International Journal of Law and Psychiatry 67 (2019) 101504

Contents lists available at ScienceDirect

International Journal of Law and Psychiatry

journal homepage: www.elsevier.com/locate/ijlawpsy

Advice-taking as a bridge between decision neuroscience and mental capacity

Elisa van der Plas[a,*], Anthony S. David[b], Stephen M. Fleming[a,c,*]

[a] Wellcome Centre for Human Neuroimaging, University College London, London WC1N 3BG, UK
[b] Institute of Mental Health, University College London, London W1T 7NF, UK
[c] Max Planck University College London, Centre for Computational Psychiatry and Ageing Research, London WC1B 5EH, UK

ARTICLE INFO

Keywords:
Capacity
Decision-making capacity
Advice-taking
Decision neuroscience

ABSTRACT

A person's capacity to process advice is an important aspect of decision making in the real world. For example, in decisions about treatment, the way patients respond to the advice of family, friends and medical professionals may be used (intentionally or otherwise) as a marker of the "use or weigh" requirement of decision-making capacity. Here we explore neuroscientific research on decision-making to identify features of advice-taking that help conceptualize this requirement. We focus on studies of the neural and computational basis of decision-making in laboratory settings. These studies originally investigated simple perceptual decisions about ambiguous stimuli, but have more recently been extended to more complex "value-based" decisions involving the comparison of subjective preferences. Value-based decisions are a useful model system for capacity-related decision-making as they do not have an objectively 'correct' answer and are instead based on subjective preferences. In this context, advice-taking can be seen as a process in which new evidence for one or other option is integrated, leading to altered behaviour or choices. We use this framework to distinguish between different types of advice-taking: private compliance consists of updating one's privately held beliefs based on new evidence, whereas in the case of public compliance, people change their behaviour at a surface level without shifting their privately-held beliefs. Importantly, both types of advice-taking may lead to similar outcomes but rely on different decision processes. We suggest that understanding how multiple mechanisms drive advice-taking holds promise for targeting decision-making support and improving our understanding of the use and weigh requirement in cases of contested capacity.

1. Introduction

The young patient "Z" has been considered by her clinicians not to have full insight into her risky online behaviour and social interactions. The court questions whether Z's abnormal susceptibility to others is an indication of her inability to 'use and weigh':

> *"…as a consequence of Z's autism, which impacts on her ability to put herself in other people's shoes and make judgments with regard to their intentions towards her, she struggles to think through the consequences of having contact with another individual who may pose a risk to her […]. As a consequence of this inability to weigh up the positives […] Z lacks capacity to make decisions regarding contact with others."* (Cobb [2016] EWCOP 4)

The legal notion of mental capacity refers to an individual's ability to make autonomous decisions about their own welfare, often referred to as *decision-making capacity* (DMC). In the UK, the concept of capacity in relation to various types of decisions was initially developed via the common law, before being incorporated into a statutory framework in the *Adults with Incapacity (Scotland) Act* (2000) and the *Mental Capacity Act* (MCA; 2005) in England and Wales. These instruments are intended to cover the circumstances in which necessary acts of caring can be administered (and decisions about these acts taken) on behalf of individuals lacking capacity to either consent to care or make their own decisions.

According to the MCA, the criteria for assessing mental capacity are twofold – diagnostic and functional. The diagnostic criteria state that a person's lack of capacity must be due to an "*impairment of, or a disturbance in the functioning of, the mind or brain*" (s. 2(1)). The functional criterion is a test for the capacity to make decisions, and states (s. 3(1)) that an individual *is unable to make a decision for himself if he is unable –*

* Corresponding authors.
E-mail addresses: elisa.plas.18@ucl.ac.uk (E. van der Plas), stephen.fleming@ucl.ac.uk (S.M. Fleming).

https://doi.org/10.1016/j.ijlp.2019.101504
Received 12 June 2019; Received in revised form 9 September 2019; Accepted 10 September 2019
Available online 26 November 2019
0160-2527/ © 2019 Published by Elsevier Ltd.

a) to understand the information relevant to the decision,
b) to retain that information,
c) to use or weigh that information as part of the process of making the decision, or
d) to communicate his decision (whether by talking, using sign language or any other means).

While a majority of capacity cases are uncontroversial (e.g. Ruck Keene, Kane, Kim & Owen, 2019, estimate that 95% of applications to the Court of Protection fall into the category of uncontentious applications determined without a hearing), in those that are contested, the most contentious point is the "using or weighing" of information which does not always have obvious outward signs (Case, 2016; Ruck Keene et al., 2019). The MCA 2005 is also clear that the application of idiosyncratic values should not bias the assessment of use or weigh, stating that *P "is not to be treated as unable to make a decision merely because he makes an unwise decision"* (s.1(4)). In other words, the law aspires to be value-liberal – supporting the right of patients to act for "*any reason, rational or irrational, or for no reason at all*" (Re MB [1997]; Re T [1992] ER 649). As an example, we can consider the recent case of "C", who refused treatment for dialysis despite good prognosis because she believed it would compromise her freewheeling lifestyle. In a statement to the court, her daughter wrote: *"Although they are not reasons that are easy to understand, I believe that they are not only fully thought through, but also entirely in keeping with both her (unusual) value system and her (unusual) personality"* (Kings College Hospital NHS Foundation Trust v C and V [2015] EWCOP 80).

The use or weigh requirement refers to the *process* of making a decision, rather than the *outcome* of a decision. Evaluating how such a decision process unfolds is typically contingent on *new* information being provided to P in the form of information or advice – either formally (in the case of interactions with health professionals) or informally (from friends and family). For instance, in the case extract with which we began this article, we see that the court's impression of Z's decision-making capacity is informed by her excessive susceptibility to the influence of others, suggesting that she did not truly use or weigh information herself.

In this article we focus on the role that decision neuroscience might play in informing our understanding of the advice-taking process, and thereby how we conceptualize the use or weigh requirement. Because the process of advice-taking can be simulated in a laboratory setting, it forms a natural bridge between the literatures on decision neuroscience (e.g. Campbell-Meiklejohn, Bach, Roepstorff, Dolan, & Frith, 2010; Campbell-Meiklejohn, Simonsen, Frith, & Daw, 2017; De Martino, Bobadilla-Suarez, Nouguchi, Sharot, & Love, 2017; Izuma & Adolphs, 2013) and mental capacity (Gomez-Beldarrain, Harries, Garcia-Monco, Ballus, & Grafman, 2004; Kennedy, Dornan, Rutledge, O'Neill, & Kennedy, 2009). Another advantage of bridging between work on DMC and advice-taking is that decision neuroscience offers the promise of a deeper and more nuanced understanding of *why* certain decision processes may be impaired due to "*an impairment of the mind or brain*".

The outline of this article is as follows. First, we start with briefly reviewing the general information-processing view of the mind in cognitive science that grounds a majority of work in decision neuroscience. Second, we review studies of the cognitive and neural basis of simple decisions, before exploring how these frameworks can be extended to accommodate advice-taking and so-called *post-decisional* processing. We then consider different types of advice-taking behaviour (such as how public compliance may differ from private belief change) and outline the contribution of metacognition (or insight) and mentalising in advice-taking. Finally, we explore how these lines of work in the neurosciences may inform an understanding of the psychological basis of the use or weigh requirement.

Before we begin, we should point out that the *information-processing* view of most current neuroscientific studies of decision-making is simplistic compared to the often complex, emotionally-charged decisions about which capacity is typically questioned. Further, it is not known how the core principles of decision-making elucidated in the lab generalise to these situations. For example, decisions that usually give rise to capacity assessments do not have an objectively 'correct' answer (e.g. treatment decisions, financial decisions, sexual consent), an issue we will return to towards the end of the article. On balance, however, we believe that there are benefits from adopting a decision neuroscience perspective, not least in providing a value-liberal framework with which to begin to quantify the process, rather than the content, of a decision – i.e. the internal workings of the mind of an individual who is making the decision and not what preferences or values constitute this process.

An example of value-based practise in health care is that of a patient who decides not to engage in treatment because the treatment reduces the patient's ability to engage in activity X (e.g. work as a primary school teacher). Good medical practise is to judge this patient's DMC in a way that is unbiased by the patient's personal preferences, e.g. the content of X (Petrova, Dale, & Bill, 2006); even better medical practise would be to judge the patient's DMC based on the process through which the preference for X was generated, in a way that is impartial to the content of X. Our framework does not say what people should choose, but suggests means of support to encourage people to use and weigh new information in light of their current beliefs, with a mechanistic model of decision-making as a guide.

2. Cognitive and computational approaches to decision-making

The dominant model in the cognitive science of decision-making is that a decider receives some kind of input (from the senses or from memory) and then processes these inputs to arrive at a discrete choice. An agent is thought to reduce uncertainty about which decision to make by sampling information about the potential benefits and disadvantages of the various available options (Gold & Shadlen, 2007). The goal of decision neuroscience is to work out what kind of internal processes govern this transformation of input into output.

Research in this field has focused on two different types of decision. First, research on "perceptual decision-making" (PDM) builds on classical psychophysics from the late 19th century (e.g. Peirce & Jastrow, 1884), and aims to understand how subjects discriminate different types of sensory information (such as deciding whether an object is an apple or an orange from only a brief glimpse; see Hanks & Summerfield, 2017, for a recent review). In contrast, research on "value-based decision-making" (VDM) grew out of the field of behavioural economics, and studies the processes involved in choices involving a comparison of subjective preferences (such as deciding whether to eat an apple or an orange; see Rangel, Camerer, & Montague, 2008). In both cases, the decision process is often studied under conditions of uncertainty, such as when sensory information is noisy or ambiguous, or when value-based choices are made about stimuli that deliver a variable and changing level of reward.

A key benefit of studying PDM is that an explicitly 'correct' answer exists, allowing a precise quantification of the speed and accuracy of individual choices. An influential task in studies of PDM is the random-dot motion task (Fig. 1a). Random-dot motion stimuli consist of a rapidly moving cloud of dots presented briefly on a computer screen (typically for less than a second), and a given proportion of the dots are arranged to move coherently in a particular direction (e.g. left or right; Britten, Shadlen, Newsome, & Movshon, 1992; Kim & Shadlen, 1999). On each 'trial' of the task subjects are asked to decide whether the dot cloud is mostly moving in one or other direction. It is striking that behaviour in these kinds of tasks can be precisely accommodated by computational models that assume the brain receives noisy *samples* of evidence about the world (e.g. whether the dots are moving left or right), and compares these samples to an internal decision threshold (Fig. 1c; Vickers, 1979; Luce, 1992; Ratcliff, 1978). When these samples are aggregated over time, these models can predict how long the

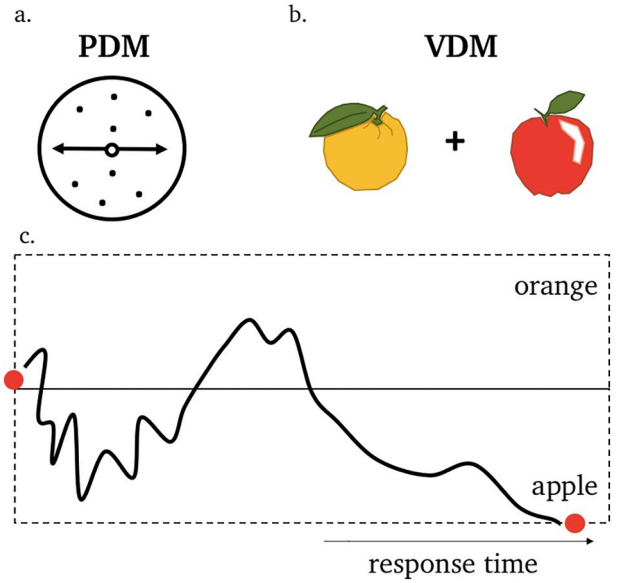

Fig. 1. A schematic example of an evidence accumulation framework for perceptual decision-making (PDM) and value-based decision-making (VDM). (a). A decision about whether a patch of dots is moving to the left or right (PDM), or (b) whether to eat an apple or an orange (VDM), are both studied under cases of uncertainty e.g. when sensory information is noisy or ambiguous or when the value of each object is not perfectly known. (c). Computational models of the decision process assume that the brain receives noisy samples of evidence about the world and compares the accumulated evidence to an internal threshold for each available option (the threshold for when accumulated value results in a selection of an apple or an orange, for instance). When these samples are aggregated over time, such models can predict both choice and how long the observer takes to make a decision (response time).

observer takes to make a decision (Smith & Ratcliff, 2004; Wald, 1947) and how confident they are in their choice (Kiani & Shadlen, 2009).

Similar evidence accumulation models have been successfully applied to model both decision time and choice in VDM (Basten, Biele, Heekeren, & Fiebach, 2010; Hutcherson, Bushong, & Rangel, 2015; Krajbich, Armel, & Rangel, 2010; Krajbich, Hare, Bartling, Morishima, & Fehr, 2015). A classic example is the food choice task, on which hungry subjects are presented with images of food items that can be selected for consumption at the end of the experiment (Fig. 1b). How much time it takes for someone to decide is an indication of the subjective difficulty of the decision (which is usually higher when the two options are more similar in value; Krajbich et al., 2010; Fig. 1c). These findings imply that subjects also have internal uncertainty about their own preferences – for instance how confident they are in preferring apples over oranges (De Martino, Fleming, Garrett, & Dolan, 2013; Manski, 1977). We will return to this important point later in the article.

In the context of value-based decision-making, decision neuroscience has also emphasised a distinction between whether individuals tend to act *habitually*; recalling the value that has often been attached to a previous action in the past, or in a way that foresees the future consequences of sequences of actions (sometimes known as *model-based* planning; Valentin, Dickinson, & Doherty, 2007). While this distinction is beyond the scope of the current article, it is plausible that the depth of planning may intersect with advice-taking in interesting ways (e.g. an individual who is myopic in their planning may be over-responsive to advice).

3. Decision-making in the brain

Following the creation of simple paradigms for studying decision-making in the lab, there has been increasing interest in the neural basis of these types of decisions (Gold & Shadlen, 2007). We refer the reader to comprehensive reviews on the neural basis of PDM and VDM (Glimcher & Fehr, 2013; Hanks & Summerfield, 2017), and here focus on a few core principles that have emerged from this work. To a rough approximation, different subregions of the cortex are engaged in different, specialized computations. For instance, the visual cortex, in the posterior temporal and occipital lobes, is specifically involved in processing inputs from the eyes to support visual perception. A series of studies in animals and humans over the past two decades has begun to bridge the gap between perception and action to understand how simple choices are made. In studies of PDM, for instance, neural populations in the association cortex – the parietal and frontal lobes – have been found to gradually increase their firing rate up to a point at which a decision is made (Heekeren, Marrett, Bandettini, & Ungerleider, 2004; O'Connell, Dockree, & Kelly, 2012; Roitman & Shadlen, 2002). The idea is that these activity patterns receive input from areas processing sensory input (for instance area V5 of the visual cortex in the case of the random dot motion task), and represent a neural instantiation of the evidence accumulation process inherent to the computational frameworks introduced above.

Intriguingly, the same circuits are also involved when people make more subjective choices in VDM paradigms (e.g. Platt & Glimcher, 1999). Rather than integrating sensory information such as direction of motion, subjective values (such as how much I prefer apples over oranges) are thought to feed into these choice circuits until a decision is made. It is thought that the medial prefrontal cortex, in conjunction with regions important for memory such as the hippocampus (Barron, Dolan, & Behrens, 2013), retrieves information about past experience to construct these subjective values (Chib, Rangel, Shimojo, & Doherty, 2009; Glimcher, 2011; Rangel et al., 2008). On these tasks, there is no objectively 'correct' answer, but difficulty can be manipulated by changing the difference in subjective value between the two choice options. In the next section we turn to how these studies of information processing in the service of simple decisions help shed light on the use or weigh process during advice-taking.

4. Post-decisional evidence processing as a model of advice-taking

A majority of work in decision neuroscience has focused on characterising the processes leading up to a value-based or perceptual choice, but less attention has been paid to how people integrate new information *after* a choice has been made, or how this post-decisional processing may lead to a change of mind. Initial work in this area characterised how evidence accumulation may continue for a short time after the choice, supporting endogenous error monitoring (Murphy, Robertson, Harty, & O'Connell, 2015; Rabbitt, 1966) and changes of mind (Resulaj, Kiani, Wolpert, & Shadlen, 2009; van den Berg et al., 2016). More recently this line of work has been extended to ask how *exogenous evidence* presented after an initial choice may lead to later changes of mind.

In these tasks, subjects first make a judgment based on some evidence (e.g. an estimation of the direction of a random-dot motion display), after which they are presented with *new* evidence (e.g. additional motion) and are asked to make a final judgment. The general finding is that people tend to update their final judgment after seeing the new evidence and that this updating is stronger when the new evidence is more reliable or stronger (Bronfman et al., 2015; Fleming, van der Putten, & Daw, 2018), or when the new evidence confirms the initial judgment (known as 'confirmation bias'; Talluri, Urai, Tsetsos, Usher, & Donner, 2018).

For example, Fleming et al. (2018) asked subjects to make a first judgment about whether a random-dot motion display with variable evidence strength was mostly moving to the left or right. After this decision, people were shown an additional display of motion in the same direction as the first display (post-decision evidence) which the subjects could use to update their confidence about whether their initial

decision was correct. The authors also scanned the brains of participants while they received the new evidence using *functional Magnetic Resonance Imaging* (fMRI), which tracks fluctuations in blood flow (a proxy of neuronal activity) in different regions of the brain. This allowed the identification of regions in which activity fluctuations tracked whether or not people changed their minds. The posterior medial frontal cortex (pMFC), a region that activates when people detect that they have made a mistake and adjust their behaviour accordingly (Dehaene, Posner, & Tucker, 1994; Ridderinkhof, van den Wildenberg, Segalowitz, & Carter, 2004), tracked the strength of post-decision evidence and signalled the need for behavioural adaptation. In contrast, activity in the anterior prefrontal cortex (aPFC), a region involved in metacognition (the ability to know whether we are right or wrong, see below) mediated the impact of new evidence on people's subjective confidence.

Together, these studies have contributed to an emerging understanding of how new evidence is processed to flexibly change an initial judgment when needed. Interestingly, these studies have also shown that the *way* in which new evidence is integrated is similar across distinct tasks (e.g. perceptual and numerical tasks; Bronfman et al., 2015; Talluri et al., 2018). Ongoing research is now building upon this finding by examining how new evidence is integrated from both non-social and social information sources, and whether non-social and social post-decision evidence provide similar information content (Olsen, Roepstorff & Bang, 2018; Pescetelli & Yeung, 2018). These paradigms ask subjects to make a first (perceptual or value-based) decision and then present the opinion of an 'adviser' (Behrens, Hunt, Woolrich, & Rushworth, 2008; Campbell-Meiklejohn et al., 2010; Campbell-Meiklejohn et al., 2017; De Martino et al., 2017; Gomez-Beldarrain et al., 2004; Sniezek & Van Swol, 2001). Interestingly, and in line with the finding that more reliable evidence elicits more changes of mind (Bronfman et al., 2015; Fleming et al., 2018; Talluri et al., 2018), the reliability of the advice is a crucial determinant of how much it engenders a change in subjects' beliefs. In advice-taking settings, this advice reliability can be communicated in the form of the confidence of the adviser (Campbell-Meiklejohn et al., 2010; Campbell-Meiklejohn et al., 2017; Gomez-Beldarrain et al., 2004; Sniezek & Van Swol, 2001), as judgments made with higher confidence are typically more likely to be correct (Fig. 2). In turn, the effects of the advice are strongest when the initial decision is ambiguous or when an agent is uncertain about their initial judgment (Cialdini & Goldstein, 2004; De Martino et al., 2017). This information-processing view of advice gives rise to useful predictions for when advice is most likely to have an impact on its recipient, in that new (sensory or social) evidence would be expected to have a larger impact on uncertain agents. Confidence is often distorted in psychiatric disorders (e.g. more or less confidence than warranted by performance; Rouault, Seow, Gillan, & Fleming, 2018; Stephan et al., 2016); this may explain why some patients have been reported to be over- or under-susceptible to advice, a topic to which we will return in the next section. Together, these studies provide initial evidence that social advice is processed similarly to post-decisional evidence in mediating changes of mind.

5. The role of metacognition in advice-taking

In the previous section we outlined how changes of mind and improvements in performance are tightly coupled: being sensitive to new evidence is especially beneficial when one's initial judgment is incorrect in order to engender a change of mind. However, even if confidence and accuracy are usually tightly linked, individuals vary in the ability to monitor their performance, known as *metacognition*. Specifically, some people are less able to notice differences in their performance, or believe themselves to be more or less accurate than they objectively are (Rouault, Seow, et al., 2018). It has recently become possible to precisely quantify a person's ability to discriminate between their own correct and incorrect judgments (known as "*metacognitive sensitivity*") using a variant of signal detection theory (Fleming & Lau, 2014). There are various ways of calculating metacognitive sensitivity, with more sophisticated methods being uncontaminated by differences in both accuracy and *metacognitive bias,* a person's overall propensity to report high confidence (Fleming, Weil, Nagy, Dolan, & Rees, 2010; Rouault, Seow, et al., 2018).

Individual differences in metacognitive sensitivity have been found to correlate with the structure and function of the aPFC, a brain area involved in self-evaluation (Christoff & Gabrieli, 2000; Fleming, Huijgen & Dolan, 2012; Hilgenstock, Weiss, & Witte, 2014; Shimamura & Squire, 1986). This region together with the pMFC (discussed above in the context of advice-taking) is reliably activated during the self-evaluation of performance (Fleming & Dolan, 2012). It is notable that areas such as pMFC and aPFC are implicated in both advice-taking and the metacognitive evaluation of recent decisions, suggesting an important link between these two processes. The general link is unsurprising given that knowing whether we are right or wrong (metacognition) is one of the features that enable us to know when new evidence (advice-taking) is beneficial. For example, if you are completely convinced about the best course of action, you may be less likely to take someone's advice or consider another approach. On the other hand, if you are unsure about your decision, you may be more likely to look at what others are doing and follow their advice (Fig. 2 illustrates the general relationship between confidence and advice-taking). In sum, as advice-taking is most beneficial when one's first judgment was wrong, it follows that metacognition is one of the features that

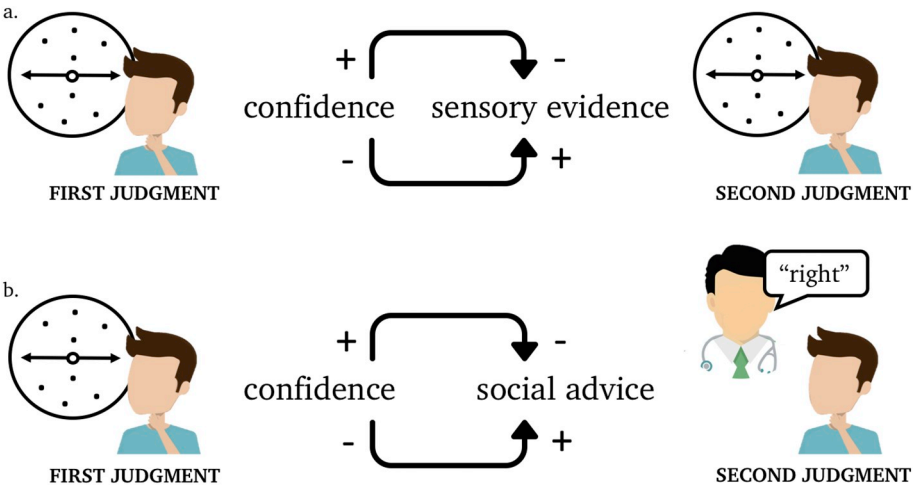

Fig. 2. People who are more certain about an initial judgment are less likely to change their minds upon being presented with new sensory evidence or social advice. A decider receives a first sample of sensory evidence on a random-dot motion task (first judgment) and is then presented with either new sensory evidence (a) or social advice (b). Based on this new evidence, the decider makes a second judgment. Confidence in the first judgment affects how likely the decider is to take on board the new evidence (integrating the new sensory evidence or following the advice). More (versus less) confident decisions are less (versus more) likely to be updated when new sensory or social evidence becomes available.

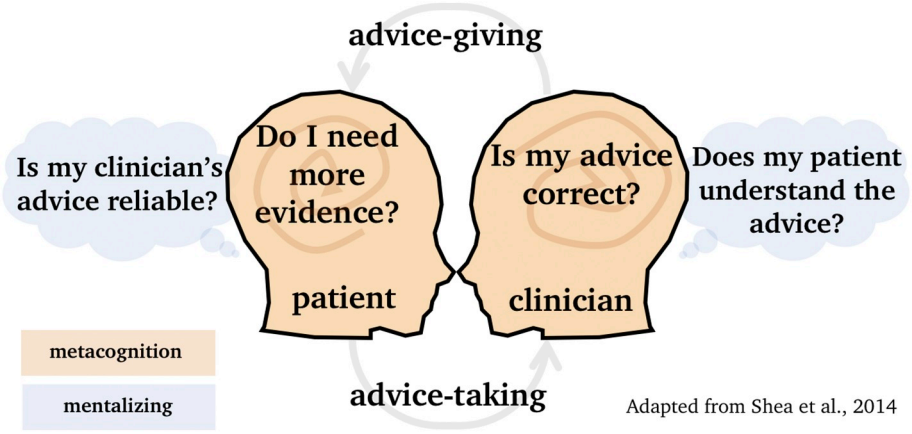

Fig. 3. A schematic illustration of how mentalising and metacognition may play a role in the advice-taking and advice-giving process between a patient and a clinician. Metacognition (indicated in orange) may be used by the patient to understand whether additional evidence is needed (how confident the patient is that their own opinion is correct and that their perceptions are 'real'). For the clinician, metacognition is involved in expressing the right level of confidence in the advice (how confident the clinician is that their advice is correct). Mentalising (indicated in blue) plays a role in understanding *why* someone is deciding one way or another, which may be relevant for a clinician to know whether a patient is thinking through the consequences of different courses of action and whether the advice needs to be communicated differently to achieve a given level of influence. For the patient, mentalising may be used to understand the fidelity of the adviser's confidence (the clinician's metacognition). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

contributes to knowing when to take on board advice (Frith, 2012; Shea et al., 2014).

In clinical settings, the term insight is used to encompass a patient's awareness that they have a disorder, their ability to reflect on specific beliefs including their response to hypothetical contradiction, and to some extent follow the clinician's advice, for example regarding treatment (Amador & David, 2004; David, 1990). There is also an association between the concept of clinical insight and metacognition (David, Bedford, Wiffen, & Gilleen, 2012; Koren et al., 2005, 2004). Patients deemed to have poor insight often show impaired decision-making capacity (Owen et al., 2009). Furthermore, patients who lack mental capacity are more likely to experience coercion (Cairns et al., 2005). To the extent to which insight depends on metacognitive mechanisms, it may be that poor insight leads to difficulty with the advice-taking process (and vice-versa).

This link between metacognition and advice-taking is supported by a recent study from Rollwage, Dolan, and Fleming (2018), who measured both metacognitive sensitivity and sensitivity to post-decision evidence, and asked how these aspects were related to a personality feature known as *dogmatism* (measured as the extent to which individuals were accepting of conflicting views on political issues). Metacognitive sensitivity predicted the extent to which subjects integrated new evidence on the perceptual task, supporting adaptive changes of mind. In turn, both of these features of decision-making were attenuated among those with higher levels of dogmatism. This finding shows that metacognitive sensitivity may promote adaptive changes of mind when new evidence becomes available. On top of that, this study suggests similar processes may govern the processing of new evidence in both low-level perceptual discrimination tasks and the broader, more subjective decisions about topics such as politics.

6. The role of mentalising in advice-taking

When a decision is difficult to make, it can be advantageous to use the full range of available evidence that is at one's disposal to reduce choice uncertainty, including the advice of others. However, as briefly mentioned at the beginning of this paper, from an information-processing perspective, advice-taking should not only be dependent on the *decider's* confidence but also on the *reliability* of the new evidence (De Martino et al., 2017). In other words, efficient advice-taking involves assessing not only the probability that *oneself* is correct (metacognition) but also the probability that *the adviser* is correct (Burke, Tobler, Baddeley, & Schultz, 2010; De Martino et al., 2017; Harvey & Fischer, 1997; Pescetelli & Yeung, 2019). Advisors' expressions of certainty are typically a useful source of information about the advisers' reliability, as people who say that they are confident are usually also more likely to be correct (Campbell-Meiklejohn et al., 2010; Campbell-Meiklejohn et al., 2017; Gomez-Beldarrain et al., 2004; Sniezek & Van Swol, 2001). This process is complicated when the fidelity of the advisers' confidence ratings is not representative of their accuracy (the *adviser's* metacognitive sensitivity). Put differently, it is often sensible to take advisers' certainty estimates with a pinch of salt (Bahrami et al., 2012, 2010; Bang et al., 2017), and learn, over the course of repeated interactions, which advisers' confidence estimates are more reliable than others (Hertz et al., 2017; Pescetelli & Yeung, 2019). Pescetelli and Yeung (2019) show that, over the course of repeated interactions with an adviser, the extent to which 'trust' relationships are built/broken predict advice-taking. In psychiatric settings, this may be related to trusting treatment providers to give the right advice.

The general set of processes involved in making inferences about others' preferences and mental states is referred to as *mentalising* or Theory of Mind (ToM; Frith & Frith, 2012, 2003). ToM and metacognition may have a shared mechanism (Carruthers, 2009; Fleming & Daw, 2017) that involves understanding that others' world views may differ from our own and using this discrepancy to reason, for example, how reliable others' advice may be (Behrens et al., 2008). This form of ToM develops relatively late in human development (around five years; Wimmer & Perner, 1983) probably through external feedback that shows that others' worldviews can be different from one's own (Gopnik & Wellman, 2010; Heyes & Frith, 2014). In cases of advice-taking about treatment decisions (which are highly dependent on personal subjective preferences) mentalising may play a crucial role in understanding *why* someone decides one way or another. A well-established link between psychiatric disorders and mentalising is supportive of this idea (Pousa et al., 2008; but see, Frith & Happé, 1994), in addition to more recent neuroimaging studies that explain clinical insight as a dysregulated interplay between mentalising and metacognition (Holt et al., 2011; Modinos, Renken, Ormel, & Aleman, 2011). For the patient, mentalising may be involved in understanding that advice can be biased by the dissimilar worldviews of others or the fidelity of the clinician's metacognitive representations. For the clinician, mentalising may be involved in understanding the patient's motives or idiosyncratic preferences. Thus, ToM reasoning in advice-taking is a two-way process – a sound understanding of another person's views is important both for the patient to assess the reliability of the clinician and for the clinician to understand the decision-process of the patient (Fig. 3).

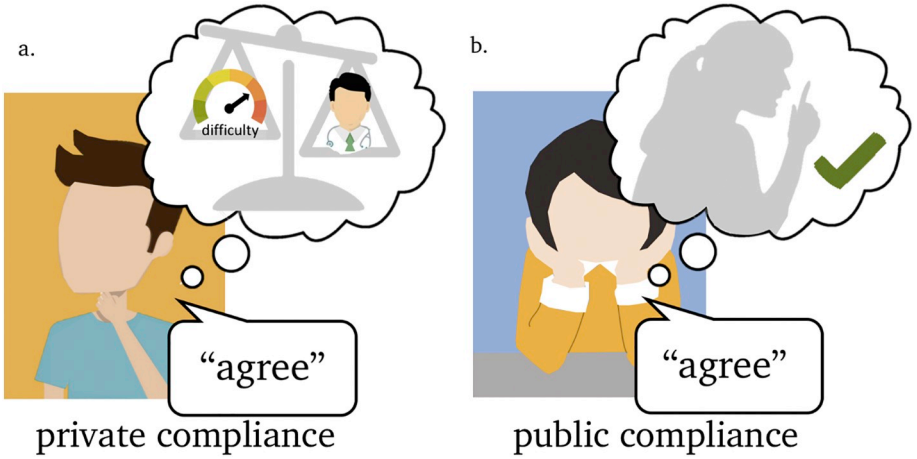

Fig. 4. A schematic illustration of private and public compliance. (a). An example of private compliance, in which the use or weigh process consists of engaging with new evidence (e.g. the reliability of the advice) leading to a shift in private beliefs. (b). An example of public compliance, where advice leads to public statements of agreement without a shift in privately held beliefs. The resulting decision outcome is similar in panels (a) and (b), i.e. in agreement with the advice, but has different consequences for behaviour.

7. Private versus public compliance

Up until this point we have considered how models of evidence integration in simple decisions can be extended to characterise the process of revising and updating a belief on the basis of new information. These *private* changes in belief are thought to affect what the subject truly feels or beliefs about the decision. In contrast, there are plenty scenarios in which subjects *publicly* comply in order to fit in with a social group (Bobek, Hageman, & Kelliher, 2011; Bond & Smith, 1996) or under the influence of an authority figure (Milgram, 1963), but do not alter their privately held beliefs. As an example, consider a clinician who advises a patient to take a particular medication and a patient who complies with this advice. It may be that the patient blindly follows the advice, or choses to do so tactically (e.g. to hasten their discharge from hospital) irrespective of whether (s)he agrees that the treatment is actually needed (Fig. 4b). Here, the outcome of the decision, i.e. following the advice, appears superficially similar to a case in which a patient complies because (s)he appreciates the need for the treatment – i.e. shows private as well as public compliance (Fig. 4a). Despite an apparently similar decision outcome in both cases, the consequences of advice-taking for behaviour may be radically different. For example, in the case of *public compliance* (i.e. the patient outwardly expresses their approval but is not genuinely convinced) a patient may be less likely to follow the advice when the adviser is removed from the situation (e.g. at home when medication needs to be taken). Yet in the case of *private compliance*, the patient's agreement is more likely to persist in time (Edelson, Sharot, Dolan, & Dudai, 2011; Izuma & Adolphs, 2013).

A classic laboratory demonstration of supposedly public compliance is the Asch study, in which subjects were asked to judge the length of lines. When asked privately, subjects had an approximately 99% chance of getting the decision correct. In contrast, when subjects made the same decision in the company of six mock participants who were instructed to give the wrong answer, 40% followed the mock participants – a classic case of succumbing to social advice even when this advice-taking results in an incorrect answer (Asch, 1955). Deutsch and Gerard shed new light on the original interpretation of this behaviour as (automatic) public compliance. They experimentally removed the social pressure from the Asch paradigm, identifying two distinct types of compliance: that which was mitigated when the social pressure was removed (*normative* conformity to social expectations/public compliance), and that which persisted even in the absence of social pressure (accepting others' views because they are *informative*/private compliance; Deutsch & Gerard, 1955). This suggests that advice-taking may be motivated both by a public willingness to conform to social rules and by private convictions. This view has been supported by psychological experiments (Freedman & Fraser, 1966; Sowden et al., 2018) and, more recently, also by computational simulations (Constant, Ramstead, Veissière, & Friston, 2019; Toelch & Dolan, 2015).

These studies suggest that the decision processes underpinning public and private compliance can be distinguished using experimental techniques, and that advice-taking often consists of combining these two types of belief change. This distinction is supported by studies that have investigated the neural basis of public versus private compliance. For example, Izuma and Adolphs (2013) asked students to update value-based decisions based on the advice of a liked or disliked group, and tested whether the updated preferences still persisted four months later. Strikingly, the researchers found that, when activity in the dorsomedial prefrontal cortex (dmPFC) was low, the subjects showed public compliance that did not lead to a persistent shift in preferences. In contrast, when dmPFC activation was high, the shift in preference was persistent — representing a neural signature of private belief change. A similar dmPFC region has also been shown to take into account the reliability of other's views about everyday products in the form of Amazon star ratings (De Martino et al., 2017) and is involved in observational learning, where it tracks the actions of others (Burke et al., 2010). Together these studies suggest a role of the dmPFC in using or weighing others' advice in a way that is informed by its reliability (Behrens et al., 2008; Campbell-Meiklejohn et al., 2010; Campbell-Meiklejohn et al., 2017), which may be more relevant to informational (private) motivations than to social (public) motivations.

8. Neuro-computational insights into using or weighing

What is meant by the ability to "use or weigh" in the context of DMC? The contentious status of use or weigh in cases in which capacity is in dispute suggests that a perspective from decision neuroscience may shed light on the kinds of processes that support the ability to use or weigh decision evidence as per the MCA requirement. In the current article we have argued that advice-taking provides a "model system" for understanding the use and weigh requirement, given that how individuals respond to (formal, informal or hypothetical) advice is often at the heart of both capacity assessments and capacity disputes.

In this last section we consider how insights from decision neuroscience can be generalised to more complex, real-world decision problems. Most studies of advice-taking have used simple PDM tasks in which there is an objectively correct answer (such as the random-dot motion task). It remains to be seen how post-decision processing and advice-taking operate in tasks such as the food choice task introduced earlier as a canonical example of VDM. More broadly, there is a substantial challenge in translating insights from laboratory decision-making paradigms into the emotionally-charged, often one-off decisions faced by individuals subject to capacity assessments. However, studies of moral decision-making suggest that similar domain-general

evidence-accumulation frameworks also hold for even the most serious choice problems (Shenhav & Greene, 2014). For instance, the classic *trolley dilemma* asks how people trade-off the motivations to save certain lives at the cost to others' (Fischer & Ravizza, 1992; Foot, 1967; Thomson, 1985). People with higher dorsolateral prefrontal cortex (dlPFC) activity are more outcome-based, in that they are more likely to consider killing one to save five morally justifiable (Greene, Nystrom, Engell, Darley, & Cohen, 2004; Greene, Sommerville, Nystrom, Darley, & Cohen, 2001), possibly because the dlPFC is involved in suppressing automatic emotional responses in complex social situations (Sanfey, Rilling, Aronson, Nystrom, & Cohen, 2013). Importantly however, when the dilemmas have actual consequences or when the act of killing is made more salient, people who first approve of the more outcome-based decision often change their preferences (Gold, Pulford, & Colman, 2015; Thomson, 1985), suggesting an important role of emotions in changes of mind about moral issues (for a possible computational basis, see: Crockett, 2013). Indeed, some aspects of the advice-taking process – such as metacognition and mentalising – may provide an enabling context that supports the ability to use or weigh decision evidence about a range of decision problems. We focus here on three aspects of metacognition that exemplify this point:

1) Metacognition is a broadly domain-general resource – having good metacognition on one task predicts having good metacognition on another, unrelated task (McCurdy et al., 2013; Song et al., 2011; Mazaniceux, Fleming, Souchay, & Moulin, 2018), and this domain-generality may be supported by abstract representations of self-performance in anterior prefrontal cortex (Morales, Lau, & Fleming, 2018; Rouault, Mcwilliams, Allen, & Fleming, 2018). Understanding how metacognition operates in laboratory situations may therefore inform how metacognition (and its failures) affect advice-taking in real-world contexts.
2) Metacognitive sensitivity is not necessarily associated with greater intelligence or intellectual ability. In a recent study, we collected data from almost 1000 individuals in a general population sample, quantifying both their metacognitive awareness of making good or bad perceptual choices, and a proxy for IQ (Rouault, Seow, et al., 2018). While there was a strong link between IQ and the speed/accuracy of the initial decision, there was no relationship between IQ and people's ability to evaluate their decision-making. Metacognitive sensitivity may therefore track an individual's awareness of their decision process independently of whether it is determined by a particular set of values or intellectual style.
3) People have awareness of the consistency or accuracy of even value-based, subjective choices. In other words, people have a sense of when they are making a good choice, and this self-knowledge may share a similar neural basis to metacognition about other types of decision (De Martino et al., 2013). In turn, people with greater metacognitive sensitivity about value-based choices tend to make more consistent decisions over time (Folke, Jacobsen, Fleming, & De Martino, 2016). This self-knowledge about our preferences may be akin to a "higher-order desire" that endorses our first-order motives – we "want to want" something. Frankfurt (1971) identifies the presence of such second-order desires with autonomy, and empirical work on metacognition about VDM may inform an emerging understanding of autonomy and agency in DMC. This perspective holds promise for integrating work on metacognition of decision-making (as indexing second-order desires) and notions of free will and autonomy that are important in relation to decision-making capacity (Zürcher, Elger, & Trachsel, 2019). Indeed, it seems that practitioners "on the ground" often inform their assessment of whether a patient is using or weighing decision-evidence according to whether an individual has insight into the relevant evidence supporting a particular choice, and can reflect on the consequences of their actions (Owen, Freyenhagen, Hotopf, & Martin, 2015).

More broadly, the interplay between metacognition, mentalising and advice-taking processes may provide a useful conceptual framework for thinking about apparent "failures" of DMC. For instance, one could usefully ask whether the lack of capacity of Z in the opening example is due to a problem with mentalising, metacognition about her own preferences, or with the processes involved in incorporating social influence with her own beliefs (either privately or publicly). Knowledge of this sort could be used to provide targeted support, e.g. by providing metacognitive training (Carpenter et al., 2019) or by seeking to reduce or remove social pressures. For instance, Owen, Freyenhagen, Richardson, and Hotopf (2009) note that in the case of delusions, patients' metacognition may be intact, in that deluded beliefs and behaviours may be entirely self-consistent. However, mentalising may be less so, in the sense that delusions by definition do not map onto a shared, social reality. A finer-grained conceptualisation of the neuro-computational processes supporting advice-taking will, we hope, provide a framework for discussing such cases.

Author contributions

All authors conceptualized the paper. E.v.d.P. and S.M.F. wrote the paper with substantial comments from A.S.D.

Declarations of Competing Interest

None.

Funding

The authors are all core researchers in the 'metacognition' workstream of the Mental Health and Justice Project, which is funded by a grant from the Wellcome Trust (203376/2/16/Z). The Wellcome Centre for Human Neuroimaging is supported by core funding from the Wellcome Trust (203147/Z/16/Z). S.M.F. is supported by a Sir Henry Dale Fellowship jointly funded by the Wellcome Trust and Royal Society (206648/Z/17/Z). A.S.D. is also supported by the National Institute of Health Research Biomedical Research Centre at UCLH.

List of included cases

WBC v Z [2016] EWCOP 4.
Kings College Hospital NHS Foundation Trust v C and V [2015] EWCOP 80.
Re MB [1997] EWCA 3093.
Re T [1992] All ER 649.

References

Amador, X. F., & David, A. S. (2004). *Insight and Psychosis: Awareness of Illness in Schizophrenia and Related Disorders.* https://doi.org/10.1093/med/9780198525684.001.0001.

Asch, S. (1955). Opinions and social pressure. *Nature, 193*(5), 31–35. https://doi.org/10.1038/1761009b0.

Bahrami, B., Olsen, K., Bang, D., Roepstorff, A., Rees, G., & Frith, C. (2012). What failure in collective decision-making tells us about metacognition. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 367*(1594), 1350–1365. https://doi.org/10.1098/rstb.2011.0420.

Bahrami, B., Olsen, K., Latham, P. E., Roepstorff, A., Rees, G., & Frith, C. D. (2010). Optimally interacting minds. *Science, 329*(5995), https://doi.org/10.1126/science.1185718.

Bang, D., Aitchison, L., Moran, R., Herce Castanon, S., Rafiee, B., Mahmoodi, A., ... Summerfield, C. (2017). Confidence matching in group decision-making. *Nature Human Behaviour, 1*(6), 1–7. https://doi.org/10.1038/s41562-017-0117.

Barron, H. C., Dolan, R. J., & Behrens, T. E. J. (2013). Online evaluation of novel choices by simultaneous representation of multiple memories. *Nature Neuroscience, 16*(10), 1492–1498. https://doi.org/10.1038/nn.3515.

Basten, U., Biele, G., Heekeren, H. R., & Fiebach, C. J. (2010). How the brain integrates costs and benefits during decision making. *Proceedings of the National Academy of Sciences, 107*(50), 21767–21772. https://doi.org/10.1016/0020-711X(72)90050-X.

Behrens, T. E. J., Hunt, L. T., Woolrich, M. W., & Rushworth, M. F. S. (2008). Associative learning of social value. *Nature, 456*, 245–249. https://doi.org/10.1038/

nature07538.
van den Berg, R., Anandalingam, K., Zylberberg, A., Kiani, R., Shadlen, M. N., & Wolpert, D. M. (2016). A common mechanism underlies changes of mind about decisions and confidence. *ELife*. https://doi.org/10.7554/elife.12192.
Bobek, D. D., Hageman, A. M., & Kelliher, C. F. (2011). The social norms of tax compliance: Scale development, social desirability, and presentation effects. *Advances in Accounting Behavioral Research, 74*(1), 49–64. https://doi.org/10.1108/S1475-1488(2011)0000014005.
Bond, R., & Smith, P. B. (1996). Culture and conformity: A meta-analysis of studies using asch's (1952b, 1956) line judgment task. *Psychological Bulletin, 119*(1), 111–137. https://doi.org/10.1037/0033-2909.119.1.111.
Britten, K. H., Shadlen, M. N., Newsome, W. T., & Movshon, J. A. (1992). The analysis of visual motion: A comparison of neuronal and psychophysical performance. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience. 12*(12), 4745–4765. https://doi.org/10.1523/JNEUROSCI.12-12-04745.1992.
Bronfman, Z. Z., Brezis, N., Moran, R., Tsetsos, K., Donner, T., Usher, M., & Usher, M. (2015). Decisions reduce sensitivity to subsequent information. *Proceedings of the Biological Sciences, 282*(1810), 1–45. https://doi.org/10.1098/rspb.2015.0228.
Burke, C. J., Tobler, P. N., Baddeley, M., & Schultz, W. (2010). Neural mechanisms of observational learning. *Proceedings of the National Academy of Sciences, 107*(32), 14431–14436. https://doi.org/10.1073/pnas.1003111107.
Cairns, R., Buchanan, A., David, A. S., Hayward, P., Richardson, G., & Szmukler, G. (2005). Prevalence and predictors of mental incapacity in psychiatric in-patients. *British Journal of Psychiatry, 187*(4), 379–385. https://doi.org/10.1192/bjp.187.4.379.
Campbell-Meiklejohn, D., Bach, D. R., Roepstorff, A., Dolan, R. J., & Frith, C. D. (2010). Report how the opinion of others affects our valuation of objects. *Current Biology, 20*(13), 1165–1170. https://doi.org/10.1016/j.cub.2010.04.055.
Campbell-Meiklejohn, D., Simonsen, A., Frith, C. D., & Daw, N. D. (2017). Independent neural computation of value from other people's confidence. *The Journal of Neuroscience, 37*(3), 673–684. https://doi.org/10.1523/JNEUROSCI.4490-15.2017.
Carpenter, J., Sherman, M. T., Kievit, R. A., Seth, A. K., Lau, H., & Fleming, S. M. (2019). Domain-general enhancements of metacognitive ability through adaptive training. *Journal of Experimental Psychology: General, 148*(1), 51–64. https://doi.org/10.1037/xge0000505.
Carruthers, P. (2009). How we know our own minds: The relationship between mind-reading and metacognition. *Behavioral and Brain Sciences, 32*(2), 121–138. https://doi.org/10.1017/S0140525X09000545.
Case, P. (2016). Negotiating the domain of mental capacity: Clinical judgment or judicial diagnosis? *Medical Law International, 16*(3-4), 174–205. https://doi.org/10.1177/0968533216674047.
Chib, V. S., Rangel, A., Shimojo, S., & Doherty, J. P. O. (2009). Evidence for a common representation of decision values for dissimilar goods in human ventromedial prefrontal cortex. *The Journal of Neuroscience, 29*(39), 12315–12320. https://doi.org/10.1523/JNEUROSCI.2575-09.2009.
Christoff, K., & Gabrieli, J. D. E. (2000). The frontopolar cortex and human cognition: Evidence for a rostrocaudal hierarchical organization within the human prefrontal cortex. *Psychobiology, 28*(2), 168–186. https://doi.org/10.3758/BF03331976.
Cialdini, R. B., & Goldstein, N. J. (2004). Social influence: Compliance and conformity. *Annual Review of Psychology, 55*, 591–621. https://doi.org/10.1146/annurev.psych.55.090902.142015.
Constant, A., Ramstead, M. J. D., Veissière, S. P. L., & Friston, K. (2019). Regimes of expectations: An active inference model of social conformity and human decision making. *Frontiers in Psychology, 10*, 679. https://doi.org/10.3389/fpsyg.2019.00679.
Crockett, M. J. (2013). Models of morality. *Trends in Cognitive Sciences, 17*(8), 363–366. https://doi.org/10.1016/j.tics.2013.06.005.
David, A. S. (1990). Insight and psychosis. *British Journal of Psychiatry, 156*, 798–808.
David, A. S., Bedford, N., Wiffen, B., & Gilleen, J. (2012). Failures of metacognition and lack of insight in neuropsychiatric disorders. *Philosophical Transactions of the Royal Society B, 267*(1594), 1379–1390.
De Martino, B., Bobadilla-Suarez, S., Nouguchi, T., Sharot, T., & Love, B. C. (2017). Social information is integrated into value and confidence judgments according to its reliability. *The Journal of Neuroscience, 37*(25), 6066–6074. https://doi.org/10.1523/jneurosci.3880-16.2017.
De Martino, B., Fleming, S. M., Garrett, N., & Dolan, R. J. (2013). Confidence in value-based choice. *Nature Neuroscience, 16*(1), 105–110. https://doi.org/10.1038/nn.3279.Confidence.
Dehaene, S., Posner, M. I., & Tucker, D. M. (1994). Localization of a neural system for error detection and compensation. *Psychological Science, 5*(5), 303–305.
Deutsch, M., & Gerard, H. B. (1955). A study of normative and informational social influences upon individual judgment. *Journal of Abnormal and Social Psychology, 51*(3), 629–636. https://doi.org/10.1037/h0046408.
Edelson, M., Sharot, T., Dolan, R. J., & Dudai, Y. (2011). Following the crowd: Brain substrates of long-term memory conformity. *Science, 333*(6038), 108–111. https://doi.org/10.1126/science.1203557.
Fischer, J. M., & Ravizza, M. (1992). *Responsibility, freedom, and reason. Ethics, problems and principles*. Forth Worth, T. X: Harcourt Brace Jovanovich College Publishers.
Fleming, S. M., & Daw, N. D. (2017). Self-evaluation of decision-making: A general Bayesian framework for metacognitive computation. *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. https://doi.org/10.1098/rstb.2011.0417.
Fleming, S. M., Huijgen, J., & Dolan, R. J. (2012). Prefrontal contributions to metacognition in perceptual decision making. *Journal of Neuroscience, 32*(18), 6117–6125. https://doi.org/10.1523/JNEUROSCI.6489-11.2012.
Fleming, S. M., & Lau, H. C. (2014). How to measure metacognition. *Frontiers in Human Neuroscience, 8*, 1–9. https://doi.org/10.3389/fnhum.2014.00443.
Fleming, S. M., van der Putten, E. J., & Daw, N. D. (2018). Neural mediators of changes of mind about perceptual decisions. *Nature Neuroscience, 21*(4), 617–624. https://doi.org/10.1038/s41593-018-0104-6.
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. https://doi.org/10.1126/science.1191883.
Folke, T., Jacobsen, C., Fleming, S. M., & De Martino, B. (2016). Explicit representation of confidence informs future value-based decisions. *Nature Human Behaviour, 1*, 17–19.
Foot, P. (1967). The problem of abortion and the doctrine of the double effect. *Virtues and vices. Oxford review*. Blackwell.
Frankfurt, H. G. (1971). Freedom of the will and the concept of a person. *The Journal of Philosophy*. https://doi.org/10.2307/2024717.
Freedman, J. L., & Fraser, S. C. (1966). Compliance without pressure: The foot-in-the-door technique. *Journal of Personality and Social Psychology, 68*(1), 5–20. https://doi.org/10.1037/h0023552.
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. https://doi.org/10.1098/rstb.2012.0123.
Frith, C. D., & Frith, U. (2012). Mechanisms of social cognition. *Annual Review of Psychology, 63*, 287–313. https://doi.org/10.1146/annurev-psych-120710-100449.
Frith, U., & Frith, C. D. (2003). Development and neurophysiology of mentalising. *Philosophical Transactions of the Royal Society, B: Biological Sciences, 358*(1431), 459–473. https://doi.org/10.1098/rstb.2002.1218.
Frith, U., & Happé, F. (1994). Autism: Beyond "theory of mind". *Cognition, 50*(1–3), 115–132. https://doi.org/10.1016/0010-0277(94)90024-8.
Glimcher, P. W. (2011). Understanding dopamine and reinforcement learning: The dopamine reward prediction error hypothesis. *Proceedings of the National Academy of Sciences, 108*(3), 15647–15654. https://doi.org/10.1073/pnas.1014269108.
Glimcher, P. W., & Fehr, E. (2013). *Neuroeconomics: Decision making and the brain* (2nd ed.). https://doi.org/10.1016/C2011-0-05512-6.
Gold, J. I., & Shadlen, M. N. (2007). The neural basis of decision making. *Annual Review of Neuroscience, 30*(1), 535–574. https://doi.org/10.1146/annurev.neuro.29.051605.113038.
Gold, N., Pulford, B. D., & Colman, A. M. (2015). Do as I say, Don't do as I do: Differences in moral judgments do not translate into differences in decisions in real-life trolley problems. *Journal of Economic Psychology, 47*, 50–61. https://doi.org/10.1016/j.joep.2015.01.001.
Gomez-Beldarrain, M., Harries, C., Garcia-Monco, J. C., Ballus, E., & Grafman, J. (2004). Patients with right frontal lesions are unable to assess and use advice to make predictive judgments. *Journal of Cognitive Neuroscience, 16*(1), 74–89. https://doi.org/10.1162/089892904322755575.
Gopnik, A., & Wellman, H. M. (2010). The theory theory. *Mapping the mind*https://doi.org/10.1017/cbo9780511752902.011.
Greene, J. D., Nystrom, L. E., Engell, A. D., Darley, J. M., & Cohen, J. D. (2004). The neural bases of cognitive conflict and control in moral judgment. *Neuron, 44*, 389–400. https://doi.org/10.1016/j.neuron.2004.09.027.
Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., & Cohen, J. D. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science, 239*, 2105–2108. https://doi.org/10.1126/science.1062872.
Hanks, T. D., & Summerfield, C. (2017). Perceptual decision making in rodents, monkeys, and humans. *Neuron, 93*(1), 15–31. https://doi.org/10.1016/j.neuron.2016.12.003.
Harvey, N., & Fischer, I. (1997). Taking advice: Accepting help, improving judgment, and sharing responsibility. *Organizational Behavior and Human Decision Processes, 70*(2), 117–133.
Heekeren, H. R., Marrett, S., Bandettini, P. A., & Ungerleider, L. G. (2004). A general mechanism for perceptual decision-making in the human brain. *Nature, 431*(7010), 859–862. https://doi.org/10.1038/nature02966.
Hertz, U., Palminteri, S., Brunetti, S., Olesen, C., Frith, C. D., & Bahrami, B. (2017). Neural computations underpinning the strategic management of influence in advice giving. *Nature Communications, 8*, 2191. https://doi.org/10.1038/s41467-017-02314-5.
Heyes, C. M., & Frith, C. D. (2014). The cultural evolution of mind reading. *Science, 344*(6190), 1243091. https://doi.org/10.1126/science.1243091.
Hilgenstock, R., Weiss, T., & Witte, O. W. (2014). You'd better think twice: Post-decision perceptual confidence. *NeuroImage, 99*, 323–331. https://doi.org/10.1016/j.neuroimage.2014.05.049.
Holt, D. J., Cassidy, B. S., Andrews-Hanna, J. R., Lee, S. M., Coombs, G., Goff, D. C., ... Moran, J. M. (2011). An anterior-to-posterior shift in midline cortical activity in schizophrenia during self-reflection. *Biological Psychiatry, 69*(5), 415–423. https://doi.org/10.1016/j.biopsych.2010.10.003.
Hutcherson, C. A., Bushong, B., & Rangel, A. (2015). A Neurocomputational model of altruistic choice and its implications. *Neuron, 87*(2), 451–463. https://doi.org/10.1016/j.neuron.2015.06.031.
Izuma, K., & Adolphs, R. (2013). Social manipulation of preference in the human brain. *Neuron, 78*(3), 563–573. https://doi.org/10.1016/j.neuron.2013.03.023.
Kennedy, M., Dornan, J., Rutledge, E., O'Neill, H., & Kennedy, H. G. (2009). Extra information about treatment is too much for the patient with psychosis. *International Journal of Law and Psychiatry, 32*(6), 369–376. https://doi.org/10.1016/j.ijlp.2009.09.006.
Kiani, R., & Shadlen, M. N. (2009). Representation of confidence associated with a decision by neurons in the parietal cortex. *Science, 324*(5928), 759–764. https://doi.org/10.1126/science.1169405.
Kim, J., & Shadlen, M. N. (1999). Neural correlates of a decision in the dorsolateral prefrontal cortex of the macaque. *Nature Neuroscience, 2*(2), 176–185.
Koren, D., Poyurovsky, M., Seidman, L. J., Goldsmith, M., Wenger, S., & Klein, E. M. (2005). The neuropsychological basis of competence to consent in first-episode

schizophrenia: A pilot metacognitive study. *Biological Psychiatry, 57*(6), 609–616. https://doi.org/10.1016/j.biopsych.2004.11.029.
Koren, D., Seidman, L. J., Poyurovsky, M., Goldsmith, M., Viksman, P., Zichel, S., & Klein, E. (2004). The neuropsychological basis of insight in first-episode schizophrenia: A pilot metacognitive study. *Schizophrenia Research, 70*, 195–202. https://doi.org/10.1016/j.schres.2004.02.004.
Krajbich, I., Armel, C., & Rangel, A. (2010). Visual fixations and the computation and comparison of value in simple choice. *Nature Neuroscience, 13*, 1292–1298. https://doi.org/10.1038/nn.2635.
Krajbich, I., Hare, T., Bartling, B., Morishima, Y., & Fehr, E. (2015). A common mechanism underlying food choice and social decisions. *PLoS Computational Biology, 11*(10), 1–24. https://doi.org/10.1371/journal.pcbi.1004371.
Luce, D. R. (1992). Response Times. *Oxford Psychology Series*. https://doi.org/10.1108/eb006536.
Manski, C. F. (1977). The structure of random utility models. *Theory and Decision*. https://doi.org/10.1007/BF00133443.
Mazaniceux, A., Fleming, S. M., Souchay, C., & Moulin, C. J. A. (2018). *Retrospective confidence judgments across tasks: Domain-general processes underlying metacognitive accuracy*. https://doi.org/10.31234/osf.io/dr7ba.
McCurdy, L. Y., Maniscalco, B., Metcalfe, J., Liu, K. Y., de Lange, F. P., & Lau, H. (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. doi:https://doi.org/10.1523/JNEUROSCI.1890-12.2013.
Milgram, S. (1963). Behavioral study of obedience. *Journal of Abnormal and Social Psychology, 67*, 371–378. https://doi.org/10.1037/h0040525.
Modinos, G., Renken, R., Ormel, J., & Aleman, A. (2011). Self-reflection and the psychosis-prone brain: An fMRI study. *Neuropsychology, 25*(3), 295–305. https://doi.org/10.1037/a0021747.
Morales, J., Lau, H., & Fleming, S. M. (2018). Domain-general and domain-specific patterns of activity supporting metacognition in human prefrontal cortex. *The Journal of Neuroscience, 38*(14), 3534–3546. https://doi.org/10.1523/JNEUROSCI.2360-17.2018.
Murphy, P. R., Robertson, I. H., Harty, S., & O'Connell, R. G. (2015). Neural evidence accumulation persists after choice to inform metacognitive judgments. *ELife*. https://doi.org/10.7554/elife.11946.
O'Connell, R. G., Dockree, P. M., & Kelly, S. P. (2012). A supramodal accumulation-to-bound signal that determines perceptual decisions in humans. *Nature Neuroscience, 15*(12), 1729–1735. https://doi.org/10.1038/nn.3248.
Olsen, K., Roepstorff, A., & Bang, D. (2019). Knowing whom to learn from: individual differences in metacognition and weighting of social information. *PsyArXiv Preprints*. https://doi.org/10.31234/osf.io/jqheu.
Owen, G. S., Freyenhagen, F., Richardson, G., & Hotopf, M. (2009). Mental capacity and decisional autonomy: An interdisciplinary challenge. *Inquiry, 52*(1), 79–107. https://doi.org/10.1080/00201740802661502.
Owen, G. S., Richardson, G., David, A. S., Szmukler, G., Hayward, P., & Hotopf, M. (2009). Mental capacity, diagnosis, and insight in psychiatric inpatients: A cross sectional study. *Psychological Medicine, 39*(8), 1389–1398.
Owen, G. S., Freyenhagen, F., Hotopf, M., & Martin, W. (2015). Temporal inabilities and decision-making capacity in depression. *Phenom Cogn Sci, 14*, 163–182. https://doi.org/10.1007/s11097-013-9327-x.
Peirce, C. S., & Jastrow, J. (1884). On small differences in sensation. *Memoirs of the National Academy of Science, 3*, 75–83. https://doi.org/10.5860/choice.35sup-462.
Pescetelli, N., & Yeung, N. (2019). The role of decision confidence in advice-taking and trust formation. *PsyArXiv*.
Petrova, M., Dale, J., & Bill, F. (2006). Values-based practice in primary care. *British Journal of General Practice, 56*(530), 703–709.
Platt, M. L., & Glimcher, P. W. (1999). Neural correlates of decision variables in parietal cortex. *Nature, 400*(6741), 233–238. https://doi.org/10.1038/22268.
Pousa, E., Duñó, R., Brebion, G., David, A. S., Ruiz, A. I., & Obiols, J. E. (2008). Theory of mind deficits in chronic schizophrenia: Evidence for state dependence. *Psychiatry Research, 158*(1), 1–10.
Rabbitt, P. M. (1966). Errors and error correction in choice-response tasks. *Journal of Experimental Psychology, 71*(2), 264–272. https://doi.org/10.1037/h0022853.
Rangel, A., Camerer, C., & Montague, P. R. (2008). Neuroeconomics: The neurobiology of value-based decision-making. *Nature Reviews. Neuroscience, 9*(7), 545–556. https://doi.org/10.1038/nrn2357.Neuroeconomics.
Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*. https://doi.org/10.1037/0033-295X.85.2.59.
Resulaj, A., Kiani, R., Wolpert, D. M., & Shadlen, M. N. (2009). Changes of mind in decision-making. *Nature, 461*(7261), 263–266. https://doi.org/10.1038/nature08275.
Ridderinkhof, K. R., van den Wildenberg, W. P., Segalowitz, S. J., & Carter, C. S. (2004). Neurocognitive mechanisms of cognitive control: The role of prefrontal cortex in action selection, response inhibition, performance monitoring, and reward-based learning. *Brain and Cognition, 56*(2), 129–140.
Roitman, J. D., & Shadlen, M. N. (2002). Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. *The Journal of Neuroscience, 22*(21), 9475–9489.
Rollwage, M., Dolan, R. J., & Fleming, S. M. (2018). Metacognitive failure as a feature of those holding radical beliefs. *Current Biology, 28*(24), 4014–4021.e8. https://doi.org/10.1016/j.cub.2018.10.053.
Rouault, M., Mcwilliams, A., Allen, M. G., & Fleming, S. M. (2018). Human metacognition across domains : Insights from individual differences and neuroimaging. *Personality Neuroscience, 1*(E17), 1–35. https://doi.org/10.1017/pen.2018.16.
Rouault, M., Seow, T., Gillan, C. M., & Fleming, S. M. (2018). Psychiatric symptom dimentions are associated with dissociable shifts in metacognition but not task performance. *Biological Psychiatry, 84*(6), 443–451. https://doi.org/10.1016/j.biopsych.2017.12.017.
Ruck Keene, A., Kane, N. B., Kim, S. Y. H., & Owen, G. S. (2019). Taking capacity seriously? Ten years of mental capacity disputes before England's Court of Protection. *International Journal of Law and Psychiatry, 62*, 56–72. https://doi.org/10.1016/j.ijlp.2018.11.005.
Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E., & Cohen, J. D. (2013). The neural basis of economic decision making in the ultimatum game. *Science, 300*, 1755–1758. https://doi.org/10.1126/science.1082976.
Shea, N., Boldt, A., Bang, D., Yeung, N., Heyes, C., & Frith, C. D. (2014). Supra-personal cognitive control and metacognition. *Trends in Cognitive Sciences, 18*(4), 186–193. https://doi.org/10.1016/j.tics.2014.01.006.
Shenhav, A., & Greene, J. D. (2014). Integrative moral judgment: Dissociating the roles of the amygdala and ventromedial prefrontal cortex. *The Journal of Neuroscience, 34*(13), 4741–4749. https://doi.org/10.1523/JNEUROSCI.3390-13.2014.
Shimamura, A. P., & Squire, L. R. (1986). 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. https://doi.org/10.1037/0278-7393.12.3.452.
Smith, P. L., & Ratcliff, R. (2004). Psychology and neurobiology of simple decisions. *Trends in Neurosciences, 27*(3), 161–168. https://doi.org/10.1016/j.tins.2004.01.006.
Sniezek, J. A., & Van Swol, L. M. (2001). Trust, confidence, and expertise in a judge – Advisor system. *Organizational Behavior and Human Decision Processes, 84*(2), 288–307. https://doi.org/10.1006/obhd.2000.2926.
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), 1782–1792. https://doi.org/10.1016/j.concog.2010.12.011.
Sowden, S., Koletsi, S., Lymberopoulos, E., Militaru, E., Catmur, C., & Bird, G. (2018). Quantifying compliance and acceptance through public and private social conformity. *Consciousness and Cognition, 65*, 359–367. https://doi.org/10.1016/j.concog.2018.08.009.
Stephan, K. E., Manjaly, Z. M., Mathys, C. D., Weber, L. A. E., Paliwal, S., Gard, T., ... Petzschner, F. H. (2016). Allostatic self-efficacy: A metacognitive theory of Dyshomeostasis-induced fatigue and depression. *Frontiers in Human Neuroscience, 10*. https://doi.org/10.3389/fnhum.2016.00550.
Talluri, B. C., Urai, A. E., Tsetsos, K., Usher, M., & Donner, T. H. (2018). Confirmation bias through selective overweighting of choice-consistent evidence. *Current Biology, 28*, 3128–3135. https://doi.org/10.1016/j.cub.2018.07.052.
Thomson, J. J. (1985). The trolley problem. *The Yale Law Journal, 94*, 1395–1415. https://doi.org/10.1119/1.1976413.
Toelch, U., & Dolan, R. J. (2015). Informational and normative in fl uences in conformity from a neurocomputational perspective. *Trends in Cognitive Sciences, 19*(10), 579–589. https://doi.org/10.1016/j.tics.2015.07.007.
Valentin, V. V., Dickinson, A., & Doherty, J. P. O. (2007). Determining the neural substrates of goal-directed learning in the human brain. *The Journal of Neuroscience, 27*(15), 4019–4026. https://doi.org/10.1523/JNEUROSCI.0564-07.2007.
Vickers, D. (1979). *Decision processes in visual perception.* Academic Presshttps://doi.org/10.1016/c2013-0-11654-6.
Wald, A. (1947). Review: Theory of games and economic behavior. *The Review of Economic Statistics*. https://doi.org/10.2307/1925651.
Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition, 13*, 103–128. https://doi.org/10.1016/0010-0277(83)90004-5.
Zürcher, T., Elger, B., & Trachsel, M. (2019). The notion of free will and its ethical relevance for decision-making capacity. *BMC Medical Ethics, 20*(31), 3–10. https://doi.org/10.1186/s12910-019-0371-0.