M, Gilbert CD. 2000 Learning to find a shape. *Nat. Neurosci.* **3**, 264–269. (doi:10.1038/72979)
31. Garner KG, Tombu MN, Dux PE. 2014 The influence of training on the attentional blink and psychological refractory period. *Atten. Percept. Psychophys.* **76**, 979–999. (doi:10.3758/s13414-014-0638-y)
32. Verghese A, Mattingley JB, Garner KG, Dux PE. 2017 Decision-making training reduces the attentional blink. *J. Exp. Psychol. Hum. Percept. Perform.* **44**, 195–205. (doi:10.1037/xhp0000454)
33. Mack A, Rock I. 1998 *Inattentional blindness*. Cambridge, MA: MIT Press.
34. Block N. 2011 Perceptual consciousness overflows cognitive access. *Trends Cogn. Sci.* **15**, 567–575. (doi:10.1016/j.tics.2011.11.001)
35. Lamme VAF. 2010 What introspection has to offer, and where its limits lie. *Cogn. Neurosci.* **1**, 232–240. (doi:10.1080/17588928.2010.502224)
36. Sperling G. 1960 The information available in brief visual presentations. *Psychol. Monogr. Gen. Appl.* **74**, 1–29. (doi:10.1037/h0093759)
37. Wolfe JM. 1999 Inattentional amnesia. In *Fleeting memories* (ed. V Coltheart), pp. 71–94. Cambridge, MA: MIT Press.
38. Cerf M, Frady EP, Koch C. 2009 Faces and text attract gaze independent of the task: experimental data and computer model. *J. Vis.* **9**, 10.1–10.15. (doi:10.1167/9.12.10)
39. Fletcher-Watson S, Findlay JM, Leekam SR, Benson V. 2008 Rapid detection of person information in a naturalistic scene. *Perception* **37**, 571–583. (doi:10.1068/p5705)
40. Johnson MH, Dziurawiec S, Ellis H, Morton J. 1991 Newborns' preferential tracking of face-like stimuli and its subsequent decline. *Cognition* **40**, 1–19. (doi:10.1016/0010-0277(91)90045-6)
41. Cashon CH, Cohen LB. 2003 The construction, deconstruction, and reconstruction of infant face perception. In *The development of face processing in*

Downloaded from http://rstb.royalsocietypublishing.org/ on October 29, 2018

infancy and early childhood: current perspectives (eds O Pascalis, A Slater), pp. 55–68. Hauppauge, NY: Nova Science Publishers.

42. Bindemann M, Burton AM, Hooge ITC, Jenkins R, de Haan EHF. 2005 Faces retain attention. *Psychon. Bull. Rev.* **12**, 1048–1053. (doi:10.3758/BF03206442)
43. Theeuwes J, Van der Stigchel S. 2006 Faces capture attention: evidence from inhibition of return. *Vis. Cogn.* **13**, 657–665. (doi:10.1080/13506280500410949)
44. Persaud N, Navindra P, Peter M, Alan C. 2007 Post-decision wagering objectively measures awareness. *Nat. Neurosci.* **10**, 257–261. (doi:10.1038/nn1840)
45. Stoerig P, Cowey A. 2007 Blindsight. *Curr. Biol.* **17**, R822–R824. (doi:10.1016/j.cub.2007.07.016)
46. Weiskrantz L, Warrington EK, Sanders MD, Marshall J. 1974 Visual capacity in the hemianopic field following a restricted occipital ablation. *Brain* **97**, 709–728. (doi:10.1093/brain/97.1.709)
47. Debner JA, Jacoby LL. 1994 Unconscious perception: attention, awareness, and control. *J. Exp. Psychol. Learn. Mem. Cogn.* **20**, 304–317. (doi:10.1037/0278-7393.20.2.304)
48. Jiang Y, Costello P, He S. 2007 Processing of invisible stimuli: advantage of upright faces and recognizable words in overcoming interocular suppression. *Psychol. Sci.* **18**, 349–355. (doi:10.1111/j.1467-9280.2007.01902.x)
49. Merikle PM, Joordens S. 1997 Parallels between perception without attention and perception without awareness. *Conscious. Cogn.* **6**, 219–236. (doi:10.1006/ccog.1997.0310)
50. Jiang Y, Sheng H. 2006 Cortical responses to invisible faces: dissociating subsystems for facial-information processing. *Curr. Biol.* **16**, 2023–2029. (doi:10.1016/j.cub.2006.08.084)
51. Dehaene S, Changeux J-P. 2011 Experimental and theoretical approaches to conscious processing. *Neuron* **70**, 200–227. (doi:10.1016/j.neuron.2011.03.018)
52. Yang E, Brascamp J, Kang M-S, Blake R. 2014 On the use of continuous flash suppression for the study of visual processing outside of awareness. *Front. Psychol.* **5**, 724. (doi:10.3389/fpsyg.2014.00724)
53. Watson AB, Pelli DG. 1983 Quest: a Bayesian adaptive psychometric method. *Percept. Psychophys.* **33**, 113–120. (doi:10.3758/BF03202828)
54. Fleming SM, Lau HC. 2014 How to measure metacognition. *Front. Hum. Neurosci.* **8**, 443. (doi:10.3389/fnhum.2014.00443)
55. Song C, Kanai R, Fleming SM, Weil RS, Schwarzkopf DS, Rees G. 2011 Relating inter-individual differences in metacognitive performance on different perceptual tasks. *Conscious. Cogn.* **20**, 1787–1792. (doi:10.1016/j.concog.2010.12.011)
56. Stein T, Peelen MV. 2017 Object detection in natural scenes: independent effects of spatial and category-based attention. *Atten. Percept. Psychophys.* **79**, 738–752. (doi:10.3758/s13414-017-1279-8)
57. Matthews JR, Wu J, Corneille V, Hohwy J, van Boxtel JJA, Tsuchiya N. 2018 Sustained conscious access to incidental memories in RSVP. *PsyArXiv*. (doi:10.17605/OSF.IO/YSCDU)
58. VanRullen R, Reddy L, Koch C. 2004 Visual search and dual tasks reveal two distinct attentional resources. *J. Cogn. Neurosci.* **16**, 4–14. (doi:10.1162/089892904322755502)
59. Ramsøy TZ, Overgaard M. 2004 Introspection and subliminal perception. *Phenomenol. Cogn. Sci.* **3**, 1–23. (doi:10.1023/B:PHEN.0000041900.30172.e8)
60. Sandberg K, Timmermans B, Overgaard M, Cleeremans A. 2010 Measuring consciousness: is one measure better than the other? *Conscious. Cogn.* **19**, 1069–1078. (doi:10.1016/j.concog.2009.12.013)
61. Kaunitz LN, Rowe EG, Tsuchiya N. 2016 Large capacity of conscious access for incidental memories in natural scenes. *Psychol. Sci.* **27**, 1266–1277. (doi:10.1177/0956797616658869)
62. Cousineau D. 2005 Confidence intervals in within-subject designs: a simpler solution to Loftus and Masson's method. *Tutor. Quant. Methods Psychol.* **1**, 42–45. (doi:10.20982/tqmp.01.1.p042)
63. O'Brien F, Cousineau D. 2014 Representing error bars in within-subject designs in typical software packages. *Tutor. Quant. Methods Psychol.* **10**, 56–67. (doi:10.20982/tqmp.10.1.p056)
64. Li FF, VanRullen R, Koch C, Perona P. 2005 Why does natural scene categorization require little attention? Exploring attentional requirements for natural and synthetic stimuli. *Vis. Cogn.* **12**, 893–924. (doi:10.1080/13506280444000571)
65. Galvin SJ, Podd JV, Drga V, Whitmore J. 2003 Type 2 tasks in the theory of signal detectability: discrimination between correct and incorrect decisions. *Psychon. Bull. Rev.* **10**, 843–876. (doi:10.3758/BF03196546)
66. Kunimoto C, Miller J, Pashler H. 2001 Confidence and accuracy of near-threshold discrimination responses. *Conscious. Cogn.* **10**, 294–340. (doi:10.1006/ccog.2000.0494)
67. Sherman MT, Seth AK, Barrett AB, Kanai R. 2015 Prior expectations facilitate metacognition for perceptual decision. *Conscious. Cogn.* **35**, 53–65. (doi:10.1016/j.concog.2015.04.015)
68. Chen B, Mundy M, Tsuchiya N. 2016 Learning improves conscious access at the bottom, but not the top: reverse hierarchical effects in perceptual learning and metacognition. *bioRxiv* (doi:10.1101/073130)
69. Rahnev D, Maniscalco B, Graves T, Huang E, de Lange FP, Lau H. 2011 Attention induces conservative subjective biases in visual perception. *Nat. Neurosci.* **14**, 1513–1515. (doi:10.1038/nn.2948)
70. Dux PE, Tombu MN, Harrison S, Rogers BP, Tong F, Marois R. 2009 Training improves multitasking performance by increasing the speed of information processing in human prefrontal cortex. *Neuron* **63**, 127–138. (doi:10.1016/j.neuron.2009.06.005)
71. Houtkamp R, Braun J. 2010 Cortical response to task-relevant stimuli presented outside the primary focus of attention. *J. Cogn. Neurosci.* **22**, 1980–1992. (doi:10.1162/jocn.2009.21327)
72. Nieuwenstein MR, Potter MC. 2006 Temporal limits of selection and memory encoding: a comparison of whole versus partial report in rapid serial visual presentation. *Psychol. Sci.* **17**, 471–475. (doi:10.1111/j.1467-9280.2006.01730.x)
73. Nieuwenstein MR, Johnson A, Kanai R, Martens S. 2007 Cross-task repetition amnesia: impaired recall of RSVP targets held in memory for a secondary task. *Acta Psychol.* **125**, 319–333. (doi:10.1016/j.actpsy.2006.08.006)
74. VanRullen R. 2006 On second glance: still no high-level pop-out effect for faces. *Vision Res.* **46**, 3017–3027. (doi:10.1016/j.visres.2005.07.009)
75. Pastukhov A, Fischer L, Braun J. 2009 Visual attention is a single, integrated resource. *Vision Res.* **49**, 1166–1173. (doi:10.1016/j.visres.2008.04.011)
76. Peters MAK, Ro T, Lau H. 2016 Who's afraid of response bias? *Neurosci. Conscious.* **2016**, niw001. (doi:10.1093/nc/niw001)
77. Fleming SM, Weil RS, Nagy Z, Dolan RJ, Rees G. 2010 Relating introspective accuracy to individual differences in brain structure. *Science* **329**, 1541–1543. (doi:10.1126/science.1191883)
78. Cohen MA, Konkle T, Rhee JY, Nakayama K, Alvarez GA. 2014 Processing multiple visual objects is limited by overlap in neural channels. *Proc. Natl Acad. Sci. USA* **111**, 8955–8960. (doi:10.1073/pnas.1317860111)
79. Cohen MA, Nakayama K, Konkle T, Stantić M, Alvarez GA. 2015 Visual awareness is limited by the representational architecture of the visual system. *J. Cogn. Neurosci.* **27**, 2240–2252. (doi:10.1162/jocn_a_00855)
80. Cohen MA, Rhee JY, Alvarez GA. 2016 Limits on perceptual encoding can be predicted from known receptive field properties of human visual cortex. *J. Exp. Psychol. Hum. Percept. Perform.* **42**, 67–77. (doi:10.1037/xhp0000108)

Linking text between chapter 2 and 3

In Chapter 2, we established dissociations between conscious access, metacognition, and selective (focal) attention. In Experiment 1 we found that participants can differentiate face-gender with high metacognitive sensitivity despite perception taking place in the near absence of attention. In Experiment 2 we observed that participants were incapable of differentiating the orientation of coloured discs above chance-level if their attention was diverted. In Experiment 3 we accounted for the possibility that faces attract bottom-up attention by blending face and disc stimuli. Consistent with results from Experiments 1 and 2, participants were able to differentiate face-gender but not colour orientation under dual-task conditions. Moreover, participants exhibited high metacognitive sensitivity when rating the perceptual awareness of these faces supporting the notion that face-gender is accessible to conscious access in the near absence of attention.

Since peripheral stimuli remained task-relevant in the dual-task condition for all our experiments, the necessity of attention for consciousness remains an open question on the basis of this study. However, our results demonstrate that metacognition for certain categories of high-level perceptual features is possible even when the focus of attention is strongly engaged in a challenging primary task. On the basis of assumptions from the Metacognitive Model (Nelson, 1996), our results demonstrate that conscious access is possible for some stimulus features even if they are perceived with little attentional amplification and imply metacognition might be possible in the absence of attention.

In Chapter 3, we delved deeper into metacognition, conscious access, and attention by examining how these processes relate to short-term memory. Specifically, we contrasted memory capacity and metacognition for stimuli intentionally encoded and maintained in VWM versus stimuli incidentally encoded in the context of another memory task and thus not actively maintained in VWM. Examining short-term memory was important because VWM is typically regarded as severely capacity limited (Cowan, 2010; Luck & Vogel, 1997, 2013). It is conceivable that these limits also apply to conscious access and metacognition depending on how these processes relate to memory.

The effect of intentions on memorisation have been investigated for many years, particularly in visual long-term memory. The view that emerged in this time was intention to remember facilitates performance relative to incidental encoding but only insofar as this affects the depth of processing (e.g., encoding the meaning of stimuli rather than their appearance) (Craik & Lockhart, 1972; Postman, 1964). An open question is the effect of intentions and active maintenance on metacognitive performance. It is conceivable that when stimuli are encoded incidentally memorisation is supported largely by implicit rather than explicit knowledge (Graf & Birt, 1996). Explicit intentions to remember and active maintenance in VWM might then facilitate metacognitive performance relative to stimuli that are incidentally encoded or not actively maintained in VWM.

Incidental memory is almost always investigated in the context of visual search (Castelhano & Henderson, 2005; Chun & Jiang, 1998; Kaunitz et al., 2016; Williams et al., 2005). Visual search is a self-directed active process that is supported by endogenous and exogenous attentional mechanisms (Davis & Palmer, 2004). It is conceivable that these processes support memory encoding or retrieval. To account for this we used the Rapid Serial Visual Presentation paradigm to display stimuli for controlled durations at a central fixation.

We found that participants had high metacognitive sensitivity for stimuli perceived in upright and inverted orientation even if those stimuli were incidentally encoded in the context of a separate memory task. Subjects were aware that these stimuli would be tested. As such, we cannot completely exclude the possibility that subjects directed some attention to encoding these stimuli but we discouraged this behaviour by allowing only a short time period to respond to the memory task and presented a startling, feedback screen in case of memory errors.

Metacognitive access was sustained for at least 7 items in memory which is greater than the limit of 3 or 4 items typically associated with VWM (Luck & Vogel, 1997). Remarkably, memory capacity and performance was not markedly better when items were the target of intentional memorisation and explicitly maintained in VWM. We concluded that incidental short-term memory has a large capacity, remains consciously accessible, and does not require active maintenance. With respect to metacognition our findings suggest that sustained metacognitive access is possible for items that are not actively maintained in VWM. This implies that the cognitive mechanisms that support metacognition are dissociable from short-term memory, at least in some respects. Active maintenance from selective attention is not required for these memory traces to be employed in metacognitive decision-making, merely that the stimuli are perceived.

Chapter 3

Sustained conscious access to incidental memories in RSVP

Julian Matthews[1,2], Jamin Wu[2], Vanessa Corneille[2], Jakob Hohwy[1], Jeroen van Boxtel[2,3], Naotsugu Tsuchiya[2,3]

1. Cognition and Philosophy Lab, Monash University, Clayton, Australia

2. School of Psychological Sciences, Monash University, Clayton, Australia

3. Monash Institute of Cognitive and Clinical Neurosciences, Clayton, Australia

Attention, Perception, & Psychophysics
https://doi.org/10.3758/s13414-018-1600-1

Sustained conscious access to incidental memories in RSVP

Julian Matthews [1,2] · Jamin Wu [1] · Vanessa Corneille [1] · Jakob Hohwy [2] · Jeroen van Boxtel [1,3] · Naotsugu Tsuchiya [1,3]

© The Psychonomic Society, Inc. 2018

Abstract

In visual search of natural scenes, differentiation of briefly fixated but task-irrelevant distractor items from *incidental memory* is often comparable to explicit memorization. However, many characteristics of incidental memory remain unclear, including the capacity for its conscious retrieval. Here, we examined incidental memory for faces in either upright or inverted orientation using Rapid Serial Visual Presentation (RSVP). Subjects were instructed to detect a target face in a sequence of 8–15 faces cropped from natural scene photographs (Experiment 1). If the target face was identified within a brief time window, the subject proceeded to an incidental memory task. Here, subjects used incidental memory to discriminate between a probe face (a distractor in the RSVP stream) and a novel, foil face. In Experiment 2 we reduced scene-related semantic coherency by intermixing faces from multiple scenes and contrasted incidental memory with explicit memory, a condition where subjects actively memorized each face from the sequence without searching for a target. In both experiments, we measured objective performance (Type 1 AUC) and metacognitive accuracy (Type 2 AUC), revealing sustained and consciously accessible incidental memory for upright and inverted faces. In novel analyses of face categories, we examined whether accuracy or metacognitive judgments are affected by shared semantic features (i.e., similarity in *gender*, *race*, *age*). Similarity enhanced the accuracy of incidental memory discriminations but did not influence metacognition. We conclude that incidental memory is sustained and consciously accessible, is not reliant on scene contexts, and is not enhanced by explicit memorization.

Keywords Metacognition · Consciousness · Face perception · Short-term memory · Signal detection theory · Gender

Introduction

Our ability to recognize previously viewed pictures is remarkable. When tasked with memorizing large image sets, memory capacities consistently measure in the thousands, persist for great lengths of time, and exhibit remarkable specificity of content (Brady, Konkle, Alvarez, & Oliva, 2008; Konkle, Brady, Alvarez, & Oliva, 2010; Standing, 1973). However, in the bustling unpredictability of everyday life it is rarely the case that we benefit from the foresight and time for explicit memorization. Rather, critical decisions often rely on us accessing our *incidental memory* for items viewed for moments in time and outside the focus of our primary behavioral goals.

The study of incidental memory can be traced back to at least the 1930s (Silverman & Cason, 1934). Its interest in more contemporary research was sparked by the finding that implicit learning of contextual cues can guide spatial attention during visual search (Chun & Jiang, 1998, 2003). This, in turn, inspired comparisons between incidental and explicit encoding of visual scenes that revealed recognition memory capacity was sustained irrespective of memorization procedure (Castelhano & Henderson, 2005; Williams, Henderson, & Zacks, 2005). These studies concluded that the mere perception of a scene was sufficient to form detailed representations in long-term memory. However, the precise relationship between incidental memory and consciousness remains elusive. For example, if one assumes a strong association between working memory and consciousness (Baars & Franklin, 2003; Baddeley, 2003), one might expect that

Electronic supplementary material The online version of this article (https://doi.org/10.3758/s13414-018-1600-1) contains supplementary material, which is available to authorized users.

✉ Julian Matthews
julian.r.matthews@gmail.com

1 School of Psychological Sciences, Monash University, Clayton, VIC, Australia

2 Cognition and Philosophy Lab, Monash University, Clayton, VIC, Australia

3 Monash Institute of Cognitive and Clinical Neurosciences, Monash University, Melbourne, VIC, Australia

Published online: 05 October 2018

incidental memory should be strongly limited in sustained conscious access due to the severe capacity restrictions of these memory systems (Luck & Vogel, 1997; Simons & Levin, 1997).

In a previous study, we explored the relationship between incidental memory and consciousness (Kaunitz, Rowe, & Tsuchiya, 2016). We demonstrated reliable discrimination of upright distractor faces that had been briefly fixated but rejected as a target during an attention-demanding visual search. In that study, we tracked subjects' fixations while they moved their eyes to find a target face in a crowd scene as quickly as possible. If subjects found the target within a narrow time window (3–5 s), they performed an incidental memory task, discriminating between a previously unseen, *foil* face cropped from the original photograph and a *probe* face that had been fixated but rejected during the preceding visual search. We found that incidental memories for upright faces exhibit impressive capacity and clarity. These memories extend to at least seven intervening faces, are robust to the masking qualities of saccadic eye-movements, and, critically, are associated with above-chance metacognitive accuracy. Metacognitive accuracy is a marker of conscious access that highlights the functional quality of incidental memory (Matthews, Schröder, Kaunitz, van Boxtel, & Tsuchiya, 2018; Nelson, 1996). In contrast, when the same procedure was applied to inverted scenes, incidental memory for inverted, distractor faces was limited to between three and five items, within the traditional limits associated with visual short-term memory (Luck & Vogel, 1997). A critical differentiation between our previous study of incidental memory (Kaunitz et al., 2016) and those that demonstrate massive capacity for explicit memory (Brady et al., 2008; Konkle et al., 2010; Standing, 1973) was our use of the visual search paradigm. Subjects freely scanned each photograph trying to find a target face (which was never the target for the memory task). While this procedure is a natural reflection of how we might use incidental memory in everyday life, two specific features of visual search might limit us from generalizing the claim that incidental memory has sustained conscious access: (1) the self-directed nature of visual search, and (2) the semantic coherency of natural scenes.

First, visual search with overt eye movements is an inherently self-directed, active process, and benefits from preattentive processing as well as the processing associated with selective attention and conscious agency. The link between selective attention and visual working memory is documented so frequently that some have proposed they share cognitive mechanisms (Awh & Jonides, 2001; Awh, Vogel, & Oh, 2006; Gazzaley & Nobre, 2012; Theeuwes, Belopolsky, & Olivers, 2009; Woodman & Luck, 2004). These proposals extend more broadly to arguments that equate working memory with consciousness (Baars & Franklin, 2003; Baddeley, 2003). Given this, it is plausible that the self-directed nature of overt visual search reinforces incidental memories and might underlie their metacognitive access.

Second, visual search is supported by the semantic coherency of natural scenes. Humans can recognize the *gist* of a novel image with a single glance (Oliva & Torralba, 2006). While the precise depth of the image representation is debated, growing evidence suggests that human subjects accumulate information within a visual scene in long-term memory across fixational eye movements (Henderson, 2005; Hollingworth & Henderson, 2002). As a consequence, they form a detailed representation of the scene and gain enhanced recognition both for objects explicitly memorized from natural scenes as well as those incidentally viewed (Beck, Peterson, Boot, Vomela, & Kramer, 2006; Castelhano & Henderson, 2005). Given this, it is plausible that the semantic coherency of natural scenes might enhance the capacity of incidental memories during visual search.

While the distinction between incidental and explicit memory has been explored in considerable depth for visual search (Beck et al., 2006; Castelhano & Henderson, 2005; Varakin, Frye, & Mayfield, 2012; Varakin & Hale, 2014), it has only rarely been studied in passive, eye-movement constrained tasks. Early studies that projected items in sequence, obviating eye-movements, found that incidental memory discrimination was comparable to explicit memorization only when certain categories of stimuli were used such as faces or familiar objects (Bird, 1976; Bower & Karlin, 1974; Cohen, 1973). Similarly, a more recent study found that distractors related to the target of search by category or color were associated with enhanced incidental memory performance (Williams, 2010). While this body of work raises a possibility that stimulus categories might influence incidental memory, to what extent this is true for rapidly presented stimuli remains an open question.

Indeed, in these studies that removed visual search, stimuli were unmasked and presented up to several seconds at a time. And, while subjects were not informed that their incidental memory would be tested, to limit this sense of anticipation it was necessary that encoding and testing of items was separated by indistinct passages of time and hundreds of other stimuli. Likewise, those studies that demonstrate massive capacity for explicit memory also displayed stimuli for 3–10 s per image (Brady et al., 2008; Konkle et al., 2010; Standing, 1973; Standing, Conezio, & Haber, 1970). Presentation intervals such as these encourage rehearsal, rendering these studies of long-term rather than short-term memory.

Furthermore, previous studies have failed to characterize to what extent different memory types are supported by conscious access. Accurate metacognitive judgment about incidental memory (i.e., if you know you remember it) rules out the possibility that incidental memory is purely implicit (see a related argument on visual sensory memory in (Vandenbroucke et al., 2014)). Such evidence implies that incidental memory serves a functional role in everyday life through voluntary conscious access. Despite its potential importance, there have been no studies, to our knowledge, that characterize the degree of conscious access to incidental memory and examine whether this changes when subjects employ explicit memorization.

Our paper addresses these topics, examining performance and metacognitive access of incidental and explicit memory for rapidly presented upright and inverted faces at fixation. We examined faces due to their uniform size, high subcategory variability, and susceptibility to configural disruption as a result of the face inversion effect (Freire, Lee, & Symons, 2000) but return to the question of generalizability in the *General discussion*. We conducted our investigation in two parts: Firstly, with the behavioral results of psychophysical experiments and secondly, by examining the influence of stimulus categories on memory performance and metacognitive access.

In two experiments, we used the Rapid Serial Visual Presentation (RSVP) paradigm to control eye movements and precise stimulus timing. We tested incidental memory using a dual-task design to maximize the number of trials for behavioral and category analysis. Subjects searched for a target face as their primary task and, if the target face was successfully identified, discriminated between a novel, foil face and probe face that appeared as a distractor in the preceding sequence. Subjects could anticipate this test of incidental memory because it followed on every (correct) trial but attention to this task was partial at most since we enforced a brief time window (740 ms) for the primary task (ensuring it was difficult) and displayed a startling feedback screen in the case of target misses or errors (ensuring participants were motivated to perform the primary task optimally). In Experiment 1, we contrasted incidental memory for faces in upright and inverted orientation without self-directed visual search. Faces were drawn from a single crowd-scene photograph per trial. In Experiment 2, we intermixed faces from multiple crowd photographs to further reduce the influence of scene-related, semantic coherency. We then contrasted incidental memory with an explicit memory condition where target search was removed and subjects were told to remember all faces from the trial.

In our category analysis section we examined how trial-by-trial correctness and confidence are impacted by semantic features of faces. Specifically, we examined the impact of similarity, characterized here by stimuli sharing categories (e.g., *gender*, *race*, *age*). We achieved this first at the level of *item* (e.g., target and probe faces both being the same gender) and second at the level of *sequence* (e.g., the probe face being a different race from the majority of faces in the trial).

General method

Both experiments included faces cropped from natural, crowd-scene photographs used in Kaunitz et al. (2016). We employed the RSVP paradigm to control eye movements. This paradigm presents rapid sequences of visual information at a central fixation (Potter, 1976; Spence & Witkowski, 2013). The difference between experiments was the arrangement of face stimuli and task instructions. In Experiment 1, each trial included faces from a single, crowd-scene photograph in either upright or inverted orientation (Fig. 1a). Because all faces within a trial came from the same scene, we describe this stimulus condition as *within-scene*. In many cases, in a given photograph, there was a noticeable and strong correlation in facial features among the faces. For example, if a photo was from an elementary school graduation in an Asian country, the faces were almost exclusively populated by young, Asian faces. In Experiment 2, to limit the influence of semantic coherency and contextual cues we used *across-scene* stimuli. We achieved this by including faces from multiple photographs per trial, contrasting upright (Exp. 2a: Fig. 1b) and inverted faces (Exp. 2b: Fig. 1c) as a between-subjects factor.

Our incidental memory condition employed a dual-task design. Subjects searched for a target face as their primary task, requiring as fast and accurate a response as possible (see *Target task* below). Provided the primary task was successful, subjects then discriminated probe faces as a secondary task. We enforced task priority by allowing only a short time window for target detection and giving feedback on the primary but not the secondary task. Responses to probe faces served as a test for incidental memory (see *Probe task*). In Experiments 2a and 2b, we included an explicit memory condition. This was achieved by removing the Target Task and instructing subjects to remember all faces from the trial. On these trials, no target face was shown. Instead, subjects actively memorized each face and discriminated between the probe and previously unseen, foil face.

Subjects

Thirty-four subjects with normal or corrected-to-normal vision completed the experiments (18 males and 16 females, ages 19–31 years). For statistical power we collected data from 12 subjects per experiment. Subject numbers were based on our previous study (Kaunitz et al., 2016). There, a large effect size (d >0.8) was obtained when examining the difference between upright and inverted conditions even with small sample sizes (N < 4). Subjects were recruited from the staff and student body of Monash University. Each subject received $20 per 1-h session. The present study received ethical approval; and subjects gave informed, written consent in accordance with the Monash University Human Research Ethics Committee.

The same 12 subjects participated in both upright and inverted versions of Experiment 1 (Fig. 1a, b), with six subjects completing the upright version first. For Experiment 2,

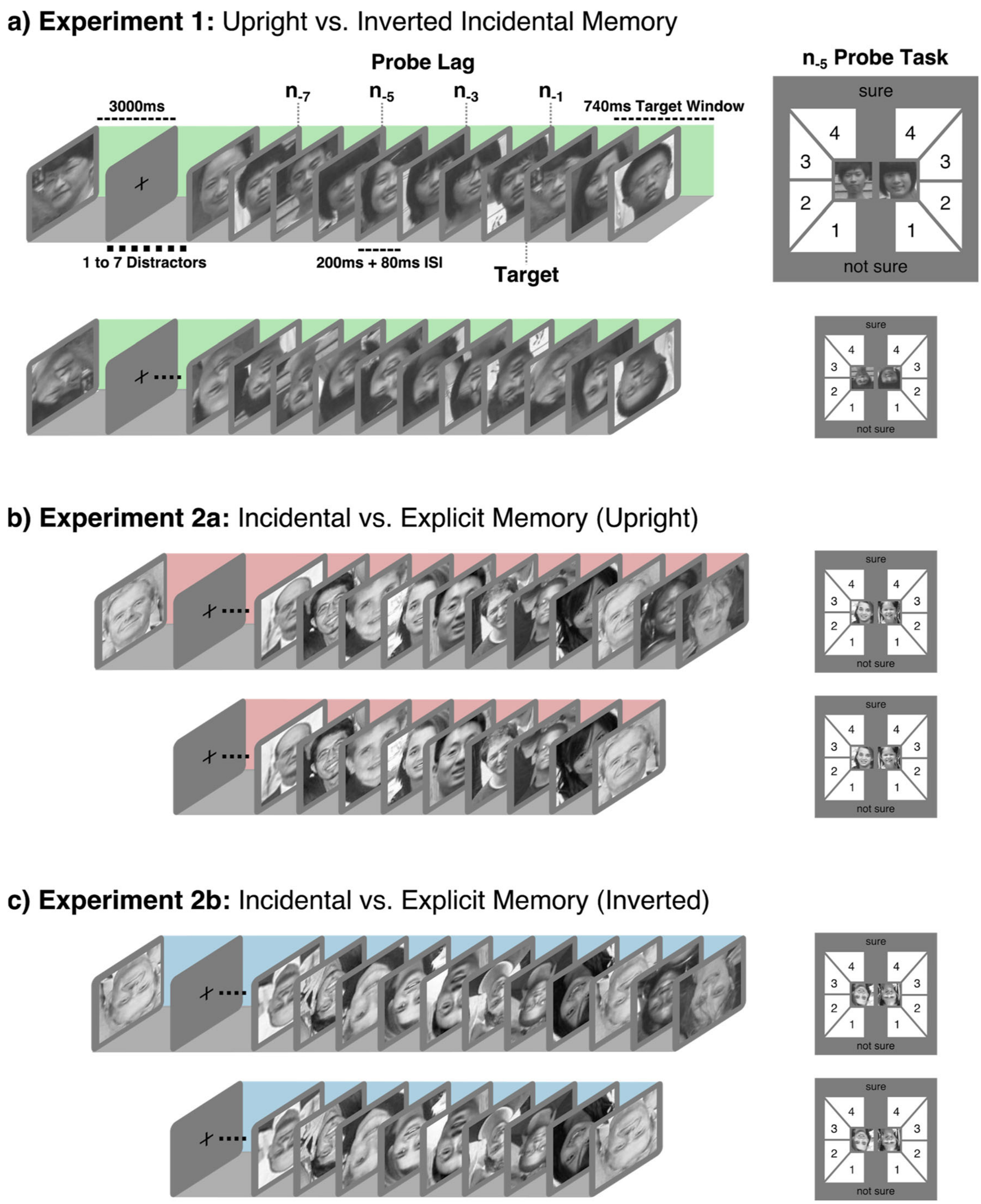

Fig. 1 Task design for all experiments. In the incidental memory condition (top row in **a**, **b**, and **c**), subjects identified a target face amongst a rapid-stream of eight to 15 distractor faces. Subjects were required to detect the target face within a 740 ms time window and received a startling feedback screen if they made a detection error or miss. If the target was successfully identified, subjects proceeded to the secondary Probe Task. Subjects selected between a *probe* face that had been presented but rejected during the stream and a novel *foil* face that had not been presented before and did not appear in subsequent trials. Subjects registered their decision and confidence with a single 8AFC mouse-click. The probe and target were separated by varying lags (n_{-1}, n_{-3}, n_{-5}, or n_{-7}); all figures in **a**, **b**, and **c** show examples of n_{-5} trials. In the explicit memory condition (**b**, **c**), subjects tried to remember all faces in the sequence and proceeded directly to the Probe Task. In the within-scene condition (**a**) faces were selected from one crowd photograph. In the across-scene condition (**b**, **c**) faces were selected from many photographs. Task-relevant faces (i.e., *targets*, *probes*, and *foils*) were only seen in a single trial per experiment.

12 subjects completed both the incidental and explicit memory tasks in upright orientation (Exp. 2a: Fig. 1c,d), with six subjects performing the explicit memory version first. Two subjects who had taken part in Experiment 1 also completed

Experiment 2a. A final 12 new subjects were recruited to complete the incidental and explicit memory tasks in inverted orientation (Exp. 2b: Fig. 1e, f), with six subjects performing the explicit version first.

Apparatus

All experiments were performed on a MacBook Pro laptop connected to a 22-in. SMI monitor located approximately 60 cm from the subject. The monitor was set with its refresh rate at 60 Hz and its screen resolution at 1,680 x 1,050 pixels. All experiments were programmed and conducted using the Psychophysics toolbox extension (Psychtoolbox-3) for MATLAB (Brainard, 1997).

Face stimuli were adapted from Kaunitz et al. (Kaunitz et al., 2016). These were 200 photographs of crowds (graduation ceremonies, bleachers, etc.) that were downloaded from the internet, gray-scaled, and rescaled into 1,100 x 768 pixels. Among 200 photos, we manually selected 160 of them that had many high-resolution unobstructed faces. All task-relevant faces (Target, Probe, and Foil) were individually cropped from the original photograph into 101 x 101 pixel squares. Task-irrelevant distractor faces were taken from the same set of 160 photos as the task-relevant faces.

Among all available faces, we selected faces that were largely unobstructed and predominantly forward facing with key features associated with visual classification (i.e., eyes, nose, mouth) visible. In some cropped face stimuli, facial features of another person could be seen but the central face dominated. Overall, we selected 4,225 faces from an initial pool of 5,006 in the 160 photos.

Procedure

Our RSVP paradigm proceeded as follows (see Fig. 1a). First, a randomly selected image of a target face was displayed for 3,000 ms. Second, a sequence of 8–15 faces was presented with each image displayed for 200 ms and followed by another one after an 80-ms blank interstimulus interval (ISI). These parameters (i.e., 200 and 80 ms) approximate the fixation and saccade duration during visual search (Kaunitz et al., 2016).

Subjects completed two sessions, with each session comprised of four blocks. Each block of 40 trials was followed by a short break. Task-relevant faces (i.e., targets, probes, and foils) were only presented on a single trial across an entire experiment. The remaining task-irrelevant faces appeared only once within a session of 160 trials (approximately 2,400 faces total) but could reappear between sessions. Thus some untested distractors used in the first session were presented in the second (approximately 20% of task-irrelevant faces). Counterbalancing of temporal positions of the probe (n_{-1}, n_{-3}, n_{-5}, and n_{-7}) and sequence length (8–15) was performed within each block. In total, subjects completed 320 trials per experiment.

Target task

Subjects were asked to click the left mouse button as fast and accurately as possible, when the target face image appeared in the RSVP sequence. To ensure sufficient attentive focus on this primary task, we imposed a narrow time window of 740 ms starting from 100 ms after the onset of the target. This time window included the target presentation, two subsequent faces, and three ISIs. We regarded mouse clicks that occurred within the target time window as “hit” and before as “false alarm” (only 137 instances of a < 100-ms response were registered across all experiments). If no click were recorded by the end of the window, we regarded the trials as “miss.”

As an extra reinforcement for the primary task, we employed a startling feedback screen if subjects made a target detection error. Upon false alarms or misses, subjects received feedback in the form of a full-screen flashing alert that cycled between yellow, red, and black at 12 Hz for approximately 500 ms. This reminded subjects that the target detection task had to be prioritized and discouraged subjects from devoting undue attention to the probe task. We did not collect a response to the probe task if subjects made a target-detection error. Instead, subjects skipped the probe task and continued to the next trial.

Probe task

If the target was successfully identified within the time window, subjects proceeded to the probe task; a test of incidental memory. This task involved discriminating between a previously unseen “foil” face image and a “probe” face image that was displayed during the sequence leading up to the target. The probe was the face displayed either n_{-1}, n_{-3}, n_{-5}, or n_{-7} faces prior to the target in the incidental memory condition and relative to the final face in the trial sequence in the explicit memory condition.

On the response screen, the probe and foil face images were displayed centrally with their presentation side (i.e., left or right) randomized in each trial. A response square surrounded the probe and foil faces. The response square comprised of eight evenly split segments, four each for the probe and foil (Fig. 1), which corresponded to four confidence levels. This response screen allowed subjects to register their 2AFC discrimination report as well as confidence in this decision with a single mouse click. Prior to the experiment, and during practice, subjects were verbally instructed to express their confidence on a scale from a complete guess (rating 1) to certainty (rating 4). Verbal descriptors for judgments of 2 or 3 were not made explicit; however, the experimenter encouraged subjects to fix these confidence criterions across the two sessions and use all four confidence levels. The labels “sure” and “not sure” were displayed at the top and bottom of the screen to remind subjects of the confidence scale.

Behavioral analysis

Methods

Objective performance: Type 1 AUC

We adopted signal detection theory to estimate subjects’ objective discrimination accuracy and calculate Type 1 performance (Kaunitz et al., 2016; Macmillan & Creelman, 2004; Matthews et al., 2018). To construct a Type 1 receiver operating characteristics (ROC) curve, we regarded a trial in which the probe face was presented on the left side of the response square as a signal-present trial and classified the response as a hit or miss. If the probe face appeared on the right side of the response square it was regarded as signal-absent and classified as a correct rejection or false alarm. We shifted the criterion in seven steps to obtain a seven-inflection ROC curve. For signal-present trials at the first inflection, if subjects chose a face on the left side with the highest confidence (4) we classified the response as a hit and as a miss otherwise. Likewise for signal-absent trials, if subjects chose a face on the left with the highest confidence the response was classified as a false alarm and as a correct rejection otherwise. We shifted the criterion at the second inflection, if subjects chose a face on the left side with confidence ratings of 4 or 3 they were classified as hits and false alarms. We repeated this procedure until confidence ratings from 4 to 1 on the left and 1 to 3 on the right were classified as hits and false alarms. Thus, the proportion of hits and false alarms was computed for seven possible criteria resulting in a ROC curve with seven inflection points. The area under this seven-inflection ROC curve (Type 1 AUC) was then computed to provide a non-parametric estimate of objective accuracy (Kaunitz et al., 2016; Matthews et al., 2018).

Metacognitive accuracy: Type 2 AUC

To examine whether subjects identified the probe using information accessible to consciousness, we used Type 2 signal detection to assess metacognitive accuracy. We first categorized trials as correct or incorrect. Correct responses were those where subjects made a signal-present response for a signal-present trial or a signal-absent response for a signal-absent trial, regardless of the level of confidence. Otherwise, we regarded that trial as incorrect.

Trials were then classified according to confidence using the procedure above to shift the criterion in three steps and construct a three-inflection ROC curve. First, we regarded a correct trial (regardless of the side of the probe face) with confidence rating of 4 as a (metacognitive) hit and a miss otherwise. Likewise, we regarded an incorrect trial with confidence 4 as a false alarm and correct rejection otherwise. For the second criterion, correct trials with confidence of 4 or 3 were classified as hits and incorrect trials with confidence of 4 or 3 as false alarms. In the third criterion, these conventions were applied to confidence ratings of 4, 3, or 2. The proportion of hits and false alarms was computed for three possible criteria resulting in a ROC curve with three inflection points. The area under this three-inflection ROC curve (Type 2 AUC) was then adopted as a non-parametric estimate of metacognitive accuracy (Kaunitz et al., 2016; Matthews et al., 2018).

Linear mixed effect modeling

We used linear mixed effect (LME) analysis in MATLAB to examine the effects of various factors on objective and subjective AUCs. In Experiment 1, Type 1 AUC and Type 2 AUC were each modeled as dependent variables with the fixed effects of image orientation and temporal lag of the probe relative to the target (i.e., n_{-1}, n_{-3}, n_{-5}, or n_{-7}). In Experiment 2, LME analysis was conducted separately for upright and inverted orientation. Type 1 AUC and Type 2 AUC were each modeled as a function of the fixed effects of memory condition (incidental or explicit) and probe lag (relative to the target for the incidental memory condition or the final face in the sequence for the explicit memory condition). As a significance test, we performed likelihood ratio tests between full models containing both factors and reduced models excluding each factor of interest. In all models, random intercepts were defined for each subject with random slopes for each fixed effect (Barr, Levy, Scheepers, & Tily, 2013). Examination of residual plots did not reveal any deviations from homoscedasticity or normality.

To examine the relationship between confidence, correctness, and our factors of interest we used all trials from each experiment without averaging over block or stream length. Confidence rating was modeled as a function of orientation (in Experiment 1) or memory condition (in Experiment 2) along with probe lag, correctness, and their interactions. Interaction effects were tested by examining full models including interaction against simplified models with additive effects only. In the case of significant interactions, we subset data by each factor level and performed likelihood ratio tests between full models that included the factor of interest and a null model that excluded that factor.

Results

Objective performance (Type 1 AUC)

We discarded trials in which the target face was not correctly identified (see *General methods*). In Experiment 1, this yielded an average of 80.7% (SEM=2.4%) valid trials in the upright condition and 73.1% (SEM=2.8%) in the inverted condition. In the incidental memory condition of Experiment 2, an average of 88.3% (SEM=1.8%) trials were valid in the upright faces (Experiment 2a) and 83.9% (SEM=1.7%) in the inverted faces (Experiment 2b).

Results with Type 1 AUC are summarized in Fig. 2. In Experiment 1 (Fig. 2a), objective performance (Type 1 AUC) for probe faces presented in upright orientation (M=.69, SEM=.02) was significantly greater than inverted orientation (M=.60, SEM=.02) revealed by a main effect of face orientation in our likelihood ratio analysis ($\chi^2(1)=16.5$, $p<.001$). Similarly, we observed a significant main effect of probe lag ($\chi^2(3)=12.4$, $p=.01$). Tukey-Kramer adjusted *post hoc* comparisons confirmed that for both upright and inverted faces, probe discrimination was significantly greater at n_{-1} lag than the other lags (all $p<.01$). No significant differences were observed between the other lags (all $p>.25$). Two-tailed t-tests with Holm-Bonferroni correction were used to compare Type 1 AUC performance against chance (AUC > 0.5) at each probe lag for both upright and inverted orientation. Performance remained significantly greater than chance in all instances.

In Experiment 2a (Fig. 2 in red), which used upright faces, likelihood ratio tests revealed a significant interaction between memory condition (explicit vs. implicit) and probe lag position ($\chi^2(3)=9.1$, $p=.03$). Subsetting by each probe lag revealed that objective performance did not differ significantly between memory conditions in any but the n_{-7} lag ($\chi^2(1)=4.7$, $p=.03$) suggesting that explicit and incidental memory strategies were broadly equivalent for probe discrimination. The main effect of lag was highly significant when subsetting by each memory condition (both $p<.001$). Type 1 AUC remained significantly greater than chance in all cases (all $p<.001$).

In Experiment 2b (Fig. 2 in blue), which used inverted faces, the interaction between memory condition and probe lag did not reach significance ($\chi^2(3)=3.8$, $p>.25$) and objective performance did not differ significantly between the memory conditions ($\chi^2(1)=3.5$, $p=.06$). In contrast, the main effect of lag was highly significant ($\chi^2(3)=24.8$, $p<.001$) with performance significantly greater than chance in all cases (all $p<.01$).

Confidence

Results of confidence rating are summarized in Fig. 3. In Experiment 1 (Fig. 3a and d), likelihood ratio tests revealed that confidence levels were significantly influenced by probe lag ($\chi^2(3)=13.4$, $p<.01$) and correctness of the response ($\chi^2(1)=15.8$, $p<.001$) (in Fig. 3, filled symbols for correct and empty symbols for incorrect trials). However, they were not influenced by face orientation ($\chi^2(1)=.1$, $p>.25$) (in Fig. 3, upper (**a**, **b**, **c**) and lower (**d**, **e**, **f**) panels show the results for upright and inverted faces, respectively). The interaction between orientation and correctness was significant ($\chi^2(1)=6.0$, $p=.01$), with confidence ratings being more separated between correct and incorrect responses in the upright condition versus the inverted condition. Further, the interaction between lag and correctness was also significant ($\chi^2(3)=8.9$, $p=.03$), with confidence ratings being more separated between correct and incorrect responses for shorter lags. In contrast, the interactions between orientation and lag ($\chi^2(3)=7.6$, $p=.05$) and the 3-way interaction ($\chi^2(6)=8.9$, $p=.18$) did not reach significance.

In Experiment 2a (Fig. 3b and c) with upright faces, we observed significant main effects of lag ($\chi^2(3)=13.9$, $p<.01$) and correctness ($\chi^2(1)=20.9$, $p<.001$) but not memory condition ($\chi^2(1)=.6$, $p>.25$). A significant interaction between lag and correctness was observed ($\chi^2(3)=11.7$, $p<.01$) such that shorter lags were associated with greater separation between correct and incorrect responses. To examine the nature of the interaction, we subset data by correctness or lag. For each lag we

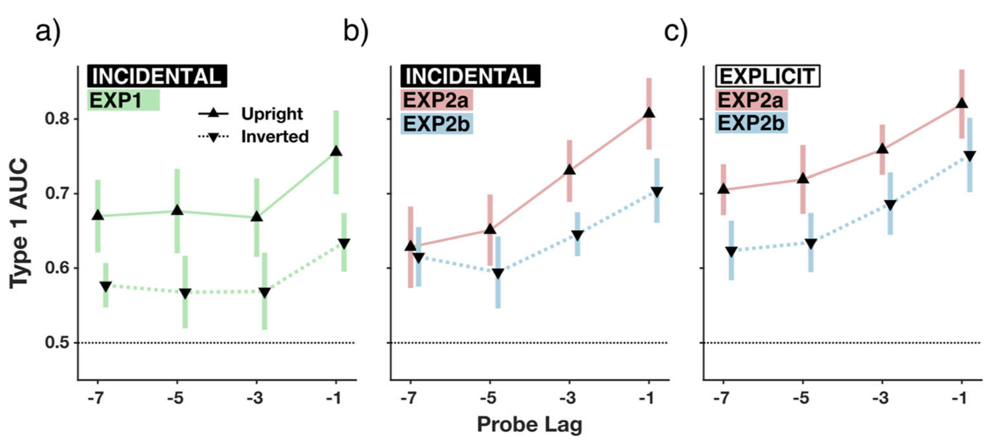

Fig. 2 Objective performance (Type 1 AUC) as a function of orientation, probe lag, and memory condition for Experiments 1 and 2. Upright triangles with solid lines reflect trials with upright faces (Exp. 1 and Exp. 2a). Inverted triangles with dotted lines represent trials with inverted faces (Exp. 1 and Exp. 2b). Error bars are 95% confidence intervals

Springer

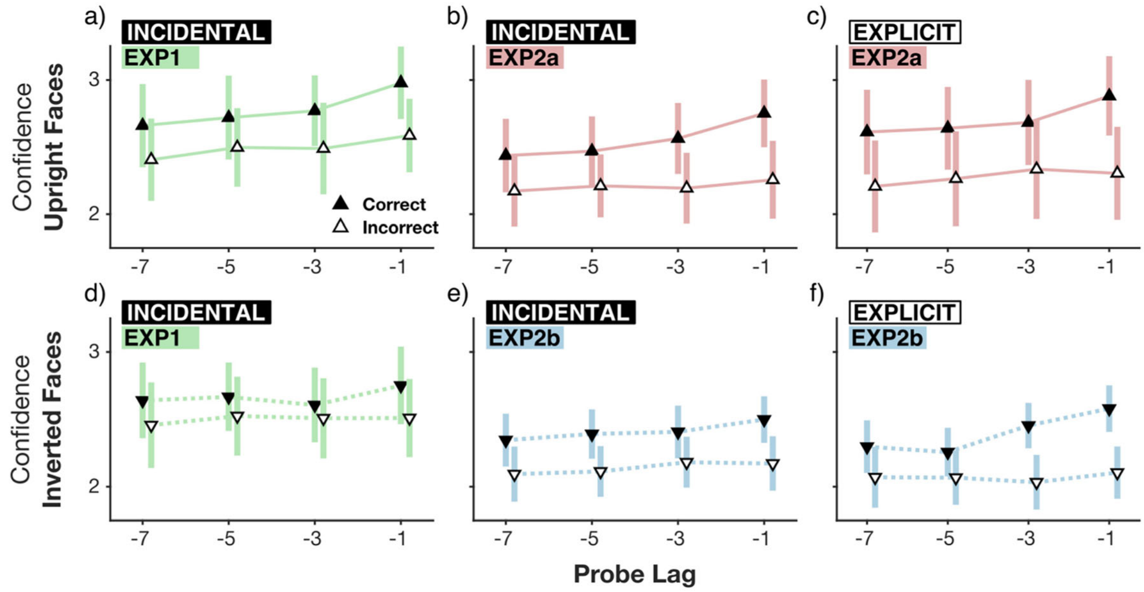

Fig. 3 Confidence as a function of face orientation, probe lag, memory condition, and correctness for Experiments 1 and 2. Filled triangles reflect correct discrimination and unfilled triangles incorrect discrimination. Upright triangles and solid lines represent trials with upright faces, while inverted triangles and dotted lines reflect trials with inverted faces. Error bars are 95% confidence intervals.

observed a significant main effect of correctness (all p<.001) such that correct judgments were associated with higher confidence. In contrast, subsetting by correct and incorrect judgments revealed a significant main effect of lag for correct judgments ($\chi^2(3)$=87.6, p<.001) but not incorrect judgments ($\chi^2(3)$=3.7, p>.25). The remaining 2-way interactions and the 3-way interaction did not reach significance (all p>.25).

We observed the same pattern of results for Experiment 2b (Fig. 3e and f) with inverted faces; significant main effects of lag ($\chi^2(3)$=17.9, p<.001) and correctness ($\chi^2(1)$=25.0, p<.001) but not memory condition ($\chi^2(1)$=.01, p>.25). The interaction between lag and correctness was significant ($\chi^2(3)$=10.9, p=.01), which comes from a significant main effect of correctness such that correct judgments were associated with higher confidence at each lag (all p<.001). Subsetting by correctness, lag had a significant effect on confidence ratings for correct judgments ($\chi^2(3)$=45.9, p<.001) but not incorrect judgments ($\chi^2(3)$=2.4, p>.25). The remaining interactions did not reach significance (all p>.07).

Overall, confidence ratings were higher for correct than incorrect trials, implying accurate metacognition, through all probe lags. Interestingly, more recent faces were rated more confidently only in the context of correct discrimination.

Metacognitive accuracy: Type 2 AUC

Results of metacognitive accuracy are summarized in Fig. 4. In Experiment 1, metacognitive accuracy was above chance (i.e.,

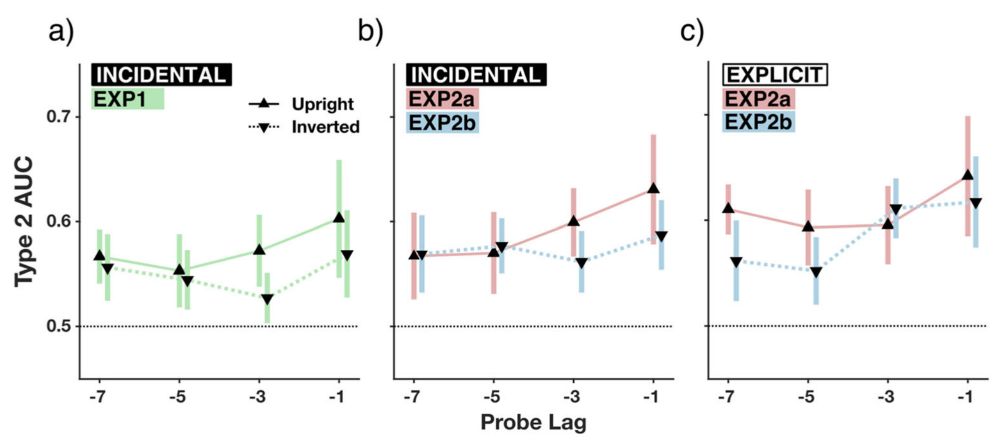

Fig. 4 Metacognitive accuracy (Type 2 AUC) as a function of face orientation, probe lag, and memory condition for Experiments 1 and 2. Upright triangles with solid lines represent upright face trials (Exp.1 and Exp. 2a). Inverted triangles and dotted lines are inverted face trials (Exp. 1 and Exp. 2b). Error bars are 95% confidence intervals

Type 2 AUC = 0.5) for all probe lags and face orientations (each $p<.05$) except n_{-3} lag for inverted faces ($t(11)=1.6$, $p=.13$). We observed no main effects of orientation or probe lag nor any interaction (all $p>.15$) which suggests metacognitive accuracy was broadly equivalent between the conditions.

In Experiment 2a and 2b we observed similar results; metacognitive accuracy was significantly greater than chance for all probe lags, face orientations and memory conditions (all $p<.05$). No main effect of memory condition was found for either upright or inverted faces ($\chi^2(1)=0.7$, $p>.25$ and $\chi^2(1)=2.0$, $p=.16$ respectively). The main effect of lag was not significant for upright faces ($\chi^2(3)=6.7$, $p=.08$), however, it did reach significance for inverted faces ($\chi^2(3)=8.7$, $p=.03$). Tukey-Kramer adjusted *post hoc* comparisons for inverted faces pooling across memory conditions confirmed that metacognitive accuracy at n_{-1} lag was significantly greater than n_{-5} ($p=.02$) but the remaining contrasts did not reach significance (all $p>.05$).

Summary of behavioral analysis

In Experiment 1, we built upon our previous study (Kaunitz et al., 2016) using the RSVP paradigm to remove the influence of self-directed visual search and overt eye movements. On each trial, faces were presented from a single, crowd scene photograph and face orientation was contrasted within-subjects. Our findings confirm that a large capacity of conscious, incidental memory for faces is not reliant on overt visual search.

In Experiment 2 we further removed the influence of semantic coherency by intermixing faces across scenes. Objective performance (Fig. 2a, b), confidence (Fig. 3a, b, d, e), and metacognitive accuracy (Fig. 4a, b) did not differ markedly between Experiments 1 and 2, despite differences in semantic coherency. Thus we conclude that semantic coherency among faces is not critical for incidental memory. Although face inversion was associated with a minor cost to objective performance (see Fig. 2), both objective performance and metacognitive accuracy remained significantly greater than chance for almost all probe lags in both upright (Exp. 2a) and inverted (Exp. 2b) orientation.

The second purpose of Experiment 2 was contrasting incidental memory with explicit memorization. Whether subjects were instructed to remember all faces from a given trial (*explicit memory*) or were engaged in a concurrent target detection task (*incidental memory*) made almost no difference for their capacity to discriminate probe faces. Objective performance, confidence, and metacognitive accuracy were broadly equivalent between memory conditions.

Category analysis

How is it possible that incidental memory is so similar to explicit memory in RSVP, as shown in Experiment 2? Williams (2010) examined this question in a series of experiments on long-term memory. Objects were presented in sequence at a central fixation and subjects either tried to memorize all of the objects or to search for a specific target (by counting its occurence). After presentation of 2 x 32 trial sequences (384 unique objects displayed twice for a total of 768 stimulus presentations), a series of 2AFC discrimination tests were performed. These tests were either unprompted or announced to the subject ahead of time. In their study, knowledge that a test would follow had no effect on memory. However, a secondary finding from this study was that searching for specific targets (e.g., a white car) impaired the memory of distractors unrelated to that target (e.g., a red door). Conversely, searching for a target did not affect memory for distractors that shared a category with the search target. Putting aside questions of face expertise (see *General discussion*) in the following analysis, we tested if our finding of equivalent performance between incidental and explicit memory may be partly explained by effects of similarity and dissimilarity between faces. A related consideration is the influence of similarity between the probe face and the sequence of distractor faces on each trial. In visual search, target discrimination can be predicted purely as a function of shared semantic features between targets and distractors in the scene (Mohan & Arun, 2012). Likewise, target stimuli (such as objects) in RSVP are known to receive enhanced encoding if preceded by semantically related items (Harris, Benito, & Dux, 2010; Maki, Frigen, & Paulson, 1997). Given this, it is possible that successful probe discrimination may be partly explained by similarity between the probe and the sequence of distractors on each trial.

We performed a category analysis to examine the influence of similarity on trial-by-trial correctness and confidence. First, we considered *item effects*; how probe discrimination is influenced when the target of search, probes, and foils either share or do not share semantic features (e.g., the target and probe being the same gender). Second, we considered *sequence effects*; how probe discrimination is influenced by the number of items in the preceding sequence that are similar or dissimilar to the probe or foil.

Methods

Ratings

Independent raters categorized each of the 4,225 faces used in our behavioral experiments. Categories included perceived *gender* (female; male; other), *race* (Asian; Black; White), and *age* (youth; adult; senior). A low-level feature of each face was also rated, *size* (small; medium; large), reflecting the scale of the face relative to the image. These categories were selected due to their visual salience.

A custom script was programmed in MATLAB to expedite the rating procedure. Each face was displayed in randomised order on a gray background. Raters were tasked with categorizing a single feature at a time (e.g., *gender*) and input their rating using the number keys 0, 1, or 2. Once a rating was made, the sequence progressed. In case of typographical errors, pressing the backspace key displayed the preceding face in the sequence. Raters were encouraged to prioritise the speed of categorization rather than accuracy, reflecting the implicit categorizations subjects might make in our experiments. Inter-rater reliability was high across all four categories; *gender* (94.6%, three raters), *race* (91.5%, two raters), *age* (79.6%, two raters), and *size* (82.1%, two raters). To obtain average ratings for the ordinal categories *age* and *size* we computed a mean rating between raters (disagreement results in scores that differ by 0.5 as in Fig. 5a). For *gender* we used the mode of the three raters and for *race* we subset those with rating disagreement into a fourth level: "mixed race".

Modeling item and sequence effects

We operationally defined our *item* level effect as the influence on probe task performance due to *dissimilarity* between the target and probe face categories. For our nominal categories *gender* and *race*, difference could vary from 0 to 1. The *age* and *size* categories were ordinal so varied from 0 to 2 (i.e., a senior target face vs. a youth probe face was a difference of 2). This method was repeated for dissimilarity between the target and foil as well as probe and foil.

We defined our *sequence* level effect as the influence on probe task performance due to statistics accumulated across the sequence of faces in each trial. For that purpose, we computed the negative log likelihood of each category in the trial. Specifically, we counted the appearances of a particular category (n) and the total number of faces in a given trial (N), then took the negative natural log of this value ($-\log_e(n/N)$); which we call *surprise*. Surprise varied from 0 (that is, n=N and $-\log_e(1) = 0$: a case where all the faces in the trial are in the same category and perfectly expected) to approximately 3 (that is, n/N=1/15, $-\log_e(1/15) = 2.7$: a case where a face had a unique category in the trial and was highly unexpected). We computed surprise for the probe and the foil for each category in each trial.

See Supplemental Fig. 1 for an example of face category rating plus dissimilarity and surprise analysis.

Mixed effect modeling

Item and *sequence* effects on trial-by-trial correctness were examined using generalized mixed effect modeling. This

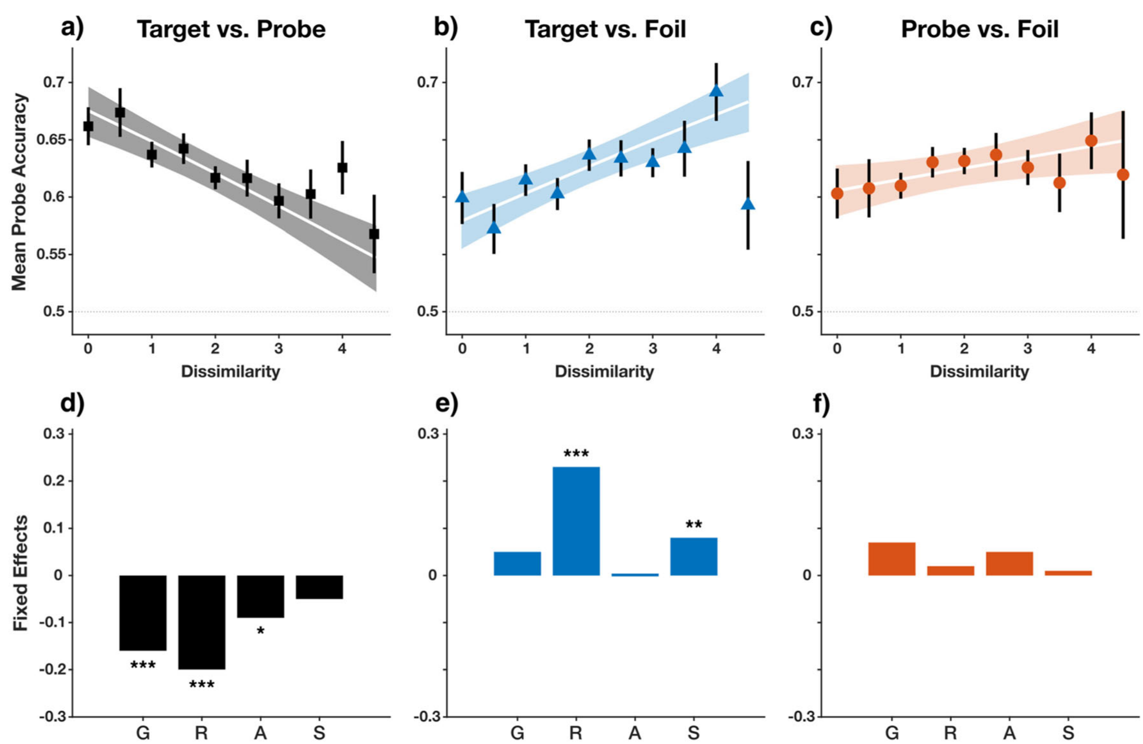

Fig. 5 Effects of item dissimilarity on accuracy of probe discrimination. (**a-c**) Mean probe accuracy as a function of dissimilarity (i.e., a dissimilarity score of 0 reflects categorically identical stimuli) for cumulative difference between the (**a**) target-probe, (**b**) target-foil, and (**c**) probe-foil. Error bars reflect standard error of the mean between-subjects. Lines are predicted response accuracy according to GLME models with their 95% confidence interval indicated by the shaded region. (**d-f**) Bars reflect the fixed effect estimates for each category (*gender*, *race*, *age*, and *size*) in the *discrete model* of dissimilarity on trial-by-trial correctness. Significance is determined by likelihood ratio tests (*=p<.05, **=p<.01, ***=p<.001)

was achieved with binomial logistic regression within the LME4 package in R (Bates, Mächler, Bolker, & Walker, 2015). Trial-by-trial confidence ratings were first subset by correctness and then examined using linear mixed effect modeling also in R.

Since the item effect *dissimilarity* was characterized as the difference between targets, probes or foils, we performed the analysis on valid trials only (i.e., the trials with a correct target detection within the time window; 12,508 trials in total). We did not analyze the trials from the explicit memory condition in Experiment 2 because there was no target in this condition. As the dependent variables (DVs), we used trial-by-trial correctness or confidence rating subset by accuracy. Two models were computed for each DV; a *cumulative model* to test effects of overall dissimilarity between targets, probes, and foils, and a *discrete model* to test the significance of each category independently. In the cumulative model, the unweighted sum of dissimilarity for all categories was computed to determine overall dissimilarity between target versus probe, target versus foil, plus probe versus foil (three fixed effects in total: see Supplemental Fig. 1). In the discrete model, dissimilarity in *gender*, *race*, *age*, and *size* between the target versus probe, target versus foil, as well as the probe versus foil were included (12 fixed effects in total).We computed the sequence effect *surprise* for the probe and foil using all valid trials of the incidental memory condition and all trials in the explicit memory condition (20,188 in total). Trial-by-trial correctness and confidence subset by accuracy were each modeled using *cumulative* and *discrete* models. Random intercepts were included per subject for both *dissimilarity* and *surprise* effects. Assumptions of multicollinearity were not violated for the above analyses as measured by variance inflation factors. Significance was determined using likelihood ratio tests against models excluding each factor of interest.

Results

Item dissimilarity effect

For trial-by-trial correctness in our *cumulative model*, we observed a significant negative correlation with target-probe dissimilarity (Fig. 5a: $\chi^2(1)=37.7$, $p<.001$), a significant positive correlation with target-foil dissimilarity (Fig. 5b: $\chi^2(1)=26.1$, $p<.001$), and a weaker positive correlation with probe-foil dissimilarity (Fig. 5c: $\chi^2(1)=4.6$, $p=.03$). No interactions were observed (all $p>.25$). These results imply that a probe was more likely to be selected if it was a similar category to the target, dissimilar to the foil, or if the foil itself was dissimilar to the target.

For our *discrete model* of trial-by-trial correctness, significant negative correlations were found for target-probe dissimilarity (Fig. 5d): *gender* ($\chi^2(1)=16.8$, $p<.001$), *race* ($\chi^2(1)=22.1$, $p<.001$), and *age* ($\chi^2(1)=4.1$, $p=.04$). Significant positive effects for target-foil dissimilarities were also found (Fig. 5e): *race* ($\chi^2(1)=30.9$, $p<.001$) and *size* ($\chi^2(1)=7.6$, $p=.01$). No significant effects of discrete probe-foil dissimilarity were observed (Fig. 5f). These results imply that a probe face was more likely to be selected if it was a similar gender, race, or age to the target face. Conversely, a probe face was less likely to be selected if the target and foil face were a similar race or size.

For response confidence subset by accuracy in our *cumulative model*, no significant effects of target-probe dissimilarity were found for either correct or incorrect discriminations, (Fig. 6a: each $p>.25$). Likewise, no significant effects of target-foil dissimilarity were observed (Fig. 6b: each $p>.25$). However, we observed a significant positive effect of probe-foil dissimilarity for correct judgments (Fig. 6c: $\chi^2(1)=5.5$, $p=.02$). This result implies that when subjects correctly selected the probe, their confidence in that choice was higher on average if the probe and foil face were dissimilar.

Our *discrete model* of confidence subset by accuracy elucidated this result. No effect was significant for target-probe or target-foil dissimilarity (all $p>.05$, Fig. 6d and e). Significant positive correlations of probe-foil dissimilarity were observed for *race* when correct ($\chi^2(1)=6.6$, $p=.01$) but also when incorrect ($\chi^2(1)=4.1$, $p=.04$) and for *gender* for correct responses ($\chi^2(1)=6.5$, $p=.01$). These results imply that when subjects correctly selected the probe, their confidence in that choice was higher on average if the probe and foil faces were a different gender. Moreover, independent of response accuracy, subjects' confidence in their choice was higher on average if the probe and foil faces were a different race.

Sequence surprise effect

For trial-by-trial correctness in our *cumulative* model, we found significant negative correlation with probe surprise (Fig. 7a: $\chi^2(1)=31.7$, $p<.001$) and positive correlation with foil surprise (Fig. 7b: $\chi^2(1)=52.7$, $p<.001$). These results imply that a probe face was less likely to be selected if it was dissimilar to many faces in the trial sequence surrounding it (i.e., the probe was *surprising*). Conversely, the probe face was more likely to be selected if the foil face was surprising.

For trial-by-trial correctness in our *discrete* model, we found significant negative correlations with probe surprise (Fig. 7c) in *gender* ($\chi^2(1)=22.2$, $p<.001$), *race* ($\chi^2(1)=11.0$, $p<.001$), and *size* ($\chi^2(1)=7.2$, $p<.01$) and positive correlations with foil surprise (Fig. 7d) in *gender* ($\chi^2(1)=25.1$, $p<.001$), *race* ($\chi^2(1)=27.1$, $p<.001$), and *size* ($\chi^2(1)=16.7$, $p<.001$). These results imply that a probe face was less likely to be selected if its gender, race, or size was surprising. Conversely, a probe face was more likely to be selected if the gender, race, or size of the foil face was surprising.

For confidence subset by accuracy in our *cumulative model*, we found a significant negative correlation with foil

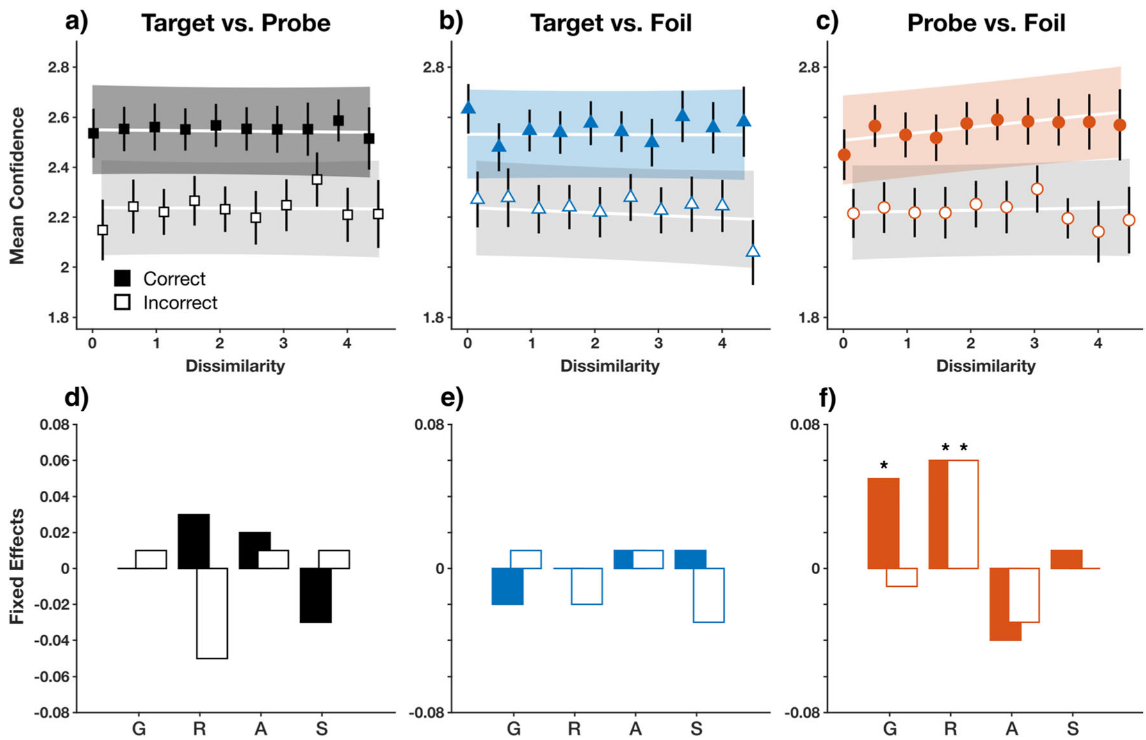

Fig. 6: Effects of item dissimilarity on confidence in discrimination subsetting by response accuracy. (**a-c**) Mean confidence subset by accuracy (filled=correct, empty=incorrect) as a function of dissimilarity for cumulative difference between (**a**) target-probe, (**b**) target-foil, and (**c**) probe-foil. Error bars reflect standard error of the mean between subjects. Lines are predicted confidence according to LME models, with their 95% confidence interval indicated by the shaded region. (**d-f**) Bars reflect the fixed effect estimates for each category (*gender*, *race*, *age*, and *size*) in the *discrete models* of dissimilarity on confidence for correct and incorrect trials (filled and empty respectively). Significance is determined by likelihood ratio tests (*=p<.05, **=p<.01, ***=p<.001)

surprise (Fig. 8b) for incorrect trials ($\chi^2(1)=10.3$, p=.001) but not for correct trials, and no significant effects for probe surprise (Fig. 8a and b, all p>.25). These results imply that when subjects' probe response was incorrect (i.e., they selected the foil), their confidence in that choice was lower on average if the foil was surprising.

In our *discrete model* of confidence subset by accuracy, we found significant negative correlation with probe surprise for *gender* in correct trials (Fig. 8c: $\chi^2(1)=5.3$, p=.02) and with foil surprise (Fig. 8d) for *gender* ($\chi^2(1)=8.8$, p<.01) and *race* ($\chi^2(1)=4.1$, p=.04) in incorrect trials. These results imply that when subjects' probe response was correct, their confidence in that choice was lower on average if the gender of the probe was surprising. Conversely, when subjects' probe response was incorrect (i.e., they selected the foil), their confidence in that choice was lower on average if the gender or race of the foil was surprising.

General discussion

We reveal that incidental memory for rapidly presented faces is consciously accessible and sustained, while not dependent on overt visual search or the semantic coherency of features within a single photograph. Across two experiments, we used RSVP to minimize overt eye movements and contextual whole-scene cues, plus limited stimulus viewing times by presenting stimuli for only 200 ms. Although the incidental memory test could be anticipated, we required that subjects engage their attention in a demanding target detection task (Fig. 1) and only analyzed incidental memory when target detection was successful. We tested memory performance and metacognitive accuracy for incidentally viewed probe items and, in the same subjects, contrasted this with an explicit memory condition where subjects actively memorized each item using full attention. To get insight into what mechanisms support memory, our category analysis examined how trial-by-trial accuracy and confidence is affected by faces sharing semantic features (i.e., gender, race, age, size). We examined the influence of similarity between the probe, target, and foil faces (the *item effect*) and the influence of similarity between the probe (or the foil) and the sequence of distractor faces that surround it (the *sequence effect*).

Our key results were: (1) subjects exhibit sustained objective memory and metacognitive access for upright and inverted faces seen for the first time and presented for only 200 ms up to seven items prior to an unrelated target; (2) in terms of measures of performance, confidence, and

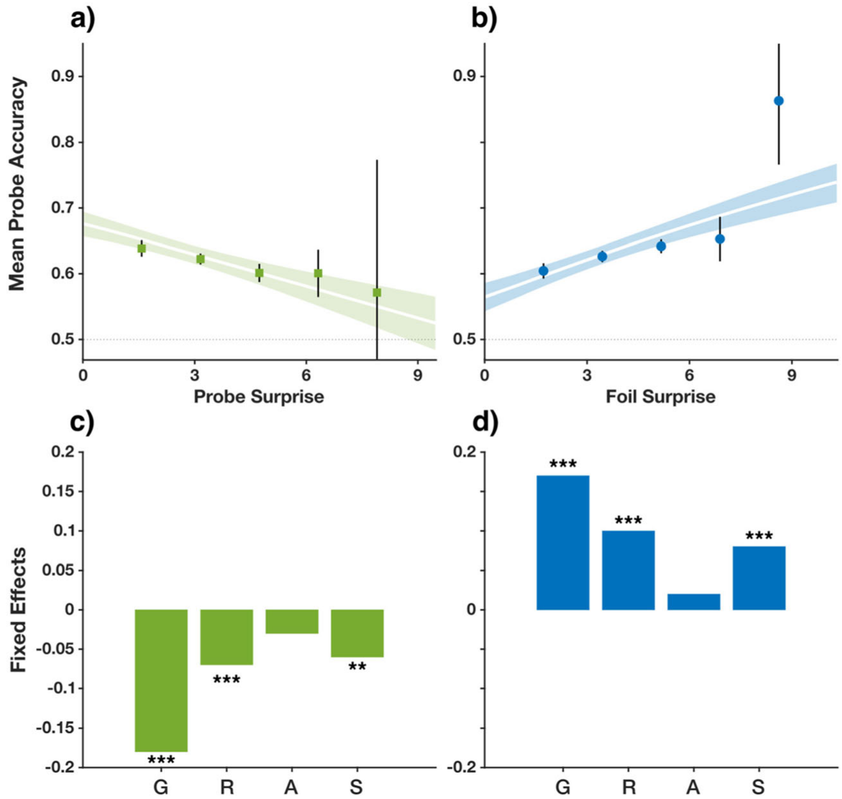

Fig. 7: Effects of surprise on trial-by-trial correctness. (**a**, **b**) Mean probe accuracy as a function of binned cumulative surprise for the (**a**) probe, and (**b**) foil. Error bars reflect standard error of the mean between subjects. Lines are predicted accuracy from each GLME model with their 95% confidence interval shaded. (**c**,**d**) Bars reflect the fixed effect estimates for each category (*gender*, *race*, *age*, and *size*) in the discrete models of surprise on trial-by-trial correctness. Significance is determined by likelihood ratio tests (*=p<.05, **=p<.01, ***=p<.001)

metacognitive accuracy, incidental memory and explicit memorization were broadly equivalent (Figs. 2, 3, and 4); and, (3) that trial-by-trial accuracy is strongly influenced by both *item* and *sequence* effects (Figs. 5 and 7), but these same effects have very little or no influence on trial-by-trial confidence ratings (Figs. 6 and 8).

Sustained conscious memory for incidental probes in RSVP

The act of self-directed, visual search involves two processes that reinforce incidental memory. First, incidental memory during active eye movements might be enhanced by overlapping neural mechanism for overt attention and working memory (Awh & Jonides, 2001; Awh et al., 2006; Gazzaley & Nobre, 2012). Second, incidental memory during visual search for an object in a scene can also be enhanced through semantic coherence of that scene, which is known to improve detection speed and recognition both for targets and incidentally-viewed objects (Beck et al., 2006; Castelhano & Heaven, 2010; Castelhano & Henderson, 2005). Given this, it was conceivable that the remarkable capacity and metacognitive access observed for incidentally-viewed probes in our previous visual search study was contingent on the act of visual search itself (Kaunitz et al., 2016). However, even when we eliminated overt eye movements using RSVP and minimized the influence of within-scene contextual cues, we still demonstrate that short-term incidental memory is sustained and consciously accessible.

There was one major difference between the results of our studies. Previously we found that incidental memory for inverted scenes had a limited capacity, while here we found sustained incidental memory for inverted faces (although upright faces had an advantage). Inversion of faces is known to disrupt the locations that people fixate upon them (Barton, Radcliffe, Cherkasova, Edelman, & Intriligator, 2006; Hills, Cooper, & Pake, 2013; Hills, Sullivan, & Pake, 2012). Although inversion of scenes (rather than faces) is much less studied, recent research has demonstrated that scene inversion disrupts subjects' capacity to foveate targets and expected target locations (Koehler & Eckstein, 2015). Given this, we surmise that fixations were not optimized for inverted scenes in our previous study (Kaunitz et al., 2016) while it was optimized in our RSVP design (due to the fixed location of stimulus presentation), which may explain the discordance in incidental memory for inverted faces. Moreover, our current study examined the effects of orientation and lag position within-subjects. This design is especially important for

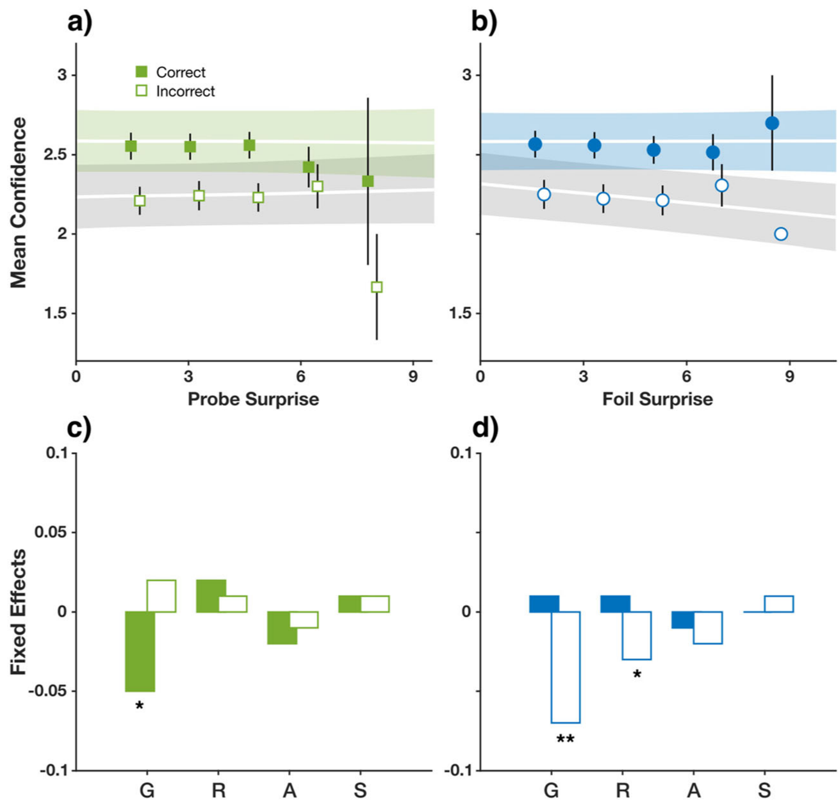

Fig. 8 Effects of surprise on trial-by-trial confidence subset by accuracy. (**a**, **b**) Mean confidence subset by accuracy (filled=correct, empty=incorrect) as a function of binned cumulative surprise for the (**a**) probe, and (**b**) foil. Error bars reflect standard error of the mean between subjects. Lines are predicted confidence by the LME models with their 95% confidence interval shaded. (**c**,**d**) Bars reflect the fixed effect estimates for each category (gender, race, age, and size) in the discrete models of surprise on trial-by-trial confidence subset by accuracy (filled=correct, empty=incorrect). Significance is determined by likelihood ratio tests (*=p<.05, **=p<.01, ***=p<.001)

analyzing confidence ratings and metacognitive accuracy as these measures are vulnerable to individual differences (Fleming, Weil, Nagy, Dolan, & Rees, 2010; Song et al., 2011).

What is the likelihood that our results generalize for stimuli other than faces? Incidental memory has been demonstrated for a range of objects in long-term memory and visual search (Castelhano & Henderson, 2005; Williams, 2010). The finding that short-term incidental memory capacity is sustained for upright and even inverted faces suggests incidental memory does not rely on holistic processing and might generalize beyond faces. In extending this paradigm to other categories of stimuli, the possibility remains that examining memory for one type of object might induce effects that would not be found otherwise. One such example is *recognition-induced forgetting* (Maxcey & Woodman, 2014). Here, when different visual stimuli are encoded into long-term memory, practice with a particular subset of those items impairs memory for related but unpracticed items when compared to unrelated *baseline* stimuli from the same encoding period. Recognition-induced forgetting has been demonstrated for race categories in faces such that rehearsing white faces impairs recognition for unrehearsed white faces but not those in the baseline group (black faces) (Rugo, Tamler, Woodman, & Maxcey, 2017). Future research might examine these factors in short-term incidental memory by balancing categories across the stimulus set. Moreover, these studies might employ our metacognitive measures to examine whether effects such as recognition-induced forgetting are consciously accessible.

Incidental and explicit memorization are broadly equivalent in RSVP

Incidental and explicit memorization instructions have been explored using the visual search paradigm. These studies find that explicit memorization guides spatial attention but makes little difference for recognition from long-term memory (Castelhano & Henderson, 2005; Varakin et al., 2012; Varakin & Hale, 2014) and only marginally improves recognition in tests of short-term memory (Beck et al., 2006). Our study here extends these results to a situation where eye movements are constrained and stimuli are only briefly presented (200 ms). In this situation, our measures of performance and metacognitive accuracy were broadly equivalent in the same subjects between memorization conditions (Figs. 2, 3, and 4). In particular, our finding of above-chance metacognitive accuracy suggests these representations are accessible to consciousness without explicit attentional amplification.

One caveat to this claim is our subjects were aware that their memory for incidental probes might be tested. This is quite unlike previous studies of incidental, long-term memory that presented subjects with an unexpected memory test after all stimuli had been viewed and contrasted performance with an intentional memory experiment where different subjects were informed that memory tests would follow (Bird, 1976; Castelhano & Henderson, 2005; Williams, 2010). While the above studies have certain advantages (e.g., removing the anticipation of a memory test), this design severely limits the number of test trials and examines memory conditions between-subjects. In contrast, our RSVP design generated a very large set of within-subject data which allowed us to perform detailed analysis of the categories of targets, probes and distractors plus account for individual differences in performance, confidence ratings, and metacognitive accuracy. Though subjects in our design could anticipate questions about probe items, they could not direct full attention to these items because we imposed a short time window for target detection and provided startling feedback if subjects made target detection errors. Further, we only analyzed incidental probe memory on trials where subjects correctly detected the target. Thus, our results can be taken to reveal the nature of incidental, short-term memory for items that are anticipated, at most, but not explicitly memorized. Future research employing our design might further reduce this sense of anticipation by only testing incidental memory on a limited subset of trials.

Shared semantic features affect trial-by-trial accuracy but not confidence

Semantic features related to the target of visual search are known to guide saccades (Pomplun, 2006; Shen, Reingold, Pomplun, & Williams, 2003; Wu, Wick, & Pomplun, 2014) and improve accuracy for the rapid categorization of natural scenes (Fabre-Thorpe, 2011). The influence of semantic features have also been found when eye movements are constrained; recognition is enhanced for objects that share categories with a target (Bower & Karlin, 1974; Williams, 2010). Given this, it was possible that searching for a specific target face in our incidental memory task might enhance the encoding of probe faces from similar categories. This is an important consideration because a corresponding effect on metacognitive accuracy raises a possibility that semantic association may also enhance sustained conscious access to incidental memory.

Our *item effect* category analysis supports the claim that searching for a specific target improves recognition. Probes or foils that shared categories with the target (e.g., gender or race) were selected as a response alternative more frequently (Fig. 5). However, this *item effect* had very little to no impact on trial-by-trial confidence ratings (Fig. 6). Were metacognitive judgments supported by similarity with a target, we would expect confidence to be enhanced in the case of similarity and diminished in the case of dissimilarity. With respect to the cognitive structure of incidental memory and its relationship to consciousness, this result implies that it may not be necessary to consciously reflect on certain stimulus features for them to be encoded, processed, and then accurately retrieved by declarative conscious access. This position is supported by recent evidence that metacognitive conscious access is possible for certain complex stimuli with little or even no selective attention (Matthews et al., 2018). Thus, we conclude that sustained conscious access to incidental memory is not related to those mechanisms that drive semantic association and likely arises as a natural product of perception.

Relatedly, it is possible that trial sequences dominated by a particular category of face may reinforce the encoding of probes from that category. This point is informed by behavioral and neuroscientific studies that demonstrate enhanced processing of target faces if they are preceded by semantically related faces; an effect tied to associative priming (Schweinberger, Pfütze, & Sommer, 1995; Vladeanu, Lewis, & Ellis, 2006). Our *sequence effect* category analysis revealed that the categories of faces presented in a trial does influence performance. Specifically, trial-by-trial correctness is affected by how surprising a probe or foil is relative to the sequence of faces that precede it; correctness is lower for surprising probes and higher when the foil is surprising (Fig. 7).

This result supports the claim that trial sequences dominated by particular categories can associatively prime incidental memory for items of a related category. However, this *sequence effect* had very little impact on trial-by-trial confidence ratings (Fig. 8). Only in the case of incorrect judgments (i.e., selecting a foil) were subjects' confidence ratings lower as a function of surprise (e.g., a male foil preceded by a sequence of predominantly female faces; see Fig. 8b,d). Were metacognitive judgments biased by associative priming from trial sequences, we would expect the effects observed for trial-by-trial correctness to correspond with confidence ratings but this was not the case. Thus, we conclude that shared semantic features in a trial sequence have little, if any, influence on sustained conscious access to faces.

Conclusion

Using RSVP we reveal that incidental memory for upright and inverted faces is sustained and consciously accessible independent from self-directed visual search, explicit memorization, or the influences of semantic priming from targets and distractors.

It is appealing to think that our subjective efforts to attend and memorize the world are an important feature in us remembering it. Indeed, some researchers have suggested that short-term memory is strongly dependent or even subsumed by attentional mechanisms (Awh & Jonides, 2001; Awh et al., 2006; Gazzaley & Nobre, 2012). However, our findings build on evidence from visual long-term memory (Varakin et al., 2012; Varakin & Hale, 2014) that the mere act of perception may be enough to form sustained and consciously accessible memories with no or little attentional amplification (Block, 2011; Lamme, 2016). Attentional amplification might prove important for the massive capacities of visual long-term memory (Brady et al., 2008; Konkle et al., 2010; Standing, 1973) but in elucidating the mechanics of conscious experience, future research might be best served prioritizing the remarkable properties of perception.

References

Awh, E., & Jonides, J. (2001). Overlapping mechanisms of attention and spatial working memory. *Trends in Cognitive Sciences*, *5*(3), 119–126.

Awh, E., Vogel, E. K., & Oh, S.-H. (2006). Interactions between attention and working memory. *Neuroscience, 139*(1), 201–208.

Baars, B. J., & Franklin, S. (2003). How conscious experience and working memory interact. *Trends in Cognitive Sciences*, *7*(4), 166–172.

Baddeley, A. (2003). Working memory: Looking back and looking forward. *Nature Reviews. Neuroscience*, *4*(10), 829–839.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3). https://doi.org/10.1016/j.jml.2012.11.001

Barton, J. J. S., Radcliffe, N., Cherkasova, M. V., Edelman, J., & Intriligator, J. M. (2006). Information processing during face recognition: The effects of familiarity, inversion, and morphing on scanning fixations. *Perception, 35*(8), 1089–1105.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*(1). https://doi.org/10.18637/jss.v067.i01

Beck, M. R., Peterson, M. S., Boot, W. R., Vomela, M., & Kramer, A. F. (2006). Explicit memory for rejected distractors during visual search. *Visual Cognition, 14*(2), 150–174.

Bird, J. E. (1976). Effects of intentional and incidental instructions on picture recognition. *Perceptual and Motor Skills, 42*(2), 555–561.

Block, N. (2011). Perceptual consciousness overflows cognitive access. *Trends in Cognitive Sciences, 15*(12), 567–575.

Bower, G. H., & Karlin, M. B. (1974). Depth of processing pictures of faces and recognition memory. *Journal of Experimental Psychology, 103*(4), 751–757.

Brady, T. F., Konkle, T., Alvarez, G. A., & Oliva, A. (2008). Visual long-term memory has a massive storage capacity for object details. *Proceedings of the National Academy of Sciences of the United States of America, 105*(38), 14325–14329.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10*(4), 433–436.

Castelhano, M., & Heaven, C. (2010). The relative contribution of scene context and target features to visual search in scenes. *Attention, Perception & Psychophysics, 72*(5), 1283–1297.

Castelhano, M., & Henderson, J. (2005). Incidental visual memory for objects in scenes. *Visual Cognition, 12*(6), 1017–1040.

Chun, M. M., & Jiang, Y. (1998). Contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology, 36*(1), 28–71.

Chun, M. M., & Jiang, Y. (2003). Implicit, long-term spatial contextual memory. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 29*(2), 224–234.

Cohen, G. (1973). How are pictures registered in memory? *The Quarterly Journal of Experimental Psychology, 25*(4), 557–564.

Fabre-Thorpe, M. (2011). The characteristics and limits of rapid visual categorization. *Frontiers in Psychology, 2,* 243.

Fleming, S. M., Weil, R. S., Nagy, Z., Dolan, R. J., & Rees, G. (2010). Relating introspective accuracy to individual differences in brain structure. *Science, 329*(5998), 1541–1543.

Freire, A., Lee, K., & Symons, L. A. (2000). The face-inversion effect as a deficit in the encoding of configural information: Direct evidence. *Perception, 29*(2), 159–170.

Gazzaley, A., & Nobre, A. C. (2012). Top-down modulation: Bridging selective attention and working memory. *Trends in Cognitive Sciences, 16*(2), 129–135.

Harris, I. M., Benito, C. T., & Dux, P. E. (2010). Priming from distractors in rapid serial visual presentation is modulated by image properties and attention. *Journal of Experimental Psychology. Human Perception and Performance, 36*(6), 1595–1608.

Henderson, J. (2005). Introduction to real-world scene perception. *Visual Cognition, 12*(6), 849–851.

Hills, P. J., Cooper, R. E., & Pake, J. M. (2013). First fixations in face processing: The more diagnostic they are the smaller the face-inversion effect. *Acta Psychologica, 142*(2), 211–219.

Hills, P. J., Sullivan, A. J., & Pake, J. M. (2012). Aberrant first fixations when looking at inverted faces in various poses: The result of the centre-of-gravity effect? *British Journal of Psychology , 103*(4), 520–538.

Hollingworth, A., & Henderson, J. M. (2002). Accurate visual memory for previously attended objects in natural scenes. *Journal of Experimental Psychology. Human Perception and Performance, 28*(1), 113–136.

Kaunitz, L. N., Rowe, E. G., & Tsuchiya, N. (2016). Large capacity of conscious access for incidental memories in natural scenes. *Psychological Science, 27*(9), 1266–1277.

Koehler, K., & Eckstein, M. P. (2015). Scene inversion slows the rejection of false positives through saccade exploration during search. In D. C. Noelle, R. Dale, A. S. Warlaumont, J. Yoshimi, T. Matlock, C. D. Jennings, & P. P. Maglio (Eds.), *Proceedings of the 37th Annual Meeting of the Cognitive Science Society* (pp. 1141–1146). Austin, TX: Cognitive Science Society.

Konkle, T., Brady, T. F., Alvarez, G. A., & Oliva, A. (2010). Scene memory is more detailed than you think: The role of categories in visual long-term memory. *Psychological Science, 21*(11), 1551–1556.

Lamme, V. (2016). The crack of dawn: Perceptual functions and neural mechanisms that mark the transition from unconscious processing to conscious vision. In T. Metzinger & J. W. Windt (Eds.), *Open MIND*. Frankfurt am Main: MIND Group.

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature, 390*(6657), 279–281.

Macmillan, N. A., & Creelman, C. D. (2004). *Detection theory: A user's guide*. New York: Psychology Press.

Maki, W. S., Frigen, K., & Paulson, K. (1997). Associative priming by targets and distractors during rapid serial visual presentation: Does word meaning survive the attentional blink? *Journal of Experimental Psychology. Human Perception and Performance, 23*(4), 1014–1034.

Matthews, J., Schröder, P., Kaunitz, L., van Boxtel, J. J. A., & Tsuchiya, N. (2018). Conscious access in the near absence of attention: Critical extensions on the dual-task paradigm. *Philosophical Transactions of* *the Royal Society of London. Series B, Biological Sciences, 373*(1755). https://doi.org/10.1098/rstb.2017.0352

Maxcey, A. M., & Woodman, G. F. (2014). Forgetting induced by recognition of visual images. *Visual Cognition, 22*(6), 789–808.

Mohan, K., & Arun, S. P. (2012). Similarity relations in visual search predict rapid visual categorization. *Journal of Vision, 12*(11), 19.

Nelson, T. O. (1996). Consciousness and metacognition. *The American Psychologist, 51*(2), 102.

Oliva, A., & Torralba, A. (2006). Building the gist of a scene: The role of global image features in recognition. *Progress in Brain Research, 155,* 23–36.

Pomplun, M. (2006). Saccadic selectivity in complex visual search displays. *Vision Research, 46*(12), 1886–1900.

Potter, M. C. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology: Human Learning and Memory*, *2*(5), 509–522.

Rugo, K. F., Tamler, K. N., Woodman, G. F., & Maxcey, A. M. (2017). Recognition-induced forgetting of faces in visual long-term memory. *Attention, Perception & Psychophysics, 79*(7), 1878–1885.

Schweinberger, S. R., Pfütze, E.-M., & Sommer, W. (1995). Repetition priming and associative priming of face recognition: Evidence from event-related potentials. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 21*(3), 722–736.

Shen, J., Reingold, E. M., Pomplun, M., & Williams, D. E. (2003). Saccadic selectivity during visual search. In J. Hyönä, R. Radach, & H. Deubel (Eds.), *The mind's eye: Cognitive and applied aspects of eye movement research* (pp. 65–88). Amsterdam: Elsevier Science Publishers.

Silverman, A., & Cason, H. (1934). Incidental memory for pleasant, unpleasant, and indifferent words. *The American Journal of Psychology, 46*(2), 315–320.

Simons, D. J., & Levin, D. T. (1997). Change blindness. *Trends in Cognitive Sciences, 1*(7), 261–267.

Song, C., Kanai, R., Fleming, S. M., Weil, R. S., Schwarzkopf, D. S., & Rees, G. (2011). Relating inter-individual differences in metacognitive performance on different perceptual tasks. *Consciousness and Cognition, 20*(4), 1787–1792.

Spence, R., & Witkowski, M. (2013). *Rapid serial visual presentation: Design for cognition*. Springer.

Standing, L. (1973). Learning 10000 pictures. *The Quarterly Journal of Experimental Psychology, 25*(2), 207–222.

Standing, L., Conezio, J., & Haber, R. N. (1970). Perception and memory for pictures: Single-trial learning of 2500 visual stimuli. *Psychonomic Science, 19*(2), 73–74.

Theeuwes, J., Belopolsky, A., & Olivers, C. N. L. (2009). Interactions between working memory, attention and eye movements. *Acta Psychologica, 132*(2), 106–114.

Vandenbroucke, A. R. E., Sligte, I. G., Barrett, A. B., Seth, A. K., Fahrenfort, J. J., & Lamme, V. A. F. (2014). Accurate metacognition for visual sensory memory representations. *Psychological Science, 25*(4), 861–873.

Varakin, D. A., Frye, K. M., & Mayfield, B. (2012). Intentional memory instructions do not improve visual memory. *International Journal of Brain and Cognitive Sciences*, *1*(3), 18–25.

Varakin, D. A., & Hale, J. (2014). Intentional memory instructions direct attention but do not enhance visual memory. *SAGE Open*, *4*(4), 2158244014553588.

Vladeanu, M., Lewis, M., & Ellis, H. (2006). Associative priming in faces: Semantic relatedness or simple co-occurrence? *Memory & Cognition, 34*(5), 1091–1101.

Williams, C. C. (2010). Incidental and intentional visual memory: What memories are and are not affected by encoding tasks? *Visual Cognition, 18*(9), 1348–1367.

Williams, C. C., Henderson, J. M., & Zacks, R. T. (2005). Incidental visual memory for targets and distractors in visual search. *Perception & Psychophysics, 67*(5), 816–827.

Woodman, G. F., & Luck, S. J. (2004). Visual search is slowed when visuospatial working memory is occupied. *Psychonomic Bulletin & Review, 11*(2), 269–274.

Wu, C.-C., Wick, F. A., & Pomplun, M. (2014). Guidance of visual attention by semantic information in real-world scenes. *Frontiers in Psychology*, *5*, 54.

Linking text between chapter 3 and 4

Chapter 2 and 3 explored the link between metacognition, conscious access, selective attention, and short-term memory. Results from these chapters suggest that metacognition might be dissociable from the core cognitive functions that support conscious access and perception more broadly. That is, metacognition is possible for decisions involving stimuli that are perceived with little or no attentional amplification and that are not actively maintained in memory. I will expand upon these points in my concluding remarks.

My final chapter turns to metacognition and cognitive function in abnormal brains (Matthews, Nagao, et al., 2018). Specifically, we examined perception in functional and organic motor disorders. This seems like an irrational subject of inquiry but a tight link between perception and action is proposed by the predictive processing framework (Clark, 2013, 2016; Friston, 2010; Friston, Daunizeau, Kilner, & Kiebel, 2010; Hohwy, 2013, 2016; Keller & Mrsic-Flogel, 2018). Predictive processing is an increasingly influential theory of brain function that offers persuasive descriptions of the symptoms that characterise many neurological disorders including psychosis and autism (Friston et al., 2014; Horga et al., 2014; Palmer et al., 2017; Powers et al., 2017; Van de Cruys et al., 2014) but also functional motor and sensory disorders (Edwards et al., 2012).

We drew upon predictive processing and contemporary visual psychophysics to contrast 4 cognitive domains implicated in the pathogenesis of FMD—attention, expectations, metacognition, and perceptual sensitivity. We achieved this using a dual-task paradigm augmented with several features pioneered in Chapter 2 (Matthews, Schröder, et al., 2018; Sherman et al., 2015). Our extended dual-task paradigm allowed us to contrast the 4 domains within one experimental design. We tested objective performance and metacognition in patients with functional motor disorders, healthy age-matched controls, and patients with phenotypically-matched organic motor disorders.

We found that attention, expectations, and metacognition function normally in FMD when compared to healthy controls. However, FMD patients had significant impairments in perceptual sensitivity to visual contrast. The same degree of sensory impairment was identified in organic patients but the organic group exhibited differences in attention and expectations that were not found in FMD and controls. We concluded that the distinctive behavioural profile of FMD may arise from an impairment in perceptual sensitivity but attentional, expectational and metacognitive processes remain intact.

The results from Chapter 4 imply that the mechanisms that support metacognition are likely to be distinct from the processes that underlie selective attention and expectations but might be related to those processes involved in perception. I reach this conclusion because metacognitive sensitivity and efficiency was broadly equivalent between the groups once underlying differences in perceptual sensitivity were accounted for. Also, we observed sharp differences in metacognitive sensitivity and efficiency that were contingent on the type of perceptual report (i.e., judging whether the stimulus was present or absent). Specifically, we found decisive evidence that metacognition was best for judgments of stimulus presence. This finding supports the notion that the evidence that drives high metacognitive performance (i.e., conscious knowledge) derives at least in part from the contents of perception (i.e., conscious perception).

Chapter 4

Impaired perceptual sensitivity with intact attention and metacognition in functional motor disorder

Julian Matthews[1,2], Kanae Nagao[3], Catherine Ding[3], Rachel Newby[3], Peter Kempster[3,4], Jakob Hohwy[1]

1. Cognition and Philosophy Lab, Monash University, Clayton, Australia

2. School of Psychological Sciences, Monash University, Clayton, Australia

3. Neurosciences Department, Monash Medical Centre, Clayton, Australia

4. Department of Medicine, Monash University, Clayton, Australia

Abstract

Functional motor disorders (FMDs) are distinguished by signs that lack congruence with recognised patterns of organic disease and show inconsistency over time. Their pathophysiology is poorly understood, but there is evidence that irregularities in perceptual and cognitive processing lie at the heart of these conditions.

Here, we draw on a predictive processing account of functional neurological disorders to study perceptual decision-making in three groups: 20 patients with FMDs (14 with functional movements and 6 with functional weakness), 20 with phenotypically-matched organic motor disorders, and 20 age-matched healthy controls. We examine four cognitive domains with putative roles in FMD pathogenesis: attention, expectations, sensory processing (perceptual sensitivity), and metacognition (introspective evaluation of performance). We augmented a dual-task visual decision-making paradigm to examine each of these domains within a single experimental design and employed Bayesian statistics to test the strength of evidence for each hypothesis.

With sensory input (stimulus contrast) psychometrically adjusted to threshold performance at a fixed level for all groups, the FMD group exhibited statistically equivalent attentional, expectational and metacognitive processing to healthy controls. However, FMD subjects required higher contrast strength to reach these performance thresholds. This was statistically equivalent to the contrast strength required by the organic group, and could not be accounted for by medication use or comorbid psychopathology. Those with organic motor disorders showed differences in processing of attention and expectations for perceptual decisions that were not observed in either healthy controls or the functional group.

The distinctive behavioural profile of FMDs may arise from abnormalities in basic sensory processing, while higher attentional, expectational and metacognitive mechanisms remain intact. Conceptualising functional neurological disorders under the predictive processing account may consolidate and refine existing pathophysiological theories about them.

Introduction

Functional motor disorders (FMDs) are common neurological conditions (Stone et al., 2010) with distinctive phenomenological features that separate them from their organic counterparts. A poverty or excess of motor output—manifesting as negative (functional weakness), or positive (such as functional tremor, dystonia or chorea) symptoms—disrupts goal-directed actions, leaving more reflexive movements intact. Levels of disability and psychological distress reported in FMDs are typically equivalent to, or exceed, those observed in comparable organic neurological diseases (K. E. Anderson et al., 2007).

Functional movements possess an apparent voluntariness. They frequently require attention to manifest and demonstrate distractibility—attenuation or extinction when attentional resources are directed elsewhere. Functional tremor may display entrainability (a tendency to synchronise with rhythmic voluntary movements). Functional myoclonus is preceded by a cortical Bereitschaftspotential, or 'readiness potential', an electroencephalographic signature of the preparation of voluntary movement. But these findings conflict with reports from FMD patients, who describe their motor symptoms as involuntary. A fundamental loss of agency (the sense of ownership over one's actions) is thus implied.

Psychological theories of FMD seek to resolve this conflict by drawing an arbitrary line between conscious and unconscious thought, with abnormal motor output formulated as a product of psychological transactions within the subconscious mind. The inability of these theories satisfactorily to explain FMDs has motivated a shift towards more nuanced biopsychosocial frameworks. One such approach, the *predictive processing* model of brain function, helps to unravel questions about the interface between voluntary and involuntary movement.

This paper adopts the assumption that a predictive processing account of brain function holds the key to a better understanding of functional neurological disorders (Edwards et al., 2012). We use contemporary psychophysics to analyse four cognitive domains implicated in the predictive processing model—*attention*, *expectations*, *sensory perception* (perceptual sensitivity), and *metacognition* (introspective access to decision-making performance).

The predictive processing account takes as its starting point Bayes' rule for the conditional probability of events, where expectations—*predictive beliefs*, or *prior probabilities*—are updated in response to new evidence. In the brain, Bayesian inference is approximated by a continuous process of prediction error minimisation between anticipated and recorded sensory input, across all levels of the cortical hierarchy (Clark, 2013, 2016; Friston, 2010; Hohwy, 2013). Perception and cognition are then processes for inferring the causes of sensory information received from the external world—*perceptual inference*, in shorthand. Prediction error minimisation can in addition be achieved by changing sensation through action to make it fit with existing expectations—*active inference*. Perception and action therefore have a common basis in the sense that both serve to minimise prediction error.

Attentional processes play a key role in maintaining an optimal balance between prior beliefs and input for both perceptual and active inference. Each prediction error signal is afforded a certain precision weighting. Those with higher expected precision have greater modulatory access to prior probabilities encoded at higher levels—they can drive associative learning at a higher rate. A salient environmental signal will attract more attentional resources (it will receive greater precision up-weighting) and thus have a greater capacity to modify predictive beliefs. Attention—the selective focus of the mind on single items to the exclusion of others—is thus cast in Bayesian terms as the brain's optimisation of expected precisions. When functioning effectively, attentional processes filter sensory input so that the most reliable and relevant data has the greatest capacity to refine predictive beliefs. Similarly, in active inference, prediction error that is expected to be precise will have a greater propensity for eliciting action. Bayesian belief, updated through precision-weighting, is then the brain's store of prior probabilities, its expectations. Only a subset enter conscious awareness.

One attempt to explain FMDs under predictive processing describes how functional symptoms might arise from the combination of abnormal attention (precision optimisation), and faulty predictive beliefs (prior probabilities) (Edwards et al., 2012). Cognitive representations of simple motor behaviours (e.g., tremor, dystonic contraction) are encoded in intermediate levels of the sensorimotor hierarchy. In patients with FMD, abnormal attention is directed at representations in this level of the hierarchy. These representations are then afforded abnormally high precision relative to sensory input. To account for this imbalance, prediction error is minimised through active inference—autonomous neural activity that culminates in the activation of spinal reflex arcs and produces the very motor behaviour associated with those representations (Newby et al., 2016). However, because this activity was not predicted by the highest (personal) levels of the cortical hierarchy, this results in a secondary, faulty belief at the high level—misattributing the cause of the behaviour to some pathology. Sense of agency is then diminished and behaviours typically generated in a voluntary way are perceived as involuntary.

An alternative perspective that draws on the same conceptual framework comes from Stenner and Haggard (2016). Their proposal accommodates the fluctuating symptomology of FMDs over time. In this version, the brain predicts a higher level of conscious access to the motoric detail (second-by-second proprioceptive changes) of movement than it is calibrated to deliver. When this expectation is not met, attentional resources are directed towards channels carrying proprioceptive information—that is, the precision of motor signals is boosted. Consequently, sensorimotor noise that would ordinarily be filtered out is instead misperceived as abnormal movement. Through active inference, this motor 'hallucination' (anomalous motor expectation) is then translated into real action. Since sensorimotor noise is inherently variable, it aligns with different (lay) beliefs about illness at different times and produces both positive and negative motor symptoms. This conclusion is supported by clinical evidence. Weakness and hyperdynamic movements frequently coexist in FMDs (Factor, Podskalny, & Molho, 1995) and have been described as two sides of the same coin, both signifying a loss of normal voluntary motor control (Janet, 1901).

Individual components of the predictive model of FMDs have been explored experimentally. Disruption of normal attentional processing is suggested by the finding that motor performance in FMDs is impaired in tasks involving explicit cueing, whereas responses to implicit visuomotor cues are intact (Pareés et al., 2013). Another study (McIntosh, McWhirter, Ludwig, Carson, & Stone, 2017) examined exogenous and endogenous attentional cueing in unilateral functional weakness, and found that responses to cues were broadly intact for detection of visual stimuli but a selective impairment was noted for tactile stimuli on the functionally weak side. This was interpreted as a consequence of allocating attention in the context of faulty predictive beliefs (i.e. the expectation of motor weakness and numbness), rather than an impairment of attentional processing *per se*. Certain cognitive biases in FMD may predispose sufferers towards the formation of aberrant motor beliefs. These include 'jumping to conclusions' when decision-making in probabilistic learning tasks (with a tendency to be swayed by disconfirmatory evidence) (Pareés et al., 2012), and selective motor impulsivity that manifests in errors of commission in go/no-go paradigms (Voon et al., 2013).

There is also evidence that processing of sensory inputs is altered in FMDs, and this may be shaping the elusive subjectivity that attends functional movements or weakness. Sensory attenuation, the ability to selectively filter out (down-weight) the sensory consequences of one's own actions, is impaired in FMDs (Pareés et al., 2014). Accordingly, strong attentional focus on motor feedback may occur at the expense of processing power for other sensory modalities. For instance, reduced perceptual sensitivity in FMD has been shown for both interoceptive (impaired heartbeat detection) (Ricciardi et al., 2016) and exteroceptive (higher temporal discrimination thresholds for tactile stimuli) inputs (Morgante et al., 2011).

Alternatively, the persistence of these counterproductive attentional biases could be explained in terms of metacognitive deficits—reduced awareness of how precision is assigned to incoming sensory streams. Metacognitive impairment is suggested when actigraphic records are compared with self-reported duration of tremor. While patients with organic motor disorders tend to overestimate the frequency of their movement symptoms, this mismatch is much greater in functional tremor (Pareés et al., 2011). Morgante et al. (2018) observed normal pain thresholds and an increase in pain tolerance in subjects with functional dystonia. Pain percepts may arise autonomously (analogous to 'action possibilities' in the FMD model described above) as a result of impaired introspective evaluation of sensory information, a metacognitive failure.

The deficits in *attention*, *expectations*, *sensation*, and *metacognition* implicated in these studies can all be construed within the predictive processing framework. But these processes overlap and show inter-dependence, making it difficult to attribute primacy to attentional, expectational, sensory or metacognitive abnormalities when interpreting an experimental study of FMDs. Expectation and attention interact (Summerfield & Egner, 2009), and should be manipulated orthogonally in the same paradigm. Moreover, sensory input must be controlled to avoid confounding expectational, attentional, or metacognitive processes. Likewise, metacognition is a critical consideration because effects might otherwise be explained by differences in participants' decision criteria (i.e., their willingness or unwillingness to report a perceptual experience (Barrett et al., 2013; Fleming & Lau, 2014)). To overcome these experimental challenges, we employed an extended version of the psychophysical dual-task paradigm (Matthews, Schröder, et al., 2018; Sherman et al., 2015) to study all four processes at once in FMDs. In addition to healthy controls, we selected a comparison group with matched organic motor disorders to control for the effects of motor disability alone.

In the visual dual-task paradigm, participants direct their attention to a central stimulus whilst a secondary stimulus is briefly presented in the periphery. Attention is manipulated by comparing a condition where only the peripheral stimulus is task-relevant (*full attention*) against a condition where both the central and peripheral stimuli are task-relevant (*diverted attention*). Expectations are manipulated by varying the likelihood that the peripheral stimulus will appear. Sensory processing is accounted for by using psychometric staircasing to adjust the visual contrast strength of stimuli in each attention condition to achieve a predefined performance threshold. Consequently, the effect of attention on perception is validated by examining to what extent stimulus contrast must increase to compensate for diversion of attention between conditions. Moreover, by balancing task difficulty, this staircasing procedure accounts for differences in each participant's ability and minimises ceiling and floor effects in objective performance. This allows us to gauge the full extent of attentional and/or expectational effects on behaviour.

Finally, metacognitive processing is examined by measuring metacognitive sensitivity. This approach instructs participants both to register a perceptual decision and to rate their confidence in that decision on each trial. By computing the correspondence between objective task performance and subjective confidence ratings, metacognitive sensitivity (metacognition henceforth), should provide a bias-free estimate of the degree of conscious insight a participant has into their decision-making (cf. type 1 versus type 2 signal detection theory) (Fleming & Lau, 2014; Maniscalco & Lau, 2012).

On the basis of the predictive processing account of FMD, we hypothesise that the relationship between attention, expectations, metacognition, and sensory processing will differ systematically in a population with FMDs when compared with healthy controls. We pose conditional hypotheses for each cognitive domain—1) if the basis of FMD is attentional, then the FMD group may be characterised by impairments in attention allocation and consequently require higher contrast to reach performance thresholds in the diverted attention condition; 2) if the basis is expectational, then the effects of target expectations on measures of performance and metacognition will be exaggerated in FMD; 3) if metacognitive, then we will observe an overall dampening in type 2 measures but no change in type 1 measures or overall confidence; 4) if sensory, then the FMD group will require higher contrast to reach performance thresholds in both attention conditions but will exhibit no significant differences in other measures. We find support for hypothesis four.

Materials and methods

Participants

Twenty patients with FMDs were recruited. Fourteen had functional movement disorders (9 with functional tremor, 5 with functional dystonia) and 6 had functional weakness. All fulfilled the Diagnostic and Statistical Manual V (DSM-V) criteria for Functional Neurological Symptom Disorder. Twenty healthy subjects, matched for age and education with the FMD patients, composed one control group. An additional 20 subjects with phenotypically matched organic motor disorders were recruited to control for the presence of motor symptoms. This group comprised patients with benign essential tremor, adult-onset focal dystonia and neuromuscular weakness (chosen because these diagnostic categories are not associated with macroscopic, microscopic or molecular pathological changes in the central nervous system). Potential participants were excluded if they had significant cognitive or visual impairment. Sample size was derived from previous research employing the dual-task paradigm and studying functional motor disorders (McIntosh et al., 2017; Pareés et al., 2014; Sherman et al., 2015). Participants were recruited at Monash Medical Centre, Melbourne or from the Australian Dystonia Support Group. The study was approved by the Monash Health Human Research Ethics Committee and all subjects gave written, informed consent. See **Table 1** and **Table 2** for participant demographics and characteristics.

Table 1. Participant demographics

	Healthy control (n=20)	Organic motor disorder (n=20)	Functional motor disorder (n=20)
Female sex (%)	14 (70)	8 (40)	13 (65)
Age - yrs mean ± SD (range)	41.7 ± 16.9 (21-68)	63.7 ± 10.9 (34-82)	45.6 ± 17.7 (20-69)
Ethnicity %Caucasian	50	85	90
Education Year 12 and above	17	11	12
Occupation Retired Unemployed Employed/student	 3 0 17	 13 2 5	 2 9 9
Family history of neurological disorder	3	8	9
Psychiatric disorder	0	10	16
Current psychoactive medication use	1	10	15
Phenomenology Tremor Dystonia Weakness	 N/A N/A N/A	 9 5 6	 9 5 6
Median Duration of motor symptoms (yrs)	N/A	6.5	1.6

HADS	7 (SD 5.1)	13 (SD 5.6)	17 (SD 9.9)
MoCA	28 (SD 1.6)	26 (SD 2.8)	26 (SD 2.5)
PDI	14 (SD 16.6)	25 (SD 28.5)	31 (SD 29)

HADS - Hospital Anxiety and Depression Scale

MoCA - Montreal Cognitive Assessment

PDI - Peter’s Delusion Inventory

Table 2. Characteristics of FMD cohort

Age and gender	Dominant functional motor disorder	Onset and time course	Other functional neurological symptoms	Symptom duration (yrs)	Psychoactive medications
64 M	Upper limb tremor	Gradual/ progressive	N/A	50	SN
27 M	Dystonia of neck, trunk	Acute/ fluctuating	Tremor	6	TE
54 F	Right limb weakness	Acute/ persistent	Visual, speech, sensory	31	SN
62 F	Head, trunk and limb tremor	Acute/ persistent	Speech, cognitive, balance, pain	3	TR, AE
69 F	Dystonia of left limbs	Acute/ paroxysmal	Speech, cognitive, pain	5	TR, AE
66 F	Hand dystonia	Acute/ improving	Visual, cognitive, tremor	1.3	SS
20 F	Limb tremor	Acute/ paroxysmal	Speech, disequilibrium	0.8	nil
24 F	Leg weakness	Gradual/relaps ing	Speech, sensory	0.1	SS
20 F	Distal leg weakness	Acute/ paroxysmal	Sensory	0.3	SS
32 M	Bilateral limb tremor	Acute/ fluctuating	Sensory, speech, cognitive, pain	0.03	B
35 F	Bilateral leg weakness	Acute/ fluctuating	Speech, cognitive, sensory, pain	1.5	SN, B, AE, Ba
62 M	Limbs, truncal tremor	Acute/ paroxysmal	Speech, cognitive, balance, pain	0.1	nil
65 F	Upper limb and head tremor	Gradual onset Progressive	Swallowing, balance, weakness	1.6	nil

43 F	Quadriparesis	Acute/ paroxysmal	Mutism, sensory	1.6	B
25 F	Facial and limb dystonia	Acute/ paroxysmal	Speech, tremor	1.7	SS, AE
25 M	Left leg weakness	Acute/ persistent	Cognitive, sensory, pain	2.5	SN, AE, O, THC
48 M	Head, trunk, and limb tremor	Gradual/ relapsing	Speech, cognitive, weakness, balance, pain	1.0	SN, AE, TR
61 F	Facial and limb dystonia	Acute/ paroxysmal	Speech	46	nil
52 M	Limb tremor	Gradual/ fluctuating	Speech, cognitive, swallowing, balance, weakness, jaw twitching	0.7	SS, O, THC, AP, AE
57 F	Vocal, head and limb tremor	Acute/ persistent	Balance	26	nil

SN - SNRI antidepressant; SS - SSRI antidepressant; TE - tetracyclic antidepressant; TR - tricyclic antidepressant: B - benzodiazepine; O - opiate; AP - atypical antipsychotic; AE - anti-epileptic, THC - cannabinoid; Ba - baclofen

Protocol

All subjects completed a Montreal Cognitive Assessment (MoCA), Hospital Anxiety and Depression Scale (HADS; (Zigmond & Snaith, 1983)), Peters' Delusions Inventory (E. Peters, Joseph, Day, & Garety, 2004) and the dual attention task (see **Design and procedure**).

Apparatus

The experiment was programmed and conducted using the psychophysics toolbox in MatLab. Stimuli were presented using a Dell XPS13 laptop connected to a 22 inch Dell E2216HV monitor (resolution 1920 x 1080 pixels) with refresh rate fixed at 60 Hz. Subjects were tested individually sitting approximately 60 cm from the screen.

Design and procedure

The experimental design was adapted from Sherman et al. (Sherman et al., 2015) incorporating elements from Matthews et al. (Matthews, Schröder, et al., 2018) (**Figure 1**). In the Gabor Task, subjects reported the presence or absence of a near-threshold gabor patch. In the Letter Task, subjects identified the presence or absence of a target (the uppercase character 'T') within a cluster of distractor letters (uppercase 'L').

Trials began with a black fixation cross (0.38° x 0.38° visual angle), presented at the centre of a grayscale screen for a random duration between 500-1500 ms. On gabor present trials, this was followed by presentation of a gabor patch (spatial frequency 2c/°, Gaussian SD=2°) in one of the four quadrants of the screen (approximately 8.5° x 7° eccentricity from fixation with a randomly generated jitter up to 1.24° x 0.66° visual angle). Contrast of the gabor was staircased for each subject using QUEST (Watson & Pelli, 1983) to achieve a discrimination threshold of 79.4% (see **QUEST staircasing**). Presentation interval for the gabor was fixed at 23 frames (approximately 383 ms) with a gradual onset and offset.

On all trials, presentation of the central stimulus followed immediately after the fixation cross (simultaneous with the gabor on Gabor Present trials). The central letter stimulus consisted of four white letters in Helvetica typeface, each 1.43° x 1.43° in size. Letters were arranged around the fixation point at 0°, 90°, 180°, and 270° with each character randomly rotated from 0° to 359°. On Target Present trials the stimulus consisted of 3 'L's and a 'T' whereas on Target Absent trials the stimulus was just letter 'L's. Target presence was counterbalanced such that the target appeared on 50% of trials within each block.

The presentation interval of the letter stimulus was staircased for each subject using QUEST to achieve a discrimination threshold of 79.4%. The letter stimulus was backward masked by four letter 'F's which remained on screen for 18 frames (300 ms). Following presentation of the trial stimuli, subjects registered their discrimination response and confidence (see **Supplementary Material**). Once response collection was complete subjects were presented with a final screen with the message 'click to continue' where they could take a short break if desired and prepare for the next trial.

Expectations and Attention manipulation

Expectations and attention were manipulated in the Gabor Task using parameters adapted from Sherman et al. (Sherman et al., 2015). Attention was manipulated over blocks of trials by altering the task relevance of the central letter stimulus. Since the appearance of the central letter stimulus was identical between attention conditions this meant that in the Full Attention condition subjects ignored the letter stimulus and responded to the Gabor Task only. In the Diverted Attention condition, subjects responded both to the Gabor Task and the Letter Task. Subjects were instructed to prioritise the Letter Task in the Diverted Attention condition. To reinforce this instruction, if performance on the Letter Task dropped below 60% within a block, subjects received feedback in the form of a full-screen flashing alert that cycled between yellow, red, and black at 2.5 Hz for approximately 3000 ms. This reminded subjects that the Letter task was a priority and discouraged subjects from devoting undue attention to the Gabor Task.

Expectations were manipulated within each block of trials by varying the probability of Gabor presence (25%, 50%, or 75% probability of presence). An expectation of Gabor presence was induced in the 75% condition since the Gabor was relatively likely to appear. The 50% presence condition served as a control and the 25% condition induced an expectation of Gabor absence. Before each block began, an on-screen prompt informed subjects of the Attention and Expectations condition in writing and with a visual cue. An additional written prompt regarding the expectations condition appeared after each trial. Subjects completed 36 blocks in total (6 of each of the 6 conditions counterbalanced between-subjects using a Latin Square design). Each block contained 12 trials resulting in a total of 432 experimental trials per subject.

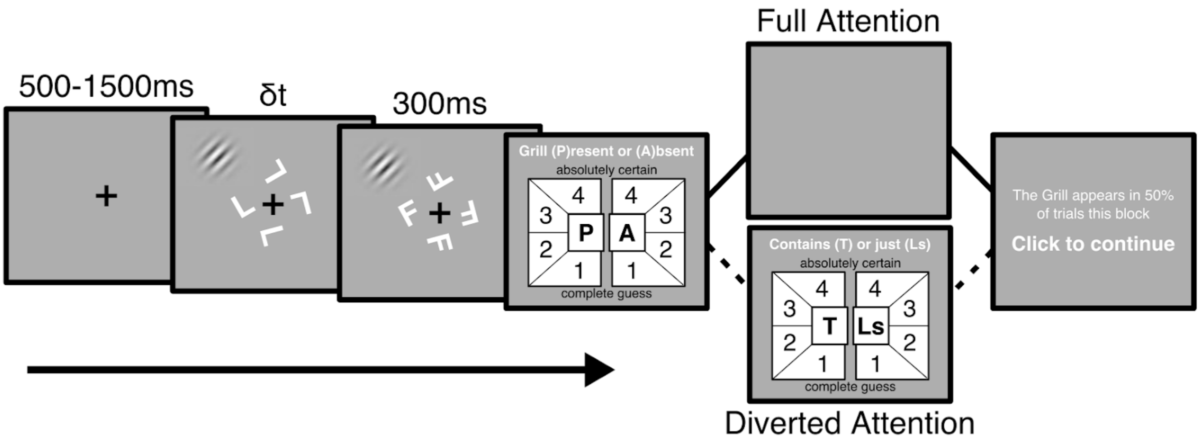

Figure 1. Task design

Method for letter and gabor task including 8AFC response to Gabor presence (named 'grill' in the experiment to aid description for subjects). To manipulate attention, participants performed the Gabor Task alone (Full Attention condition) or in conjunction with a Letter Task (Diverted Attention condition). To manipulate expectations the presence of the peripheral gabor was altered between blocks (25%, 50%, and 75% likelihood). Subjects were instructed before each block of trials and after each trial about the probability of gabor presence. To examine perceptual sensitivity (i.e., sensation) we measured the contrast strength of the peripheral gabor that was required to reach 79.4% performance thresholds in each attention condition (see **QUEST staircasing**). To examine metacognition we quantified the correspondence between trial-by-trial accuracy and confidence ratings. δt is stimulus-onset asynchrony (SOA), equivalent to the time between presentation of the letter stimulus and mask. In order to standardise the difficulty of the Letter Task, SOA timing was adjusted psychometrically for each participant during training (see **QUEST staircasing**).

QUEST staircasing

To examine the sensory input required to equate performance between subjects and attention conditions, stimulus alpha contrast was adjusted on a trial-by-trial basis using the Quick Estimate of Threshold (QUEST) adaptive staircase procedure (Watson & Pelli, 1983). We set the β parameter for QUEST to be 2 and the standard deviation to be 70%. On each trial during training we updated the peripheral alpha contrast such that discrimination performance was fixed at 79.4% correct (approximate type 1 d'=2). This was achieved separately for the full and diverted attention conditions. To ensure a tight control of performance throughout the experiment we continued to use QUEST during the 50% Expectations blocks.

In addition to the gabor task, we employed QUEST to staircase performance on the central letter task. The parameters listed above were employed to adjust stimulus onset asynchrony (SOA) to achieve 79.4% correct per block. This meant training in our experiments involved the same number of trials for each participant: 8 blocks of gabor discrimination under full attention, 8 blocks of the central letter task, and 8 blocks of peripheral gabor task with attention diverted to the central task; a total of 288 training trials.

Statistical analysis

Our analysis pipeline was adapted from Sherman et al. (Sherman et al., 2015) augmented with Bayesian statistics as well as between-subjects group effects and interactions. Objective performance was assessed using type 1 signal detection theoretic (Green & Swets, 1966) measures *detection sensitivity* (type 1 d') and *decision criterion* (type 1 c). A negative/positive c reflects bias towards reporting target presence/absence. Likewise, metacognitive performance was assessed using type 2 signal detection theoretic measure *metacognitive sensitivity* (type 2 area under the curve; AUC). Metacognitive sensitivity examines the correspondence between accuracy of judgments and confidence ratings (Fleming & Lau, 2014). Metacognitive hits are correct judgments with high confidence, metacognitive misses are correct judgments with low confidence. Conversely, false alarms are incorrect judgments with high confidence and correct rejections are incorrect judgments with low confidence. A type 2 receiver operating characteristic curve is then constructed by shifting the criterion between low and high confidence from a liberal criterion (low confidence for ratings of '1'; high confidence for ratings of '2','3','4') to a conservative criterion (low = ratings '1','2','3'; high = rating '4').

To assess the robustness of our metacognitive results, we used freely available MatLab code to measure response-conditional *meta-d'* (Barrett et al., 2013; Maniscalco & Lau, 2012). Meta-d' is computed using the same type 2 signal conventions above and reflects the expected type 1 detection sensitivity of an optimal observer given the observed confidence ratings. The ratio of meta-d'/d', termed *metacognitive efficiency*, reflects the degree of type 1 information subjects' use in their type 2 decision-making.

For statistical analyses, we report not only conventional p-values but also *Bayes factors* computed using JASP (JASP Team, 2018). In contrast to p-values, Bayes factors permit the quantification of evidence in favour of both the alternative and null hypotheses; important for clarifying group differences or lack thereof. For instance, when the Bayes factor $BF_{10}=9$, the observed data are nine times more likely to have occurred under the alternative hypothesis (H_1) than under the null hypothesis (H_0). Conversely, when $BF_{10}=1/4=.25$, the observed data are four times more likely to have occurred under the null than the alternative hypothesis. An important consideration when using Bayesian statistics is justifying the model of H_1 (i.e., what does our theory predict given the scientific context) (Dienes, 2014). There is no standardised approach (Aczel et al., 2018) however cases are made for a weakly informative default that approximates conventional frequentist decision thresholds (Gelman, 2006; Gelman, Jakulin, Pittau, & Su, 2008; Polson & Scott, 2012). Unless specified, a default Cauchy prior of .707 was used for all tests (Rouder, Speckman, Sun, Morey, & Iverson, 2009). Calculations were performed using a fixed effect size of Cohen's d = 0.7. This approach to specifying the model of H_1 favours the null hypothesis when compared to other techniques (Dienes, 2008b) but we deemed it appropriate given the exploratory nature of investigating perception in FMD.

A counterintuitive scenario that can occur when using both Bayesian and frequentist approaches is a test giving conflicting results for certain prior distributions. This scenario has become known as *Lindley's Paradox* and continues to be a topic of discussion in philosophy of statistics (Cousins, 2017; Jeffreys, 1939; Lindley, 1957; Nickerson, 2000). Frequentist p-values measure the probability of obtaining a test statistic as large as the one obtained conditional on the null hypothesis being true and assuming a sufficiently powered sample. A p-value does not provide evidence for or against the null hypothesis per se nor does it provide information about the probability of obtaining the test statistic conditional on the alternative hypothesis being true. In contrast, Bayes factors provide a likelihood ratio between the alternative and null hypothesis and reflect the evidence for each hypothesis versus the other. Arguments for both statistical approaches can be made but a major advantage of the Bayesian approach is it constitutes a formal procedure for accumulating evidence of an effect across studies over time (Nickerson, 2000).

Data availability

We have made core experiment and analytical code used in our study publicly available (github.com/julian-matthews/fmd-public-repository). This repository also contains an archive of all data used for our analyses and conclusions including all 43,200 behavioural trials and summary statistics following signal detection theoretic processing. Identifying information has been removed for confidentiality purposes.

Results

Lower contrast sensitivity in FMD and organic patients when compared to healthy controls

The relationship between attention and group on gabor contrast threshold was examined using a Bayesian mixed ANOVA. Although assumptions of homoscedasticity were not significantly violated, we observed the same pattern of results when this analysis was repeated using log-transformed contrast thresholds. We report untransformed contrast analysis to aid interpretation. A substantial main effect of attention was observed such that higher gabor contrast was required to reach threshold performance when attention was diverted ($BF_{10}=1.2 \times 10^8$, $F(1,57)=67.8$, $p<.001$, $p\eta^2=.54$). This main effect was found in all groups which indicates our dual-task paradigm successfully diverted attention. The effect of attention interacted with group ($BF_{10}=23.2$, $F(2,57)=5.9$, $p=.005$, $p\eta^2=.17$) such that substantially greater contrast was required to maintain performance under diverted attention for the organic group ($BF_{10}=1.4 \times 10^4$, $t(19)=7.0$, $p<.001$, 95% CI [.103 .192]) compared to FMD ($BF_{10}=41.8$, $t(19)=4.0$, $p<.001$, 95% CI [.044 .142]) and healthy controls ($BF_{10}=10.0$, $t(19)=3.2$, $p=.005$, 95% CI [.017 .082]) (**Figure 2a**).

We also observed a significant main effect of group on contrast thresholds ($BF_{10}=16.4$, $F(2,57)=4.0$, $p=.023$, $p\eta^2=.12$). When compared to healthy controls, Holm-corrected post hoc comparisons revealed that overall gabor contrast thresholds were higher in FMD ($BF_{10}=8.1$, $p_{holm}=.066$) and organic motor disorder ($BF_{10}=138.6$, $p_{holm}=.030$) but did not differ between the patient groups ($BF_{10}<.33$, $p_{holm}>.25$) (posterior odds were corrected for multiple comparisons by fixing the prior probability that the null hypothesis holds to .5 across all comparisons (Westfall, Johnson, & Utts, 1997)) (**Figure 2b**). This pattern of results was also found in supplementary analysis that accounted for the effects of medication use and comorbid psychopathologies (see **Supplementary Material**).

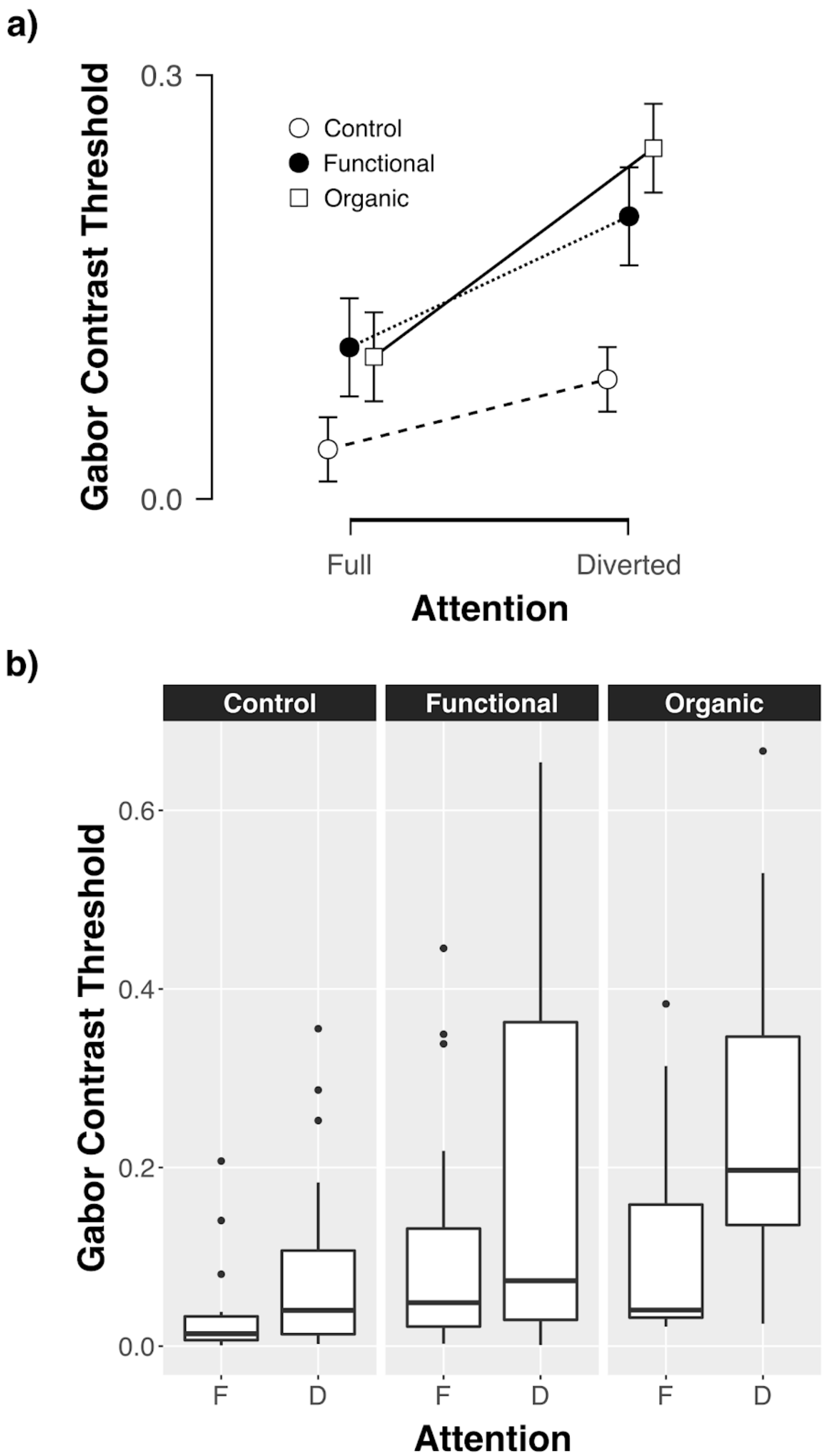

Figure 2. Perceptual sensitivity

a) Mean gabor contrast required for 79.4% performance thresholds in each attention condition and group. Error bars reflect 95% confidence interval. b) Boxplots of gabor contrast required for 79.4% performance thresholds in each attention condition and group. 'F' and 'D' refer to Full and Diverted attention respectively.

Detection sensitivity is equivalent for FMD and healthy controls

To examine detection sensitivity (type 1 d') on the gabor task we conducted a Bayesian mixed ANOVA on mean detection sensitivity relating the within-subjects factors attention and expectations between groups. We observed a main effect of expectations (BF_{10}=4.0, $F(2,114)=11.30$, $p<.001$, $p\eta^2=.17$) that did not interact with other effects (each $BF_{10}<.33$). Holm-corrected post hoc comparisons revealed significant differences ($BF_{10}>3.0$, $p_{holm}<.05$) between each level of expectation with definitive evidence that detection sensitivity in the low expectation conditions (M=2.13, SD=.70) was greater than the high expectation conditions (M=1.88, SD=.65) ($BF_{10}=1.4 \times 10^3$, $p_{holm}<.001$).

A main effect of attention was observed ($BF_{10}=25.5$, $F(1,57)=3.85$, $p=.055$, $p\eta^2=.06$) moderated by an interaction between attention and group ($BF_{10}=4.6$, $F(2,57)=2.10$, $p=.132$, $p\eta^2=.07$). Subsetting by group, paired sample t-tests revealed anecdotal evidence that detection sensitivity was higher for full attention (M=2.34, SD=.71) than diverted attention (M=1.99, SD=.61) in the organic group ($BF_{10}=1.8$, $t(19)=2.2$, $p=.037$, 95% CI [.030 .878]). In contrast, evidence favoured the null hypothesis that detection sensitivity was unchanged between full and diverted attention conditions in FMD (full: M=2.03, SD=.91, and diverted: M=1.87, SD=.81; $BF_{10}=0.3$, $t(19)=.97$, $p>.25$, 95% CI [-.185 .507]) and healthy controls (full: M=1.96, SD=.71, and diverted: M=1.99, SD=.61; $BF_{10}=0.2$, $t(19)=-.27$, $p>.25$, 95% CI [-.321 .248]).

Decision criterion is equivalent for FMD and healthy controls but differs in organic patients

The tendency to judge gabors as 'present' or 'absent' (decision criterion–type 1 c) was examined using a Bayesian mixed ANOVA. The main effect of expectations on decision criterion was definitive ($BF_{10}=2.3 \times 10^4$, $F(2,114)=29.3$, $p<.001$, $p\eta^2=.34$) and did not interact with group ($BF_{10}<.33$, $p>.25$). Decision criterion became progressively more liberal as expectations of gabor presence increased, revealed by Holm-corrected post hoc comparisons (each $BF_{10}>30$, $p_{holm}<.01$). This result demonstrates that our expectations manipulation was successful.

A substantial main effect of attention was observed (BF_{10}>9.9 x 10^9, $F(1,57)=18.4$, $p<.001$, $p\eta^2=.24$) that interacted with group ($BF_{10}=7.8$, $F(2,57)=2.0$, $p=.139$, $p\eta^2=.07$). Subsetting by group, paired sample t-tests revealed that participants' decision criterion was more conservative with full attention than diverted attention for both FMD and healthy controls (each BF_{10}>25, $p<.001$). Conversely, the organic group remained uniformly conservative with evidence favouring the null hypothesis that decision criterion was unchanged between attention conditions ($BF_{10}=0.3$, $t(19)=0.7$, $p>.25$, 95% CI [-.201 .399]) (**Figure 3**).

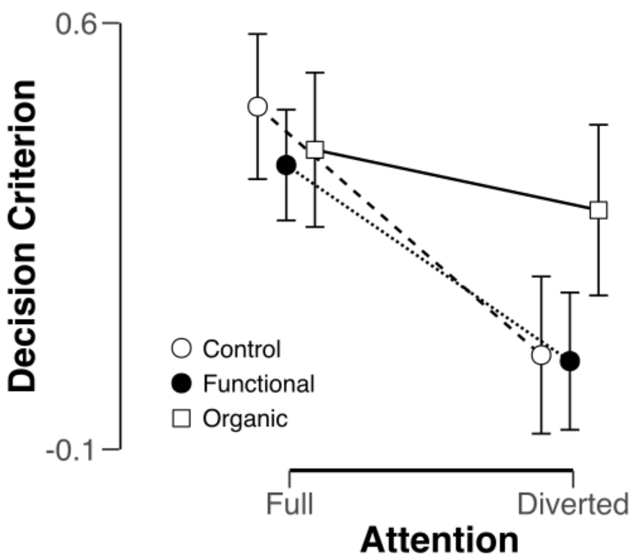

Figure 3. Decision criterion

Mean Type 1 decision criterion as a function of attention and group identity collapsing over expectation conditions. Criterion values above zero are *conservative* and reflect a tendency to report Gabor Absent. Error bars reflect 95% confidence interval.

Metacognition is equivalent between groups but differs as a function of report

A Bayesian mixed ANOVA examined metacognitive sensitivity (type 2 AUC) as a function of report, expectations, attention and the between-subjects factor group. Reporting target presence was associated with higher metacognitive sensitivity revealed by a definitive main effect of report (BF_{10}>9.9 x 10^9, $F(1,57)=191.9$, $p<.001$, $p\eta^2=.69$). This result demonstrates that metacognition was successfully manipulated by our task design. No remaining main effects or higher-order interactions were observed. Notably, there was strong evidence to accept the null hypothesis for the main effect of group (BF_{10}=0.1, F(2,57)=0.1, p>.25, pη²<.01). That is, metacognitive sensitivity was broadly equivalent between groups.

Results for metacognitive efficiency (meta-d'/d') aligned with those for metacognitive sensitivity. We observed definitive evidence for a main effect of report (BF_{10}>9.9 x 10^9, F(1,45)=102.9, p<.001, pη²=.70) and evidence against a main effect of group (BF_{10}=0.3, F(2,45)=2.5, p=.091, pη²=.10). An interaction between report and expectations was also observed (BF_{10}=203.1, F(2,90)=8.6, p<.001, pη²=.16). Subsetting by report, paired samples t-tests revealed differences in metacognitive efficiency as a function of expectations only when reporting gabor absence. Metacognitive efficiency was higher in the 75% target presence condition (M=.91, SD=.94) than either the 25% presence condition (M=.28, SD=.47; BF_{10}=3.5 x 10^3, t(59)=5.0, p<.001, 95% CI [.368 .926]) and 50% presence condition (M=.43, SD=.51; BF_{10}=51.6, t(59)=3.7, p<.001, 95% CI [.208 .742]) (**Figure 4**).

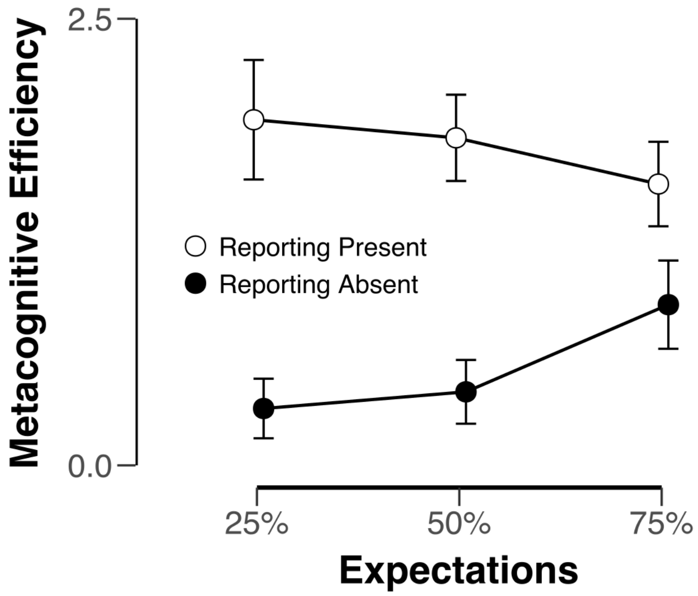

Figure 4. Metacognitive efficiency

Mean metacognitive efficiency (meta-d'/d') as a function of expectations of target presence (25% to 75% likelihood) and report (Gabor Present or Gabor Absent). Results collapsing over attention condition and group identity. Higher metacognitive efficiency reflects greater insight into one's detection sensitivity. Error bars reflect 95% confidence interval.

Discussion

In this study, we elaborate on the predictive processing account of FMDs, in which action is inextricably tied to perception. It follows that the abnormal motor symptoms that characterise FMDs may be understood by examining the mechanisms that underlie perceptual and active inference in the brain. Accordingly, we employed a variant of the dual-task paradigm to delineate four perceptual and cognitive domains implicated in FMD: attention, expectations, sensory processing, and metacognition. In our perceptual tasks, we found strong evidence that attentional processing, expectational processing and metacognitive decision-making were equivalent between a cohort of patients with FMDs and a matched sample of healthy controls. However, this equivalence was contingent on adjusting stimulus contrast to correct for a broad impairment in perceptual sensitivity in FMDs. This impairment necessitated that stimulus contrast was boosted to levels equivalent to those required by a group of patients with organic motor disorders (**Figure 2b** and **Figure 5**). Our findings suggest that attentional, expectational and metacognitive mechanisms are intact in patients with FMD, and it is impairments in basic sensory processing that conceivably underlie this cohort's unique symptomatology.

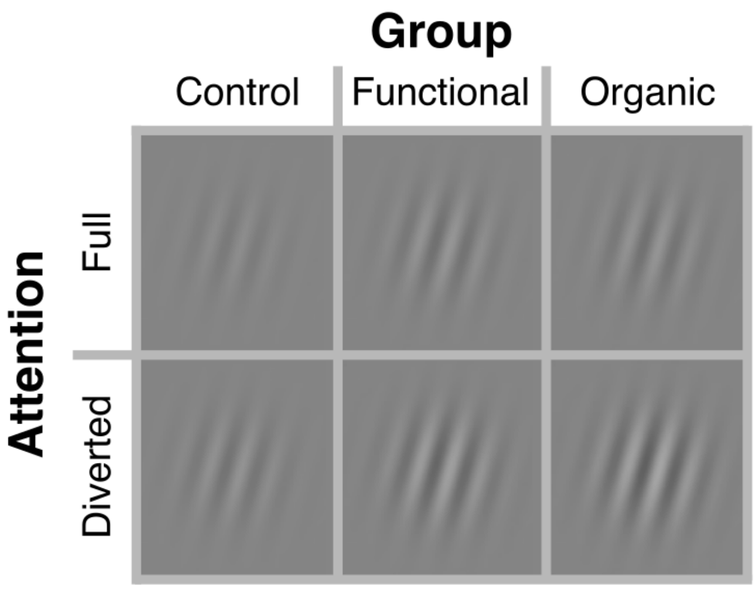

Figure 5. Perceptual thresholds

Mean gabor contrast required to achieve 79.4% performance threshold under Full and Diverted attention for each group.

Our paradigm psychometrically adjusted stimulus contrast using the QUEST adaptive staircase procedure (Watson & Pelli, 1983). We can be confident that attention was successfully manipulated by the dual-task paradigm due to the clear within-subjects difference in contrast required to maintain peripheral performance thresholds at ~79.4% accuracy in each attention condition (**Figure 2a**). Unbeknownst to participants, this procedure yielded stable performance for both the FMD group and healthy controls, permitting us to gauge the effects of attention and expectations on decision criteria and metacognition independent of underlying differences in basic sensory thresholds. We observed decisive evidence for independent effects of attention and expectations on decision criterion setting. However, the profile of these effects was statistically equivalent for the FMD and healthy control cohorts. Participants in both groups adopted a uniformly conservative criterion when their attention was fully focussed (**Figure 3**) and they progressively liberalised that criterion as the underlying expectations of stimulus presence increased. Likewise, metacognitive sensitivity and efficiency, as measured by signal detection theoretic type 2 AUC and meta-d', differed markedly as a function of report type and was responsive to our expectations manipulation; the pattern of these effects were statistically equivalent between groups (**Figure 4**). Collectively, these results imply that attentional, expectational, and metacognitive processes operate normally in FMDs, at least for the function of visual perceptual decision-making. However, the FMD cohort differs from healthy controls in the strength of sensory input required to achieve these benchmarks.

Implications for the predictive processing account of FMD

It is clear that these results do not easily fit the existing neurobiologic theory of FMDs exclusively based on disturbed attention, faulty predictive beliefs and resultant 'unconscious' motor execution through active inference (Edwards et al., 2012). Yet predictive processing, in which action is inextricably tied to perception, retains its power to explain basic cognitive functions including perceptual sensitivity. Other studies of FMDs support this theory (Edwards, 2016; McIntosh et al., 2017). Accordingly, there may be alternative ways to interpret a broad impairment in sensory sensitivity in the absence of organic neurological disease within the predictive processing model.

Since we did not observe a difference in FMD participants as a result of manipulating expectations, we can largely rule out a generalised alteration in modulation of prediction error weightings as priors get stronger or weaker. No differences in the ability to allocate attention (Full vs Diverted) were detected in the FMD group, suggesting that precision optimisation remains intact in the sense of allocating relative weights that sum to one. Metacognitive measures were comparable between FMD and control subjects, indicating that internal assessment of precisions is likely to be unaffected. This points to a more chronic, context-insensitive lowering of weight on all sensory prediction errors, which could be explained in terms of lower expected precision for sensory input in general. This would decrease the gain on low level sensory input and thereby lower the learning rate in perceptual inference. In turn, this could explain the need to increase the signal strength on the sensory input for FMD participants to obtain healthy control performance benchmarks. One possibility is that general lowering of expected precision relates to past precision learning. If attending to sensory input has historically resulted in poor inferences (i.e., those which produce less long-term average prediction error minimisation), then down-weighting signals from these channels may be seen as an adaptive response to inferential underperformance. Conceptually, this impairment would relate to exogenous attentional processing (Hohwy, 2012). Under active inference, this lowered afferent weighting could then lead to lack of action, as proprioceptive prediction error is missed (functional weakness); or failed attempts at minimisation of proprioceptive prediction error—when the weakened neural signals encoding these prediction errors intermittently exceed the threshold for expected noise—giving rise to functional movement (tremor or dystonia).

An alternative proposition is that down-weighting of external sensory input is the consequence of abnormal body-directed attentional focus, with elevated proprioceptive precision occurring at the expense of processing power for other modes of sensory feedback[6]. This would align with the observation that patients with FMDs perform better than controls in force-matching tasks (implying loss of sensory attenuation, i.e. an inability to selectively 'tune out' proprioceptive feedback) and exhibit impairments in spatial attentional shifting in cases where tactile cues are presented near their affected limbs (McIntosh et al., 2017). If this were the case, it might go some way to explaining the recalcitrance of abnormal predictive beliefs in FMD, since proprioceptive feedback cannot easily be 'reality-checked' against other sensory streams.

Both of these propositions show how attention might still have an important role in FMD pathogenesis, although not simply in its commonly understood form as a domain-general and endogenously driven spotlight. Comprehensive treatment of these positions is outside the scope of our empirical study. Yet it seems clear that perception is a fruitful domain of enquiry in motor disorders. Further research is needed to explore how perception relates to action in FMDs, computationally, in terms of aetiology, and in terms of how it might relate to and modify symptoms. We will consider evidence for the positions we outline above and their ramifications for FMD in a future theoretical review.

What then justifies a predictive processing account of perception over traditional views? An exhaustive account is outside the scope of this work but the foundational texts of predictive processing and, more broadly, the free energy principle (Friston, 2010; Friston et al., 2010; Friston, Kilner, & Harrison, 2006) were built on accounts of perceptual inference in the visual system—*predictive coding* (Rao & Ballard, 1999; Srinivasan, Laughlin, & Dubs, 1982). These accounts emphasised cortical feedback (carrying higher-level predictions) as an efficient alternative to purely feedforward

[6] A similar mechanism is raised in a recent account of hypnosis under the predictive processing framework (J.-R. Martin & Pacherie, 2019). According to their model, motor suggestions trigger abnormal body-directed attentional focus resulting in elevated proprioceptive precision. When expected sensory evidence is compared to actual sensory evidence, the highly precise prediction errors that result demand explanation. Motor suggestions supply this explanation— a prior of non-agency for the subject.

models of natural image encoding. A similar process of efficient information transfer was proposed to describe how spatiotemporal filters might be used to maximise information flow in the human visual system. The most utility could be extracted from the eye if one accounted for its limitations and predicted that it typically delivers noisy signals with limited dynamic range (Van Hateren, 1993). The sensitivity of these filters approximated the sensitivity of visual contrast sensitivity and this model mapped onto several psychophysical phenomena in vision. An advantage then of the predictive processing view is that it can describe perceptual phenomena building from a foundation in visual perception, it can link perception to action via active inference, and it can fit this account into a unifying framework of cognition that already describes many cases of brain dysfunction.

Study limitations

Three quarters of our FMD group and one half of organic motor disorder patients were using psychoactive medication at the time of testing. Some of these drugs have visual or ocular adverse effects (Fraunfelder & Fraunfelder, 2004). Although this raises a possibility that medication use might be contributing to contrast impairment, there are two studies of depressive disorder that have observed no discernible medication effects on contrast sensitivity (Bubl, Ebert, Kern, van Elst, & Bach, 2012; Bubl, Tebartz Van Elst, Gondan, Ebert, & Greenlee, 2009). While we cannot exclude the possibility that medication had some influence on our results, when medication use and psychopathological comorbidity were factored into our statistical modelling we still observed strong evidence for group-level differences in contrast sensitivity (see **Supplementary Material**). Furthermore, given its broad influence on mood and cognition, it seems unlikely that psychoactive medication would strongly impair sensory processing in FMD patients but have no measurable effect on attentional, expectational, and metacognitive assessments.

Although we endeavoured to match the age of participants between groups, our patients with organic motor disorders were significantly older than the other groups (each $p<.001$). While this does not detract from our main result with respect to FMD, it is possible that deficits in perceptual contrast sensitivity in organic patients might be associated with their comparatively advanced age. Aging is associated with deficits in contrast sensitivity that manifest from approximately the age of 60 but evidence suggests these impairments are most apparent for spatial frequencies of 4 cycles per degree and higher (Derefeldt, Lennerstrand, & Lundh, 1979; Elliott, 1987; Owsley, Sekuler, & Siemsen, 1983). We employed gabor patch stimuli with spatial frequencies of only 2 cycles per degree and would not expect substantial impairments in sensitivity for these stimuli contingent on age. Supplementary Bayesian analysis including age as an ordinal covariate revealed inconclusive evidence for a main effect of age or interaction between age and group but we still observed strong evidence for group-level differences in sensitivity. Further, for participants in the most advanced age categories (70s and 80s, all organic patients), the impairments in contrast sensitivity they exhibited were substantially greater than sensitivity benchmarks established in healthy individuals of the same age using stimuli with spatial frequencies of 2 cycles per degree (Owsley et al., 1983).

Although the QUEST performance thresholding procedure was successful in stabilising performance for each attention condition in our FMD and healthy control groups, patients with organic motor disorders did not reach performance benchmarks when their attention was diverted. Perceptual contrast sensitivity and attentional processing are known to be impaired in neurological motor disorders such as Parkinson’s disease (Botha & Carr, 2012; Kemps, Szmalec, Vandierendonck, & Crevits, 2005; Lin et al., 2015). Given this, it is not unexpected that trial numbers and parameters adapted from studies of the healthy population might be insufficient to threshold performance in a group with organic motor disorders. Future studies of perceptual decision-making in organic motor disorder might specify more lenient initial contrast levels and parameters for contrast adjustment in order to stabilise performance thresholds.

Impairments in perceptual contrast sensitivity were identified in both functional and organic patient groups. We cannot rule out the possibility that reduced perceptual sensitivity accompanies any motor deficit (be it functional or organic) rather than being unique to FMDs. Further studies are needed to fully distinguish these groups, though we note that the sensory deficit was especially strong in the organic patient group, and that these patients exhibited significant changes in detection sensitivity and decision criterion that were not observed in either the FMD group or healthy controls. These results do speak to fundamental differences in perception between organic and functional motor disorders that might be able to inform diagnosis.

Conclusion

We have demonstrated evidence for an impairment in perceptual contrast sensitivity in patients with functional and organic motor disorders compared with healthy controls. Attentional, expectational, and metacognitive processing was broadly intact in the FMD group, once sensory input was boosted to account for the underlying impairment in sensory perceptual sensitivity. Organic motor disorders were distinguished from FMD and healthy controls by differences in the use of attention and expectations for perceptual decision-making. We relate our findings to an account of functional neurological disorders under predictive processing. The symptomatology of FMDs is conceivably grounded in a chronic and context-insensitive lowering in expected precision for sensory input, rather than broad impairments in predictive beliefs, precision weighting, or the internal assessment of precisions during perceptual inference. Conceptualising functional neurological disorders as disorders of perceptual and active inference thus leads to a consolidation and refinement of the core pathophysiologies associated with these disabling conditions and provides a unifying framework to understand them better.

Acknowledgements

We thank Bryan Paton for his advice regarding mixed statistical designs. We acknowledge Olivia Carter for her insightful comments on a preliminary presentation of this work.

Supplementary material

8AFC response screen

Perceptual decisions and confidence ratings were registered using an eight-alternative forced-choice (8AFC) response screen (Matthews, Schröder, et al., 2018; Matthews, Wu, et al., 2018). In the Full Attention condition (Gabor Task only), subjects were presented with a single response screen. This screen consisted of the printed question 'Was the Grill (P)resent or (A)bsent' and a response square split into eight segments, four each for judgements of Present and Absent. Each segment corresponded to one of four confidence levels. With a single mouse click, subjects could register their 2AFC discrimination report as well as confidence in this decision. Prior to the experiment, and during training, subjects were verbally instructed to express their confidence on a scale from complete guess (rating 1) to certainty (rating 4). Verbal descriptors for judgements of 2 or 3 were not made explicit, however, the experimenter encouraged subjects to fix these confidence criterions across the two sessions and use all four confidence levels. The labels 'absolutely certain' and 'complete guess' were displayed at the top and bottom of the screen to remind subjects of the confidence scale.

In the Diverted Attention condition (Gabor Task plus Letter Task) a Letter response screen was also presented. This consisted of a similar response square with the question 'Did the Letter (T) appear or (Ls)?'. Subjects registered their discrimination report and confidence with a single mouse click.

Group differences in contrast sensitivity are not explained by depressive traits or medication

Depression, along with other psychiatric comorbidities, are known to be more prevalent in functional motor disorder (FMD) than the general population (Factor et al., 1995; Gelauff, Stone, Edwards, & Carson, 2014). This presents a limitation for our contrast results as an association between depressive states and impaired perceptual contrast sensitivity has been suggested (Bubl et al., 2012, 2009). A one-way ANOVA on overall scores on the Hospital Anxiety and Depression Scale (HADS) as a function of group revealed that depression and anxiety traits differed significantly between the groups (BF_{10}=76.2, F(2,57)=9.0, p<.001, $p\eta^2$=.24); HADS scores were higher in FMD and organic motor disorder than in healthy controls (each BF_{10}>50.0, p<.05) but did not differ between the patient groups (BF_{10}=0.6, p>.25). A related limitation is the effect that medication might have on contrast sensitivity. Use of psychoactive medication was more prevalent in the FMD (n=15) and organic (n=10) groups than healthy controls (n=1).

To jointly account for the impact of medication and depressive traits on our findings, we repeated the Bayesian mixed ANOVA on log-transformed contrast thresholds including scores on the HADS depression subscale as a covariate and medication as an additional between-subjects factor. Despite controlling for these effects, results mirrored our previous findings with independent main effects of attention (BF_{10}>9.9 x 10^9), group (BF_{10}=16.8), and an attention by group interaction (BF_{10}=3.8). Conversely, evidence favoured exclusion of the HADS covariate (BF_{10}=0.6, F(1,57)<0.1, p>.25), medication factor (BF_{10}=0.9, F(1,57)=0.2, p>.25), and all higher-order interactions (each BF_{10}<1.5, p>.25) from the model that best fit our data. This pattern of results was also found using untransformed contrast thresholds.

Conclusion

Metacognition is a functional cognitive process that was prominent enough to appear in David Chalmers' list of 'easy problems' of consciousness: "the ability of a system to access its own internal states" (Chalmers, 1995, p. 200). The topic of metacognition is especially important when discussing consciousness because conscious awareness, that introspective feature of consciousness, is metacognitive in nature and encompasses conscious experience so completely that the terms awareness and consciousness are often used interchangeably. It is critical for consciousness science, philosophy of mind, and basic as well as applied neurocognitive research that we understand metacognition and how it relates to not only consciousness but the overall function of the brain.

An important step in this process is to situate metacognition relative to the core mechanisms that support perception and cognition in both healthy and disordered minds. My thesis has achieved this by exploring perceptual decision-making using visual psychophysics and contemporary behavioural analysis. I consolidate diverse threads of knowledge on attention, memory, and perception to examine how the mechanisms that support these processes relate to our human capacity for metacognition and conscious access. My thesis reflects progress in knowledge but it also appeals to progress seen in the wider scientific community—the empirical work that constitutes this document is characterised by a commitment to many of the principles of Open science and I have purposively sought to employ statistical and analytical techniques that are emerging as some of the most robust currently available.

Metacognition has much to tell us about how the brain functions. The overall view that emerges from this thesis is that metacognitive processes can be dissociated from attention and working memory in several key ways. A dissociation between attention and metacognition was theorised almost 20 years ago (Rosenthal, 2000). However, early neuroscientific models of metacognition included metacognitive processes within the neural mechanisms involved in selective attention (Fernandez-Duque et al., 2000; Shimamura, 2000) so a concerted effort to dissociate these processes has not been attempted.

Chapter 2 demonstrated that certain stimulus features that are perceived and encoded into memory with little or possibly no attentional amplification nevertheless remain available for metacognitive report. Despite minimal training, items perceived in the near absence of selective attention were differentiated with high metacognitive sensitivity. Metacognition was only slightly improved when those same items were perceived with full attentional focus—it is conceivable that with more training metacognitive performance would be statistically equivalent between attention conditions. The results from Chapter 2 imply that the cognitive mechanisms that support metacognition are likely to be dissociable from selective attention. These results also build on evidence that consciousness may be doubly dissociable from selective attention (Chennu et al., 2013; Hohwy, 2012; Jennings, 2015; Koch & Tsuchiya, 2007a, 2007b; Koivisto, Revonsuo, & Salminen, 2005; Lamme, 2003; Pitts et al., 2018; Tallon-Baudry, 2011; Tsuchiya & Koch, 2008; van Boxtel et al., 2010a; van Boxtel, Tsuchiya, & Koch, 2010b) although I discuss some caveats for linking metacognition to consciousness below.

Chapter 3 lends further support to a distinction between metacognitive processing and attention but also working memory. We found sustained metacognitive access to incidental memories. Memory traces were encoded into memory in the context of a separate, attention-demanding task so were perceived but could not be actively maintained—a prerequisite for working memory according to prominent theories of memory function (Baddeley, 2003; Luck & Vogel, 2013). Consistent with another study of metacognition and working memory (Samaha, Barrett, Sheldon, LaRocque, & Postle, 2016), the results from Chapter 3 suggest it is not necessary for memory traces to be actively maintained in order for them to be accessible for metacognitive processing. This conclusion is also supported by our finding that metacognition was no better and subjective confidence ratings no higher when participants focused their attention on memorising each stimulus. Recent discussion of latent working memory theorizes that a high capacity store of 'low-energy' memory might be available that does not require active maintenance (Sergent, 2018). This theory cannot account for our findings as we demonstrated high metacognitive sensitivity for faces that had never been seen before and found sustained access whether faces were presented in upright or inverted orientation. Our findings are consistent with the emerging view of non-capacity limited working memory (Gross, 2018; Kaunitz et al., 2016). We conclude that the mere act of perception is sufficient to encode detailed and sustained memories that remain available for conscious access and metacognitive processing without active maintenance by attention.

Chapter 4 identified further dissociations between metacognitive function and the mechanisms that underlie selective attention as well as expectations. We found equivalent metacognitive sensitivity between healthy controls and two clinical groups once we accounted for underlying differences in perceptual sensitivity. Further support for an association between perception and metacognition was found—report type had a strong influence on metacognitive performance. When participants from each group assessed their confidence in reports of stimulus presence they exhibited considerably higher metacognitive sensitivity and efficiency than when they reported stimulus absence. This finding is supported by many other studies that have used present versus absent perceptual judgments and observe superior metacognitive sensitivity for stimulus present reports (Kanai, Walsh, & Tseng, 2010; Meuwese, van Loon, Lamme, & Fahrenfort, 2014; M. Peters et al., 2016; Ruby, Giles, & Lau, 2017). The results from Chapter 4 highlight that metacognition is supported by information that is contained in visual perception and that determines subjective perceptual thresholds. I discuss below why metacognition is unlikely to draw on this perceptual evidence alone and what this means for the relationship between metacognition and consciousness.

An emerging view from this thesis is that extra care must be taken when employing metacognition as a method for differentiating conscious from unconscious mental states. Many acknowledge that metacognition is distinct from conscious awareness (Dienes, 2004; Dienes & Perner, 2004; Fleming & Lau, 2014; Maniscalco & Lau, 2012; Rosenthal, 2000; Timmermans, Schilbach, Pasquali, & Cleeremans, 2012) but there is still widespread belief that metacognitive measures (particularly those based on signal detection theory) offer unparalleled access to consciousness and the conscious status of perception (Faivre, Arzi, Lunghi, & Salomon, 2017; Odegaard & Lau, 2016; Persaud, McLeod, & Cowey, 2007; Rausch, Müller, & Zehetleitner, 2015). Metacognitive processes are unlikely to be domain-general (McCurdy et al., 2013; Rouault, McWilliams, Allen, & Fleming, 2018; Ruby et al., 2017) and evidence is converging that metacognitive judgments cannot derive from perceptual evidence alone (Bang & Fleming, 2018; Boldt, de Gardelle, & Yeung, 2017; Fleming & Daw, 2017; Grimaldi et al., 2015; Moran, Teodorescu, & Usher, 2015). However, there is still considerable explanatory power in the Metacognitive Model for determining whether decisions are formed on the basis of conscious or unconscious knowledge (Nelson, 1996). This is exemplified in the zero-correlation and guessing criteria (Chan, 1992; Cheesman & Merikle, 1984; Dienes, 2004; Dienes et al., 1995; Dienes & Berry, 1997; Dienes & Perner, 2004). It might not be possible to trace conscious knowledge to perception *per se*, but perceptual evidence clearly influences the strength of metacognitive judgments (Baruch, Kimchi, & Goldsmith, 2014; Meuwese et al., 2014), an effect that might be dissociable from selective attention, working memory, and expectations (Matthews, Nagao, et al., 2018; Matthews, Schröder, et al., 2018; Matthews, Wu, et al., 2018).

Going forward, the influence that conscious and unconscious knowledge has on metacognitive operations will be vital for the future of neurocognitive research on metacognition and conscious awareness. It is also important that we factor contemporary research on the neural basis of subjective confidence and certainty into models of metacognitive processing that, for too long, have assumed a very tight or even absolute association between perceptual evidence and metacognitive reports (Baranski & Petrusic, 1998; Clarke et al., 1959; Galvin et al., 2003; Kunimoto et al., 2001). Identifying to what extent conscious knowledge is necessary to form confidence judgments or other metacognitive reports will aid in identifying how perceptual evidence and other cognitive factors contribute to metacognition. This might reveal the contribution that certain forms of conscious knowledge have on decision-making over and above unconscious knowledge which could highlight possible functions of conscious awareness—curiosity and communication being two examples from the wider literature on metacognition (Frith, 2012; Litman, 2009; Metcalfe et al., 2017).

Theories and neurobiological models of cognitive function should build on the finding that metacognition might be distinct from selective attention and working memory. This thesis highlights that selective attention is unlikely to be necessary to *encode* or *maintain* perceptual representations for metacognitive processing. However, an important point for future research will be to assess whether attention is necessary or sufficient to *retrieve* these representations for a metacognitive report. There exists skepticism that conscious recollection is possible in the absence of the internal focus of attention (De Brigard, 2012). It remains to be seen whether this skepticism should apply for conscious as well as unconscious access, and whether attentional amplification alone is sufficient for metacognitive processing. Distinctions such as these will aid in framing neurocognitive models of metacognition and introspection that might then be differentiated from related cognitive processes and the minimal neural mechanisms sufficient for consciousness.

The study of metacognition is vital for understanding the behavioural and neurophysiological functions of the brain. It distinguishes complex dynamics between memory and attention that underlie cognition and highlights how these processes and others are employed for internal monitoring and cognitive control. Metacognition is also important for understanding dysfunction in the brain. It is a marker of psychopathology but also a tool in treatment, therapy, and applied research. Crucially, metacognition elucidates conscious awareness. It is a behavioural measure for categorising perception and knowledge, a cognitive mechanism for investigating possible functions of the mind, and an essential property of introspection. Metacognition is not the conclusion to the quest for consciousness but by studying its role in cognitive function and reflecting on how it is distinguished, philosophy and cognitive science has identified a pathway forwards.

References

Aczel, B., Hoekstra, R., Gelman, A., Wagenmakers, E.-J., Kluglist, I. G., Rouder, J., … van Ravenzwaaij, D. (2018). *Expert opinions on how to conduct and report bayesian inference*. *PsyArXiv*. https://doi.org/10.31234/osf.io/23m7f

Aitchison, L., Bang, D., Bahrami, B., & Latham, P. E. (2015). Doubly Bayesian analysis of confidence in perceptual decision-making. *PLoS Computational Biology*, *11*(10). https://doi.org/10.1371/journal.pcbi.1004519

Alvarez, G. A., & Cavanagh, P. (2004). The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychological Science*, *15*(2), 106–111.

Anderson, J. R., Matessa, M., & Lebiere, C. (1997). ACT-R: a theory of higher level cognition and its relation to visual attention. *Human-Computer Interaction*, *12*(4), 439–462.

Anderson, K. E., Gruber-Baldini, A. L., Vaughan, C. G., Reich, S. G., Fishman, P. S., Weiner, W. J., & Shulman, L. M. (2007). Impact of psychogenic movement disorders versus Parkinson's on disability, quality of life, and psychopathology. *Movement Disorders: Official Journal of the Movement Disorder Society*, *22*(15), 2204–2209.

Awh, E., Barton, B., & Vogel, E. K. (2007). Visual working memory represents a fixed number of items regardless of complexity. *Psychological Science*, *18*(7), 622–628.

Awh, E., & Jonides, J. (2001). Overlapping mechanisms of attention and spatial working memory. *Trends in Cognitive Sciences*, *5*(3), 119–126.

Baars, B. J. (1988). *A cognitive theory of consciousness*. Cambridge, UK: Cambridge University Press.

Baars, B. J. (1997). In the theatre of consciousness. Global workspace theory, a rigorous scientific theory of consciousness. *Journal of Consciousness Studies*, *4*(4), 292–309.

Baars, B. J. (2005). Global workspace theory of consciousness: toward a cognitive neuroscience of human experience. *Progress in Brain Research*, *150*, 45–53.

Baars, B. J., & Franklin, S. (2003). How conscious experience and working memory interact. *Trends in Cognitive Sciences*, *7*(4), 166–172.

Baddeley, A. D. (1986). *Working memory*. Oxford, UK: Clarendon Press.

Baddeley, A. D. (2000). The episodic buffer: a new component of working memory? *Trends in Cognitive Sciences*, *4*(11), 417–423.

Baddeley, A. D. (2003). Working memory: looking back and looking forward. *Nature Reviews. Neuroscience*, *4*(10), 829–839.

Bang, D., & Fleming, S. M. (2018). Distinct encoding of decision confidence in human medial prefrontal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(23), 6082–6087.

Baranski, J. V., & Petrusic, W. M. (1998). Probing the locus of confidence judgments: experiments on the time to determine confidence. *Journal of Experimental Psychology. Human Perception and Performance*, *24*(3), 929–945.

Barrett, A. B., Dienes, Z., & Seth, A. K. (2013). Measures of metacognition on signal-detection theoretic models. *Psychological Methods*, *18*(4), 535–552.

Baruch, O., Kimchi, R., & Goldsmith, M. (2014). Attention to distinguishing features in object recognition. *Visual Cognition*, *22*(9-10), 1184–1215.

Beare, J. I. (2010). On memory and reminiscence Aristotle (ca. 350 b.c.). *Annals of Neurosciences*, *17*(2), 87–91.

Beck, M. R., Peterson, M. S., Boot, W. R., Vomela, M., & Kramer, A. F. (2006). Explicit memory for rejected distractors during visual search. *Visual Cognition*, *14*(2), 150–174.

Bergström, F., & Eriksson, J. (2014). Maintenance of non-consciously presented information engages the prefrontal cortex. *Frontiers in Human Neuroscience*, *8*, 938.

Block, N. (1995). On a confusion about a function of consciousness. *The Behavioral and Brain Sciences*, *18*(2), 227–247.

Block, N. (2007a). Consciousness, accessibility, and the mesh between psychology and neuroscience. *The Behavioral and Brain Sciences*, *30*(5-6), 481–499; discussion 499–548.

Block, N. (2007b). Overflow, access, and attention. *The Behavioral and Brain Sciences*, *30*(5-6), 530–548.

Block, N. (2011a). Perceptual consciousness overflows cognitive access. *Trends in Cognitive Sciences*, *15*(12), 567–575.

Block, N. (2011b). The higher order approach to consciousness is defunct. *Analysis*, *71*(3), 419–431.

Block, N. (2014). Rich conscious perception outside focal attention. *Trends in Cognitive Sciences*, *18*(9), 445–447.

Bloom, P. A., Friedman, D., Xu, J., Vuorre, M., & Metcalfe, J. (2018). Tip-of-the-tongue states predict enhanced feedback processing and subsequent memory. *Consciousness and Cognition*, *63*, 206–217.

Boldt, A., de Gardelle, V., & Yeung, N. (2017). The impact of evidence reliability on sensitivity and bias in decision confidence. *Journal of Experimental Psychology. Human Perception and Performance*, *43*(8), 1520–1531.

Boly, M., Massimini, M., Tsuchiya, N., Postle, B. R., Koch, C., & Tononi, G. (2017). Are the neural correlates of consciousness in the front or in the back of the cerebral cortex? Clinical and neuroimaging evidence. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *37*(40), 9603–9613.

Bor, D., & Seth, A. K. (2012). Consciousness and the prefrontal parietal network: insights from attention, working memory, and chunking. *Frontiers in Psychology*, *3*, 63.

Botha, H., & Carr, J. (2012). Attention and visual dysfunction in Parkinson's disease. *Parkinsonism & Related Disorders*, *18*(6), 742–747.

Braun, J., & Julesz, B. (1998). Withdrawing attention at little or no cost: detection and discrimination tasks. *Perception & Psychophysics*, *60*(1), 1–23.

Brown, J., Lewis, V. J., & Monk, A. F. (1977). Memorability, word frequency and negative recognition. *The Quarterly Journal of Experimental Psychology*, *29*(3), 461–473.

Brown, R. (2014). Consciousness doesn't overflow cognition. *Frontiers in Psychology*, *5*, 1399.

Brown, R. (2015). The HOROR theory of phenomenal consciousness. *Philosophical Studies*, *172*(7), 1783–1794.

Brown, R., & McNeill, D. (1966). The "tip of the tongue" phenomenon. *Journal of Verbal Learning and Verbal Behavior*, *5*(4), 325–337.

Bubl, E., Ebert, D., Kern, E., van Elst, L. T., & Bach, M. (2012). Effect of antidepressive therapy on retinal contrast processing in depressive disorder. *The British Journal of Psychiatry: The Journal of Mental Science*, *201*, 151–158.

Bubl, E., Tebartz Van Elst, L., Gondan, M., Ebert, D., & Greenlee, M. W. (2009). Vision in depressive disorder. *The World Journal of Biological Psychiatry: The Official Journal of the World Federation of Societies of Biological Psychiatry*, *10*(4 Pt 2), 377–384.

Büchel, C., Geuter, S., Sprenger, C., & Eippert, F. (2014). Placebo analgesia: a predictive coding perspective. *Neuron*, *81*(6), 1223–1239.

Bush, G., Vogt, B. A., Holmes, J., Dale, A. M., Greve, D., Jenike, M. A., & Rosen, B. R. (2002). Dorsal anterior cingulate cortex: a role in reward-based decision making. *Proceedings of the National Academy of Sciences*, *99*(1), 523–528.

Carruthers, P. (2000). *Phenomenal consciousness: a naturalistic theory*. Cambridge, UK: Cambridge University Press.

Carruthers, P. (2015). *The centered mind: what the science of working memory shows us about the nature of human thought*. Oxford University Press.

Carter, C. S., Botvinick, M. M., & Cohen, J. D. (1999). The contribution of the anterior cingulate cortex to executive processes in cognition. *Reviews in the Neurosciences*, *10*(1), 49–57.

Castelhano, M., & Henderson, J. (2005). Incidental visual memory for objects in scenes. *Visual Cognition*, *12*(6), 1017–1040.

Cerf, M., Frady, E. P., & Koch, C. (2009). Faces and text attract gaze independent of the task: experimental data and computer model. *Journal of Vision*, *9*(12), 10.1–15.

Chalmers, D. J. (1995). Facing up to the problem of consciousness. *Journal of Consciousness Studies*, *2*(3), 200–219.

Chalmers, D. J. (1996). *The conscious mind: in search of a fundamental theory*. Oxford University Press.

Chan, C. (1992). *Implicit cognitive processes: theoretical issues and applications in computer systems design*. University of Oxford. Retrieved from http://ethos.bl.uk/OrderDetails.do?uin=uk.bl.ethos.303636

Cheesman, J., & Merikle, P. M. (1984). Priming with and without awareness. *Perception & Psychophysics*, *36*(4), 387–395.

Chennu, S., Finoia, P., Kamau, E., Monti, M. M., Allanson, J., Pickard, J. D., … Bekinschtein, T. A. (2013). Dissociable endogenous and exogenous attention in disorders of consciousness. *NeuroImage. Clinical*, *3*, 450–461.

Chun, M. M., & Jiang, Y. (1998). Contextual cueing: implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology*, *36*(1), 28–71.

Chun, M. M., & Potter, M. C. (1995). A two-stage model for multiple target detection in rapid serial visual presentation. *Journal of Experimental Psychology. Human Perception and Performance*, *21*(1), 109–127.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *The Behavioral and Brain Sciences*, *36*(3), 181–204.

Clark, A. (2016). *Surfing uncertainty: prediction, action, and the embodied mind*. Oxford University Press.

Clarke, F. R., Birdsall, T. G., & Tanner, W. P. (1959). Two types of ROC curves and definitions of parameters. *The Journal of the Acoustical Society of America*, *31*(5), 629–630.

Cleeremans, A. (2008). Consciousness: the radical plasticity thesis. *Progress in Brain Research*, *168*, 19–33.

Cleeremans, A. (2011). The radical plasticity thesis: how the brain learns to be conscious. *Frontiers in Psychology*, *2*, 86.

Cohen, M. A., & Dennett, D. C. (2011). Consciousness cannot be separated from function. *Trends in Cognitive Sciences*, *15*(8), 358–364.

Cohen, M. A., Dennett, D. C., & Kanwisher, N. (2016). What is the bandwidth of perceptual experience? *Trends in Cognitive Sciences*, *20*(5), 324–335.

Cousins, R. D. (2017). The Jeffreys–Lindley paradox and discovery criteria in high energy physics. *Synthese*, *194*(2), 395–432.

Cowan, N. (1998). Visual and auditory working memory capacity. *Trends in Cognitive Sciences*, *2*(3), 77.

Cowan, N. (2010). The magical mystery four: how is working memory capacity limited, and why? *Current Directions in Psychological Science*, *19*(1), 51–57.

Cowey, A. (2010). The blindsight saga. *Experimental Brain Research*, *200*(1), 3–24.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: a framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*(6), 671–684.

Crick, F., & Koch, C. (1990). Towards a neurobiological theory of consciousness. *Seminars in the Neurosciences*, *2*, 263–275.

Davis, E. T., & Palmer, J. (2004). Visual search and attention: an overview. *Spatial Vision*, *17*(4-5), 249–255.

Daw, N. D., O’Doherty, J. P., Dayan, P., Seymour, B., & Dolan, R. J. (2006). Cortical substrates for exploratory decisions in humans. *Nature*, *441*(7095), 876–879.

Debner, J. A., & Jacoby, L. L. (1994). Unconscious perception: attention, awareness, and control. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *20*(2), 304–317.

De Brigard, F. (2012). The role of attention in conscious recollection. *Frontiers in Psychology*, *3*, 29.

De Brigard, F., & Prinz, J. (2010). Attention and consciousness. *Wiley Interdisciplinary Reviews. Cognitive Science*, *1*(1), 51–59.

Dehaene, S. (2014). *Consciousness and the brain: deciphering how the brain codes our thoughts*. New York: Viking Press.

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.

Dehaene, S., Lau, H. C., & Kouider, S. (2017). What is consciousness, and could machines have it? *Science*, *358*(6362), 486–492.

Del Cul, A., Dehaene, S., Reyes, P., Bravo, E., & Slachevsky, A. (2009). Causal role of prefrontal cortex in the threshold for access to consciousness. *Brain: A Journal of Neurology*, *132*(Pt 9), 2531–2540.

De Martino, B., Fleming, S. M., Garrett, N., & Dolan, R. J. (2013). Confidence in value-based choice. *Nature Neuroscience*, *16*(1), 105–110.

Deneve, S. (2012). Making decisions with unknown sensory reliability. *Frontiers in Neuroscience*, *6*, 75.

Dennett, D. (2001). Are we explaining consciousness yet? *Cognition*, *79*(1-2), 221–237.

Derefeldt, G., Lennerstrand, G., & Lundh, B. (1979). Age variations in normal human contrast sensitivity. *Acta Ophthalmologica*, *57*(4), 679–690.

Deroy, O., Spence, C., & Noppeney, U. (2016). Metacognition in multisensory perception. *Trends in Cognitive Sciences*, *20*(10), 736–747.

Dickinson, C. A., & Zelinsky, G. J. (2007). Memory for the search path: evidence for a high-capacity representation of search history. *Vision Research*, *47*(13), 1745–1755.

Dienes, Z. (2004). Assumptions of subjective measures of unconscious mental states: higher order thoughts and bias. *Journal of Consciousness Studies*, *11*(9), 25–45.

Dienes, Z. (2008a). Subjective measures of unconscious knowledge. *Progress in Brain Research*, *168*, 49–64.

Dienes, Z. (2008b). *Understanding psychology as a science: an introduction to scientific and statistical inference*. UK: Palgrave Macmillan.

Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in Psychology*, *5*, 781.

Dienes, Z., Altmann, G., Kwan, L., & Goode, A. (1995). Unconscious knowledge of artificial grammars is applied strategically. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *21*(5), 1322.

Dienes, Z., & Berry, D. (1997). Implicit learning: below the subjective threshold. *Psychonomic Bulletin & Review*, *4*(1), 3–23.

Dienes, Z., & Perner, J. (1996). Implicit knowledge in people and connectionist networks. In G. Underwood (Ed.), *Implicit knowledge* (pp. 227–255). Oxford University Press.

Dienes, Z., & Perner, J. (1999). A theory of implicit and explicit knowledge. *The Behavioral and Brain Sciences*, *22*(5), 735–755.

Dienes, Z., & Perner, J. (2004). Assumptions of a subjective measure of consciousness: three mappings. In R. J. Gennaro (Ed.), *Higher-order theories of consciousness: an anthology* (pp. 173–199). Amsterdam: John Benjamins Publishing Company.

Dilsaver, S. C. (1992). Differentiating organic from functional psychosis. *American Family Physician*, *45*(3), 1173–1180.

Drew, T., Horowitz, T. S., Wolfe, J. M., & Vogel, E. K. (2011). Delineating the neural signatures of tracking spatial position and working memory during attentive tracking. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *31*(2), 659–668.

Drew, T., & Vogel, E. K. (2008). Neural measures of individual differences in selecting and tracking multiple moving objects. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *28*(16), 4183–4191.

Dunlosky, J., & Nelson, T. O. (1994). Does the sensitivity of judgements of learning (JOLs) to the effects of various study activities depend on when the JOLs occur? *Journal of Memory and Language*, *33*, 545–545.

Edwards, M. J. (2016). Chapter 12 - Neurobiologic theories of functional neurologic disorders. In M. Hallett, J. Stone, & A. Carson (Eds.), *Handbook of Clinical Neurology* (Vol. 139, pp. 131–137). Elsevier.

Edwards, M. J., Adams, R. A., Brown, H., Pareés, I., & Friston, K. J. (2012). A Bayesian account of "hysteria." *Brain: A Journal of Neurology*, *135*(Pt 11), 3495–3512.

Einhäuser, W., Koch, C., & Makeig, S. (2007). The duration of the attentional blink in natural scenes depends on stimulus category. *Vision Research*, *47*(5), 597–607.

Elliott, D. B. (1987). Contrast sensitivity decline with ageing: a neural or optical phenomenon? *Ophthalmic & Physiological Optics: The Journal of the British College of Ophthalmic Opticians* , *7*(4), 415–419.

Eng, H. Y., Chen, D., & Jiang, Y. (2005). Visual working memory for simple and complex visual stimuli. *Psychonomic Bulletin & Review*, *12*(6), 1127–1133.

Evans, K. K., & Treisman, A. (2005). Perception of objects in natural scenes: is it really attention free? *Journal of Experimental Psychology. Human Perception and Performance*, *31*(6), 1476–1492.

Evans, S., & Azzopardi, P. (2007). Evaluation of a “bias-free” measure of awareness. *Spatial Vision*, *20*(1), 61–77.

Factor, S. A., Podskalny, G. D., & Molho, E. S. (1995). Psychogenic movement disorders: frequency, clinical profile, and characteristics. *Journal of Neurology, Neurosurgery, and Psychiatry*, *59*(4), 406–412.

Faivre, N., Arzi, A., Lunghi, C., & Salomon, R. (2017). Consciousness is more than meets the eye: a call for a multisensory study of subjective experience. *Neuroscience of Consciousness*, *2017*(1), nix003.

Fazekas, P., & Overgaard, M. (2018). Perceptual consciousness and cognitive access: an introduction. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *373*(1755). https://doi.org/10.1098/rstb.2017.0340

Fei-Fei, L., Iyer, A., Koch, C., & Perona, P. (2007). What do we perceive in a glance of a real-world scene? *Journal of Vision*, *7*(1), 10.

Fernandez-Duque, D., Baird, J. A., & Posner, M. I. (2000). Executive attention and metacognitive regulation. *Consciousness and Cognition*, *9*(2 Pt 1), 288–307.

Flanagan, O. J. (1992). *Consciousness reconsidered*. Cambridge, MA: MIT Press.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *The American Psychologist*, *34*(10), 906–911.

Flavell, J. H., & Wellman, H. M. (1977). Metamemory. In R. V. Kail & J. W. Hagen (Eds.), *Perspective on the development of memory and cognition* (pp. 3–33). Hillsdale, NJ: Erlbaum.

Fleming, S. M. (2017). HMeta-d: hierarchical Bayesian estimation of metacognitive efficiency from confidence ratings. *Neuroscience of Consciousness*, *2017*(1). https://doi.org/10.1093/nc/nix007

Fleming, S. M., & Daw, N. D. (2017). Self-evaluation of decision-making: a general Bayesian framework for metacognitive computation. *Psychological Review*, *124*(1), 91–114.

Fleming, S. M., & Dolan, R. J. (2012). The neural basis of metacognitive ability. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *367*(1594), 1338–1349.

Fleming, S. M., Huijgen, J., & Dolan, R. J. (2012). Prefrontal contributions to metacognition in perceptual decision making. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *32*(18), 6117–6125.

Fleming, S. M., & Lau, H. C. (2014). How to measure metacognition. *Frontiers in Human Neuroscience*, *8*. https://doi.org/10.3389/fnhum.2014.00443

Fleming, S. M., Weil, R. S., Nagy, Z., Dolan, R. J., & Rees, G. (2010). Relating introspective accuracy to individual differences in brain structure. *Science*, *329*(5998), 1541–1543.

Fodor, J. A. (1981). The mind-body problem. *Scientific American*, *244*(1), 114–123.

Folke, T., Jacobsen, C., Fleming, S. M., & De Martino, B. (2016). Explicit representation of confidence informs future value-based decisions. *Nature Human Behaviour*, *1*, 0002.

Fougnie, D., & Alvarez, G. A. (2011). Object features fail independently in visual working memory: evidence for a probabilistic feature-store model. *Journal of Vision*, *11*(12). https://doi.org/10.1167/11.12.3

Fraunfelder, F. W., & Fraunfelder, F. T. (2004). Adverse ocular drug reactions recently identified by the National Registry of Drug-Induced Ocular Side Effects. *Ophthalmology*, *111*(7), 1275–1279.

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature Reviews. Neuroscience*, *11*(2), 127–138.

Friston, K., Daunizeau, J., Kilner, J., & Kiebel, S. J. (2010). Action and behavior: a free-energy formulation. *Biological Cybernetics*, *102*(3), 227–260.

Friston, K., Kilner, J., & Harrison, L. (2006). A free energy principle for the brain. *Journal of Physiology, Paris*, *100*(1-3), 70–87.

Friston, K., Stephan, K. E., Montague, R., & Dolan, R. J. (2014). Computational psychiatry: the brain as a phantastic organ. *The Lancet. Psychiatry*, *1*(2), 148–158.

Frith, C. D. (2012). The role of metacognition in human social interactions. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *367*(1599), 2213–2223.

Gallagher, R., Suddendorf, T., & Arnold, D. (2018). *Confidence as a diagnostic tool for perceptual aftereffects*. *bioRxiv*. https://doi.org/10.1101/270280

Galvin, S. J., Podd, J. V., Drga, V., & Whitmore, J. (2003). Type 2 tasks in the theory of signal detectability: discrimination between correct and incorrect decisions. *Psychonomic Bulletin & Review*, *10*(4), 843–876.

García-Gutiérrez, A., Aguado, L., Romero-Ferreiro, V., & Pérez-Moreno, E. (2017). Discrimination of face gender and expression under dual-task conditions. *Attention, Perception & Psychophysics*, *79*(2), 614–627.

Gelauff, J., Stone, J., Edwards, M. J., & Carson, A. (2014). The prognosis of functional (psychogenic) motor symptoms: a systematic review. *Journal of Neurology, Neurosurgery, and Psychiatry*, *85*(2), 220–226.

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis* , *1*(3), 515–534.

Gelman, A., Jakulin, A., Pittau, M. G., & Su, Y.-S. (2008). A weakly informative default prior distribution for logistic and other regression models. *The Annals of Applied Statistics*, *2*(4), 1360–1383.

Graf, P., & Birt, A. R. (1996). Explicit and implicit memory retrieval: intentions and strategies. In L. M. Reder (Ed.), *Implicit memory and metacognition* (pp. 25–44). Hillsdale, NJ: Erlbaum.

Graham, G., & Neisser, J. (2000). Probing for relevance: what metacognition tells us about the power of consciousness. *Consciousness and Cognition*, *9*(2 Pt 1), 172–177; discussion 193–202.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics* (Vol. 1). New York: Wiley.

Grimaldi, P., Lau, H. C., & Basso, M. A. (2015). There are things that we know that we know, and there are things that we do not know we do not know: confidence in decision-making. *Neuroscience and Biobehavioral Reviews*, *55*, 88–97.

Gross, S. (2018). Perceptual consciousness and cognitive access from the perspective of capacity-unlimited working memory. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *373*(1755). https://doi.org/10.1098/rstb.2017.0343

Hallett, M. (2016). Functional (psychogenic) movement disorders - Clinical presentations. *Parkinsonism & Related Disorders*, *22 Suppl 1*, S149–S152.

Hart, J. T. (1965). Memory and the feeling-of-knowing experience. *Journal of Educational Psychology*, *56*(4), 208–216.

Hart, J. T. (1967a). Memory and the memory-monitoring process. *Journal of Verbal Learning and Verbal Behavior*, *6*(5), 685–691.

Hart, J. T. (1967b). Second-try recall, recognition, and the memory-monitoring process. *Journal of Educational Psychology*, *58*(4), 193–197.

Hauser, T. U., Allen, M., NSPN Consortium, Rees, G., & Dolan, R. J. (2017). Metacognitive impairments extend perceptual decision making weaknesses in compulsivity. *Scientific Reports*, *7*(1), 6614.

Hilgenstock, R., Weiss, T., & Witte, O. W. (2014). You'd better think twice: post-decision perceptual confidence. *NeuroImage*, *99*, 323–331.

Hohwy, J. (2012). Attention and conscious perception in the hypothesis testing brain. *Frontiers in Psychology*, *3*, 96.

Hohwy, J. (2013). *The predictive mind*. Oxford University Press.

Hohwy, J. (2016). The self-evidencing brain. *Noûs*, *50*(2), 259–285.

Hohwy, J., Roepstorff, A., & Friston, K. (2008). Predictive coding explains binocular rivalry: an epistemological review. *Cognition*, *108*(3), 687–701.

Horga, G., Schatz, K. C., Abi-Dargham, A., & Peterson, B. S. (2014). Deficits in predictive coding underlie hallucinations in schizophrenia. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *34*(24), 8072–8082.

Hutto, D. D. (2018). Getting into predictive processing's great guessing game: bootstrap heaven or hell? *Synthese*, *195*(6), 2445–2458.

James, W. (1890). *The principles of psychology, Vol I*. New York: Henry Holt and Co.

Janet, P. (1901). *The mental state of hystericals: a study of mental stigmata and mental accidents*. GP Putnam's sons.

Janowsky, J. S., Shimamura, A. P., & Squire, L. R. (1989). Memory and metamemory: comparisons between patients with frontal lobe lesions and amnesic patients. *Psychobiology* , *17*(1), 3–11.

Jarvstad, A., Hahn, U., Warren, P. A., & Rushton, S. K. (2014). Are perceptuo-motor decisions really more optimal than cognitive decisions? *Cognition*, *130*(3), 397–416.

JASP Team. (2018). JASP (Version 0.9). Retrieved from https://jasp-stats.org/

Jeffreys, H. (1939). The Theory of Probability, 1st. Oxford: Oxford University Press.

Jennings, C. D. (2015). Consciousness without attention. *Journal of the American Philosophical Association*, *1*(2), 276–295.

Joseph, J. S., Chun, M. M., & Nakayama, K. (1997). Attentional requirements in a "preattentive" feature search task. *Nature*, *387*(6635), 805.

Kanai, R., Walsh, V., & Tseng, C.-H. (2010). Subjective discriminability of invisibility: a framework for distinguishing perceptual and attentional failures of awareness. *Consciousness and Cognition*, *19*(4), 1045–1057.

Kaunitz, L. N., Rowe, E. G., & Tsuchiya, N. (2016). Large capacity of conscious access for incidental memories in natural scenes. *Psychological Science*, *27*(9), 1266–1277.

Keller, G. B., & Mrsic-Flogel, T. D. (2018). Predictive processing: a canonical cortical computation. *Neuron*, *100*(2), 424–435.

Kemps, E., Szmalec, A., Vandierendonck, A., & Crevits, L. (2005). Visuo-spatial processing in Parkinson's disease: evidence for diminished visuo-spatial sketch pad and central executive resources. *Parkinsonism & Related Disorders*, *11*(3), 181–186.

Knill, D. C., & Pouget, A. (2004). The Bayesian brain: the role of uncertainty in neural coding and computation. *Trends in Neurosciences*, *27*(12), 712–719.

Koch, C. (2004). *The quest for consciousness: a neurobiological approach*. Englewood, CO: Roberts and Company.

Koch, C., Massimini, M., Boly, M., & Tononi, G. (2016). Neural correlates of consciousness: progress and problems. *Nature Reviews. Neuroscience*, *17*(5), 307–321.

Koch, C., & Tsuchiya, N. (2007a). Attention and consciousness: two distinct brain processes. *Trends in Cognitive Sciences*, *11*(1), 16–22.

Koch, C., & Tsuchiya, N. (2007b). Phenomenology without conscious access is a form of consciousness without top-down attention. *The Behavioral and Brain Sciences*, *30*(5-6), 509–510.

Koivisto, M., Revonsuo, A., & Salminen, N. (2005). Independence of visual awareness from attention at early processing stages. *Neuroreport*, *16*(8), 817–821.

Koriat, A. (1993). How do we know that we know? The accessibility model of the feeling of knowing. *Psychological Review*, *100*(4), 609–639.

Koriat, A. (2007). Metacognition and consciousness. In P. D. Zelazo, M. Moscovitch, & E. Thompson (Eds.), *The Cambridge Handbook of Consciousness* (pp. 289–326). Cambridge, MA: Cambridge University Press.

Koster-Hale, J., & Saxe, R. (2013). Theory of mind: a neural prediction problem. *Neuron*, *79*(5), 836–848.

Ko, Y., & Lau, H. C. (2012). A detection theoretic explanation of blindsight suggests a link between conscious perception and metacognition. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *367*(1594), 1401–1411.

Krakauer, J. W., Ghazanfar, A. A., Gomez-Marin, A., MacIver, M. A., & Poeppel, D. (2017). Neuroscience needs behavior: correcting a reductionist bias. *Neuron*, *93*(3), 480–490.

Kristensen, S. B., Sandberg, K., & Bibby, B. M. (2018). *Novel regression methods for metacognition*. *bioRxiv*. https://doi.org/10.1101/423947

Kunimoto, C., Miller, J., & Pashler, H. (2001). Confidence and accuracy of near-threshold discrimination responses. *Consciousness and Cognition*, *10*(3), 294–340.

Lamme, V. A. F. (2003). Why visual attention and awareness are different. *Trends in Cognitive Sciences*, *7*(1), 12–18.

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 of the United States of America*, *103*(49), 18763–18768.

Lau, H. C., & Rosenthal, D. M. (2011). Empirical support for higher-order theories of conscious awareness. *Trends in Cognitive Sciences*, *15*(8), 365–373.

Laureys, S., & Schiff, N. D. (2012). Coma and consciousness: paradigms (re)framed by neuroimaging. *NeuroImage*, *61*(2), 478–491.

LeDoux, J. E., & Brown, R. (2017). A higher-order theory of emotional consciousness. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(10), E2016–E2025.

Lee, D. K., Koch, C., & Braun, J. (1999). Attentional capacity is undifferentiated: concurrent discrimination of form, color, and motion. *Perception & Psychophysics*, *61*(7), 1241–1255.

Li, F. F., VanRullen, R., Koch, C., & Perona, P. (2002). Rapid natural scene categorization in the near absence of attention. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(14), 9596–9601.

Lindley, D. V. (1957). A Statistical Paradox. *Biometrika*, *44*(1/2), 187–192.

Lin, T. P., Rigby, H., Adler, J. S., Hentz, J. G., Balcer, L. J., Galetta, S. L., … Adler, C. H. (2015). Abnormal visual contrast acuity in Parkinson's disease. *Journal of Parkinson's Disease*, *5*(1), 125–130.

Litman, J. A. (2009). Curiosity and metacognition. *Metacognition: New Research Developments*, 105–116.

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*(6657), 279–281.

Luck, S. J., & Vogel, E. K. (2013). Visual working memory capacity: from psychophysics and neurobiology to individual differences. *Trends in Cognitive Sciences*, *17*(8), 391–400.

Luria, R., & Vogel, E. K. (2011). Visual search demands dictate reliance on working memory storage. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *31*(16), 6199–6207.

Lycan, W. G. (1996). *Consciousness and experience*. Cambridge, MA: MIT Press.

Lysaker, P. H., Carcione, A., Dimaggio, G., Johannesen, J. K., Nicolò, G., Procacci, M., & Semerari, A. (2005). Metacognition amidst narratives of self and illness in schizophrenia: associations with neurocognition, symptoms, insight and quality of life. *Acta Psychiatrica Scandinavica*, *112*(1), 64–71.

Lysaker, P. H., Hamm, J. A., Hasson-Ohayon, I., Pattison, M. L., & Leonhardt, B. L. (2018). Promoting recovery from severe mental illness: implications from research on metacognition and metacognitive reflection and insight therapy. *World Journal of Psychiatry*, *8*(1), 1–11.

Lysaker, P. H., Warman, D. M., Dimaggio, G., Procacci, M., Larocco, V. A., Clark, L. K., … Nicolò, G. (2008). Metacognition in schizophrenia: associations with multiple assessments of executive function. *The Journal of Nervous and Mental Disease*, *196*(5), 384–389.

Mack, A., & Rock, I. (1998). *Inattentional blindness* (Vol. 33). Cambridge, MA: MIT Press.

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: a user's guide*. Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Maniscalco, B., & Lau, H. C. (2012). A signal detection theoretic approach for estimating metacognitive sensitivity from confidence ratings. *Consciousness and Cognition*, *21*(1), 422–430.

Maril, A., Simons, J. S., Mitchell, J. P., Schwartz, B. L., & Schacter, D. L. (2003). Feeling-of-knowing in episodic memory: an event-related fMRI study. *NeuroImage*, *18*(4), 827–836.

Maril, A., Simons, J. S., Weaver, J. J., & Schacter, D. L. (2005). Graded recall success: an event-related fMRI comparison of tip of the tongue and feeling of knowing. *NeuroImage*, *24*(4), 1130–1138.

Maril, A., Wagner, A. D., & Schacter, D. L. (2001). On the tip of the tongue: an event-related fMRI study of semantic retrieval failure and cognitive conflict. *Neuron*, *31*(4), 653–660.

Martin, J.-R., & Pacherie, E. (2019). Alterations of agency in hypnosis: a new predictive coding model. *Psychological Review*, *126*(1), 133–152.

Martin, K. (2004). Time waits for no man. *Nature*, *429*(6989), 243–244.

Masson, M. E. J., & Rotello, C. M. (2009). Sources of bias in the Goodman-Kruskal gamma coefficient measure of association: implications for studies of metacognitive processes. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *35*(2), 509.

Matthews, J., Nagao, K., Ding, C., Newby, R., Kempster, P., & Hohwy, J. (2018). *Impaired perceptual sensitivity with intact attention and metacognition in functional motor disorder*. *PsyArXiv*. https://doi.org/10.31234/osf.io/fz3j2

Matthews, J., Schröder, P., Kaunitz, L., van Boxtel, J. J. A., & Tsuchiya, N. (2018). Conscious access in the near absence of attention: critical extensions on the dual-task paradigm. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *373*(1755). https://doi.org/10.1098/rstb.2017.0352

Matthews, J., Wu, J., Corneille, V., Hohwy, J., van Boxtel, J., & Tsuchiya, N. (2018). Sustained conscious access to incidental memories in RSVP. *Attention, Perception & Psychophysics*. https://doi.org/10.3758/s13414-018-1600-1

Maxwell, J. P., Masters, R. S. W., & Eves, F. F. (2003). The role of working memory in motor learning and performance. *Consciousness and Cognition*, *12*(3), 376–402.

McCurdy, L. Y., Maniscalco, B., Metcalfe, J., Liu, K. Y., de Lange, F. P., & Lau, H. C. (2013). Anatomical coupling between distinct metacognitive systems for memory and visual perception. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *33*(5), 1897–1906.

McGlynn, S. M., & Kaszniak, A. W. (1991). When metacognition fails: impaired awareness of deficit in Alzheimer's disease. *Journal of Cognitive Neuroscience*, *3*(2), 183–187.

McIntosh, R. D., McWhirter, L., Ludwig, L., Carson, A., & Stone, J. (2017). Attention and sensation in functional motor disorder. *Neuropsychologia*, *106*, 207–215.

Menary, R. (2015). What? Now. Predictive coding and enculturation. In T. Metzinger & J. M. Windt (Eds.), *Open MIND*. Frankfurt am Main: MIND Group.

Merikle, P. M., & Joordens, S. (1997). Parallels between perception without attention and perception without awareness. *Consciousness and Cognition*, *6*(2-3), 219–236.

Metcalfe, J., & Schwartz, B. L. (2016). The ghost in the machine: self-reflective consciousness and the neuroscience of metacognition. In J. Dunlosky & S. Tauber (Eds.), *The Oxford Handbook of Metamemory* (pp. 407–424). Oxford University Press.

Metcalfe, J., Schwartz, B. L., & Bloom, P. A. (2017). The tip-of-the-tongue state and curiosity. *Cognitive Research: Principles and Implications*, *2*(1), 31.

Metcalfe, J., & Shimamura, A. P. (Eds.). (1994). *Metacognition: knowing about knowing*. Cambridge, MA: MIT Press.

Meuwese, J. D. I., van Loon, A. M., Lamme, V. A. F., & Fahrenfort, J. J. (2014). The subjective experience of object recognition: comparing metacognition for object detection and object categorization. *Attention, Perception & Psychophysics*, *76*(4), 1057–1068.

Modirrousta, M., & Fellows, L. K. (2008). Medial prefrontal cortex plays a critical and selective role in "feeling of knowing"meta-memory judgments. *Neuropsychologia*, *46*(12), 2958–2965.

Mole, C. (2008). Attention and consciousness. *Journal of Consciousness Studies*, *15*(4), 86–104.

Moran, R., Teodorescu, A. R., & Usher, M. (2015). Post choice information integration as a causal determinant of confidence: novel data and a computational account. *Cognitive Psychology*, *78*, 99–147.

Morgante, F., Matinella, A., Andrenelli, E., Ricciardi, L., Allegra, C., Terranova, C., … Tinazzi, M. (2018). Pain processing in functional and idiopathic dystonia: An exploratory study. *Movement Disorders: Official Journal of the Movement Disorder Society*. https://doi.org/10.1002/mds.27402

Morgante, F., Tinazzi, M., Squintani, G., Martino, D., Defazio, G., Romito, L., … Berardelli, A. (2011). Abnormal tactile temporal discrimination in psychogenic dystonia. *Neurology*, *77*(12), 1191–1197.

Moritz, S., & Woodward, T. S. (2007). Metacognitive training for schizophrenia patients (MCT): a pilot study on feasibility, treatment adherence, and subjective efficacy. *German Journal of Psychiatry*, *10*(3), 69–78.

Moritz, S., Woodward, T. S., & Balzan, R. (2016). Is metacognitive training for psychosis effective? *Expert Review of Neurotherapeutics*, *16*(2), 105–107.

Naccache, L. (2018). Why and how access consciousness can account for phenomenal consciousness. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *373*(1755). https://doi.org/10.1098/rstb.2017.0357

Nagel, T. (1974). What is it like to be a bat? *The Philosophical Review*, *83*(4), 435–450.

Nelson, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin*, *95*(1), 109–133.

Nelson, T. O. (1990). Metamemory: a theoretical framework and new findings. In G. H. Bower (Ed.), *Psychology of Learning and Motivation* (Vol. 26, pp. 125–173). New York: Academic Press.

Nelson, T. O. (1996). Consciousness and metacognition. *The American Psychologist*, *51*(2), 102.

Nelson, T. O., & Narens, L. (1994). Why investigate metacognition. In Metcalfe Janet Shimamura A (Ed.), *Metacognition: knowing about knowing* (pp. 1–25). Cambridge, MA: MIT Press.

Nelson, T. O., & Rey, G. (2000). Metacognition and consciousness: A convergence of psychology and philosophy. *Consciousness and Cognition*, *9*(2), 147–148.

Newby, R., Alty, J., & Kempster, P. (2016). Functional dystonia and the borderland between neurology and psychiatry: New concepts. *Movement Disorders: Official Journal of the Movement Disorder Society*, *31*(12), 1777–1784.

Nickerson, R. S. (2000). Null hypothesis significance testing: a review of an old and continuing controversy. *Psychological Methods*, *5*(2), 241–301.

Noë, A. (2004). *Action in perception*. Cambridge, MA: MIT Press.

Noë, A. (2009). *Out of our heads: why you are not your brain, and other lessons from the biology of consciousness*. London, UK: Macmillan.

Norman, E., & Price, M. C. (2015). Measuring consciousness with confidence ratings. *Behavioural Methods in Consciousness Research*, 159–180.

Odegaard, B., Knight, R. T., & Lau, H. C. (2017). Should a few null findings falsify prefrontal theories of conscious perception? *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *37*(40), 9593–9602.

Odegaard, B., & Lau, H. C. (2016). Methodological considerations to strengthen studies of peripheral vision. *Trends in Cognitive Sciences*, *20*(9), 642–643.

Otto, A. R., Fleming, S. M., & Glimcher, P. W. (2016). Unexpected but incidental positive outcomes predict real-world gambling. *Psychological Science*, *27*(3), 299–311.

Overgaard, M. (2012). Blindsight: recent and historical controversies on the blindness of blindsight. *Wiley Interdisciplinary Reviews. Cognitive Science*, *3*(6), 607–614.

Owsley, C., Sekuler, R., & Siemsen, D. (1983). Contrast sensitivity throughout adulthood. *Vision Research*, *23*(7), 689–699.

Palmer, C. J., Lawson, R. P., & Hohwy, J. (2017). Bayesian approaches to autism: towards volatility, action, and behavior. *Psychological Bulletin*, *143*(5), 521–542.

Pannu, J. K., Kaszniak, A. W., & Rapcsak, S. Z. (2005). Metamemory for faces following frontal lobe damage. *Journal of the International Neuropsychological Society: JINS*, *11*(6), 668–676.

Pareés, I., Brown, H., Nuruki, A., Adams, R. A., Davare, M., Bhatia, K. P., … Edwards, M. J. (2014). Loss of sensory attenuation in patients with functional (psychogenic) movement disorders. *Brain: A Journal of Neurology*, *137*(Pt 11), 2916–2921.

Pareés, I., Kassavetis, P., Saifee, T. A., Sadnicka, A., Bhatia, K. P., Fotopoulou, A., & Edwards, M. J. (2012). “Jumping to conclusions” bias in functional movement disorders. *Journal of Neurology, Neurosurgery, and Psychiatry*, *83*(4), 460–463.

Pareés, I., Kassavetis, P., Saifee, T. A., Sadnicka, A., Davare, M., Bhatia, K. P., … Edwards, M. J. (2013). Failure of explicit movement control in patients with functional motor symptoms. *Movement Disorders: Official Journal of the Movement Disorder Society*, *28*(4), 517–523.

Pareés, I., Saifee, T. A., Kassavetis, P., Kojovic, M., Rubio-Agusti, I., Rothwell, J. C., … Edwards, M. J. (2011). Believing is perceiving: mismatch between self-report and actigraphy in psychogenic tremor. *Brain: A Journal of Neurology*, *135*(1), 117–123.

Pastukhov, A., Fischer, L., & Braun, J. (2009). Visual attention is a single, integrated resource. *Vision Research*, *49*(10), 1166–1173.

Pavlopoulou, C., & Yu, S. X. (2010). Classification and feature selection with human performance data. In *IEEE International Conference on Image Processing* (pp. 1557–1560).

Persaud, N., McLeod, P., & Cowey, A. (2007). Post-decision wagering objectively measures awareness. *Nature Neuroscience*, *10*(2), 257–261.

Persuh, M., LaRock, E., & Berger, J. (2018). Working memory and consciousness: the current state of play. *Frontiers in Human Neuroscience*, *12*, 78.

Peters, E., Joseph, S., Day, S., & Garety, P. (2004). Measuring delusional ideation: the 21-item Peters et al. Delusions Inventory (PDI). *Schizophrenia Bulletin*, *30*(4), 1005–1022.

Peters, M., Kentridge, R. W., Phillips, I., & Block, N. (2017). Does unconscious perception really exist? Continuing the ASSC20 debate. *Neuroscience of Consciousness*, *2017*(1), nix015.

Peters, M., Ro, T., & Lau, H. C. (2016). Who's afraid of response bias? *Neuroscience of Consciousness*, *2016*(1). https://doi.org/10.1093/nc/niw001

Pezzulo, G. (2012). An Active Inference view of cognitive control. *Frontiers in Psychology*, *3*, 478.

Phillips, I. (2018). The methodological puzzle of phenomenal consciousness. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *373*(1755). https://doi.org/10.1098/rstb.2017.0347

Pitts, M. A., Lutsyshyna, L. A., & Hillyard, S. A. (2018). The relationship between attention and consciousness: an expanded taxonomy and implications for "no-report"paradigms. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *373*(1755), 20170348.

Polson, N. G., & Scott, J. G. (2012). On the half-Cauchy prior for a global scale parameter. *Bayesian Analysis* , *7*(4), 887–902.

Postman, L. (1964). Short-term memory and incidental learning. In A. W. Melton (Ed.), *Categories of human learning* (pp. 145–201). New York: Academic Press.

Potter, M. C. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology. Human Learning and Memory*, *2*(5), 509–522.

Potter, M. C., Wyble, B., Hagmann, C. E., & McCourt, E. S. (2014). Detecting meaning in RSVP at 13 ms per picture. *Attention, Perception & Psychophysics*, *76*(2), 270–279.

Powers, A. R., Mathys, C., & Corlett, P. R. (2017). Pavlovian conditioning-induced hallucinations result from overweighting of perceptual priors. *Science*, *357*(6351), 596–600.

Prinz, J. (2012). *The conscious brain*. Oxford University Press.

Proust, J. (2013). *The philosophy of metacognition: mental agency and self-awareness*. Oxford University Press.

Quak, M., London, R. E., & Talsma, D. (2015). A multisensory perspective of working memory. *Frontiers in Human Neuroscience*, *9*, 197.

Rahnev, D., & Denison, R. N. (2018). Suboptimality in perceptual decision making. *The Behavioral and Brain Sciences*, 1–107.

Rao, R. P., & 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.

Rausch, M., Müller, H. J., & Zehetleitner, M. (2015). Metacognitive sensitivity of subjective reports of decisional confidence and visual experience. *Consciousness and Cognition*, *35*, 192–205.

Rausch, M., & Zehetleitner, M. (2017). Should metacognition be measured by logistic regression? *Consciousness and Cognition*, *49*, 291–312.

Raymond, J. E., Shapiro, K. L., & Arnell, K. M. (1992). Temporary suppression of visual processing in an RSVP task: an attentional blink? *Journal of Experimental Psychology. Human Perception and Performance*, *18*(3), 849–860.

Reddy, L., Reddy, L., & Koch, C. (2006). Face identification in the near-absence of focal attention. *Vision Research*, *46*(15), 2336–2343.

Reddy, L., Wilken, P., & Koch, C. (2004). Face-gender discrimination is possible in the near-absence of attention. *Journal of Vision*, *4*(2), 106–117.

Rees, G., Kreiman, G., & Koch, C. (2002). Neural correlates of consciousness in humans. *Nature Reviews. Neuroscience*, *3*(4), 261–270.

Rensink, R. A., O'Regan, J. K., & Clark, J. J. (1997). To see or not to see: the need for attention to perceive changes in scenes. *Psychological Science*, *8*(5), 368–373.

Ricciardi, L., Demartini, B., Crucianelli, L., Krahé, C., Edwards, M. J., & Fotopoulou, A. (2016). Interoceptive awareness in patients with functional neurological symptoms. *Biological Psychology*, *113*, 68–74.

Rosenthal, D. M. (1986). Two concepts of consciousness. *Philosophical Studies*, *49*(3), 329–359.

Rosenthal, D. M. (2000). Metacognition and higher-order thoughts. *Consciousness and Cognition*, *9*(2 Pt 1), 231–242.

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

Rosenthal, D. M. (2012). Higher-order awareness, misrepresentation and function. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *367*(1594), 1424–1438.

Rosenthal, D. M. (2018). Consciousness and confidence. *Neuropsychologia*. https://doi.org/10.1016/j.neuropsychologia.2018.01.018

Rouault, M., McWilliams, A., Allen, M. G., & Fleming, S. M. (2018). Human metacognition across domains: insights from individual differences and neuroimaging. *Personality Neuroscience*, *1*. https://doi.org/10.1017/pen.2018.16

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, *16*(2), 225–237.

Ruby, E., Giles, N., & Lau, H. C. (2017). *Finding domain-general metacognitive mechanisms requires using appropriate tasks*. *bioRxiv*. https://doi.org/10.1101/211805

Salmela, V. R., Moisala, M., & Alho, K. (2014). Working memory resources are shared across sensory modalities. *Attention, Perception & Psychophysics*, *76*(7), 1962–1974.

Samaha, J., Barrett, J. J., Sheldon, A. D., LaRocque, J. J., & Postle, B. R. (2016). Dissociating perceptual confidence from discrimination accuracy reveals no influence of metacognitive awareness on working memory. *Frontiers in Psychology*, *7*, 851.

Sampanes, A. C., Tseng, P., & Bridgeman, B. (2008). The role of gist in scene recognition. *Vision Research*, *48*(21), 2275–2283.

Schacter, D. L., Moscovitch, M., Tulving, E., McLachlan, D. R., & Freedman, M. (1986). Mnemonic precedence in amnesic patients: an analogue of the AB error in infants? *Child Development*, *57*(3), 816–823.

Schnyer, D. M., Verfaellie, M., Alexander, M. P., LaFleche, G., Nicholls, L., & Kaszniak, A. W. (2004). A role for right medial prefontal cortex in accurate feeling-of-knowing judgements: evidence from patients with lesions to frontal cortex. *Neuropsychologia*, *42*(7), 957–966.

Schwartz, B. L. (2001). *Tip-of-the-tongue states: phenomenology, mechanism, and lexical retrieval*. New York: Psychology Press.

Schwartz, B. L., & Cleary, A. M. (2016). Tip-of-the-tongue states, déjà vu experiences, and other odd metamemory experiences. In J. Dunlosky & S. Tauber (Eds.), *The Oxford Handbook of Metamemory* (p. 95). Oxford University Press.

Schwartz, B. L., & Metcalfe, J. (1994). Methodological problems and pitfalls in the study of human metacognition. In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: knowing about knowing* (pp. 93–113). Cambridge, MA: MIT Press.

Schwartz, B. L., & Metcalfe, J. (2011). Tip-of-the-tongue (TOT) states: retrieval, behavior, and experience. *Memory & Cognition*, *39*(5), 737–749.

Scott, R. B., Dienes, Z., Barrett, A. B., Bor, D., & Seth, A. K. (2014). Blind insight: metacognitive discrimination despite chance task performance. *Psychological Science*, *25*(12), 2199–2208.

Searle, J. R., Dennett, D. C., & Chalmers, D. J. (1997). *The mystery of consciousness*. New York Review of Books.

Sergent, C. (2018). The offline stream of conscious representations. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *373*(1755). https://doi.org/10.1098/rstb.2017.0349

Seth, A. K. (2014). A predictive processing theory of sensorimotor contingencies: explaining the puzzle of perceptual presence and its absence in synesthesia. *Cognitive Neuroscience*, *5*(2), 97–118.

Seth, A. K., Dienes, Z., Cleeremans, A., Overgaard, M., & Pessoa, L. (2008). Measuring consciousness: relating behavioural and neurophysiological approaches. *Trends in Cognitive Sciences*, *12*(8), 314–321.

Sherman, M. T., Seth, A. K., & Barrett, A. B. (2018). *Quantifying metacognitive thresholds using signal-detection theory*. *bioRxiv*. https://doi.org/10.1101/361543

Sherman, M. T., Seth, A. K., Barrett, A. B., & Kanai, R. (2015). Prior expectations facilitate metacognition for perceptual decision. *Consciousness and Cognition*, *35*, 53–65.

Shimamura, A. P. (2000). Toward a cognitive neuroscience of metacognition. *Consciousness and Cognition*, *9*(2 Pt 1), 313–323; discussion 324–326.

Shimamura, A. P. (2008). A neurocognitive approach to metacognitive monitoring and control. In J. Dunlosky & R. A. Bjork (Eds.), *Handbook of metamemory and memory* (pp. 373–390). New York: Psychology Press.

Shimamura, A. P., & Squire, L. R. (1986a). Korsakoff's syndrome: a study of the relation between anterograde amnesia and remote memory impairment. *Behavioral Neuroscience*, *100*(2), 165–170.

Shimamura, A. P., & Squire, L. R. (1986b). Memory and metamemory: a study of the feeling-of-knowing phenomenon in amnesic patients. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *12*(3), 452–460.

Shimamura, A. P., & Squire, L. R. (1988). Long-term memory in amnesia: cued recall, recognition memory, and confidence ratings. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *14*(4), 763–770.

Simons, D. J., & Levin, D. T. (1997). Change blindness. *Trends in Cognitive Sciences*, *1*(7), 261–267.

Sorabji, R. (1972). *Aristotle on memory*. Duckworth.

Soto, D., Mäntylä, T., & Silvanto, J. (2011). Working memory without consciousness. *Current Biology: CB*, *21*(22), R912–R913.

Soto, D., & Silvanto, J. (2014). Reappraising the relationship between working memory and conscious awareness. *Trends in Cognitive Sciences*, *18*(10), 520–525.

Spence, R., & Witkowski, M. (2013). *Rapid serial visual presentation: design for cognition*.

Sperling, G., & Dosher, B. A. (1986). Strategy and optimization in human information processing. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of Perception and Human Performance* (Vol. 1, pp. 1–65). New York: Wiley.

Srinivasan, M. V., Laughlin, S. B., & Dubs, A. (1982). Predictive coding: a fresh view of inhibition in the retina. *Proceedings of the Royal Society of London. Series B, Containing Papers of a Biological Character. Royal Society* , *216*(1205), 427–459.

Stenner, M.-P., & Haggard, P. (2016). Voluntary or involuntary? A neurophysiologic approach to functional movement disorders. *Handbook of Clinical Neurology*, *139*, 121–129.

Stone, J., Carson, A., Duncan, R., Roberts, R., Warlow, C., Hibberd, C., ... Sharpe, M. (2010). Who is referred to neurology clinics? The diagnoses made in 3781 new patients. *Clinical Neurology and Neurosurgery*, *112*(9), 747–751.

Summerfield, C., & Egner, T. (2009). Expectation (and attention) in visual cognition. *Trends in Cognitive Sciences*, *13*(9), 403–409.

Swan, L. (2001). Unilateral spatial neglect. *Physical Therapy*, *81*(9), 1572–1580.

Tallon-Baudry, C. (2011). On the neural mechanisms subserving consciousness and attention. *Frontiers in Psychology*, *2*, 397.

Terrace, H. S., & Metcalfe, J. (2005). *The missing link in cognition: origins of self-reflective consciousness*. Oxford University Press.

Timmermans, B., Schilbach, L., Pasquali, A., & Cleeremans, A. (2012). Higher order thoughts in action: consciousness as an unconscious re-description process. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *367*(1594), 1412–1423.

Tononi, G. (2008). Consciousness as integrated information: a provisional manifesto. *The Biological Bulletin*, *215*(3), 216–242.

Tononi, G., Boly, M., Massimini, M., & Koch, C. (2016). Integrated information theory: from consciousness to its physical substrate. *Nature Reviews. Neuroscience*, *17*(7), 450–461.

Tse, P. U. (2004). Mapping visual attention with change blindness: new directions for a new method. *Cognitive Science*, *28*(2), 241–258.

Tsuchiya, N., & Koch, C. (2008). The relationship between consciousness and attention. In S. Laureys, O. Gosseries, & G. Tononi (Eds.), *The neurology of consciousness: Cognitive neuroscience and neuropathology* (pp. 63–78). New York: Academic Press.

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.

Vallar, G. (2001). Extrapersonal visual unilateral spatial neglect and its neuroanatomy. *NeuroImage*, *14*(1 Pt 2), S52–S58.

Vallar, G. (2007). Spatial neglect, Balint-Homes' and Gerstmann's syndrome, and other spatial disorders. *CNS Spectrums*, *12*(7), 527–536.

van Boxtel, J. J. A., & Lu, H. (2013). A predictive coding perspective on autism spectrum disorders. *Frontiers in Psychology*, *4*, 19.

van Boxtel, J. J. A., Tsuchiya, N., & Koch, C. (2010a). Consciousness and attention: on sufficiency and necessity. *Frontiers in Psychology*, *1*, 217.

van Boxtel, J. J. A., Tsuchiya, N., & Koch, C. (2010b). Opposing effects of attention and consciousness on afterimages. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(19), 8883–8888.

Van de Cruys, S., Evers, K., Van der Hallen, R., Van Eylen, L., Boets, B., de-Wit, L., & Wagemans, J. (2014). Precise minds in uncertain worlds: predictive coding in autism. *Psychological Review*, *121*(4), 649–675.

Van Hateren, J. H. (1993). Spatiotemporal contrast sensitivity of early vision. *Vision Research*, *33*(2), 257–267.

VanRullen, R., Reddy, L., & Koch, C. (2004). Visual search and dual tasks reveal two distinct attentional resources. *Journal of Cognitive Neuroscience*, *16*(1), 4–14.

Vogel, E. K., Woodman, G. F., & Luck, S. J. (2006). The time course of consolidation in visual working memory. *Journal of Experimental Psychology. Human Perception and Performance*, *32*(6), 1436–1451.

Voon, V., Ekanayake, V., Wiggs, E., Kranick, S., Ameli, R., Harrison, N. A., & Hallett, M. (2013). Response inhibition in motor conversion disorder. *Movement Disorders: Official Journal of the Movement Disorder Society*, *28*(5), 612–618.

Watson, A. B., & Pelli, D. G. (1983). QUEST: a Bayesian adaptive psychometric method. *Perception & Psychophysics*, *33*(2), 113–120.

Watzl, S. (2011a). The nature of attention. *Philosophy Compass*, *6*(11), 842–853.

Watzl, S. (2011b). The philosophical significance of attention. *Philosophy Compass*, *6*(10), 722–733.

Weiskrantz, L. (1986). *Blindsight: a case study and implications*. Oxford University Press.

Weiskrantz, L. (1997). *Consciousness lost and found: a neuropsychological exploration*. Oxford University Press.

Wells, A. (2011). *Metacognitive therapy for anxiety and depression*. Guilford Press.

Werth, R., Von Cramon, D., & Zihl, J. (1986). Neglect: phänomene halbseitiger vernachlässigung nach hirnschädigung. *Fortschritte Der Neurologie Psychiatrie*, *54*(1), 21–32.

Westfall, P. H., Johnson, W. O., & Utts, J. M. (1997). A Bayesian perspective on the Bonferroni adjustment. *Biometrika*, *84*(2), 419–427.

Whiteley, L., & Sahani, M. (2008). Implicit knowledge of visual uncertainty guides decisions with asymmetric outcomes. *Journal of Vision*, *8*(3), 2.1–15.

Williams, C. C. (2010). Incidental and intentional visual memory: what memories are and are not affected by encoding tasks? *Visual Cognition*, *18*(9), 1348–1367.

Williams, C. C., Henderson, J. M., & Zacks, R. T. (2005). Incidental visual memory for targets and distractors in visual search. *Perception & Psychophysics*, *67*(5), 816–827.

Wundt, W. M. (1874). *Grundzüge der physiologischen psychologie*. Leipzig, Germany: Wilhelm Engelmann.

Yokoyama, O., Miura, N., Watanabe, J., Takemoto, A., Uchida, S., Sugiura, M., … Nakamura, K. (2010). Right frontopolar cortex activity correlates with reliability of retrospective rating of confidence in short-term recognition memory performance. *Neuroscience Research*, *68*(3), 199–206.

Zigmond, A. S., & Snaith, R. P. (1983). The hospital anxiety and depression scale. *Acta Psychiatrica Scandinavica*, *67*(6), 361–370.