Confidence and Information Seeking in Decision Making Under Uncertainty

Linda E. L. Wei

Magdalen College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy in Experimental Psychology

Trinity 2021

Words: 53,023

Abstract

Confidence and information-seeking are aspects of metacognition that help people make better choices under conditions of uncertainty. However, previous research has found that these metacognitive functions can vary greatly between individuals and can become maladaptively decoupled or decalibrated in psychiatric disorders. This thesis explores the relationship between confidence and information-seeking behaviour when making decisions under uncertainty, and individual differences in these measures.

This work begins with the development of two decision tasks that provide parallel measures of confidence and information-seeking in the domains of perceptual judgment and probabilistic reasoning. I then investigate whether psychiatric dimensions of interest might be reliably associated with certain patterns of confidence and information-seeking. I present data showing that, while trait anxiety tended to be associated with a small underconfidence bias across domains, psychiatric dimensions on the whole fail to be consistently predictive of confidence or information-seeking.

I then proceed to explore specific aspects of confidence and information-seeking. While previous literature and preliminary findings had suggested that people might employ strategies other than information-seeking to manage uncertainty during decision-making, I present results suggesting that participants do not typically engage in time-effective strategies like seeking feedback or opting out of decisions, even when they are available as alternatives to information-seeking.

I then look more closely into the relationship between choice and confidence in the decision process. I report a robust increase in confidence at the point of decision, even after accounting for differences in the objective probability of being correct. After investigating possible mechanisms for this effect, I suggest that the act of committing to a decision evokes a boost in confidence, but that this decision-evoked confidence boost fails to have lasting effects on subsequent confidence, information-seeking, or choices.

Overall, my results suggest that confidence and information-seeking generally help us make more adaptive decisions under uncertainty. The thesis concludes with a brief discussion of the broader implications of the findings for psychiatric and cognitive research.

Acknowledgements

My first debt of gratitude is to my primary supervisor Nick Yeung, who has been there with me at every step of this four-year odyssey. His attention to detail, his insightful approach to complex problems, and his unwavering patience have helped me become a better scientific thinker and researcher, and they are qualities that I hope to take with me wherever my career path might lead.

My second debt of gratitude is to my secondary supervisor David Clark, who brought an alternative perspective to my D.Phil. research with warmth, wisdom, and gentle humour, and who always had time in his busy schedule to meet and discuss my research. This D.Phil. journey has often taken me in unforeseen directions; while writing this thesis, I was struck by what it could have been, had our research paths overlapped more.

My third debt of gratitude is to Magdalen College, for being a home away from home ever since I came to the UK. My heartfelt thanks to Mark Pobjoy and other members of the Tutorial Board for their belief in me from the start, and to them and the generous Magdalen alumni over the years for making my D.Phil. research possible through the Senior Mackinnon and Fleming Family Scholarships.

Lastly – to my family and friends, who have gotten me through the day to day of this odyssey. To my parents and extended family for supporting me through the D.Phil. To Matt J. for being the best lab mate one could ask for and for helping me pick up three essential life skills alongside my degree: coding, driving, and manoeuvring furniture through tight spaces. To Maja, Mira, Mirela, Ron, and other members of my lab and the Experimental Psychology department for the warmth of conversations and companionship in this academic journey. To Leeho, Emily T., and other members of start@ox for inspiring me to reconsider what's possible. To Christina, Elly, Aimée, Linda L., Emily G., Andrea, Grace, and all my other friends outside of the Oxford bubble who have laughed with me, cried with me, reminded me of what life is like in other corners of this vast world, and reminded me of who I was and who I continue to be. And finally, to Michelle, for being the light in my life in some of the darkest of times, and for instilling in me hope and confidence about what lies ahead.

Ah how shameless—the way these mortals blame the gods.
From us alone, they say, come all their miseries, yes,
but they themselves, with their own reckless ways,
compound their pains beyond their proper share.

- Homer, *Odyssey*

Table of Contents

1
INTRODUCTION

1.1 Preface

Confidence and information-seeking are fundamental aspects of human decision-making, and they work in tandem to help people make better choices under conditions of uncertainty. The relationship between confidence and information-seeking can be illustrated in the morning routines of two fictional Oxford students: Jane and John. Both live in Oxford, a city where the probability of rain on any given day of the year is quite high, and both hate the idea of being caught in the rain without an umbrella on their morning commute.

Every morning before she goes into the office, Jane decides whether to pack an umbrella in her bag based on how the weather looks outside her window. If the grey skies outside fail to assuage her concerns about being caught in the rain during her walk to the office, she might decide to look up the probability of rain on her phone's weather app. After seeing the app's predictions of no rain, Jane might feel confident enough about the prospect of a dry morning commute to forego packing her umbrella.

That same morning, John too must decide whether to bring an umbrella with him. Like Jane, he looks at the grey skies outside his window and checks his phone's weather app. However, John fails to be assured by the app's prediction of a dry morning commute. He might feel that the app does not have a great track record of accurately predicting the weather in Oxford. John might proceed to look up additional weather reports from what he thinks are more reliable sources, such as BBC Weather or the Met Office. All the reports support the weather app's prediction. Despite all credible information suggesting no rain is to be expected, John still does not feel convinced and decides to bring an umbrella with him — just in case.

The similarities in the decision processes of Jane and John highlight the close association between confidence and information-seeking: when low in confidence, people tend to seek more information about their choices, and with more information, people tend to be more confident and more likely to reach a decision. This relationship between confidence and information-seeking is observed in decisions as trivial as considering whether to pack an umbrella in the morning, or as consequential as selecting who to vote for in an upcoming election. However, as the differences between Jane and John demonstrate, there may be variability between individuals in how evidence translates into confidence, as well as in the level of evidence or confidence an individual needs to reach a decision. At the extreme end of the spectrum, like in the case of some psychiatric disorders, confidence and information-seeking may become maladaptively decalibrated or decoupled. For instance, individuals with psychosis may maintain high confidence in delusional beliefs despite significant evidence to the contrary (Garety & Freeman, 1999), and individuals with obsessive-compulsive disorder may engage in compulsive, repetitive checking behaviour that fails to raise their level of confidence (Strauss et al., 2020).

In this thesis, I will explore the relationship between confidence and information-seeking behaviour when making decisions under uncertainty, and the role that individual differences can play in these aspects of the decision process. In this introductory chapter, I will review key theories and empirical research on confidence and information-seeking, and discuss the implications of variability and disturbances in these measures for various psychiatric disorders. The chapter will finish with an outline the work reported in this thesis, which builds on existing work in and addresses key questions about how people metacognitively monitor and regulate their behaviour and cognition.

1.2 Confidence in decision-making

Confidence is an intrinsic part of human cognition. Whether we are making decisions or recalling information from memory, we are usually able to report a subjective sense of certainty or uncertainty about the accuracy of these cognitive processes. While confidence is thought to reflect the balance of the evidence regarding this judgment, there are many instances where confidence may reflect or be influenced by factors other than evidence strength – as a salesperson trying to flog a dodgy car knows all too well. In this section, we will explore definitions and formalisations of confidence, highlight various factors that may influence confidence, and discuss possible dissociations between reported and internal confidence.

1.2.1 Confidence and metacognition

Defining metacognition

The word "metacognition" has roots in Greek and Latin: "meta-", from the Greek prefix *μετα-*, meaning "after" or "beyond", and "cognition", from the Latin verb *cognoscere*, meaning "to know". *Cognoscere* is one of two verbs in Latin meaning "to know" (the other being *scire*): it refers, more specifically, to the type of knowing that is derived from personal or firsthand experience, in contrast to the knowing of a factual or semantic nature that is denoted by *scire*. Thus, from a direct transliteration, "metacognition" seems to suggest processes that operate or understanding that arises beyond the initial acquisition and processing of information. While John Flavell may not have had these classical etymologies in mind when he coined the term "metacognition" for psychological discourse in the

1970s, Flavell's seminal (1979) definition of metacognition as "knowledge and cognition about cognitive phenomena" retains some of the character of that transliteration, describing cognitive processes and knowledge that extend beyond the initial level of information processing.

The distinction that Flavell mentioned between metacognitive knowledge and processes persists in the current literature about metacognition. Metacognitive knowledge refers to knowledge that we have about our cognitive abilities, strategies, tasks, and such, while metacognitive regulation refers to processes that coordinate our cognition (Fernandez-Duque et al., 2000). The latter category is further broken down into bottom-up monitoring processes (e.g., error detection, source monitoring) and top-down control processes (e.g., conflict resolution, error correction, inhibitory control, planning, resource allocation) (Nelson & Narens, 1990; Fernandez-Duque et al., 2000). Nelson and Narens (1990) formalised these ideas about metacognitive regulation into a dual-level framework consisting of "object-level" and "meta-level" processes [Figure 1.1]. The object-level concerns itself with functionally distinct aspects of cognition, such as object recognition, semantic processing, and the like (Shimamura, 2008). The meta-level, on the other hand, is where a dynamic model of the object-level is housed, and is responsible for regulating (i.e., monitoring and controlling) these object-level processes (Nelson & Narens, 1990).

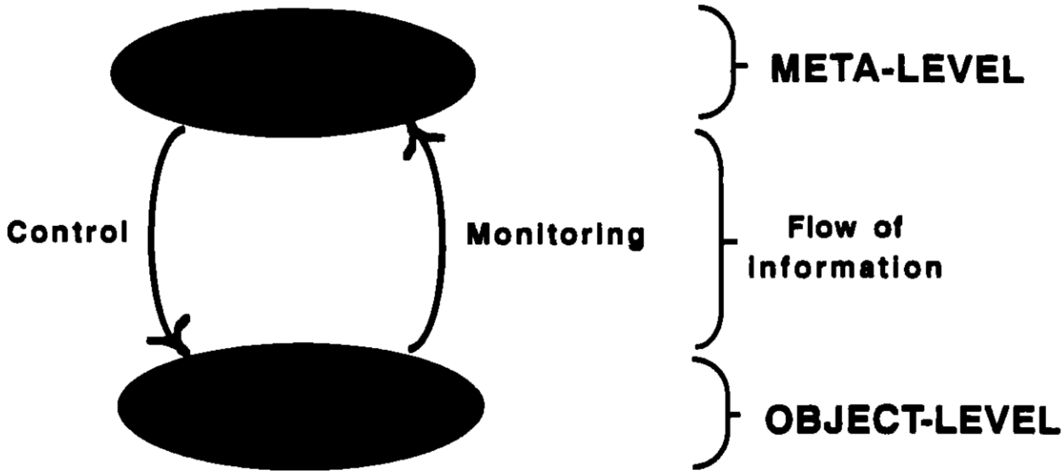


Figure 1.1. Interactions between the object-level and meta-level. From Nelson and Narens (1990).

Flavell initially described metacognition as consisting of four components: self-reflective feelings and judgments during cognitive activity ("metacognitive experiences"), deliberate actions taken to control cognition ("metacognitive strategies"), a general understanding of the cognitive abilities and strategies of oneself and others ("metacognitive knowledge"), and the overarching objectives of cognitive activity ("metacognitive goals") (Flavell, 1979; Moritz & Lysaker, 2018). However, since Flavell's coinage of metacognition, there has been debate about how metacognition differs from other terms like cognitive control and executive function, which have also been used to refer to higher-level cognitive processes and which all share similar neural localisations in the prefrontal cortex (Roeber, 2017; Shea et al., 2014; Shimamura, 2000).

In the subsequent developmental literature, distinctions have been drawn between executive functions and metacognition. Roeber (2017) suggested that executive functions and metacognition are separate but highly interdependent processes: while both underpin deliberate, goal-directed, and self-regulated information processing, metacognitive monitoring and control generally refer to higher-order functions that are slower, longer-lasting, and possibly more domain-specific. Executive function is assumed to have a more defined tri-factor structure (cf. Miyake et al., 2000), appears to operate at a more immediate and basic task level than metacognitive control, and is seen as necessary but not sufficient for metacognitive control (Diamond, 2013; Roeber, 2017).

Subtle distinctions can also be drawn between cognitive control and metacognition. Shea et al. (2014) defined cognitive control as the "cognitive mechanisms responsible for guiding thought and behaviour in accordance with current goals and intentions". Metacognition, on the other hand, was proposed by the authors to refer more specifically to monitoring and control processes that employ "one or more metacognitive representations, that is, representations of a property of a cognitive process," and which are necessary for communicating uncertainty and coordinating strategies between multiple agents engaged in joint decision-making. The general consensus in the literature appears to be that metacognition distinguishes itself from similar higher-order cognitive processes through its emphasis on monitoring processes, its close ties with longer-term knowledge and belief structures, and its role in communication during joint decision-making.

Metacognition is essential for decision-making because it is part of that network of higher-order functions that help guide behaviour and cognition to be both goal-directed and adaptive for the current environment. Going back to our earlier examples of John and Jane, we see that, during the decision process to determine how to best prepare for the weather, both of our students engaged in self-reflective feelings and judgments (e.g., feelings of confidence about the certainty of rain given the weather in the morning) and deliberate actions to control and direct this process (e.g., seeking more information about the weather via weather reports), and included a general understanding of their own and others' cognitive abilities in the final decision (Jane is generally quite sure of her own judgments, while John often doubts his and others' judgments, at least regarding the British weather). Confidence and information-seeking are therefore empirically observable measures that might be able to give us some insight into the metacognitive abilities underlying adaptive human decision-making.

Defining confidence

The definition of confidence varies widely between academic and colloquial usage, and even between different academic disciplines. Eleven discrete definitions of confidence are currently listed in the Oxford English Dictionary. They can roughly divided into two clusters: those referring to a sense of trustworthiness or self-assuredness, often in an interpersonal context (e.g., "taking someone into your confidence", "con game", "con man", etc.), and those denoting a more impersonal sense of certainty or likelihood (e.g., confidence intervals in statistics) (OED Online, 2021).

The normative understanding of confidence differs between fields of psychology. In social and clinical psychology, confidence is typically understood more in the context of the former cluster of definitions and is mentioned alongside concepts such as self-esteem and self-efficacy (Campbell, 1990; Mann et al., 2004; Zuckerman, 1989). In cognitive psychology, however, the definition of confidence leans into the latter cluster: confidence is usually defined and formalised as a probabilistic judgment about a cognitive state or mental representation, as in, for example, a "feeling of knowing" (cf. Hart, 1965). In memory and decision-making research, confidence tends to refer more specifically to the subjective probability that a judgment, problem solution, or retrieved memory is accurate or appropriate — namely, the perceived probability of being correct given the evidence (Kepecs & Mainen, 2012; Pouget et al., 2016; Yeung & Summerfield, 2012; Zylberberg et al., 2016).

Recent evidence suggests that the conflation between the two types of confidence in the English language might unwittingly reflect their close and mutually reinforcing relationship. The more long-term beliefs that people hold about their abilities might be constructed over time from more local confidence estimates about decisions (Rouault et al., 2019; Lee et al., 2021). Rouault et al.

(2019) suggested that local decision confidence is integrated over numerous instances of similar task experiences to form global confidence or self-performance estimates, which then play a role in guiding subsequent motivation, effort, and choice.

In this thesis, we were interested more specifically in these local measures of decision confidence and how they might influence or be influenced by other factors. Going forward, we will use the term confidence to refer to local or decision confidence exclusively unless specifically denoted otherwise.

1.2.2 Formalising confidence

Confidence is a measure that gives us insight into metacognitive monitoring during cognitive activities. However, as an inherently subjective construct, it has been difficult to study systematically. To tackle this, recent research has taken a more computational approach to the study of confidence, formalising decision confidence mathematically as an objective statistical quantity and deriving more precise mathematical measures of metacognitive signal from "raw" confidence reports.

Measuring confidence

The prevalent method for eliciting a "raw" measure of confidence is a simple confidence rating, where people are asked to explicitly rate or report their confidence on a scale. Confidence scales can appear in a number of ways: as a series of ordered verbal confidence categories (e.g., "very sure" to "not at all confident"), as ordinal numerical bins mapped onto verbal category labels in the instructions (e.g., "1 = not at all confident, 7 = very sure"), or as visual analogue scales. The number of steps in the scale does not appear to be particularly important once there are more than two steps: while sensitivity is higher for non-dichotomous scales (i.e., scales with more than two steps), finer-grained scales are not necessarily more sensitive, only appearing to be useful when task difficulty is low (Overgaard & Sandberg, 2012).

The key concern that many researchers have regarding simple confidence ratings is that the confidence being reported might not reflect "true" confidence (Massoni et al., 2014; Overgaard & Sandberg, 2012). As Overgaard and Sandberg (2012) state:

> *"A report is [...] an intended communication by a participant and may be delivered verbally or by any other kind of controlled behaviour (signs, button presses or whatever). We may have full metacognitive or introspective knowledge of some mental event but choose not to report it, to lie about it or to report just one aspect of it, while not mentioning another."*

This has resulted in a preference, especially in perceptual decision-making, towards alternative methods of measuring confidence, such as post-decision wagering (cf. Persaud et al., 2007) or no-loss gambling (cf. Dienes & Seth, 2010).

These alternative methods typically involve a betting aspect or some other monetary incentive for participants to report their “true” confidence (Massoni et al., 2014).

A signal detection theory framework of confidence

To better understand why some methods of measuring confidence are regarded as “truer” measures of metacognition, we need to understand the framework of signal detection theory as applied to metacognition.

Originally conceived and used in psychophysics, the original (“type 1”) signal detection theory (SDT) conceptualises perception as a process of discriminating the presence of the target stimulus (signal) from the absence of that stimulus or from the presence of other stimuli (Green & Swets, 1966). Whether the target stimulus is successfully detected depends on two factors. The first, known as sensitivity or discriminability, is how hard or easy it is to detect the target stimulus given the noise, which can come from signal corruption (e.g., by the perceptual system during processing) or non-target stimuli that are similar to the target stimulus. The second, known as bias, is where the criterion for detection — the threshold at which the external stimulus is interpreted as a target signal rather than noise — is set: the placement of the criterion can be an inherent individual difference but can also be situationally motivated (e.g., by the reward or cost associated with a particular response) (Azzopardi & Cowey, 1998).

Metacognition can be thought of as a kind of second-order ("type 2") signal detection process: instead of discriminating between events defined independently of the observer, metacognition involves the observer discriminating their own correct and incorrect decisions about these first-order events [Figure 1.2] (Galvin et al., 2003; Maniscalco & Lau, 2012). Like perceptual reports, metacognitive reports like confidence are also determined by sensitivity and bias independently.

Metacognitive bias is the participant's tendency to give higher or lower confidence ratings irrespective of actual task performance or decision accuracy (Fleming & Lau, 2014; Rouault et al., 2018). In Figure 1.2, bias is indicated by where on the X-axis an individual sets his or her response criterion to the type 2 task. In lab settings, metacognitive bias is commonly quantified as the mean trial-by-trial confidence rating (Lichtenstein et al., 1982; Rouault et al., 2018). Another way of quantifying metacognitive bias is by calculating calibration, the degree to which reported confidence matches actual task performance. Higher confidence than performance indicates an overconfidence bias, whereas lower confidence than performance indicates an underconfidence bias; both types of bias indicate poor calibration.

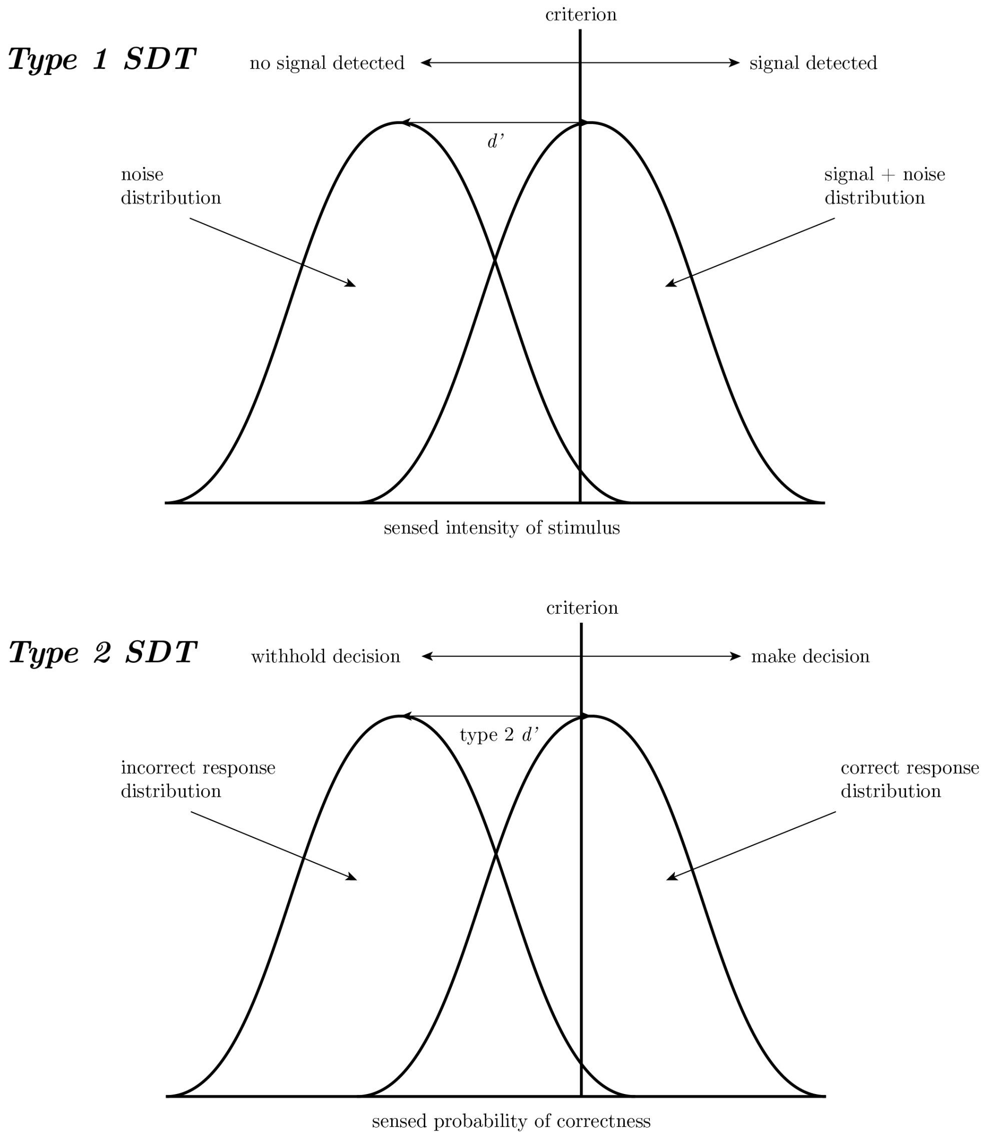


Figure 1.2. Comparing type 1 versus type 2 SDT.

A number of factors are known to affect metacognitive bias. While participants generally tend to be overconfident in their self-reports (Lichtenstein et al., 1982), large individual differences exist between individuals in where they naturally set this response criterion (see Section 1.4.1 for a fuller discussion of this) (Ais et al., 2016; Blais et al., 2005). Metacognitive bias is sensitive to task difficulty: people tend to be more overconfident on more difficult tasks than easier tasks, in what is known as the "hard-easy" effect (Lichtenstein & Fischhoff, 1977; Brenner, 2003). Like in type 1 tasks, the criterion in type 2 tasks is also sensitive to incentives, with people adopting a higher or lower criterion in response to different reward contingencies for different classes of metacognitive judgments, such as false alarms (indicating having made a mistake despite responding correctly) (Steinhauser & Yeung, 2010).

Metacognitive sensitivity or accuracy, on the other hand, refers to the precision or reliability of confidence estimation. This can be understood as how well a subject is able to discriminate between good and bad task performance in their confidence ratings: a reliable estimate of confidence (from a metacognitive agent with high sensitivity) should be high when performance is good and low when performance is poor. Metacognitive sensitivity is dissociable from metacognitive bias, as illustrated in Figure 1.3, but the two are often conflated in measures of confidence.

A key debate in the metacognition literature over the years has been around identifying the best measure of sensitivity, that is, the measure that is least contaminated by factors such as task difficulty, task performance, or metacognitive bias. One measure that is commonly encountered in early metamemory and decision-making literature is the Goodman-Kruskal gamma coefficient (Goodman & Kruskal, 1954; Nelson, 1984). Gamma is more specifically a measure of resolution, or the degree to which a metacognitive rating discriminates between a person's own correct versus incorrect responses (Higham & Higham, 2019). It is a simple and relative measure of sensitivity that builds on the intuitive understanding that accuracy and confidence should be highly correlated. However, gamma and other relative measures of sensitivity have been found to be contaminated by metacognitive bias (Masson & Rotello, 2009; Fleming & Lau, 2014).

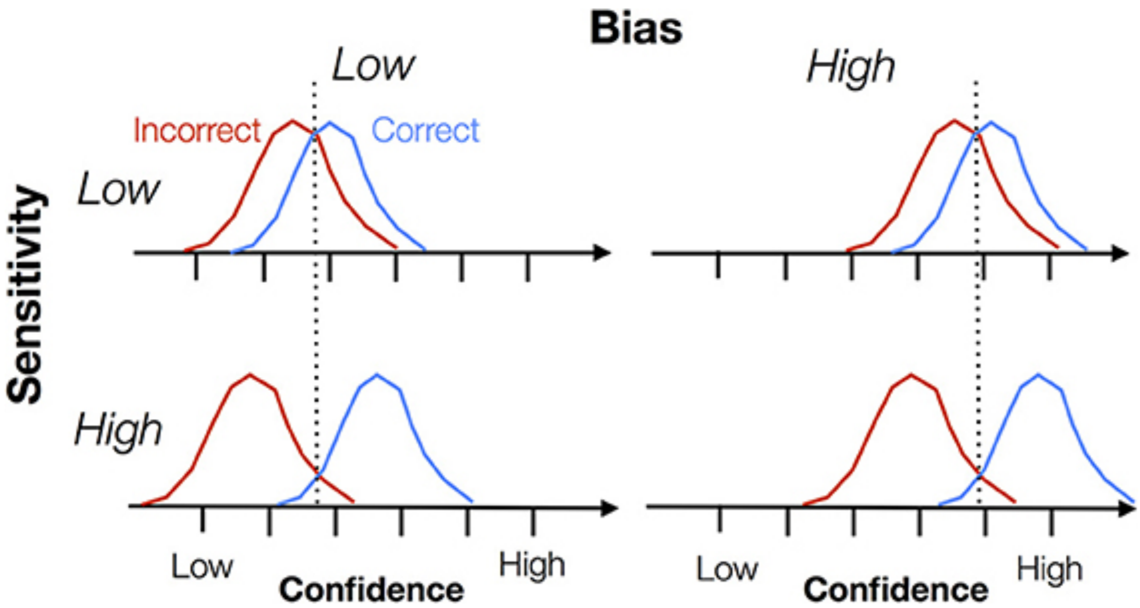


Figure 1.3. Bias versus sensitivity in type 2 (metacognitive) judgments. From Fleming & Lau (2014).

This led to several attempts to develop bias-free measures of sensitivity that took inspiration from SDT. In a Type 1 task, to remove the influence of response bias in measures of perceptual ability, a bias-free measure known as the discriminability index (d') can be used; this gives an estimate of the signal strength (i.e., the distance between the peaks of the signal and no-signal distributions) based on the hit rate and false alarm rate. Applying the SDT framework to Type 2 tasks, we can derive Type 2 *d'*, correspondingly illustrated in Figure 1.2 as the distance between the peaks of the distributions for correct and incorrect responses (Kunimoto et al., 2001). However, the issue with Type 2 *d'* is that it assumes that the internal signals for correct versus incorrect trials are Gaussian distributions with equal variance, when this is often not the case (Galvin et al., 2003; Fleming & Lau, 2014).

An alternative solution proposed is to use a non-parametric analysis based on the receiver operating characteristic (ROC) analysis used in type 1 SDT. For the type 1 task, based on manipulations of the response criterion, we can plot the hit rate (on the Y-axis) versus false alarm rate (on the X-axis) at different levels of response bias and then link the points together with a curved line of best fit (i.e., the ROC curve). The area under the ROC curve is a measure of sensitivity that is free from the equal-variance Gaussian assumption, with higher area under the curve indicating higher sensitivity. Applying this analysis to the type 2 task — calculating proportions of correctly classified correct and incorrect responses as a function of levels of confidence — can produce a corresponding type 2 ROC, with the type 2 area under the ROC curve (AUROC2) serving as a non-parametric measure of metacognitive sensitivity (Clarke et al., 1959; Galvin et al., 2003). The issue with AUROC2 is that it relies on parameters derived from task performance (i.e., type 1 *d'* and response criterion). It is thus strongly influenced by task difficulty: this measure of metacognitive sensitivity drops when task difficulty increases (Fleming & Lau, 2014). Intuitively, this makes sense, as when a task is easy, participants tend to respond correctly and be more aware of when they have made an error, but as the task difficulty increases, participants tend to guess more and be correspondingly unsure about whether they have responded correctly or made an error.

Leveraging this tight relationship between type 1 and type 2 performance, Maniscalco and Lau (2012; 2014) developed a measure of sensitivity known as meta-*d'*. Meta-*d'* is the underlying type 1 sensitivity (in units of type 1 *d'*) that has been inferred from the subject's actual type 2 performance data, assuming that the subject is metacognitively ideal. According to Fleming and Lau (2014), meta-*d'* can be understood as "the sensory evidence available for metacognition in signal-to-noise ratio units, just as type 1 *d'* is the sensory evidence available for decision-making in signal-to-noise ratio units". Because it is in units of *d'*, it is more intuitively comprehensible than AUROC2 and makes explicit and computable the link between task performance and metacognitive sensitivity.

To control for the effect of task performance, Fleming and Lau (2014) advocated dividing meta-*d'* by *d'* to derive the M-ratio, a relative value they subsequently defined as metacognitive efficiency — namely, metacognitive ability in a particular domain, controlling for task performance and free from metacognitive bias. A M-ratio of 1 would indicate that the observer is an ideal observer, while a M-ratio lower than 1 (e.g., 0.8) indicates how much sensory evidence available for decision has been lost when making metacognitive judgments. In recent years, the log of the M-ratio (i.e., log(meta-d')/d') has been more commonly used because it corrects for the non-normality of ratio measures (Howell, 2010; Rouault et al., 2018).

Computing decision confidence

A central focus of research on metacognition has been on understanding the information that is used to make these judgments. In contrast to metamemory research, where there has been a great deal of exploration of how multiple cues can contribute to metacognitive evaluations (cf. Nelson et al., 1984; Koriat & Levy-Sadot, 2001), research in confidence in perceptual decision-making has used more computational approaches to understand how internal confidence can be directly derived from the decision process.

The standard model of decision confidence in perceptual decision-making takes inspiration from both the SDT framework for type 1 tasks and the Bayesian approach to statistics: it proposes that decision confidence can be formalised as the posterior probability of being correct given the evidence, and can be estimated in computational models of decision-making as the distance between the stimulus (evidence) and bound (decision criterion) (Kepecs et al., 2008; Pouget et al., 2016; Rausch & Zehetleitner, 2019; Zylberberg et al., 2016). The standard model predicts that confidence in correct choices increases as a function of the discriminability of the stimulus while confidence in incorrect choices decreases; the two function lines cross over when the balance of evidence is neutral towards both choice options, in what is known as a folded X-pattern (Kepecs & Mainen, 2012). Many studies of perceptual decision-making, especially in non-human animals, use the folded X-pattern as a statistical signature of confidence (Kepecs et al., 2008; Komura et al., 2013; Sanders et al., 2016; Rausch & Zehetleitner, 2019).

The standard model gives us a sense of how confidence might be computed based on a static amount of evidence. However, in many decision scenarios, the decision process unfolds over a period of time based on the integration of evidence acquired over that period [Figure 1.4]. Evidence accumulation models, such as the drift diffusion model (cf. Pleskac & Busemeyer, 2010), extend this approach to think about how confidence unfolds dynamically over time. The general idea of this class of models is that a decision variable (DV), representing evidence for one of the choices, accumulates noisily during the information sampling phase until it reaches a threshold value, at which point a decision is made (Fleming & Daw, 2017; Lee & Pezzulo, 2021). During the evidence accumulation phase of the decision process, the decision agent should continuously monitor and assess the DV and evaluate whether a decision (choosing one option over the others) is warranted by the evidence they have gathered. Decision confidence reflects the balance of evidence, the strength of the evidence for the chosen option relative to other choices (Zylberberg et al., 2016; Desender et al., 2021).

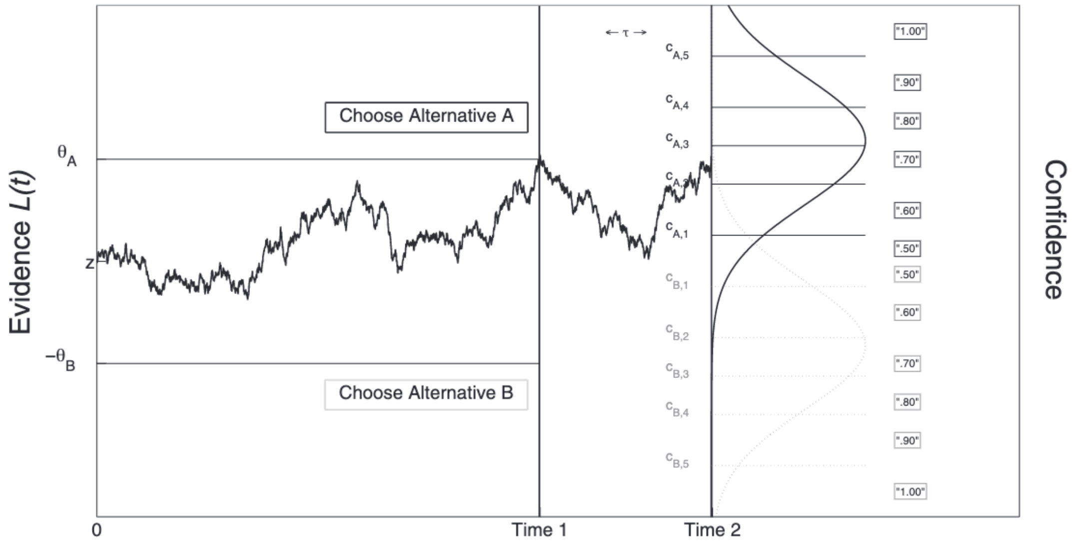


Figure 1.4. Evidence accumulation a drift diffusion model. From Pleskac & Busemeyer (2010).

Postdecisional locus theories of decision confidence tend to frame confidence as a subsequent output from the DV — realised only some time after the DV reaches a certain threshold and a choice is made (Pleskac & Busemeyer, 2010; Moran et al., 2015). Confidence is commonly described in these models as a continuous variable proportional to the interaction of the quantity of evidence at decision point and the rate at which that evidence had accumulated (Fleming & Daw, 2017; Kiani & Shadlen, 2009; Kiani et al., 2014; Yeung & Summerfield, 2012). The way that confidence plays a secondary role in these models of the decision process may partly reflect the way that confidence has been studied in perceptual decision-making: many of the classic perceptual decision-making paradigms show evidence before the decision and only ask participants to report confidence after making a decision (Pleskac & Busemeyer, 2010).

However, recent studies that look at the extended decision-making episode show that decision agents use information that is available to them from both before and after the immediate decision (Charles & Yeung, 2019). As predicted by the Bayesian framework, previous experiences (e.g., explicit confidence representations from a previous decision) can influence subsequent decisions. This can be by biasing the initial starting point of confidence (Folke et al., 2016; Murphy et al., 2015), the placement of the decision criterion (van den Berg et al., 2016), or even the selection of the decision task (Carlebach & Yeung, 2020). Confidence is also sensitive to evidence acquired after a decision (Yeung & Summerfield, 2012; Murphy et al., 2015). Evidence accumulation models suggest that the decision variable continues to track evidence following a decision to facilitate the detection of errors, which is consistent with findings of confidence being predictive of error detection and changes of mind (Folke et al., 2016; Murphy et al., 2015; Pleskac & Busemeyer, 2010; Yeung & Summerfield, 2012; Yu et al., 2015).

Increasingly, research in decision confidence has moved beyond postdecisional locus theories. Adaptive and successful decision-making under uncertainty demands a close correspondence between confidence and evidence: self-reported confidence should ideally match the objective likelihood of the decision being correct, and low confidence should motivate compensatory processes such as additional effort or information-seeking prior to decision. Because confidence influences metacognitive control processes that are active during the decision process before a choice is made, it follows that confidence is not simply an output of the decision process. Instead, confidence is more likely computed during and plays an active role in evidence accumulation and choice selection (Petrusic & Baranski, 2003).

A number of recent models have suggested that framing confidence as the DV directly may be more parsimonious and a better fit for the data (Balsdon et al., 2020; Lee & Pezzulo, 2021). As the DV, confidence might be assumed to accumulate steadily but noisily in a manner that is proportional to the evidence, dynamically reflecting the probability of being correct given both the evidence accumulated as well as any choice made (Desender et al., 2021). This is supported by findings from neural studies where a ramping up of neural signals associated with confidence has been observed during the evidence accumulation phase (Murphy et al., 2015).

1.2.3 Factors affecting decision confidence

The standard model of decision confidence focuses on the relationship between evidence and confidence. While confidence tracks accumulated evidence during the decision process, reported confidence often deviates from the objective probability of being correct given the evidence, P(correct). Work in the perceptual decision-making literature over the years has shown that factors other than the decision variable can affect confidence: confidence varies with these factors even after evidence strength is matched or taken into account. In this section, we will explore these factors that influence the confidence estimate and that might explain why reported confidence deviates from P(correct).

Stimulus characteristics

Characteristics of the perceptual stimulus other than its identity (as the target stimulus or not) have been found to influence decision confidence. For instance, Rausch et al. (2018) noted that confidence in masked orientation judgments is influenced by both the strength of the sensory evidence about the identity of the stimulus and the quality of the sensory evidence (i.e., the visibility of the stimulus).

A characteristic that is often studied in relation to confidence is stimulus variance, or evidence reliability. While the mean of stimulus strength (i.e., objective difficulty) is typically considered the main source of evidence for a perceptual decision, de Gardelle and Mamassian (2015) found that manipulating the variance of stimulus strength in a perceptual task has effects on confidence that are dissociable from the effects of performance or difficulty. However, the direction of these effects varied across individuals: some participants reported greater confidence in high variability trials than in low variability trials at equal performance levels, while others reported lower confidence in high variability trials. These individual preferences were stable over time, replicating within individuals across repeated sessions. Subsequent authors (e.g., Boldt et al., 2017; Desender et al., 2018) have leveraged the effect that variance has on confidence, independent of performance, to experimentally manipulate confidence levels. Based on the results from one such study, Boldt et al. (2017) suggested that evidence reliability has a stronger impact on confidence than evidence strength itself, specifically affecting the mapping or conversion of feelings of certainty to explicitly reported confidence.

The influence of factors like stimulus visibility and stimulus variance on confidence, above and beyond task performance or difficulty, is something that many computational models of decision-making have difficulty accounting for (Rausch et al., 2018). Studies like Rausch et al. (2018) and Boldt et al. (2017) seem to suggest that metacognition in perceptual domains may be more like metamemory than previously thought. Instead of being driven exclusively by evidence strength, decision confidence might be better accounted for by a multi-cue model like those proposed in metamemory research (cf. Koriat & Levy-Sadot, 2001).

Decision time & urgency

Another category of factors that seem to influence decision confidence is time, specifically, time taken to form the decision, and time pressure or urgency. Kiani et al. (2014) found that when participants were asked to simultaneously report choice and confidence on a random dot motion task, reported confidence correlated directly with stimulus strength, as expected, but also inversely with reaction times. This inverse relationship between reaction time and confidence persisted even in a subsequent experimental manipulation where decision times were increased without affecting choice accuracy. Kiani et al. (2014) suggested that the time to decision is factored into internal computations of decision confidence because it tends to be a good proxy for task difficulty.

Additionally, time pressure may be a moderating factor that amplifies the effects of other factors on confidence. Based on an urgency-gating model of decision-making, Thura et al. (2012) reported that decision confidence simulated from available sensory evidence in a perceptual task decreased more steeply with increasing decision time when there were external time constraints than when there were not. This suggests that time pressure may exacerbate the collapse of decision bounds for confidence over the course of a decision. However, the degree to which time pressure exerts this effect may vary greatly between individuals. Rastegary and Landy (1993) noted a distinction between time pressure (i.e., externally imposed time constraints) and time urgency (i.e., internally imposed time constraints) on decisions. While internal feelings or signals of urgency strongly correlate with extrinsic constraints on time, the effect of a stressor (e.g., time constraint) on experienced strain (e.g., perceived urgency) and consequent performance is moderated by individual differences; not everyone responds to stressors in the same way (Eysenck, 1983; Rastegary & Landy, 1993).

Choice-induced effects

One of the key principles of confidence in the normative or formal models of confidence outlined in above sections is that confidence should reflect the balance of evidence between the choice options. However, a number of recent papers have found instances where confidence overweighs evidence in favour of the chosen option (positive evidence) or underweighs decision-incongruent (negative) evidence, suggesting that the act of choice commitment — typically construed as the endpoint of the confidence accretion process — may itself influence confidence (Navajas et al., 2016; Peters, Thesen, et al., 2017; Samaha & Denison, 2020).

The effect of choice commitment on human cognition and behaviour has been explored before. In a classic finding in the judgment and decision-making literature, Brehm (1956) presented his participants with eight consumer objects and asked them to rate the desirability of the items individually before and after choosing one of them. He found that, after making a choice, people tended to rate the option they had chosen as more desirable, and the unchosen options as less desirable, compared to predecisional ratings of desirability.

Choice-induced preference has traditionally been explained in terms of cognitive dissonance theory: people who feel uncertain about their choice might try to convince themselves, consciously or not, that the choice they made is better than initially estimated and consequently inflate the internal degree of certainty associated with that option in a kind of self-fulfilling prophecy (Bem, 1967; Brehm, 1956; Harmon-Jones & Harmon-Jones, 2007). However, Lee and Daunizeau (2020) suggested, alternatively, that uncertainty during the decision process, before a choice is made, may lead to more detailed and deliberate consideration, which then boosts confidence in the ultimate choice relative to a choice resulting from relatively more straightforward and less considered decision process.

The relationship between preference and consideration is likely bidirectional, as people have been found to prefer thinking about and seeking information that confirms their existing beliefs (Schulz et al., 2020). Several recent studies have suggested that evidence accumulation during the post-decisional phase — namely, the time between registering a choice and a confidence judgment about the choice — may be overweighted in favour of the chosen option (Navajas et al., 2016; Rollwage et al., 2020; Rollwage & Fleming, 2021). A confirmation bias† in post-decisional information-seeking and evidence accumulation would likely reinforce the presence of a positive evidence bias for confidence.

Choice-induced biases in confidence and information-seeking may be related to the idea of perceived agency. Jiwa et al. (2021), for instance, found that both confidence and the subjective value of information can be influenced by the "illusion of control": participants report higher confidence and valued non-instrumental information about lottery outcome more on trials where they chose the lottery that was being played than trials where they were assigned a lottery. The effect of choice agency on confidence in this study was independent of the actual probability of winning, suggesting an effect of perceived agency on confidence above and beyond objective likelihood or performance.

The pragmatics of confidence

As mentioned in the earlier section in this chapter on how we measure confidence in experimental settings, a central methodological concern in the confidence literature is that explicitly reported confidence differs from the internal confidence used for decision-making. It is plausible that, in explicit

† Somewhat confusingly, in some of the recent literature, confirmation bias has been used somewhat interchangeably to refer to similar and related effects (preference/bias for positive/confirmatory evidence) across confidence and information-seeking (cf. Rollwage et al., 2020; Rollwage & Fleming, 2021). In the rest of this paper, we will use positive evidence bias to refer to the effect of confirmatory evidence on confidence specifically and confirmation bias to refer to the effect of confirmatory evidence on information-seeking.

communication, participants might be withholding or misrepresenting, intentionally or unintentionally, the actual metacognitive or introspective knowledge they have about the decision (Massoni et al., 2014; Overgaard & Sandberg, 2012). As an instance of communicated linguistic meaning, the pragmatics of confidence — the context in which confidence is communicated — is an important factor to consider.

Internal and external confidence presumably serve different functions. The previous section highlighted the essential role of unreported, internal confidence in intrapersonal metacognitive control, contributing to error detection, changes of mind, and the regulation of behavioural strategies such as information-seeking. External, reported confidence, on the other hand, may have more of a social function, supporting suprapersonal decision-making by allowing individuals to make their private mental states and information accessible to other decision agents (Shea et al., 2014; Heyes et al., 2020). External confidence is important for facilitating decision-making in social contexts, since accurate communication of metacognitive representations in group-decision helps improve the overall accuracy of the group's decisions (Bahrami et al., 2010; Bang et al., 2017).

Explicit confidence also serves other social motivations. Bang et al. (2020) noted that a key feature of human cognition is the division (and the ability that enables the division) between private states of mind and public actions and communications. People may deliberately express a public confidence that is different from what they might privately use to inform their decision-making and cognitive control. For instance, people might be inclined to publicly express a higher level of confidence because of the social benefits associated with high confidence: confident people are generally more persuasive in group settings (Pulford et al., 2018) and may be perceived by others as more credible (Whitley & Greenberg, 1986), competent (Schlenker & Leary, 1982), and attractive (Buunk et al., 2002). In these situations, "faking" a higher confidence in public actions and communications is strategic and advantageous, perhaps even adaptive.

More generally, circling back to our original enquiry into how confidence is defined, it appears that confidence can take on different meanings in different contexts. Empirical research over the years has generally distinguished between global self-performance-based confidence and more trial-by-trial local decision confidence, partly because of differences in how they are measured (cf. Rouault et al., 2019). However, recently there has been a growing realisation of and empirical interest in how explicit reports of the more narrowly defined decision confidence can still communicate different concepts. Navajas et al. (2017), for example, found that, in about half of their participants, reported confidence reflected two separate quantities: the perceived probability of being correct and the perceived uncertainty in the estimated variable. Individual differences between participants in the meaning of their confidence reports were stable across repeated sessions and across different types of decision tasks. These individual differences in confidence reporting will be discussed further in Chapter 1.4.

1.3 Information-seeking in decision-making

People seek information all the time. Every day, people read books and articles, watch videos about current affairs or cat antics, scroll their phones for updates on their social circle and their interests, and talk to friends and colleagues about the latest developments in their community. Sometimes, information-seeking is associated with and in aid of an imminent goal, for instance, searching online for weather reports to decide if one should pack an umbrella or not. Other times — like searching online for explanations of how a magic trick works — it is less obvious why people might seek information beyond an intrinsic personal desire to do so. In this section, we will explore how and why people might seek information, how the value of information-seeking might be formalised, and why people might in fact favour alternatives to information-seeking in certain cases.

1.3.1 Types of information-seeking

Instrumental information-seeking

One of the primary motivations for seeking information is to help us make a decision about how to act. We look both ways before we decide whether it is safe to cross the street; we (hopefully) try to learn what political candidates are advocating for before we cast our vote. In a world full of uncertainty and serious consequences for mistaken judgments, seeking more information before committing to a decision can be an adaptive behavioural strategy. Information-seeking thus has an instrumental value, as it helps us achieve our goals — at the most basic level, obtaining reward and avoiding harm (Sharot & Sunstein, 2020; Stigler, 1961).

Normative Bayes-optimal frameworks for decision-making suggest that decision agents operating in an uncertain environment should act strategically to reduce the degree of uncertainty and make the external environment more predictable (Mushtaq et al., 2011; Pasquini et al., 2019; Peters, McEwen, & Friston, 2017; Schulz et al., 2020). Confidence is the output of a metacognitive process that monitors how uncertain we are about a decision, and thus plays an important role in guiding control strategies like information search during the decision process, especially in determining when information-seeking should start and stop. In Bayes-optimal decision agents, low levels of internal confidence about the choice should motivate information-seeking to reduce uncertainty and increase confidence enough to precipitate a decision (Schulz et al., 2020).

Animal studies often involve experiments that build on this intuitive link between confidence and information-seeking. Because it is not possible to ask for verbal reports of confidence from non-human subjects, information-seeking behaviour is often used as a proxy for confidence in studies of decision-making under uncertainty. Rhesus macaque monkeys and chimpanzees have been found to spontaneously seek further information about the location of a reward that had been hidden before making a final decision about the reward's location, and they generally sought more confirmatory visual information before making a choice when they could not observe where the reward had been placed (Call & Carpenter, 2001; Rosati & Santos, 2016). Rhesus monkeys have also been found to ask for hints when solving a difficult problem, at the cost of receiving a less desirable reward for the correct response, and steadily asked for fewer hints as their performance without hints on the task improved (Kornell et al., 2007). These results suggest that low internal confidence motivates exploratory behavioural strategies before decision commitment, and that these search strategies are generally adaptive rather than excessive or inefficient.

The empirical literature for humans generally confirms the idea that confidence is predictive of information-seeking. Using a perceptual decision-making task where the experimenters orthogonally manipulated the strength and variability of evidence to make participants systematically less confident at equal levels of objective difficulty, Desender et al. (2018) found that confidence was a significant predictor of strategic information-seeking behaviour. Participants were more likely to choose to see an easier version of the stimulus again before decision when they were less confident, relative to a matched condition where they were equally accurate. On a similar perceptual decision-making task where participants were offered the chance to hear an advisor's judgment after their initial decision but before their final decision, Pescetelli et al. (2021) found that participants were more likely to seek advice before finalising their choice when they were less confident about their initial judgment. This suggests that the relationship between confidence and information-seeking in humans holds true regardless of whether that information comes from firsthand or secondhand sources.

Evidence suggesting a causal relationship of confidence on information-seeking comes from metamemory research. Metcalfe and Finn (2008) found that participants' choices of which materials to restudy closely tracked their confidence judgments about future recall for learned materials rather than their actual or historical recall performance. Bjork et al. (2013) concluded, based on their review of research on self-regulated learning, that confidence plays an important and overarching role in determining how people self-regulate their learning, especially what behavioural strategies they employ and when these strategies are initiated and terminated.

Overall, the evidence across different species and different task domains suggests that confidence directly regulates metacognitive control mechanisms like information-seeking to optimise performance. When external uncertainty is high, low internal confidence should motivate decision agents to seek more information to increase their confidence. Conversely, as more information is acquired and confidence increases, we would expect information-seeking to lessen and eventually stop, as further information results in increasingly marginal reductions in uncertainty.

Non-instrumental information-seeking

Although information-seeking often has instrumental value, humans and other animals engage in a great deal of information-seeking — sometimes at great cost — that may not be directly related to a goal involving reward gain or harm reduction (Brydevall et al., 2018; FitzGibbon et al., 2020). Humans and other species have shown willingness to pay for information about future outcomes, even when these have no impact on the actual outcome, or for information that has no relationship to monetary rewards (Bennett et al., 2016; Kang et al., 2009; Rodriguez Cabrero et al., 2019).

Curiosity is often cited as the motivator for this kind of non-instrumental information-seeking, but there are differing accounts as to what curiosity actually means. Early theories about curiosity attempted to establish it as a primary behavioural drive akin to hunger or fear (e.g., Dashiell, 1925; Harlow et al., 1950). Berlyne (1954) proposed that the motivational force behind curiosity was that of stimulus conflict or incongruity, where stimuli that are complex, novel, or surprising can elevate arousal and trigger a curiosity drive that motivates information-seeking behaviour. Magic tricks, for example, commonly arouse people's curiosity because they contain all of the above qualities.

Following from these drive theories of curiosity, one idea is that information itself might be motivationally salient because it possesses some kind of hedonic value. van Lieshout et al. (2021) found that people tend to be attracted to non-instrumental information that can make them feel good and avoid such information that might make them feel bad, suggesting that non-instrumental information-seeking may be affectively rewarding in some way. FitzGibbon et al. (2020) postulate that this hedonic value of information may be learned: in a highly uncertain world, people may very strongly desire to reduce that uncertainty, and that through reinforcement, may learn to associate the acquisition of information with a kind of hedonic experience. This experience might be inherently pleasurable, or may simply be relief from a negative affective state associated with the awareness of missing information (FitzGibbon et al., 2020; Loewenstein, 1994). This idea of hedonic value may explain why people might choose to search for trivial information like how a magic trick works: one might experience pleasure in understanding how to perform a magic trick, or feel a certain relief on resolving questions about how a magician appeared to defy the laws of physics.

However, there are many cases of non-instrumental information-seeking where the hedonic value of information is less clear. FitzGibbon et al. (2020) found several studies where people were willing to seek non-instrumental information even when it had a negative valence, came at significant cost, or would otherwise make them feel bad. This suggests that non-instrumental information-seeking is rewarding or high in utility beyond its hedonic value. van Lieshout et al. (2021) recently reported findings that self-reported curiosity increases with increasing uncertainty about outcomes in a gambling task, whether they are positive or negative. The authors proposed that people find uncertainty unpleasant and that curiosity might reflect an aversive drive to reduce this unpleasant state of uncertainty.

These finding seem to suggest a convergence between non-instrumental and instrumental information-seeking: the underlying motivation in both cases appears to be a basic desire to reduce uncertainty about the world (Kagan, 1972). Many psychologists throughout the years have proposed that curiosity might reflect a human desire to make sense of the world and that curiosity is triggered by violated expectations — in essence, a kind of prediction error (cf. Hebb, 1955; Hunt, 1965; Piaget, 1952). Individuals might be motivated to seek information because even if the information we acquire is not useful for reducing uncertainty in relation to an immediate goal, it may prove to be useful in the future. More generally, it may help us refine our mental representations of the world and reduce uncertainty about the external environment.

1.3.2 The value of information

If the general motivation behind information-seeking is uncertainty reduction, whether it is in a more local (decision oriented) or global (cognitive representational) sense, then it would be adaptive for individuals to first assess whether their information search will actually result in a meaningful reduction of uncertainty before engaging in information-seeking. The core of this initial predictive assessment would be a cost-benefit analysis of information-seeking, namely, whether the benefits of the information acquired will outweigh the costs of acquiring that information.

Formalising informational value

Classical information theory suggests that, if the primary value of information is in its ability to reduce uncertainty, we can formalise informational value in terms of how much surprise it generates (Goodfellow et al., 2016). Shannon, in his seminal 1948 paper on information theory, proposed quantifying the amount of information h contained in a discrete event x that has occurred (i.e., the event's informativeness $h(x)$) as a logarithmic function of its probability $P(x)$:

[Eq. 1.1]

$$h(x) = -log_2(P(x))$$

In this model, rarer events are more surprising and therefore higher in informational value because they would result in bigger updates to one's existing internal model of the world (Friston, 2010; Goodfellow et al., 2016; Shannon, 1948). When generalised to an uncertain event X comprising an array of discrete events x_0, x_1, x_2, ..., each with different probabilities, we can calculate an overall entropy $H(X)$ of the uncertain event – essentially, the average level of information or surprise inherent in the variable's outcomes. This is equal to the sum of the probability of each discrete event outcome multiplied by its informativeness:

[Eq. 1.2]

$$H(X) = -\sum_{i=1}^{n}(P(x_i) \cdot log_2(P(x_i)))$$

When faced with an opportunity to seek information, an entropy-minimising decision agent should assess the entropy associated with each of the two options (seeking more information versus committing to a decision based on the existing evidence). Agents should choose the information-seeking option only if it leads to a relative reduction in entropy and should continue to seek information if entropy can be reduced by new information (Quinlan, 1986; Brownlee, 2020).

Expected utility of information-seeking

The issue with the classical information theory approach is that it focuses solely on information gain, or the theoretical reduction in entropy brought about by acquiring more information. While prioritising information gain makes sense in its original context of statistics and machine learning, the singular focus on information gain is not pragmatic in many real-life decision-making contexts, as the process of acquiring new information comes with various costs and risks. In some situations, the cost of information-seeking is explicit or tangible, such as a researcher paying to access a scientific article online. More often though, the price of information is implicit or less tangible — in the case of that same researcher, perhaps the time and energy sunk into their informational search, and the opportunities they might have missed or had to pass up on because of their current endeavour.

Decision agents must also assess the risks associated with information-seeking, especially how likely it is that useful information will be acquired. Many instances of information-seeking in the real world — for instance, an animal deciding whether to venture outside of its safe home territory to find new foraging territories — involve a real risk of information acquisition not being successful, despite the initial investment of time and energy into information search, or of the information acquired being inaccurate or unreliable. Recent work published by Trudel, Scholl, and colleagues (2021) suggest that this may lead to a decision agent's information-seeking strategy changing over time, shifting from an initially exploratory preference for high-uncertainty predictors towards a more conservative preference for higher-accuracy, lower-uncertainty predictors as more information is acquired, and that this behavioural transition from a positive to negative uncertainty preference is paralleled by changes to the polarity of signals reflecting decision uncertainty in the human ventromedial prefrontal cortex.

The expected utility of the information-seeking strategy can therefore be calculated as the expected gain in utility resulting from the information gain minus the expected costs and risks associated with information-seeking. This is what we intuitively expect decision agents to assess before either committing to a decision response or seeking further information. A recent fMRI study by Kaanders et al. (2021) found that activity in the human medial frontal cortex was predictive of further information sampling before choice, over and above choice difficulty, suggesting a possible neural correlate of this assessment of expected information-seeking utility.

While theoretically straightforward and potentially paralleled in neural findings, this model is difficult to assess behaviourally because of the difficulty of quantifying the utility associated with intangible costs. Empirical assessments of simpler versions of this model focusing on purely explicit costs and rewards have found that people tend to seek less information than the optimal policy derived from an expected value maximisation principle, even when there was no explicit cost of information (Lanzetta & Kanareff, 1962). The authors suggested that participants' information-seeking strategy might be better accounted for by implicit costs such as the time spent on information acquisition and processing.

In addition, the empirical literature on information-seeking suggests that there are factors beyond information gain and information costs that influence information-seeking choice and subjective valuations of information. Information-seeking is often subject to confirmatory biases similar to those seen for confidence, with people tending to seek information that confirms what they already know or believe; these biases may be motivated by more fundamental, stable individual tendencies like the tendency to approach items associated with reward (Hunt et al., 2016). The importance of the affective value of information has been discussed in the non-instrumental information-seeking research mentioned in previous sections (cf. FitzGibbon et al., 2020; van Lieshout et al., 2021). Additionally, recent research suggests that agency over actions, specifically decision choices, also contributes to the subjective value of information, with people valuing non-instrumental information about the outcome of a choice more when the choice was self-made rather than externally imposed (Jiwa et al., 2021).

Integrative valuation frameworks for information-seeking

In part due to the discussed limitations of valuation approaches that focus exclusively on reward or uncertainty reduction, many researchers are now pushing for more integrative frameworks of information value. These frameworks tend to consider the broader implications of information-seeking and build in more latitude for the influence of individual differences and non-economic factors in the initial assessment of whether to seek information.

Kobayashi and Hsu (2019) found that information-seeking behaviour in humans can be best captured by a subjective value of information model that incorporates instrumental benefits as well as the anticipated hedonic value of the outcome (e.g., dread or savouring). Sharot and Sunstein (2020), on the other hand, have proposed that the value of information is a composite of three factors, covering the various rationales for information-seeking behaviour. The first factor is instrumental value, or the effect of information on action plans: higher instrumental value is assigned to information leading to better outcomes. The second is hedonic value, namely, the effect of information on affect, with higher hedonic value assigned to information leading to positive affective responses. The final factor is cognitive value: this is the effect information has on our cognitive worldview, with higher cognitive value assigned to information leading to more accurate mental models of the external environment. The authors suggest that individual differences are reflected in the variability in the estimates of the factor values, as well as in the weights assigned to the three factors during the calculation of the integrated value. Depending on this final integrated value, decision agents might choose to seek more information (if the value is positive), avoid further information (if the value is negative), or exhibit indifference to further information (if the value is zero). Recent experimental evaluations of this tripartite model (e.g., Cogliati Dezza et al., 2021; Kelly & Sharot, 2021) have generally supported the model and provided evidence that different aspects of informational value are being integrated in people's assessment of whether to seek or avoid further information.

Computational and neural findings support the basic idea of a common valuation mechanism for information. White et al. (2019) reported a cortico-basal ganglia mechanism involved in motivating information-seeking to reduce uncertainty. Spontaneous increases in this network of anterior cingulate cortex and basal ganglia subregions in primates predicted gaze shifts towards informational sources associated with uncertainty reduction, and pharmacological disruptions of this network reduced information-seeking behaviour (White et al., 2019). These findings are paralleled by Kobayashi and Hsu (2019), who found that the information-seeking choice induces a pattern of activity in reward learning regions such as striatum and ventromedial prefrontal cortex, similar to the activity produced by the risky reward choice. This suggests that information value may be encoded, at least neuronally, in a manner similar to more conventional indices of value and reward. However, as recent work from Jezzini et al. (2021) suggest, there may be differences in how information of different valences is encoded, with primate anterior cingulate cortex showing evidence of discrete neuronal subpopulations encoding in a valence-specific manner.

1.3.3 Alternative strategies for managing uncertainty

Not every decision made under uncertainty results in an information-seeking strategy being implemented. We have explored various factors and considerations that might bias an individual towards seeking further information or not. The assumption so far has been that information-seeking is the only strategy we can adopt to manage uncertainty regarding a decision. However, there are other strategies that decision agents can employ to manage uncertainty, and the value of information-seeking may depend as well on the availability and attractiveness of alternative strategies.

Seeking feedback

The assumption in much of the information-seeking literature, especially on the relationship between confidence and information-seeking, is that external feedback is absent or sporadic. However, in cases where it is present, feedback may have higher informational value than or be preferred to more probabilistic predecisional information about outcome. While subjective confidence in perceptual judgment paradigms is closely linked to objective properties of the stimulus, confidence in more complex judgment or learning paradigms (e.g., abstract rule learning) may be better predicted by feedback about recent accuracy. Martí et al. (2018), for example, found that recent feedback about accuracy was the best predictor of participants' reported confidence on a Boolean concept-learning task, more so than the entropy of their rule-learning model for the task.

There is substantial existing literature in organisational and educational psychology about how people actively seek and learn from feedback. People proactively seek feedback on things that are important and in situations that are new and uncertain; in organisations, this feedback-seeking is motivated by desires to improve performance and align better with organisational norms and rules (Ashford, 1986). Ashford also found that, the longer people stay in an organisation, the less frequently they seek feedback, possibly as a result of a desire to appear confident in the eyes of others. Goodman (1998) proposed a distinction between the external feedback from others (e.g., a supervisor or experimenter) that is typically investigated in these organisational settings and feedback from the task. Goodman found that external feedback, while beneficial for performance, is actually detrimental to learning, and that the benefits of task feedback for learning derive partially from how it helps train metacognitive monitoring processes like error detection and correction. Chou and Zou (2020) provide further evidence that external feedback is beneficial for refining metacognitive monitoring and control processes necessary for self-regulated learning, including implementing and monitoring learning strategies such as seeking further study.

However, direct empirical assessments of people's preference for feedback about outcome versus probabilistic predecisional information about outcome are lacking. We might speculate that, because postdecisional feedback is a more certain source of information about one's performance than probabilistic predecisional information, some individuals may prefer seeking feedback about their performance to seeking predecisional information that is only predictive of performance. This motivated us to investigate people's preferences for feedback-seeking versus information-seeking strategies in Experiment 3 [Chapter 4].

Escaping uncertainty

Another strategy that decision agents might favour over further information-seeking is escaping an uncertain decision situation, usually in favour of a less desirable but certain outcome. The traditional opt-out paradigm, prevalent in the animal metacognition literature as an implicit means of investigating uncertainty monitoring, involves a subject choosing between a high-uncertainty outcome option (choosing to make a difficult decision) and a known outcome option (choosing to opt out). In some versions of these paradigms (which we might distinguish as "sure bet" paradigms), subjects are offered a high-uncertainty, high-reward option versus a guaranteed low-reward option, with no significant punishments for incorrect responses other than a short time-out or delay (cf. Foote & Crystal, 2007; Hampton, 2001; Kiani & Shadlen, 2009). In other versions of the paradigm, subjects are offered the option to escape a highly uncertain decision offering both reward for correct responses and punishment for incorrect responses; the opt-out response in these "avoid choice" paradigms can be rewarded to a lesser degree (cf. Smith et al., 1995; 1997) or not rewarded at all (cf. Perry & Barron, 2013).

Opting out of high uncertainty decisions in both the memory and perceptual domains has been found in a number of non-human species, including dolphins (Smith et al., 1995; 1997), rhesus monkeys (Hampton, 2001; Kiani & Shadlen, 2009), rats (Foote & Crystal, 2007), and bees (Perry & Barron, 2013). In comparative studies using these paradigms (e.g., Smith et al., 1995; 1997), humans have also been found to make patterns of opt-out responses that are very similar to non-human counterparts, and to explicitly attribute their selection of opt-out responses to feelings of high uncertainty (Smith & Washburn, 2005).

Early evidence from pharmacological manipulation studies in humans has suggested a possible neurobiological mechanism for this escape preference. Salvador et al. (2020) found that people opted out of perceptual decisions substantially more frequently under the influence of ketamine than the placebo, suggesting greater decision uncertainty despite matched decision accuracy. The authors proposed a possible link between n-methyl d-aspartate (NMDA) receptor function and individual differences in how uncertainty and expected utility of available strategies might be computed.

Avoiding uncertainty

A third option that individuals might favour over further information-seeking is the avoidance of uncertainty. Avoidance can refer to two separate but similarly motivated strategies: avoiding or refusing to participate in decisions featuring high uncertainty (what we might term "uncertain choice avoidance") and avoiding further engagement with information sources that increase uncertainty (what we might call "uncertain information avoidance").

Uncertain choice avoidance is well-known. An established example of this is ambiguity aversion: humans and other primates generally and consistently favour choices with a transparent or known risk over choices featuring ambiguity, where the likelihood of risk is unknown, even in instances where the ambiguous choice may have a larger expected utility (Ellsberg, 1961; Hayden et al., 2010; Rosati & Hare, 2010). While ambiguity aversion has been previously attributed to pessimistic beliefs about the outcome of the ambiguous condition, Ahrends et al. (2019) found that people's expectations about the outcome in the ambiguous condition was appropriately calibrated, and that pessimism about outcomes generally did not affect aversion to ambiguity.

Uncertain information avoidance is also well-established. An example of this is confirmation bias, where people increase the strength of their existing beliefs by filtering out conflicting evidence (Salvador et al., 2020). In recent years, there been a resurgence of interest in understanding why people might choose to minimise the difference between the external reality and their cognitive models of reality by actively managing the reality of which they are aware instead of refining their cognitive models with more information about the external world. Researchers have tried to establish why these biases might develop and under what circumstances they might be adaptive.

Sharot and Sunstein (2020) suggest that disconfirming evidence, while helpful in refining the model for the future, can reduce people's sense of comprehension of the current situation. This can be discomfiting and involve large amounts of work rebuilding existing cognitive models, thus generating negative hedonic and cognitive value. In this case, the expected utility of information-seeking may be become negative overall, resulting in avoidance of further information despite the necessity of acquiring further information to maintain an accurate cognitive representation of the world. Several factors can exacerbate confirmation bias, including high levels of *a priori* confidence and shorter decision timeframes (Phillips et al., 2014; Salvador et al., 2020). While some have argued that confirmation bias may be adaptive in certain circumstances, more often than not it can push decision agents towards a decalibration of their beliefs about the world and the adoption of further avoidance strategies that are suboptimal or even maladaptive (Rollwage & Fleming, 2021; Schulz et al., 2020).

1.4 Individual differences in confidence and information-seeking

The research reviewed in the preceding sections has focused on general properties of confidence and information-seeking in human decision-making. However, previous research has also revealed substantial individual differences in these functions. In this section, we will explore variations in reported confidence and information-seeking behaviours that have been found in the general population, and recent research linking these individual differences to possible metacognitive disturbances in psychiatric disorders.

1.4.1 Variation in reported confidence

People vary greatly in how confident they are. These variations in confidence can also be observed both in general levels of self-confidence (i.e., global self-performance estimates), as well as in rated confidence in relation to decisions and task performance (i.e., local or online confidence) (Burns et al., 2016). On any given task, there is considerable variation in reported confidence between individuals (Ais et al., 2016; Kleitman & Stankov, 2007; Pallier et al., 2002; Stankov et al., 2014).

Within individuals, however, the reported confidence for the same task tends to fall within a distribution that is consistent across timepoints and stable across similar types of tasks in different modalities (Ais et al., 2016; Blais et al., 2005). Although there were still large variations in the shape of participants' confidence distributions, Ais et al. found the mean of their confidence distribution (i.e., metacognitive bias) was the most efficient index for discriminating between individuals, correlating strongly with a self-reported questionnaire measure of optimism bias measured a year later. This suggests that average reported confidence (i.e., metacognitive bias) may reflect stable and trait-like differences between individuals, at least within a certain task domain.

The idea of a general trait-like aspect of confidence is echoed in Pallier et al. (2002), who reported that the accuracy of confidence judgments across a wide variety of psychometric tasks across modalities is mediated by a latent factor that the authors called a "confidence trait". The Pallier et al. confidence trait was influenced to a small extent by, but remained generally relatively independent of, cognitive ability and tested personality constructs, and was stable across a range of task difficulty levels. Subsequent factor-analytic approaches by Kleitman and Stankov (2007) have yielded similar results, finding local confidence to be relatively independent of cognitive abilities but correlated with more stable and enduring traits such as self-beliefs about competency and knowledge about cognition.

The presence of marked but consistent variation between individuals in their reported confidence raises a question of what might be driving these differences. One hypothesis is that this variation reflects individual differences in where individuals might be setting their confidence threshold for decision. Jackson et al. (2016) found individual differences in decision confidence threshold that correlated well with various decision response characteristics such as decisiveness, recklessness, and optimality. The researchers, in a later study, found that this confidence threshold varied significantly between task contexts, suggesting that the confidence threshold can be influenced by contextual payoff parameters (Jackson et al., 2017).

Another hypothesis is that differences in reported confidence reflect differences in comprehension and communication. Many studies over the years have suggested that people may use the same scale in different ways based on idiosyncrasies in interpretation. People might perceive differences in distinct dimensions, but instead of isolating and only reporting the difference in the dimension of interest, they might inadvertently collapse the different dimensional ratings into a single overall rating of "differentness" and end up reporting this judgment on an inappropriate scale. Frank et al. (1993), for example, found evidence of this in judgments about taste intensity: participants in the study rated drinks to be much sweeter when they were only given a single sweetness intensity scale to discriminate drinks than when they were given multiple rating scales for each component gustatory factor (e.g., sourness, bitterness, fruitiness).

Confidence reports might also be susceptible to collapsing dimensions of judgment. While Bayesian theory proposes that confidence should only reflect the perceived probability of being correct and that this value should closely reflect the actual probability of being correct, reported confidence in some individuals has been found to reflect the perceived uncertainty in the variable being estimated as well (Navajas et al., 2017). Navajas et al. (2017) also found that the relative weight that these two interpretations have on the overall reported confidence was stable in individuals across task domains and across timepoints. Thus, responses to a decision-point confidence probe like “How sure are you of your decision?” thus might reflect the participant’s assessment of the estimated probability of the option they selected, the participant’s belief in their own decision accuracy or competence, or a mix of the two.

So far these hypotheses have assumed that differences in reported confidence are non-deliberate. But in reality, we often report confidence to communicate metacognitive information to other decision agents, which means that reported confidence can be shaped by deliberate intent and by social, cultural, and demographic factors (Shea et al., 2014; Bang et al., 2017). Publicly expressed confidence often reflects the social context of many decision processes: for instance, people might report higher or lower confidence than their internal confidence value to improve compatibility with the expressed confidence of others during group decision-making or to manage how others might perceive them more generally (Bang et al., 2017; 2020).

1.4.2 Variation in information-seeking behaviour

Individual differences can also be observed in information-seeking behaviour. Humans and other species like bees, dolphins, and chimpanzees all vary greatly in how frequently they seek information when making decisions under uncertainty (Call & Carpenter, 2001; Desender et al., 2018; Pescetelli et al., 2021; Perry & Barron, 2013). While information-seeking preferences are known to vary with factors such as confidence, information cost, and social context, this section will focus instead on the substantial differences between individuals that remain even when these factors are accounted or controlled for.

Many studies of decision behaviour in humans have pointed out large individual differences in information-seeking behaviour. Desender et al. (2018) found that participants' information-seeking frequency on a perceptual decision-making task — namely, how often they chose to see an easier version of the perceptual stimulus before committing to a final decision — had a wide range (9% to 89% of the time). Frequency of information-seeking did not correlate reliably with individual differences in measures of accuracy or reported confidence. Pescetelli et al. (2021), in a similar perceptual decision-making paradigm that offered participants the option to skip or pay for advice before committing to a final decision, also found considerable variability in advice-seeking (1% to 79% of the time). While confidence was generally predictive of advice-seeking across the participant group, Pescetelli et al. noted that the relationship between confidence and advice-seeking was less consistent at the level of individual participants. Some consistently sought advice when their confidence was low and declined advice when their confidence was high, while others displayed idiosyncratic patterns of information-seeking that were more independent of confidence. These individual preferences in information-seeking frequency appear to be relatively stable within individuals across timepoints. Call and Carpenter (2001) reported that preferences for information search before choice are similarly stable in chimpanzees and humans, and in the rare cases where a chimpanzee starts seeking more information before choice, this behavioural change is sudden and thought to be more reflective of insight or associative learning than typical trial-and-error learning.

There are several hypotheses regarding why these individual differences in information-seeking preference exist. Hausmann-Thürig & Läge (2008) have suggested that the extent of information-seeking (i.e., the point at which a participant stops seeking further information and commits to a decision) may reflect individual differences in the confidence threshold for decision. Sharot and Sunstein (2020), on the other hand, speculated that individual differences in information-seeking behaviour could be related to differences in how the component factors of information value (instrumental, hedonic, and cognitive) are estimated and weighed relative to each other in the calculation of the expected utility of information-seeking. Differences in this subjective valuation of information-seeking may also contribute to differences in how people manage uncertainty and what strategies they employ, including possible preferences for alternatives to information-seeking. In some individuals, the negative hedonic value associated with uncertain decisions (e.g., stress and discomfort) may push them towards less optimal strategies for coping with uncertainty, such as using heuristics to oversimplify the decision, avoiding further disconfirmatory information, or escaping an uncertain situation by choosing a less desirable option (Jensen et al., 2014; Pasquini et al., 2019; Salvador et al., 2020).

Many researchers have speculated about and found evidence of a link between these relatively stable behavioural differences between individuals and individual differences in certain personality or cognitive traits. For instance, Schulz et al. (2020) found that participants who rated higher in a latent factor of dogmatism were less likely to seek further information when it was offered after an initial decision in a perceptual decision-making task. Based on computational modelling of trial-by-trial confidence-driven information search, Schulz et al. suggested that this observation could reflect dogmatic individuals relying less on their internal confidence to guide their information search, especially about when they should stop or continue seeking information. Several studies (cf. Jensen et al., 2014; Salvador et al., 2020) have also speculated on a relationship between reduced information-seeking and an intolerance of uncertainty trait, which is closely associated with and has been proposed as a mechanism underlying a number of psychiatric disorders.

1.4.3 Metacognitive disturbances in psychiatric disorders

The previous sections outlined how there is a large degree of variation in the general population in confidence and information-seeking — two measures that give us some insight into underlying metacognitive monitoring and control processes, respectively. There have been attempts to link stable individual differences in these measures with personality and cognitive traits that are more commonly studied or discussed in relation to psychiatric disorders. Metacognition is important for adaptive functioning; abnormal or disrupted metacognition might be linked associatively or causally to other maladaptive symptoms exhibited in psychiatric disorders, and is thus of psychiatric interest and significance (Wells, 2000; Wells et al., 2012). In this final section, we will explore some of the associations that have been found between metacognition and psychiatric disorders or transdiagnostic dimensions of psychiatric symptoms, as manifested in their relationships with confidence and information-seeking measures.

Disturbances in confidence

Confidence helps us make predictions, learn from mistakes, and plan subsequent actions in the absence of feedback (Yeung & Summerfield, 2012). As a metacognitive monitoring process, its accuracy and precision are essential to adaptive function. Researchers in recent years have hypothesised that abnormalities of confidence underlie many psychiatric symptoms and can be seen both subclinically in the general population and clinically in the patient population (Hoven et al., 2019; Rouault et al., 2018).

Confidence bias

Several disorders have been linked with biases in reported confidence. Underconfidence biases (i.e., where reported confidence is lower than actual performance) have been found in relation to anxiety, depression, and obsessive-compulsive disorders (OCD). Within the general population, symptoms of anxiety and depression have all been found to predict an underconfidence bias in perceptual decision-making (Rouault et al., 2018). For anxiety, however, there is mixed evidence that underconfidence is seen at the disorder level, with some studies reporting anxious patients being less confident than controls in the memory domain while others finding no differences between patients and healthy controls in other domains like general knowledge or interoception (Hoven et al., 2019). Trial-by-trial confidence in memory performance has been found to track depressive severity, with clinically depressed patients exhibiting lower confidence than individuals with dysphoria, who in turn had lower confidence ratings than non-depressed controls (Fu et al., 2012). Similar findings of underconfidence in depressive patients are seen for other cognitive domains like visual and social perception (Hoven et al., 2019). Individuals with high obsessive-compulsive tendencies report lower confidence and more reliance on false feedback when evaluating muscle tension in a false bio-feedback task (Lazarov et al., 2012; Zhang et al., 2017). In clinical populations, OCD has been linked to underconfidence in memory, perception, and interoception in both neutral and OCD-relevant contexts (Hoven et al., 2019).

Overconfidence, on the other hand, has been found in connection with addiction and schizophrenia. Both clinical and subclinical schizophrenia are associated with increased confidence on error responses relative to controls; this is found in various domains from memory to social cognition to perceptual judgment (Hoven et al., 2019; Köther et al., 2012; Moritz et al., 2003; Moritz et al., 2014). For instance, Moritz et al. (2014) found that patients with schizophrenia make significantly more high-confidence errors on a perceptual judgment task than age- and education-matched OCD patient and healthy control groups. In the general population, proneness to delusional thinking, a cognitive style seen in schizophrenia, has been linked to overconfidence in probabilistic reasoning, particularly in initial decisions (McKay et al., 2006; Warman, 2008).

A few studies have found significant positive correlations between gambling severity and overconfidence at the subclinical level (Goodie, 2005; Lakey et al., 2007). Brevers et al. (2014) found evidence of disturbances in confidence calibration on an artificial grammar-learning task in gambling disorder patients: while they exhibited similar levels of confidence to healthy controls, they actually performed worse, and unlike healthy controls, their confidence generally failed to correlate with their performance. Le Berre et al. (2010) found similar disturbances in calibration with alcohol-use disorder patients in the memory domain, who performed worse than healthy controls but typically overestimated their recognition performance, leading to less accurate confidence judgments.

There have also been attempts to link confidence bias with transdiagnostic dimensions. Underconfidence, for instance, has often been linked to uncertainty intolerance. People with high uncertainty intolerance find uncertain situations and experiences to be highly threatening and distressing, and the dimension is closely associated with anxiety and depressive disorders (Buhr & Dugas, 2002; Freeston et al., 1994; Dugas et al., 2004; Jensen et al., 2016; Lauriola et al., 2018; Robinson & Freeston, 2015). Individuals with high uncertainty intolerance tend to become progressively underconfident as they make more decisions, while those with low uncertainty intolerance show the reverse pattern (Jensen et al., 2014). More recently, using a factor analysis approach in a large general-population sample, Rouault et al. (2018) found that a latent anxious-depressive symptom factor that was significantly associated with underconfidence and a latent compulsive behaviour and intrusive thought factor that was significantly associated with overconfidence in the perceptual domain. There was, however, no relationship between these latent factors and decision-level measures like accuracy and decision formation parameters.

Confidence sensitivity

Fewer studies have looked at how confidence sensitivity might be disturbed or different in psychiatric disorders. Of the studies that have, the results are mixed or inconsistent. The focus of many of these studies is schizophrenia and subclinical symptom dimensions relating to it. Hoven et al. (2019) suggested that the overconfidence in errors typically seen in clinical and subclinical schizophrenia may lead to decreased discrimination and metacognitive sensitivity. Compared to other patient groups such as OCD and post-traumatic stress disorder, patients with schizophrenia showed a specific impairment in discrimination due to higher confidence for errors on a memory task (Moritz & Woodward, 2006). In the perceptual domain, subclinical psychotic dimensions like paranoia have been found to be associated with lower confidence discrimination in perceptual decision-making (Moritz et al., 2014). Rouault et al. (2018), on the other hand, did not find any direct relationships between schizotypy and metacognitive efficiency in perceptual decision-making. Results from patient studies are equally mixed. While Bruno et al. (2012) failed to find an impairment in discrimination, Davies et al. (2018) found that psychotic patients had significantly lower metacognitive sensitivity (meta-d') relative to healthy controls despite similar performance and confidence levels. Hauser, Allen, et al. (2017) also found lower metacognitive efficiency in highly compulsive patients in a motion detection task.

Research on confidence sensitivity has been very limited in other disorders. There is some evidence of disturbances to discrimination and metacognitive efficiency associated with depression (Hancock et al., 1996) and obsessive-compulsive tendencies (Hauser, Allen, et al., 2017), but most of these findings need replication and verification in subsequent research.

Disturbances in information-seeking

In the sections above, we have discussed how abnormalities of confidence — an internal signal tracking our level of uncertainty — are observed in many psychiatric disorders. In addition to a disturbed ability to monitor uncertainty, many psychiatric disorders also feature an impaired ability to cope with uncertainty, manifesting in excessive negative affect and maladaptive behavioural strategies. Information-seeking, while normally an adaptive strategy for reducing uncertainty, can become suboptimal or pathological in some psychiatric disorders.

Elevated information-seeking

Increased information-seeking is most commonly associated with anxiety and obsessive-compulsive disorders — unsurprising, given the persistent threat-anticipation that is central to this family of disorders (Grupe & Nitschke, 2013; Rachman, 2002). In the general population, anxious and obsessive-compulsive traits have been found to be associated with elevated and suboptimal information-seeking preferences, namely, increased willingness-to-pay for non-instrumental information about uncertain future outcomes, even if it comes at the expense of future reward (Bennett et al., 2021).

Elevated information-seeking is particularly evident in OCD. One of the hallmark behaviours of OCD is compulsive checking, where individuals repetitively check or seek information about things that they are worried about (Strauss et al., 2020). Checking might be fixated on a tangible feature of the external environment, such as a stove or a door lock, or a less tangible source of information like text messages or the news (Rachman, 2002; Doron et al., 2012). In a recent meta-analysis of the compulsive checking literature, Strauss et al. (2020) found that OCD patients exhibited elevated checking behaviours relative to controls in perceptual judgments but not probabilistic reasoning, and no differences in checking were found between neutral and threatening conditions. The authors suggested that, while both types of tasks involve uncertainty and offer opportunities to seek information to reduce uncertainty, there are differences in the nature of information-seeking in the two tasks: perceptual tasks generally involve reviewing the same information, while probabilistic reasoning tasks generally involved acquiring new information. Elevated information-seeking in OCD would appear to be motivated specifically by a distrust of perceptual information, rather than a general desire to maximise uncertainty reduction; while checking temporarily reduces the distress associated with uncertainty, it can ironically increase feelings of uncertainty (van den Hout & Kindt, 2003; Strauss et al., 2020). This makes elevated information-seeking suboptimal both in its costliness and in its ineffectiveness at actually reducing uncertainty, and promotes the perpetuation of the disorder through a vicious cycle.

The transdiagnostic dimension of uncertainty intolerance attempts to capture this common acute negative affective response to uncertainty across disorders. Uncertainty intolerance has been found to be a more significant predictor of elevated information-seeking than more disorder-specific measures (Hauser, Moutoussis, et al., 2017; Rosen & Knäuper, 2009). On probabilistic reasoning tasks featuring freely determined sequential information sampling, individuals with higher scores on the self-reported Intolerance to Uncertainty Scale (IUS) tend to seek more information than, but exhibit similar final levels of task performance to, individuals with lower IUS scores (Hauser, Moutoussis, et al., 2017; Ladouceur et al., 1997). People who are more intolerant of uncertainty have also been found to engage in more feedback-seeking behaviours and may prefer feedback as a source of information in ambiguous situations (Bennett et al., 1990).

Reduced information-seeking

Reduced information-seeking before decision, or a "jumping to conclusions" (JTC) bias, is commonly discussed in relation to schizophrenia and related dimensions such as delusion proneness and paranoia. This bias in information-seeking is typically evaluated with the Beads task, a probabilistic reasoning task where participants are shown beads being drawn out of one of two jars, each containing beads of two different colours in complementary ratios (e.g., 85:15 red to green in one and 85:15 green to red in the other) (Phillips & Edwards, 1966; Huq et al., 1988; Garety et al., 1991). Participants are allowed to see as many beads as they want before guessing which jar these beads are being drawn from. Patients with delusions consistently exhibit a JTC bias, seeing significantly fewer beads seen before making a decision than control groups (Garety & Freeman, 1999). In the general population, McKay et al. (2006) found that delusion-proneness was associated with a JTC bias in the Beads task. The JTC bias has been attributed to a lowered decision threshold for confidence or increased noise in the decision process rather than an overweighting of the subjective costs of information-sampling (Averbeck et al., 2011; Moutoussis et al., 2011).

It has been suggested that individuals who are highly intolerant of uncertainty might also seek less information or failing to search for information strategically in their haste to reach a decision more quickly and reduce their exposure to an uncertain situation; this may be especially evident under time pressure (Janis, 1982; Keinan, 1987; Smock, 1955). Luhmann et al. (2011) found that people who are more intolerant of uncertainty (high IUS scoring individuals) tend to be more averse to delays to finding out the outcome of a gambling task, frequently preferring to select an immediate but less valuable reward instead. This aversion to uncertain outcomes was compounded by greater sensitivity to outcome valence: the tendency to choose the immediate but less valuable reward was increased in high-IUS participants after selecting but failing to win the more valuable, delayed reward. However, a recent study by Morein-Zamir et al. (2020) found no association between information-seeking on the Beads task and uncertainty intolerance or obsessive-compulsive symptoms in a general population sample, suggesting that decision context may ultimately determine whether uncertainty intolerance motivates increased or reduced information-seeking.

Depression, which is typically characterised in terms of a withdrawal from stimuli, has also been linked to reduced information-seeking. Depression is typically seen as the inverse of curiosity (Camp, 1986; Kaczmarek et al., 2014). While Rodrigue et al. (1987) found that participants who had been temporarily induced into a depressed mood through a mood induction procedure reported a significantly lower valuation of and desire for new information, this relationship has not always been found consistently in the empirical literature (cf. Camp, 1986).

Conclusions

Based on the empirical literature on this topic, disturbances in confidence and information-seeking are seen across a range of psychiatric disorders. Biases in confidence and abnormalities in confidence sensitivity are associated with symptoms of these disorders even at subclinical levels, and appear to track symptom severity and normalise after recovery in some disorders (Hoven et al., 2019). This suggests that individual differences in confidence may be of psychiatric interest as potential predictors of psychiatric onset or severity.

Abnormal information-seeking is seen in a range of psychiatric disorders. In some cases, most noticeably obsessive-compulsive disorders, maladaptive information-seeking can exacerbate and promote the maintenance of the disorder. Some abnormalities in information-seeking may be explained as maladaptive attempts to manage uncertainty, but may themselves be sequelae of more fundamental differences in negative affect, confidence, decision thresholds, and subjective valuation of information.

1.5 Summary & motivations

Having reviewed how confidence and information-seeking have been conceptualised and measured in the existing literature, and how individual differences in these measures have been observed in both the general and psychiatric populations, we now turn to exploring and extending some of the ideas we have encountered.

Firstly: while relationships between individual differences in psychiatric dimensions and individual differences in confidence and information-seeking have been observed in clinical settings and demonstrated in experimental settings, the effect sizes observed so far have been subtle. We aim to build on this work by looking simultaneously at complementary modes of metacognitive expression (i.e., confidence and information-seeking) and at confidence in different decision-making paradigms featuring different types of uncertainty (i.e., probabilistic reasoning and perceptual decision) to see if we can replicate and strengthen these previously observed relationships.

Secondly: existing research on information-seeking in cognitive psychology has generally framed it in terms of uncertainty reduction, although there are suggestions from other disciplines of psychology that individuals might manage uncertainty in a decision context using other methods, such as seeking feedback or opting out of uncertain decisions. We aim to explore these suggestions empirically by testing experimentally if and when people might engage in these alternative strategies for uncertainty management in preference to seeking further information.

Finally: confidence and information-seeking have predominantly been investigated in single-shot decision tasks, where participants only have one opportunity to request further information. Changes in confidence and information-seeking over the course of the decision have generally been studied with neurophysiological approaches, often in non-human subjects. We aim to extend this body of work by investigating the relationship between confidence and information-seeking in a sequential decision-making context, to better understand how confidence and information-seeking evolve over the decision process.

1.6 Thesis outline

This DPhil thesis explores the relationship between confidence and information-seeking behaviour and the question of how individual differences might influence this relationship. I use novel adaptations of two well-established decision-making paradigms, which each measure both information-seeking and subjective confidence during the decision process. Chapter 2 will introduce these two key tasks, the Information Sampling Task with Confidence (ISTc) and the Dot Discrimination Task with Information-Seeking (DDTi), in more detail and discuss the piloting and validation process of their task designs.

In the first line of experimental work, reported in Chapter 3, I look at whether individual differences along psychiatric dimensions of interest can reliably predict confidence and information-seeking tendencies in these tasks. Experiment 1 assesses whether confidence and information-seeking measures on the ISTc and DDTi correlate with four dimensions of psychiatric interest: self-reported trait anxiety, depressive severity, intolerance of uncertainty, and schizotypy. Experiment 2 attempts to replicate the main findings of Experiment 1 in a larger sample. I then conduct some statistical exploration of the relationships between the measured variables across the two experiments.

From there, in Chapters 4 and 5, the focus of the thesis work shifts from individual differences to more specific aspects of confidence and information-seeking that we observed in our initial experiments. The second line of experimentation, reported in Chapter 4, is an investigation into different strategies for managing uncertainty in the decision process, as well as the conditions that might push people towards favouring one strategy over another. Experiment 3 explores the plausibility of participants adopting a feedback-seeking strategy by testing whether the availability of feedback changes information-seeking frequency prior to decision. Experiment 4, on the other hand, explores the plausibility of opt-out preference and how it might be induced by different task parameters.

In the last line of experimental work, reported in Chapter 5, I look more closely into the relationship between choice and confidence in the decision process. Across instances of the ISTc in the task development pilots and Experiments 1 to 3, we had observed a boost in reported confidence at the point of decision relative to the pre-decisional information sampling phase, even when accounting for differences in evidence strength (i.e., objective probability of being correct). Chapter 5 delves into the relationship between likelihood and confidence by assessing possible mechanisms for this decision-evoked confidence boost effect through a combination of simulation and experimental work. I first present two simulations that test whether the confidence boost we observed merely reflected a selection bias in when participants were choosing to commit to a decision. Experiment 5 then explores whether the act of committing to a decision was contributing to this boost in confidence at the point of decision. Experiment 6 extends this idea and tests whether forcing a decision at a much lower level of evidence strength can still boost confidence at the point of decision and have lasting effects on confidence and information-seeking beyond the initial decision.

Chapter 6 concludes by summarising the work done in the previous chapters and relates our current findings to the broader psychological discourse on metacognitive phenotyping, psychopathology, and individuals' needs for control and agency. We also make suggestions for how these ideas might be developed in future work.

2
DEVELOPING NEW TASKS FOR MEASURING CONFIDENCE AND INFORMATION-SEEKING

2.1 Overview

The previous chapter reviewed the existing literature on confidence and information-seeking, and the association between these measures and dimensions of psychiatric interest. The majority of studies we reviewed used decision-making tasks that required participants to make choices between alternatives under conditions of uncertainty. Many of these studies (e.g., Rouault et al., 2018; Desender et al., 2018) use perceptual judgment tasks that allow experimenters to exert precise control over the evidence presented to the participants prior to decision and thus to manipulate perceived task difficulty and induced level of uncertainty. These tasks generally focus on sensory uncertainty, or uncertainty about the stimulus identity caused by noise in the perceptual process, and are extremely useful both for probing fundamental limitations and properties of human and animal cognition and for characterising individual and group differences (Yu, 2015).

However, in many everyday decisions, the uncertainty in decision-making is not sensory in nature. For instance, when choosing whether to pack an umbrella on a dry but cloudy morning, the uncertainty is not so much about the accuracy of your perception but more about the expected probabilities and costs associated with bringing an umbrella versus not bringing an umbrella. In other words, the uncertainty is about the consequences of the chosen option rather than about stimulus identity (Yu, 2015). This outcome uncertainty is often assessed experimentally using reinforcement learning tasks where participants have to learn the consequences of different alternatives through trial-and-error (cf. Browning et al., 2015) or risky decision-making tasks where the outcome probabilities are known but not certain (cf. Zhang et al., 2015). In the psychiatric literature, probabilistic reasoning tasks in which participants are allowed to engage freely in sequential information sampling before making a decision with outcome uncertainty (e.g., the Beads Task; Garety & Freeman, 1999) have been used to probe individual differences in dimensions of psychiatric interest, such as uncertainty intolerance and reflection impulsivity, and psychiatric disorders, like schizophrenia, OCD, and addiction (Clark et al., 2006; Garety & Freeman, 1999; Hauser, Moutoussis, et al., 2017).

Thus, during the methodology design process of the thesis, we set about looking for and refining existing task paradigms with two goals in mind. The first goal was to approach our central question about confidence, information-seeking, and the role of individual differences from multiple perspectives, by using tasks that involved complementary decision styles and focused on different types of uncertainty. We decided to pair a one-shot perceptual judgment task that manipulated sensory uncertainty (the Dot Discrimination Task, cf. Boldt & Yeung, 2015; Rouault et al., 2018) with a slower-paced, more strategic probabilistic reasoning task featuring outcome uncertainty and sequential information-seeking (the Information Sampling Task, cf. Clark et al., 2006; Hauser, Moutoussis, et al., 2017). The second goal was to modify these tasks so that each would have measures of both confidence and information-seeking. Given the intra- and inter-personal variability in how people use continuous scales to report their subjective confidence, we did not want to rely exclusively on subjective confidence reports to gain insight into metacognition. Instead, we wanted to use tasks that included measures of both confidence and information-seeking, ideally at various points throughout the decision process. By probing metacognitive monitoring and control processes together through these measures, we could see whether any convergent insights could be reached about how metacognition operates and evolves during the decision process. Each of the task paradigms we chose already had a measure of one (decision confidence in the Dot Discrimination Task and information-seeking in the Information Sampling Task), but neither had measures of both. We modified them to meet our needs, incorporating an additional information-seeking stage prior to the final decision in the Dot Discrimination Task and adding confidence ratings to capture participants' confidence at different stages up to and including the final decision to the Information Sampling Task.

In this chapter, we will provide the methodological background for the experimental work in the following chapters. Specifically, we will introduce the tasks that will be used in subsequent chapters, describe the task development process for these modified task paradigms, outline the key results of piloting work done during that development process, and discuss the basic task properties for the modified tasks.

2.2 The DDTi

2.2.1 Task design

The Dot Discrimination Task with Information-Seeking (DDTi) is a perceptual judgment task. It tests how people make decisions about basic stimulus features (e.g., identity, colour, orientation, direction of motion) under sensory uncertainty, where perceived task difficulty can be experimentally manipulated by varying the ratio of signal to noise in the stimulus presented. The DDTi was adapted from the standard dot discrimination task used in studies such as Boldt and Yeung (2015) and Rouault et al. (2018).

Trial structure

Like in the standard paradigm, participants in the DDTi were asked to judge which of two boxes contained the greater number of dots and to report their confidence in each judgement on a rating scale. On each trial, participants were presented with a fixation cross for 1000 ms before being shown two black boxes with randomly positioned white dots. One of these boxes (position counterbalanced) always contained 200 dots, while the other contained 200 + d dots, where $d \in [0, 100]$. After 300 ms, the dots disappeared; the black boxes remained on screen and a prompt and confidence rating scale appeared. Participants were asked which of the two boxes (left or right) contained more dots and indicated their combined choice and confidence using the visual analogue scale presented, where the left-hand end of the scale indicated certainty in the left-hand box containing more dots ("definitely LEFT") and the righthand end of the scale indicated certainty in the righthand box having more dots ("definitely RIGHT") [Figure 2.1].

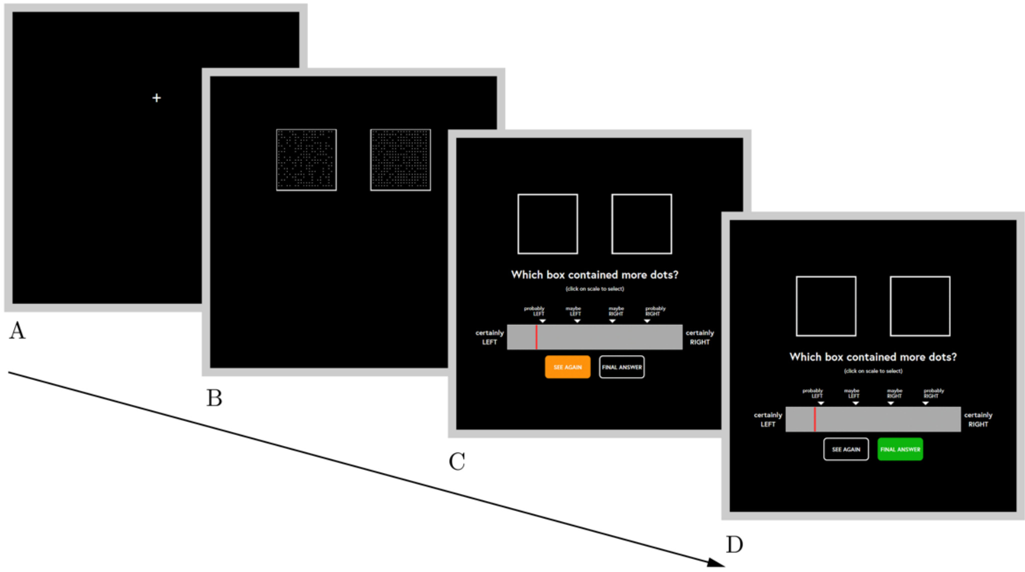


Figure 2.1. Structure of a trial on the DDTi. (A) Fixation cross presented at the start of the trial (1000 ms). (B) Stimuli presented (300 ms), with one box containing 200 dots and the other, 201 to 300 dots. (C) Participants asked to indicate which box (left or right) contained more dots as well as their confidence in that judgment on the visual analogue scale. They had two options: to submit this as a "Final Answer" or to "See Again" an easier version of the dot boxes before submitting a final answer. (D) After seeing an easier version of the stimuli (i.e., increased dot difference between boxes), participants were asked to make another confidence judgment and submit this as their "Final Answer".

Participant responses in this first-stage decision were used to calibrate a staircase procedure, which adapted the task difficulty to participant performance so as to maintain a roughly constant level of task difficulty and performance within and across participants in this first-stage decision. The staircase procedure we implemented followed the log-adaptive 2-down 1-up procedure described by Rouault et al. (2018). Step size varied on a log scale, such that steps (i.e., changes in the absolute number of dots) were larger at higher values of *d*. The step size was calculated in log-space with a starting point of 4.6 (+100 dots), changing by

± 0.4 for the first 5 trials, ± 0.2 for the next 5 trials, and ± 0.1 for the rest of the task, and used equal step sizes for steps up and down.

In the standard paradigm, participants would proceed onto the next trial after making this decision about dot count. In the DDTi, however, there was a secondary information-seeking stage following this initial (first-stage) decision, where participants could choose to see additional information ("See Again" option) before the final decision or to directly register their first-stage decision as their final decision ("Final Answer" option) [Figure 2.1]. There was no explicit cost for choosing the "See Again" option. If participants selected the "Final Answer" option after their first-stage decision, they would move on directly to the next trial after a brief blank screen (2000 ms or the amount of time they had taken on the first-stage decision on that trial, whichever was smaller). On the other hand, if participants selected the "See Again" option, a fixation cross was again presented for 1000 ms before an easier version of the same dot stimuli configuration was shown. In this easier version, the correct box remained the same, but d was increased to make it easier to discriminate. In DDTi Pilot 1, d was increased by a fixed number (10 dots), and in DDTi Pilot 2 and subsequent experiments in the thesis, d was increased by the number of dots equal to a step size of $+0.5$ in log-space. After seeing this easier version of the dot stimuli, participants would be asked to provide a final (second-stage) combined choice and confidence response. Unlike the first-stage decision, the final decision was not used to calibrate the adaptive staircase, but was instead used to determine the performance bonus reward received by the participant at the end of the experiment.

Participants were incentivised to be as accurate as they could be in the final decision. They received a flat amount for participation but were then awarded an additional fixed-amount bonus if they achieved an overall accuracy (proportion correct) of 60% or more across final decisions.

Block structure

The overall structure of the DDTi could be divided into two phases. The first was a calibration phase, where participants completed a number of practice trials: 25 with no "See Again" option, to familiarise them with the basic structure of the dot count task and confidence slider, and 25 (5 in the DDTi Pilot 1 only) with the "See Again" option enabled. During the calibration phase, trial-by-trial accuracy feedback was given to participants. The staircase procedure was initiated during the calibration phase to minimise burn-in period. Following the calibration phase, participants proceeded to the test phase, where they completed 100 trials in 4 blocks. No further trial-by-trial feedback was given during the test phase, although participants were informed of their block accuracy after each block and their overall accuracy at the end of all four blocks.

Key measures

The key dependent variables in the DDTi were participants' subjectively rated confidence and their information-seeking behaviour. We used each participant's mean confidence in the first-stage decision as their primary measure of confidence, and the proportion of trials in which they chose the "See Again" option as their primary measure of information-seeking.

Choice and confidence were registered simultaneously using a visual analogue scale ranging from "certainly LEFT" (maximum confidence that the left-hand box contained more dots) to "certainly RIGHT" (maximum confidence that the righthand box contained more dots). We converted the participant's position selection on the continuous scale into a numerical confidence value, coded in relation to the correct response on that trial. These converted confidence values ranged from 0 to 100 (excluding 50, to make the task two-alternative forced choice): 0 represented being subjectively certain but objectively incorrect (e.g., responding that the box on the left definitely contained more dots when in reality it was the righthand box), while 100 represented being subjectively certain and objectively correct (e.g., responding correctly that the righthand box contained more dots).

2.2.2 Task development

Two pilot studies were run during the development of the DDTi. Our aim in designing the DDTi was to add an option for participants to seek additional evidence prior to final decision without affecting critical aspects of the original dot discrimination task (i.e., psychophysical staircase stability) and key measures of interest (i.e., overall accuracy and confidence on the first-stage decision).

DDTi Pilot 1

In DDTi Pilot 1, after registering the first-stage decision, participants were asked to indicate whether this was their final answer ("Final Answer") or whether they wanted to see more evidence before giving a final answer ("See Again"). If they selected the "See Again" option, they were shown a version of the dot stimuli that increased the dot difference by 10 dots.

A total of 20 participants (9 female / 11 male), ranging in age from 18 to 41 years ($M = 27$, $SD = 7$) were recruited using the online recruitment platform Prolific. Participants provided informed consent in accordance with procedures approved by the University of Oxford Central University Research Ethics Committee (R56625/RE003). All participants were paid a base amount of around £5/hour plus a bonus of £1 for final-decision accuracy at or exceeding 60% correct.

The key measures from Pilot 1 are summarised in Table 2.1. Participants appeared to use the information-seeking option when they were more uncertain.

Confidence was lower for first-stage decisions where participants subsequently chose the "See Again" option (M = 63.95, SD = 12.00) than where participants subsequently registered it as the final decision (M = 73.03, SD = 15.01). Accuracy was also lower on first-stage decisions that were followed by information-seeking (M = .76, SD = .21) than on those that were not (M = .85, SD = .13). However, there were substantial individual differences in information-seeking behaviour and confidence. Although the staircase procedure was intended to ensure that first-stage accuracy was reasonably well matched across participants, average confidence for the first-stage decision still varied substantially between participants (M = 69.90, SD = 12.13). We also saw a wide range of information-seeking preferences, with some participants choosing the "See Again" option on only 3% of trials and others choosing that option on 100% of trials [Table 2.1, Figure 2.2].

In one of our manipulation checks, we found that overall accuracy did not seem to be improving between first-stage (M = .83, SD = .12) and second-stage decisions (M = .85, SD = .13), $t < 1$. After re-examining the code, we realised that an experimenter error had caused the staircasing procedure to include second-stage decisions as inputs. This led to excessively high first-stage accuracy and the staircase failing to stabilise for some participants [Appendix S2.1]. Because this pilot had found promising indications of individual differences in confidence and information-seeking between participants, we decided to fix the staircasing issue and run a second pilot study to confirm whether these individual differences would remain when the staircase was working appropriately.

Table 2.1
Behavioural performance on DDTi pilot versions

		First-stage decision			Second-stage decision		
	Range of mean "See Again" proportions	Accuracy (proportion correct)	Confidence magnitude (correct trials)	Confidence magnitude (incorrect trials)	Accuracy (proportion correct)	Confidence magnitude (correct trials)	Confidence magnitude (incorrect trials)
DDTi Pilot 1 (*N*=20)	.03 – 1.00 (*M* = .36, *SD* = .26)	.83 (.12)	76.27 (10.71)	66.89 (8.37)	.85 (.13)	77.28 (11.65)	66.25 (7.45)
DDTi Pilot 2 (*N*=20)	.00 – 1.00 (*M* = .36, *SD* = .37)	.73 (.03)	73.46 (10.81)	70.34 (11.05)	.82 (.13)	73.06 (10.34)	67.24 (9.36)

Confidence magnitude = 50 + | reported confidence – 50 |
Data presented as mean (SD) unless otherwise stipulated.

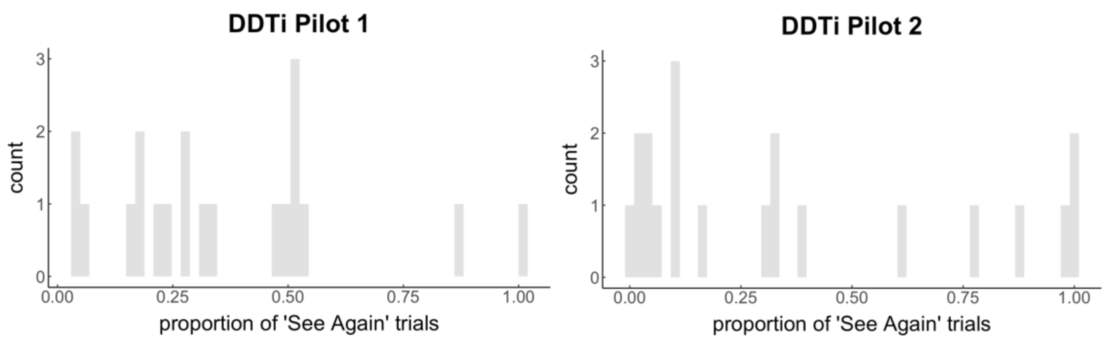


Figure 2.2. Histogram of information-seeking frequency in DDTi pilots.

DDTi Pilot 2

In DDTi Pilot 2, we restricted the inputs to the staircasing procedure to just first-stage decisions. We also increased the number of practice trials from 30 to 50 and added trial-by-trial feedback for these trials to improve stabilisation of the psychophysical staircase. To make the "See Again" option consistent with the staircase (which was calculated in log space), the dot difference boost from the "See Again" option was changed from a constant 10 dots to +0.5 in log space (i.e., so that there were larger changes in numbers of dots presented for higher initial values of *d*). This resulted in much more stable staircasing in line with the standard dot discrimination task and other existing perceptual decision-making paradigms [Appendix S2.1].

A total of 20 participants (5 female / 15 male), ranging in age from 18 to 54 years ($M = 30$, $SD = 11$) were recruited using Prolific. Participants provided informed consent in accordance with procedures approved by the University of Oxford Central University Research Ethics Committee (R56625/RE003). All participants were paid a base amount of around £5/hour plus a bonus of £1 for final-decision accuracy at or exceeding 60% correct.

The key measures from Pilot 2 are also summarised in Table 2.1. Again, confidence [$M_{seek} = 54.34$ vs. $M_{skip} = 66.52$, $p < .001$] and accuracy [$M_{seek} = .65$ vs. $M_{skip} = .80$, $p = .002$] were lower for first-stage decisions where participants chose to seek more information than for first-stage decisions where participants were satisfied with their choice. While overall accuracy for first-stage decisions was stable across participants ($M = .73$, $SD = .03$) due to the staircasing procedure, first-stage decision confidence was more variable ($M = 61.52$, $SD = 4.83$). We also continued to observe a wide range of individual preferences for information-seeking ($M = .36$, $SD = .37$), with some participants never choosing to see more information before registering a final choice and others choosing to see additional information on all trials [Table 2.1, Figure 2.2].

Looking at trials where participants chose to seek more information after the first-stage decision, we found that accuracy did significantly improve from first-stage (M = .65, SD = .18) to final decisions (M = .82, SD = .13), $t(18)$ = 4.80, $p < .001$. This suggests that the information-seeking option was being used strategically to help participants improve their performance on trials. However, stronger individual preferences for information-seeking, i.e., higher overall frequencies of choosing the "See Again" option, did not appear to be associated with lower initial accuracy or confidence. The correlation between information-seeking frequency and mean confidence on the first-stage decision was negative but non-significant in this sample, $r(18)$ = -.40, p = .077 [Figure 2.3]. The correlation between information-seeking frequency and mean accuracy on the first-stage decision, on the other hand, was positive and significant [$r(18)$ = .60, p = .005], indicating that participants who sought more information already tended to be more accurate even before seeing extra information [Figure 2.3]. This suggests that people who tended to seek more information were not less accurate or confident to begin with.

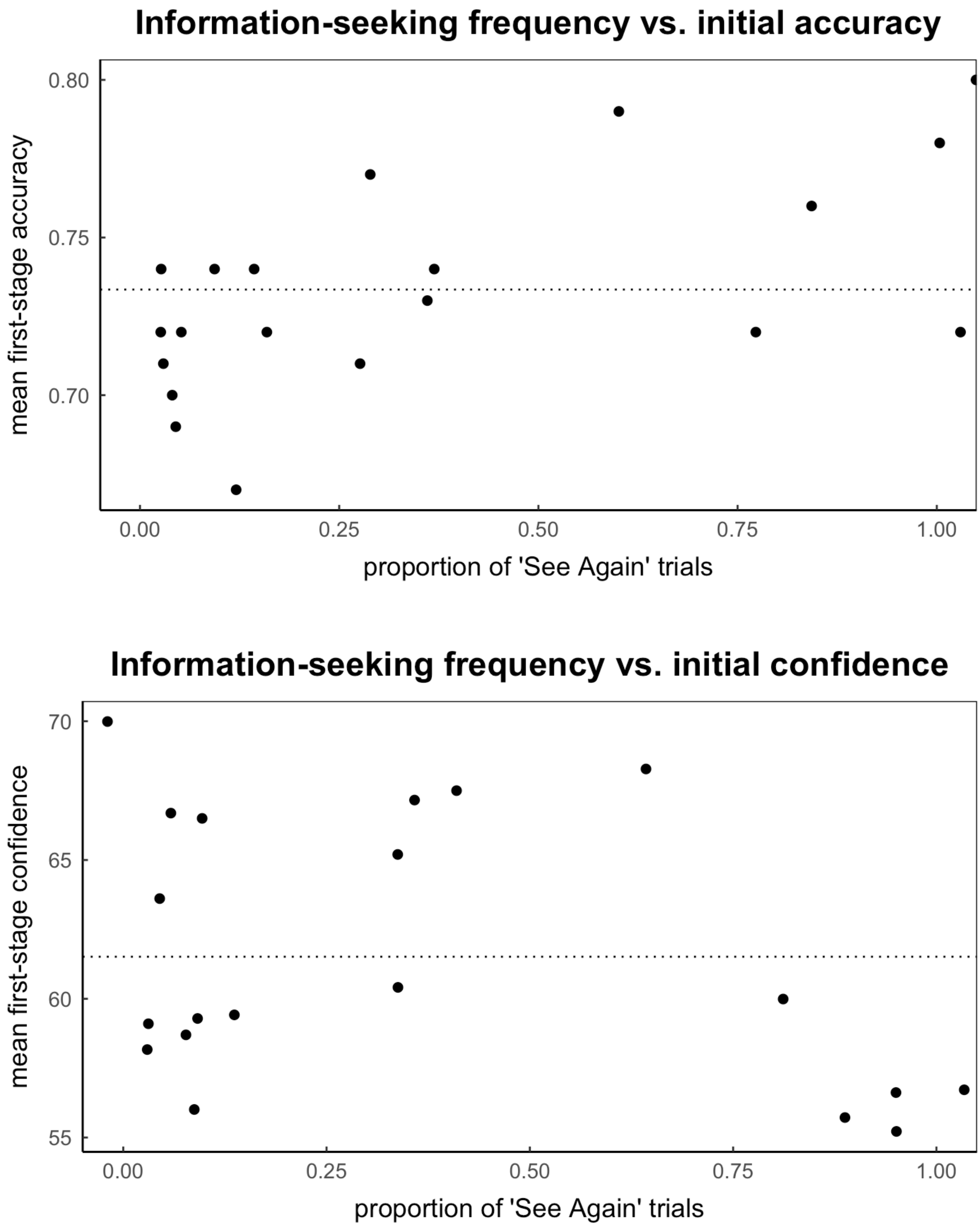


Figure 2.3. Participants' information-seeking frequency versus their mean accuracy and confidence on first-stage decisions in DDTi Pilot 2.

Discussion

The findings from the two pilot studies suggest that the DDTi shares basic task similarities with its precursor, the dot discrimination task, in the first stage of the task. However, by adding an information-seeking option before participants are expected to register their final, second-stage decision, we were able to extend the existing dot discrimination paradigm to produce a useful measure of information-seeking.

By assigning no explicit costs to the information-seeking option and rewarding participants based on the second-stage decision accuracy, the DDTi was designed to implicitly encourage participants to seek more information before committing to a final decision. We observed an interesting dichotomy in the relationship between information-seeking and accuracy and confidence measures at the trial versus participant level. At the trial level, participants generally chose to seek more information when they were less confident and less accurate in the first-stage decision. Participant performance improved after seeing the additional information, indicating that their information-seeking behaviour was generally strategic and adaptive.

However, at the participant level, we found great individual variability in both our pilot studies in how often people used the “See Again” option: 25% of participants in DDTi Pilot 2 (15% in DDTi Pilot 1) chose the “See Again” option on 5% or less of all trials, and 15% of participants in DDTi Pilot 2 (5% in DDTi

Pilot 1) chose the "See Again" option on 95% or more of all trials. Although the sample size in this pilot was not large enough to establish strong claims about relationships between these individual differences, we found that participants who tended to seek more information before reaching a final decision did not appear to be less accurate or confident on first-stage decisions than those who tended to skip seeing additional information – in fact, people who sought more information tended also to be ones that were generally more accurate to start. This suggests that individual differences in information-seeking frequency might be more influenced by factors beyond the scope of the immediate decision. The relationship between information-seeking frequency on the DDTi and individual differences along more stable trait dimensions is explored further in Chapter 3.

The changes implemented in DDTi Pilot 2 stabilised the staircase across participants relative to Pilot 1. DDTi Pilot 2 also showed comparable overall accuracy to previous findings from dot discrimination tasks (cf. Rouault et al., 2018), while retaining the substantial individual differences in "See Again" preference that we had seen in DDTi Pilot 1 [Table 2.1]. The DDTi Pilot 2 version thus appeared to be a suitable paradigm for investigating the relationship between confidence and information-seeking, and was the version we retained as the final design of the DDTi.

2.3 The ISTc

2.3.1 Task design

The Information Sampling Task with Confidence (ISTc) is a probabilistic reasoning task. It tests how people make binary decisions under outcome uncertainty, namely, where the outcomes associated with each choice option are unknown or uncertain at the start of the task. Participants can strategically reduce this outcome uncertainty by engaging in sequential information sampling before committing to a choice option. The degree of outcome uncertainty and the optimal strategy for approaching the task can also be experimentally manipulated by varying the rewards and penalties associated with correct versus incorrect decisions and the cost of sampling additional information.

Trial structure

The ISTc was adapted from the information sampling task (IST) introduced by Clark et al. (2006). Clark et al. (2006) originally developed the IST to assess reflection impulsivity, an individual's tendency to gather and evaluate information before making a decision (Kagan, 1966). The IST assesses the amount of information an individual seeks before committing to a decision. Like the Beads Task (Huq et al., 1988; Garety et al., 1991), it features free-choice sequential information sampling prior to decision, but unlike the Beads Task, the IST features a hard limit to the extent of information-seeking possible.

In the original IST, participants were shown a covered 5×5 grid of black tiles and asked to decide which of two colours was the underlying colour of the majority of the 25 tiles [Figure 2.4]. They could sequentially uncover a minimum of 1 tile and a maximum of 25 tiles without time restriction before making a final decision. After freely sampling as much information as they liked (up to the maximum), participants indicated their final decision about the majority colour by selecting the correspondingly coloured circular button below the grid.

Two score rule conditions were implemented: a 'Fixed Win' (FW) condition where participants won 100 points by declaring the correct majority colour, irrespective of the number of tiles they uncovered, and a 'Decreasing Win' (DW) condition where participants, upon declaring a correct answer, were awarded 10 points per remaining unseen (covered) tile. In both FW and DW conditions, an incorrect decision cost them 100 points. The relevant score rule for the trial was shown below the tiles throughout the trial: either "Fixed" for the FW condition or "Variable" for the DW condition. Feedback about how many points were won or lost in that trial was given for 1000 ms after each final decision, after which the next trial started immediately.

Because a central focus of the thesis was to better understand the relationship between confidence and information-seeking, we were interested in seeing how people's confidence evolved over the course of a decision in response to the build-up of evidence during this sequential information sampling phase. To experimentally assess this in the ISTc, we added measures of subjective confidence to the existing paradigm. Both during the sampling phase and at the point of decision (choice commitment), participants were asked to simultaneously report their best guess as to the majority colour based on the current evidence, as well as their confidence in this estimate ("How sure are you of the majority colour"), using the visual analogue scale presented. Each half of the scale (the sliding bar and the labels) corresponded to one of the two colours: each end of the scale indicated certainty that the majority colour was the colour associated with that side of the scale ("CERTAIN it's this colour") [Figure 2.4].

Block structure

Like the IST, the ISTc consisted of two blocks with 10 trials each: one block of Fixed Win (FW) trials and another of Decreasing Win (DW) trials. The order of these two blocks would typically be counterbalanced between participants. Prior to starting the ISTc, participants would first be walked through the task in a tutorial that would inform them of the elements of the task, how their final decision and confidence estimates could be registered, and the rewards and penalties associated with correct versus incorrect choices in the selected score rule.

Participants would then complete a number (1-3, depending on the version) of practice trials before they would be allowed to continue to the experimental blocks.

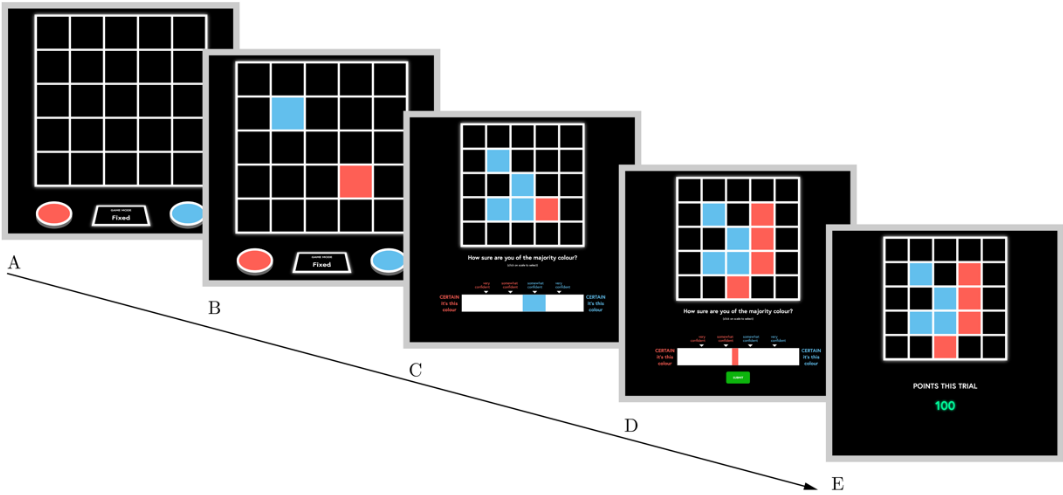


Figure 2.4. Structure of a trial on the ISTc. (A) Subjects were asked which of the two colours shown below the tiles was the underlying majority colour of the 25 covered tiles. The score rule for that trial ("Fixed" or "Variable") is shown in the box below the tiles. (B) Subjects were able to click on tiles to uncover them one at a time. (C) Pseudo-randomly during the sampling phase, participants were asked to register their confidence ("How sure are you of the majority colour?", ranging from "certain it's this colour [red]" to "certain it's this colour [blue]") by clicking on a position on the visual analogue scale. (D) Once they are ready, participants can indicate their decision by selecting the respective colour, upon which they will be asked to give another confidence rating. (E) Participants are shown feedback in the form of the number of points earned from the trial, which is coloured green for correct responses and red for incorrect responses.

Key measures

The key measured variables in the ISTc were confidence and information-seeking. We used each participant's mean confidence bias (difference between rated confidence and the objective probability of being correct given the evidence) as their primary measure of confidence, and the number of tiles they saw prior to making the final decision as the primary measure of information-seeking on the ISTc.

Confidence was registered on a two-coloured visual analogue scale ranging from "CERTAIN it's [colour A]" to "CERTAIN it's [colour B]", with each end representing maximum confidence that the colour indicated was the majority colour. We converted the participant's position selection on the visual analogue scale into a numerical confidence value, coded in relation to the correct response on that trial. These converted confidence values ranged from 0 to 100 (excluding 50, to make the task 2-alternative forced choice), with 0 representing subjectively certain but objectively incorrect and 100 representing subjectively certain and objectively correct.

The objective probability of being correct given the tiles that have been uncovered, P(correct), was calculated as per the formula in Clark et al. (2006):

[Eq. 2.1]

$$P(correct) = \frac{\sum_{z}^{k=A} \binom{z}{k}}{2^z}$$

where $z = 25$ – (number of tiles uncovered) and $A = 13$ – (number of tiles of the chosen colour). Thus, confidence bias on each trial (both during the information sampling phase as well as at the point of decision) could be calculated as:

[Eq. 2.2]

$$\text{confidence bias} = \text{reported confidence - P(correct)} \times 100$$

Confidence bias values ranged from 0 (perfectly calibrated) to ±100 (maximally uncalibrated). Positive values of confidence bias indicate overconfidence (confidence estimate stronger than objective likelihood) and negative values indicate underconfidence (confidence estimate weaker than objective likelihood).

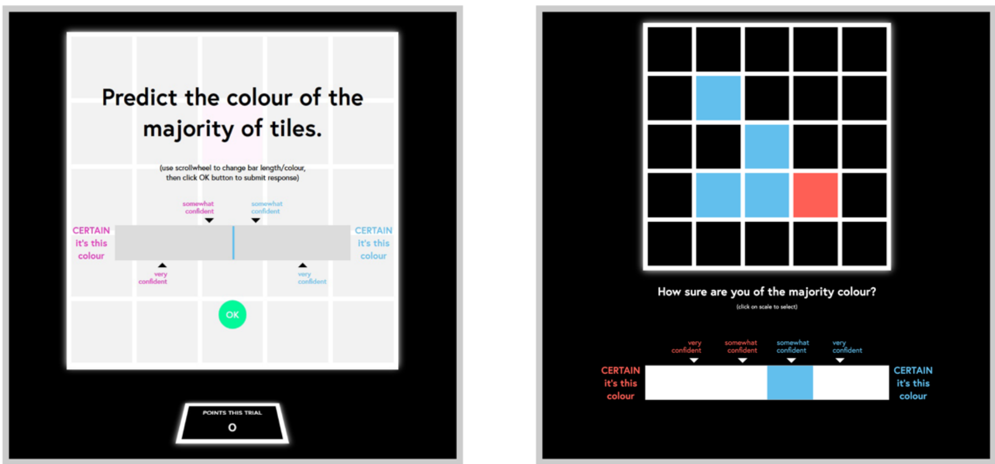
2.3.2 Task development

Three pilot studies were run during the development of the ISTc. Different variations of the confidence-rating mechanism were trialled during development of the ISTc and assessed on the user-friendliness of the interface design, the disruptiveness of the confidence rating requests, and the useability of the resulting confidence data.

ISTc Pilot 1

In ISTc Pilot 1, the confidence rating appeared as a white pop-up over the grid. The pop-up was made translucent to help remind participants of the current evidence (state of the grid). Participants used their mouse scrollwheel to change the length and colour of the confidence bar, and then pressed the OK button to continue to information sampling or final decision, if they had not made that final decision yet, or to continue to the next trial if they had indicated their final decision [Figure 2.5].

To assess the comparability of the ISTc to the IST in terms of task validity and behavioural outcomes, participants in Pilot 1 completed both the IST and the ISTc in the same session. This enabled us to benchmark key behavioural measures on the new ISTc against the IST. Participants completed 20 trials of the ISTc (10 in the Fixed Win condition and 10 in Decreasing Win condition) and 20 trials of the IST (10 in FW and 10 in DW, same order as in the ISTc). Participants always completed the IST before completing the ISTc, but the order of the scoring rule (FW then DW vs. DW then FW) was counterbalanced between participants.



A B

Figure 2.5. Confidence rating pages in the different ISTc pilot versions. (A) Confidence rating mechanism in ISTc Pilot 1. (B) Confidence rating mechanism in ISTc Pilot 2, ISTc Pilot 3, and the main experiment.

A total of 31 participants (11 female / 20 male), ranging in age from 18 to 52 years (M = 30, SD = 9) were recruited online using Prolific. Participants provided informed consent in accordance with procedures approved by the University of Oxford Central University Research Ethics Committee (R56625/RE003). All participants were paid a base amount of around £5/hour plus a bonus of up to £2.

The key measures from Pilot 1 are summarised in Table 2.2. Comparing the IST and the ISTc, we see that participants were generally more accurate [M_{IST} = .95 vs. M_{ISTc} = .87, p = .003] on the IST than the ISTc: this was true for both the FW [M_{IST} = .96 vs. M_{ISTc} = .88, p = .015] and DW [M_{IST} = .93 vs. M_{ISTc} = .86, p = .006] score conditions. This may be because participants saw more tiles before decision on the IST in both score conditions, with the difference being more pronounced for the FW condition [M_{IST} = 17.63 vs. M_{ISTc} = 11.46, $p < .001$] than the DW condition [M_{IST} = 11.85 vs. M_{ISTc} = 9.32, p = .011]. Correspondingly, participants had a higher objective probability of being correct (P(correct)) at the point of decision in the IST than the ISTc in both FW [M_{IST} = 91.44% vs. M_{ISTc} = 79.89%, $p < .001$] and DW [M_{IST} = 81.14% vs. M_{ISTc} = 74.07%, p = .002] conditions. While on FW trials, higher accuracy translated into significantly higher average score per trial on the IST [M_{IST} = 92.90 vs. M_{ISTc} = 76.13, p = .015], there was no significant difference in average score per trial between the IST and the ISTc on DW trials [M_{IST} = 112.68 vs. M_{ISTc} = 116.39, p = .633]. In fact, the average score per trial in DW on the ISTc appears to be slightly higher than on the IST, most likely because of the explicit cost of information sampling in the DW condition. Because participants generally saw fewer tiles on the ISTc than on the IST, they would also generally lose fewer points from information sampling.

We also found that participants were generally underconfident at the point of decision on the ISTc in both FW (M = -17.53, SD = 16.00) and DW (M = -13.70, SD = 14.86) conditions. While reported confidence across the two conditions correlated positively with P(correct) [$r(29) = .44$, $p = .013$] (see Appendix S2.2), mean confidence bias (i.e., the degree of underconfidence) also increased with P(correct) [$r(29) = -.57$, $p = .001$] [Appendix S2.3].

The high degree of underconfidence, combined with the significantly lower information-seeking on the ISTc compared to the IST, led us to believe that participants might be finding the confidence reporting mechanism difficult to use. The average trial response time in the IST (M = 13.3 s, SD = 7.2 s) nearly tripled in the ISTc (M = 32.3 s, SD = 19.8 s), most likely due to the addition of the confidence report after each tile selection. In practice, the ISTc has results in a noticeably slower pace and more considered feel during the information sampling phase than the IST. We suspected that participants' impatience while completing the confidence reports on the ISTc might be affecting reported confidence, as the scrollwheel mechanism meant that it would take participants longer to register more extreme confidence ratings. This might bias impatient participants towards an early termination of scrolling and a more limited range of reported confidence.

In the subsequent pilot, we improved the design of the confidence reporting mechanism and changed the frequency of confidence reporting in the ISTc to make it more comparable in pace and feel to the IST, as well as to elicit more similar levels of performance and information-seeking.

Table 2.2
Behavioural performance on ISTc pilot versions

	Accuracy (proportion correct)	Trial RT (s)	Total tiles seen	Decision-point P(correct) (%)	Decision-point reported confidence
ISTc Pilot 1 (*N* = 31)					
Fixed Win (FW)					
IST	.96 (.06)	12.4 (6.0)	17.63 (6.15)	91.44 (11.32)	n/a
ISTc	.88 (.18)	34.9 (22.9)	11.46 (6.60)	79.89 (15.82)	62.35 (15.13)
Decreasing Win (DW)					
IST	.93 (.09)	14.2 (10.0)	11.85 (5.68)	81.14 (12.73)	n/a
ISTc	.86 (.11)	29.7 (19.1)	9.32 (5.61)	74.07 (14.11)	60.38 (12.45)
ISTc Pilot 2 (*N* = 27)					
Fixed Win (FW)					
IST	.91 (.12)	14.5 (7.0)	19.36 (6.93)	88.87 (15.08)	n/a
ISTc	.87 (.15)	27.3 (22.8)	13.70 (7.86)	81.04 (17.80)	77.67 (17.03)
Decreasing Win (DW)					
IST	.84 (.13)	16.7 (17.2)	12.65 (6.94)	75.87 (16.43)	n/a
ISTc	.82 (.18)	21.0 (15.0)	10.37 (6.77)	70.69 (16.97)	69.85 (16.45)
ISTc Pilot 3 (*N* = 20)					
Fixed Win (FW)					
IST	.89 (.12)	14.5 (8.4)	17.34 (5.87)	86.31 (14.12)	n/a
ISTc	.90 (.10)	26.2 (11.5)	16.11 (5.91)	85.37 (14.75)	82.48 (14.35)
Decreasing Win (DW)					
IST	.85 (.14)	13.2 (8.5)	12.39 (4.74)	74.69 (14.01)	n/a
ISTc	.82 (.13)	24.3 (12.8)	12.31 (4.58)	74.04 (12.68)	70.59 (13.36)

Data presented as mean (SD).

ISTc Pilot 2

In ISTc Pilot 2, we made two changes to try to make the confidence mechanism more user-friendly and less disruptive to task pacing. Firstly, to make confidence reporting more efficient, we made it possible for participants to change the length and colour of the confidence bar by directly clicking on the visual analogue scale with their mouse instead of using the mouse scrollwheel. We also removed the OK button confirmation after selection to make confidence reporting a faster and more seamless process [Figure 2.5]. Secondly, to speed up the pace of the ISTc further, we reduced the frequency of confidence reporting: instead of reporting after every tile selection, we prompted participants to report confidence at the point of decision, as well as at pseudo-random intervals (approximately once every 5 trials) during the information sampling phase. This meant that participants would be asked to make a minimum of 1 (only at the point of decision) and a maximum of 6 (5 during sampling and 1 at point of decision) confidence ratings on any one trial of the ISTc.

A total of 27 participants (12 female / 15 male), ranging in age from 18 to 41 years (M = 27, SD = 6) were recruited online using Prolific. Participants provided informed consent in accordance with procedures approved by the University of Oxford Central University Research Ethics Committee (R56625/RE003). All participants were paid a base amount of around £5/hour plus a bonus of up to £2.

The key measures from Pilot 2 are summarised in Table 2.2. Comparing this version of the ISTc to the previous version, we see that the changes to the confidence rating mechanism and frequency greatly reduced average trial response times [$RT_{ver.2} = 24.1$ s vs. $RT_{ver.1} = 32.3$ s]. While the ISTc still took longer than the IST in Pilot 2, the difference in average response times between the two had narrowed [$M_{IST} = 15.6$ s vs. $M_{ISTc} = 24.1$ s, $p = .003$]. Continuing with the IST to ISTc comparison, we found that the accuracy gap between the two narrowed to the point of non-significance for both FW [$M_{IST} = .91$ vs. $M_{ISTc} = .87$, p = .253] and DW [$M_{IST} = .84$ vs. $M_{ISTc} = .82$, $p = .513$] conditions. There were also no significant differences in average score per trial in both FW [$M_{IST} = 81.48$ vs. $M_{ISTc} = 74.07$, $p = .253$] and DW [$M_{IST} = 82.52$ vs. $M_{ISTc} = 94.15$, $p = .259$] conditions, although the score on the ISTc in the DW condition trended higher, similar to what we saw in Pilot 1. Overall, this indicates that our changes to the confidence rating mechanism and frequency may have been successful in making the ISTc feel more similar to the IST.

Confidence calibration at the point of decision also improved in Pilot 2. While participants were still slightly underconfident in the FW (M = -3.37, SD = 9.52) and DW conditions (M = -.84, SD = 11.23), mean confidence bias (i.e., the degree of underconfidence) no longer significantly increased with P(correct) in the two conditions [FW: $r(25) = -.35$, $p = .077$; DW: $r(25) = -.38$, $p = .053$] [Appendix S2.3]. However, participants were still seeing fewer tiles on the ISTc than the IST: this gap was statistically significant in FW [M_{IST} = 19.36 vs. M_{ISTc} = 13.70, p < .001] and neared significance in DW [M_{IST} = 12.65 vs. M_{ISTc} = 10.37, p = .066]. Because the ISTc always followed the IST in both Pilots 1 and 2, we speculated whether the fixed task order might have biased impatient participants into rushing through the ISTc and seeing fewer tiles before decision. We thus decided to run another pilot with counterbalancing for task order to see if we could make information-seeking behaviour on the ISTc more comparable to that on the IST.

ISTc Pilot 3

In ISTc Pilot 3, to control for task order effects on information-seeking, we counterbalanced the order in which participants completed the IST and the ISTc: half the participants in the sample completed the ISTc first, while the other half completed the IST first.

A total of 20 participants (6 female / 14 male), ranging in age from 18 to 45 years (M = 27, SD = 7) were recruited online using Prolific. Participants provided informed consent in accordance with procedures approved by the University of Oxford Central University Research Ethics Committee (R56625/RE003). All participants were paid a base amount of around £5/hour plus a bonus of up to £2.

The key measures from Pilot 3 are summarised in Table 2.2. Like in Pilot 2, we continue to see non-significant differences in accuracy [M_{IST} = .87 vs. M_{ISTc} = .86, p = .470] and average score per trial [M_{IST} = 83.08 vs. M_{ISTc} = 80.25, p = .480] between the IST and ISTc. Confidence bias at the point of decision in Pilot 3 was also similar to what we observed in Pilot 2. Across both conditions, participants were still slightly underconfident (M = -3.17, SD = 6.19), but mean confidence bias did not significantly increase with P(correct) [r(18) = -.28, p = .235] [Appendix S2.3]. Most importantly, there were no significant differences between the IST and this version of the ISTc in our primary measure of information-seeking, the total number of tiles seen before decision, in both FW [M_{IST} = 17.33 vs. M_{ISTc} = 16.11, p = .261] and DW [M_{IST} = 12.39 vs. M_{ISTc} = 12.31, p = .909] conditions. This indicates that our task order counterbalancing successfully closed the gap between the two tasks on this last key measure, and suggests that the ISTc is sufficiently comparable to the IST for our purposes while providing useful measures of confidence.

Something we noticed during analyses was that, when plotting reported confidence against P(correct), participants were more underconfident on pre-decisional confidence reports than on decision-point confidence reports. The individual participant P(correct)-confidence plots in Figure 2.6 suggest that there is a significant increase in reported confidence and confidence calibration at the point of decision than before decision-point which cannot be accounted for by changes in the level of objective likelihood or uncertainty. We modelled this decision-point effect on confidence with a linear regression *confidence* ~ *pcorrect* + *decision status*, where *confidence* is the distance between the reported confidence value and the correct/incorrect crossover point of 50, *pcorrect* is the distance between 50 and the value of P(correct) scaled by a factor of 100 to match confidence, and *decision status* is a dummy-coded Boolean variable indicating the type of confidence report (pre-decision or decision-point). Each participant's pcorrect beta therefore represents how much the objective likelihood is contributing to the confidence reported, while the decision status beta represents the contribution of the decision status (i.e., choice commitment) to reported confidence. The model did not include a constant term, as we assumed that *pcorrect* would match *confidence* and *decision status* in a perfectly calibrated participant.

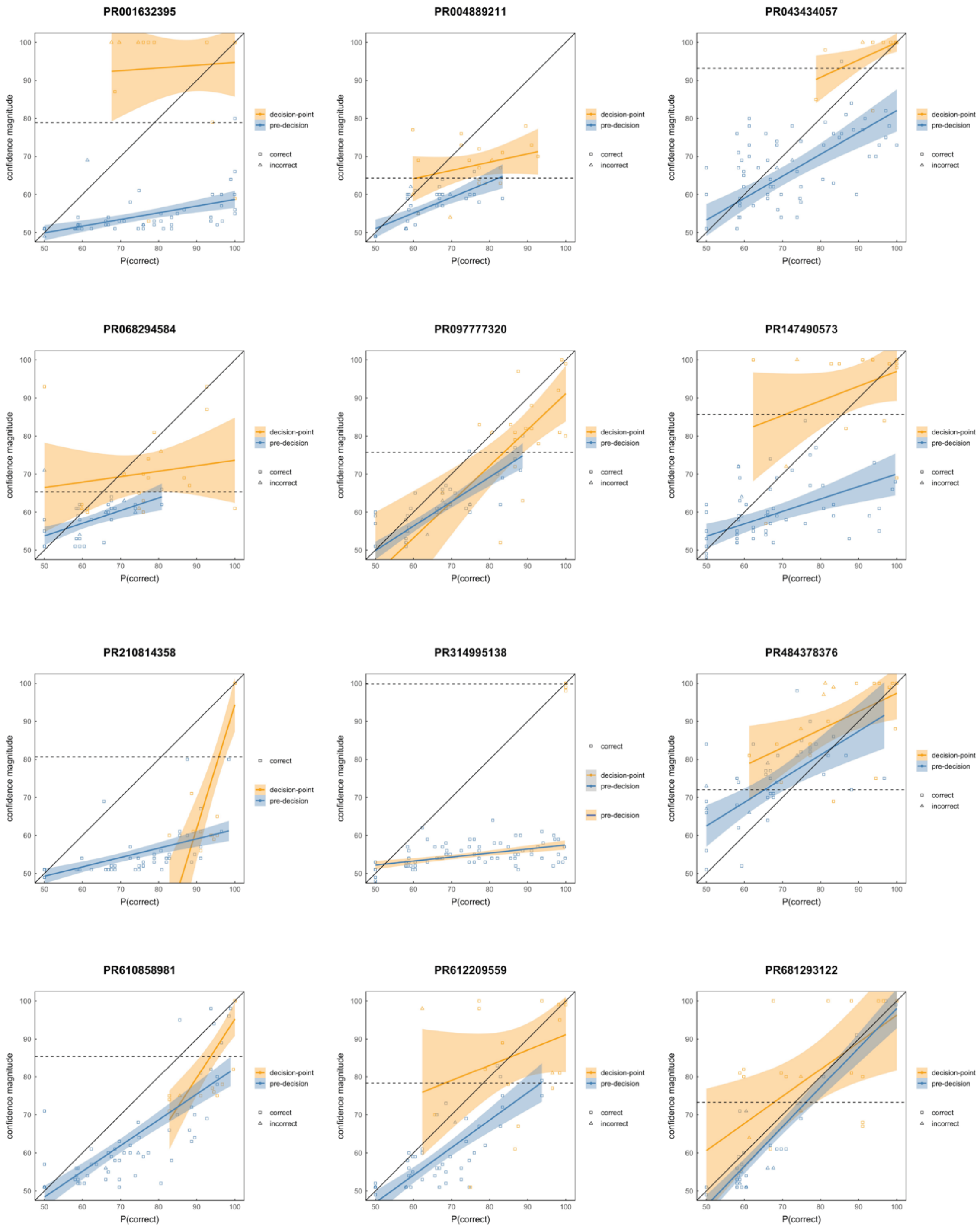


Figure 2.6. Plots of confidence magnitude (calculated as 50 plus the distance between the reported confidence value and 50) against P(correct) magnitude (calculated as 50 plus the distance between 50 and P(correct) × 100), split by pre-decisional (blue) and decision-point (orange) trials, for example participants in ISTc Pilot 3. Each plot represents all confidence reports made by a single participant. Solid diagonal line represents perfect correspondence between P(correct) and confidence. Dotted horizontal line represents the participant's average decision-point confidence.

As seen in Figure 2.7, there were large individual differences in the size of both the pcorrect beta (M = .38, SD = .19) and decision status beta (M = 16.78, SD = 11.53); for one of the participants in our sample, the decision status beta could not be calculated because they did not register a single pre-decisional confidence report during the experiment. However, the means of both betas were significantly greater than zero in one-tailed t-tests [pcorrect: $t(19) = 9.02$, $p < .001$; decision status: $t(18) = 6.35$, $p < .001$]. The mean of the decision status beta suggests that, on average, there is a 16-point confidence boost at the point of decision (measured on a 0 to 100 scale). In practice, this is a significant increase. To put that into perspective, if confidence were perfectly calibrated to P(correct), this would be approximately equivalent to the increase in P(correct) after seeing two more tiles in the same colour as the first tile selected.

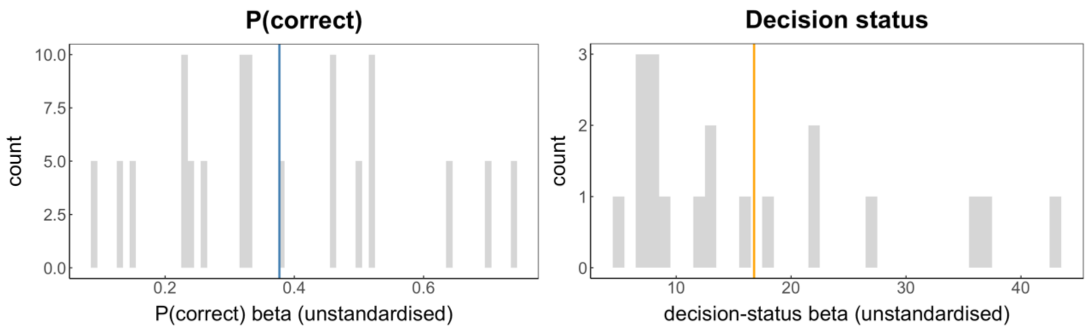


Figure 2.7. Distribution of pcorrect and decision status beta values in ISTc Pilot 3.

Discussion

The findings from the three pilot studies suggest that the ISTc shares basic task similarities with the IST, while adding an informative measure of confidence that gives us interesting insights into how confidence evolves during the decision process. Because we were able to compare people's performance on the IST and the ISTc from the same session, we could ensure that key task performance metrics, such as accuracy and the average score per trial, were consistent in our modified version of the paradigm. The ISTc also preserved differences between the Fixed Win (FW) and Decreasing Win (DW) score rule conditions in accuracy, average score per trial, and number of tiles seen before decision. While these differences were critical measures in the context of the original Clark et al. (2006) paper, they are less important for our anticipated primary analyses with the ISTc. However, preserving the contrasting score rule conditions would enable us to run secondary analyses exploring potential individual differences in how people adaptively seek information in different reward contexts. This follows from previous studies using the base IST without confidence measures (e.g., Clark et al., 2006; Hauser, Moutoussis, et al., 2017) that have found higher levels of information-seeking in the Fixed Win score rule condition among psychiatric patient samples than among normal controls.

While previous studies that have looked at the relationship between confidence and information-seeking (e.g., Desender et al., 2018; Rollwage et al.,

2020) have used two-step perceptual tasks with varying levels of evidence strength and the option to see further information after an initial decision (as seen in our DDTi), the multi-step sequential nature of information sampling in the ISTc more closely reflects the nature of evidence accumulation in many real-life decision processes, and the addition of a confidence measure throughout this information sampling phase means that we are able to probe how confidence evolves in response to the multiple changes in evidence strength preceding a final decision.

Furthermore, because participants report confidence both leading up to and immediately after the point of decision, we are also able to identify a discrepancy between these two types of confidence reports at the same level of objective likelihood. Although people seemed to be only slightly underconfident at the point of decision, the P(correct)-confidence plots and linear regression models suggest that our primary measure, the decision-point confidence bias, may hide a much stronger pre-decisional underconfidence bias followed by a decision-point boost in confidence that is not accounted for by the level of objective likelihood. The mechanisms underlying this boost, as well as the individual differences observed in the size of this boost, will be explored and discussed further in Chapters 3 and 5.

Comparing the three pilots, we found that the changes to the confidence reporting mechanism and frequency in ISTc Pilot 2 significantly improved the useability of the confidence measure and the pace of the task, while being comparable to the IST on general task performance metrics. Differences between the IST and the ISTc in our primary information-seeking measure, the total number of tiles seen before decision, were most likely due to task order effects, as these differences were no longer significant when task order counterbalancing was added in ISTc Pilot 3. Because the ISTc Pilot 3 version was comparable to the IST in all key metrics of task performance as well as in our measure of information-seeking, it appeared to be a suitable paradigm for investigating the relationship between confidence and information-seeking, and was the version we retained as the final design of the ISTc.

2.4 Summary

In this chapter, we introduced two complementary decision-making paradigms, the DDTi and the ISTc, that modified existing perceptual judgment and probabilistic reasoning tasks to yield corresponding measures of confidence (mean confidence bias in the ISTc and mean first-stage confidence in the DDTi) and information-seeking (mean number of tiles seen before decision in the ISTc and proportion of "See Again" trials in the DDTi).

The pilot studies we conducted during the development of these two tasks suggest that our measures of confidence and information-seeking meaningfully reflect how confidence evolves in response to information acquired, and how people strategically employ information-seeking to resolve uncertainty. The contrasting nature of the two tasks means that our measures of confidence and information-seeking are usefully different: while the ISTc provides insight into how objective likelihood or evidence might be converted into reported confidence, the DDTi gives us precise methods of controlling for task difficulty and quantifying the effect of information-seeking on confidence.

In the subsequent chapters, we will use these task paradigms to investigate different aspects of the relationship between confidence and information-seeking. In Chapter 3, we will use both the ISTc and the DDTi to investigate how dimensions of psychiatric interest might predict individual differences in confidence and information-seeking measures on the two tasks. In Chapters 4 and 5, we will modify the ISTc further to explore alternatives to information-seeking for uncertainty management and to delve into the mechanisms of the decision-point confidence boost we identified in our pilot data.

3
INVESTIGATING PSYCHIATRIC PREDICTORS OF CONFIDENCE AND INFORMATION-SEEKING

3.1 Introduction

Individuals with mental health disorders often hold problematic beliefs about their own cognition and exhibit differences in self-evaluative processes (Cartwright-Hatton & Wells, 1997; Wells, 2000). These metacognitive differences might be able to explain variability in patient data above and beyond cognitive factors specified in existing models (Nordahl & Wells, 2017; Wells, 2000; Wells et al., 2012). Theoretical models also suggest that disturbances in metacognitive monitoring and control strategies may contribute to the development and maintenance of various mental health disorders. For instance, in obsessive-compulsive disorder, persistent underconfidence in the perceptual domain can motivate compulsive checking behaviours that temporarily reduce the distress associated with uncertainty but ultimately perpetuate feelings of uncertainty, resulting in a vicious cycle that maintains the disorder (Hoven et al., 2019; Strauss et al., 2020; Wells, 2000). Some evidence suggests that addressing metacognitive deficits through therapeutic intervention, for instance, metacognitive therapy, might meaningfully improve patient recovery in some difficult-to-treat cases involving treatment resistance or high psychiatric comorbidity (Johnson & Hoffart, 2016; Johnson et al., 2017; Wells et al., 2012).

Several psychiatric dimensions have been found to be associated with changes in subjective confidence. The most relevant set of results for the work outlined in this chapter comes from Rouault et al. (2018), who conducted two experiments relating task performance and confidence ratings in a perceptual decision-making task to self-reported psychopathology via a wide range of psychiatric disorder inventories. Based on a large sample drawn from a normal population on a staircased (i.e., difficulty and performance-matched) dot discrimination task, the authors found anxiety to be the psychiatric symptom dimension that was most consistently associated with lower reported trial-by-trial confidence, with more severe symptomatology linked to a stronger underconfidence bias. Another psychiatric symptom dimension of interest they identified was depressive severity, which was also significantly associated with underconfidence; this corroborates previous findings of lower trial-by-trial and global post-test confidence in the memory domain among dysphoric and clinically depressed individuals (Fu et al., 2012). Interestingly, though, Rouault et al. did not find a significant association between schizotypy and confidence bias, even though patients with schizophrenia often exhibit overconfidence on error responses relative to controls for memory, social cognition, and perceptual judgment (Köther et al., 2012; Moritz et al., 2003; Moritz et al., 2014).

Given the high degree of symptom overlap and comorbidity between psychiatric disorders, it is plausible that transdiagnostic dimensions, or symptom dimensions that cut across disorder categories, might be more strongly associated with disturbances in confidence. For instance, in Rouault et al. (2018), factor analysis of items across psychiatric disorder inventories revealed a three-factor structure, where two of the latent factors identified ("Anxious-Depression" and "Compulsive Behavior and Intrusive Thought") were more robustly associated with underconfidence and overconfidence bias, respectively, than component questionnaires. One transdiagnostic dimension that we are particularly interested in is uncertainty intolerance, which reflects how unacceptably threatening and distressing an individual finds uncertain situations and experiences. Uncertainty intolerance is commonly quantified using the 27-item Intolerance of Uncertainty Scale (IUS), developed by Freeston et al. (1994) and translated into English by Buhr and Dugas (2002). IUS scores are strongly correlated with scores on psychiatric inventories, particularly with anxiety, and to a lesser extent with depression (Boswell et al., 2013; Dugas et al., 2004; Jensen et al., 2016; Lauriola et al., 2018; Robinson & Freeston, 2015). The relationship between uncertainty tolerance and confidence has been less frequently explored in the literature, but a previous study (Jensen et al., 2014) looking at this relationship in a repeated decision-making task found that individuals with high IUS scores became less confident in their decisions over time, whereas those with low IUS scores tended to become more confident in their decisions over the course of the experiment. The authors speculated that this decrease in confidence may motivate high-IUS participants to seek more information before making a decision, as previously observed by Ladouceur et al. (1997) in a probabilistic reasoning paradigm.

The relationship between confidence and information-seeking is well-established in the literature. Previous research from our lab and others has found confidence to be a strong predictor of information-seeking prior to decision (Desender et al., 2018; Hauperich et al., 2017; Schulz et al., 2020). Hausmann-Thürig & Läge (2008) have suggested that the extent of information-seeking (i.e., the point at which a participant stops seeking further information and commits to a decision) may reflect individual differences in the confidence threshold for decision. Individuals who are anxious or depressed may have a higher decision threshold and thus might seek more affirmative evidence before committing to a decision. This suggests that, to understand the relationship between metacognitive disturbance and psychopathology, we might look by proxy at the relationships between psychiatric symptom dimensions, confidence, and information-seeking.

However, the few direct experimental investigations of the relationship between specific psychiatric dimensions and information-seeking have been less than conclusive. For example, higher trait anxiety — a relatively stable individual difference in the tendency to perceive stressful situations as dangerous and threatening (Chambers et al., 2004; Spielberger et al., 1983) — has been linked to greater information-seeking in patients when it comes to information about their illness (Handelzalts et al., 2016), but not in consumers when it comes to information about a product before purchase (Locander & Hermann, 1979). This suggests that higher trait anxiety does not predict greater information-seeking in all contexts. More broadly, while the intuition is that people would seek more information when they are less confident, individuals who are intolerant of uncertainty might find uncertain situations and experiences to be stressful and distressing, and might prefer to escape or avoid uncertain situations instead of seeking more information (Luhmann et al., 2011; Smock, 1955). It is thus unclear whether there are reliable patterns of association between various psychiatric dimensions and information-seeking, as we saw for confidence in Rouault et al. (2018), and whether confidence mediates that relationship, as Jensen et al. (2014) and Hausmann-Thürig and Läge (2008) have suggested.

In summary, existing literature has supported an association between psychiatric symptoms and changes in metacognition, particularly changes to confidence bias. Transdiagnostic dimensions like uncertainty intolerance might be stronger predictors of changes to confidence bias than disorder-specific psychiatric dimensions. Because of the close relationship between confidence and information-seeking, these changes in confidence bias might also be seen in corresponding changes in information-seeking frequency. In this first experimental line of the thesis, we were interested in investigating the relationship between psychiatric symptom dimensions, confidence, and information-seeking: specifically, whether individual differences along psychiatric dimensions can reliably predict specific patterns of confidence bias and information-seeking frequency. Experiment 1 looked at changes in confidence and information-seeking in relation to four dimensions of psychiatric interest — trait anxiety, depressive severity, schizotypy, and uncertainty intolerance — using the two decision-making tasks we developed and described in Chapter 2 (the ISTc and the DDTi) in a general population sample. Experiment 2 attempted to replicate the key novel findings in Experiment 1 with a narrower focus on the specific psychiatric dimensions and measures of interest from the first experiment.

3.2 Experiment 1

3.2.1 Aims & hypotheses

Experiment 1 focused on how trait anxiety and other related psychiatric symptom dimensions in a general population sample might predict both decision confidence and information-seeking prior to decision. We also looked at whether uncertainty intolerance might explain the relationship between trait anxiety, confidence, and information-seeking and whether it could be a better predictor of decision-point confidence bias and information-seeking frequency than trait anxiety. Specifically, we preregistered the hypotheses that:

1. trait anxiety would predict lower decision-point confidence and more information-seeking behaviour;
2. the relationship between trait anxiety and decision-point confidence would be mediated by uncertainty intolerance, and;
3. the relationship between trait anxiety and information-seeking frequency would be mediated by decision-point confidence.

3.2.2 Methods

Pre-registration

The question and hypotheses, methods, data exclusion criteria, measures of interest, and anticipated analyses of Experiment 1 were preregistered with the

Open Science Framework (OSF). The preregistration document is publicly available online [Appendix S3.1], and overviews of preregistered data exclusion criteria are provided in Appendix S3.2.

Participants & exclusions

Our pre-registration specified that our intended sample size was 136 participants after exclusions. This was determined from a power analysis to detect a potential correlation of .30 or greater between our variables of interest with $\alpha = .025$ after Bonferroni correction and 90% power.

A total of 153 people were recruited using the online recruitment platform Prolific. Participants provided informed consent in accordance with procedures approved by the University of Oxford Central University Research Ethics Committee (R56625/RE003). All participants were paid a base amount of around £5/hour plus a bonus of up to £1.52.

Participants were prescreened on Prolific by self-reported age (18 to 60 years) and visual acuity (normal or corrected-to-normal vision). After completing the PHQ-9, participants were precluded from going further if they indicated suicidal or self-harm thoughts at a frequency of "nearly every day" on the questionnaire, as we would be unable to provide adequate safety monitoring for these participants in our experimental design. All 153 participants in our collected dataset passed these prescreening measures. Based on the exclusion criteria listed in Appendix S3.2, 18 people were excluded from the final analysis, leaving a final dataset size of 135 people ranging in age from 18 to 59 years (M = 28, SD = 9). 63 participants self-reported as female and 72 as male. In the final dataset, 51 trials on the ISTc (0.49% of task total) and 621 trials of the DDTi (2.06% of task total) were further removed from analysis, as they met the exclusion criteria for individual trials [Appendix S3.2].

As the focus of the main experiment was on the continuous variation of psychopathology in the general population, no information about psychiatric diagnosis or medication was recorded, so there is a possibility that some participants may have qualified for a psychiatric diagnosis and been in treatment with psychotropic medication during the study.

Design & measures

To support the idea that any alterations in confidence and information-seeking behaviour are systematic rather than limited to a specific task or domain, as has often been the case in the existing literature, I used confidence and information-seeking measures derived from two different decision-making paradigms, the DDTi and the ISTc (see Chapter 2 for task details). There were four main measures of interest across the two tasks. The two main measures of interest for confidence were confidence bias — the difference between reported confidence and P(correct) — at the point of decision on the ISTc and reported confidence for first-stage decisions on the DDTi. The two main measures of interest for information-seeking were the average number of tiles seen prior to decision on the ISTc and the proportion of trials where participants chose the "See Again" option prior to their final decision on the DDTi. Additionally, all participants completed four self-report English language questionnaires in the following order: the Patient Health Questionnaire-9 (Kroenke et al., 2001), State-Trait Anxiety Inventory - Form Y2 (Spielberger et al., 1983), Short Scales for Measuring Schizotypy (Mason et al., 2005), and Intolerance of Uncertainty Scale (Buhr & Dugas, 2002).

Procedure

Participants were asked for their demographics (age, gender, handedness) before completing the four self-report questionnaires and the two decision-making tasks. They always completed the questionnaires in the order specified before (PHQ, STAI, SSMS, IUS) commencing the decision-making tasks. Participants were allocated by random assignment to one of two task order conditions (ISTc then DDTi, or DDTi then ISTc) and, within the ISTc, one of two score rule conditions (Fixed Win then Decreasing Win, or Decreasing Win then Fixed Win), resulting in four possible procedural combinations.

At the beginning of each decision-making task, participants were given a walkthrough of the task aim, controls, and options available to them. They were asked to demonstrate their understanding of the task through completion of a number of practice trials, as specified in Chapter 2. In the ISTc, participants completed 20 trials, 10 with the Fixed Win score rule and 10 Decreasing Win score rule. In the DDTi, participants completed a total of 100 trials spread across 4 blocks of 25 trials each; participants' staircase progress was maintained throughout blocks. At the end of each task, participants were informed of how many points they had earned from the task and the total number of points they have earned across the experiment. After both tasks, participants were given a chance to give feedback about the experimental design and to report any technical issues before being debriefed. Participants took on average around 50 minutes to complete the experiment.

The entire experimental procedure was coded in JavaScript with jsPsych version 6.0.2 (de Leeuw, 2015) and run as a browser-based web application (Chrome and Firefox compatible).

3.2.3 Results

Questionnaire scores

Distributions of scores in the four self-report questionnaire measures are shown in Figure 3.1 and summarised in Table 3.1. Participants' trait anxiety scores were quite normally distributed around a mean score of 46.5. A slight right skew was seen in the depressive severity scores; using the common clinical significance cut-off score of 10 for the PHQ-9 (Kroenke et al., 2001; Manea et al., 2012), approximately 22% of our sample (30 participants) reported moderate to severe depression.

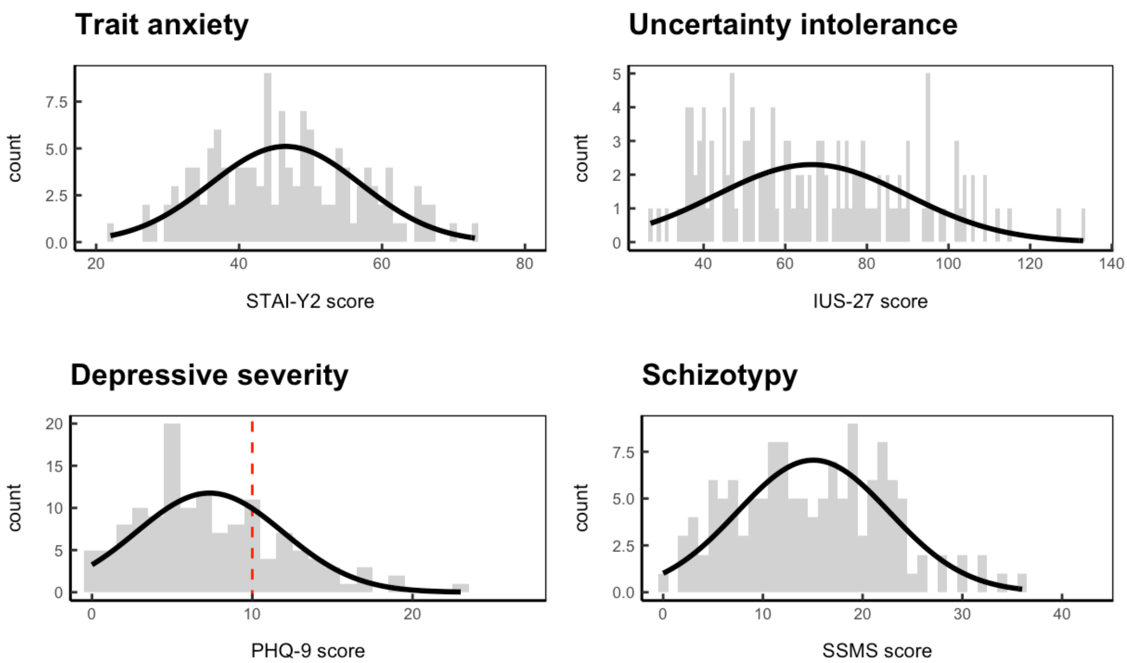


Figure 3.1. Self-report psychiatric questionnaire score distributions in Experiment 1. STAI-Y2 (top left), IUS-27 (top right), PHQ-9 (bottom left), and SSMS (bottom right). The red line on the PHQ-9/depressive severity distribution indicates the clinical cut-off score of 10.

Table 3.1
Questionnaire scores in Experiment 1

	Mean (SD)	Median	Range
Trait anxiety (STAI-Y2)	46.5 (10.5)	46.0	22 – 73
Depressive severity (PHQ-9)	7.3 (4.6)	7.0	0 – 23
Schizotypy (SSMS)	15.1 (7.6)	15.0	0 – 36
Uncertainty intolerance (IUS-27)	66.5 (23.4)	63.0	27 – 133

General task performance

Participant performance on the ISTc and DDTi was broadly consistent with findings from previous studies using predecessor variations of these tasks. Participants generally completed the tasks in a sensible and strategic manner.

ISTc scores were approximately normally distributed, and participants scored on average 1552 points across 20 trials. This average score exceeded the at-chance mean score of 700 (midpoint of the minimum possible score of -2000 and the maximum possible score of 3400), suggesting that participants were performing the task strategically. This is corroborated by generally high trial accuracy ($M = .83$, $SD = .14$) and lengthy trial RTs ($M = 24.5$ s, $SD = 15.7$ s).

Participants tended to take longer [$M_{FW} = 26.3$ s vs. $M_{DW} = 22.8$ s, $t(134) = 2.79$, $p = .006$], see more tiles before decision [$M_{FW} = 14.35$ vs. $M_{DW} = 11.19$, $t(134) = 7.61$, $p < .001$], and be slightly more accurate [$M_{FW} = .86$ vs. $M_{DW} = .81$, $t(134) = 3.19$, $p = .002$] in the Fixed Win score condition than in the Decreasing Win score condition. This is sensible given the scoring differences between the two score rule conditions. Because the Decreasing Win condition reduces participants' possible awarded score for each tile they see, it encourages a choice strategy based on less information-seeking. However, there was no significant difference in confidence bias at the point of decision [$M_{FW} = -1.94$ vs. $M_{DW} = -1.99$, $t < 1$].

DDTi performance was also consistent with expectations. Due to our staircasing procedure, all participants reached a first-stage decision accuracy (proportion correct) above .67 ($M = .73$, $SD = .03$). Participants were on average more confident [$M_{skip} = 66.74$ vs. $M_{seek} = 52.99$, $t(116) = 13.48$, $p < .001$] and more accurate [$M_{skip} = .80$ vs. $M_{seek} = .60$, $t(116) = 12.37$, $p < .001$] when they indicated their first answer to be final and skipped the second-stage decision than when they sought further information before a decision. Eight participants in our sample always indicated their first-stage decision to be final. For the remainder of our participants, second-stage decisions were more accurate (M = .85, SD = .14), $t(126)$ = 9.99, p < .001, Consistent with second-stage decision being based on more evidence, participants were generally more confident in their second-stage decisions (M = 67.67, SD = 11.89) than in their first-stage decisions (M = 61.31, SD = 5.51), $t(126)$ = 8.30, p < .001.

Some task order effects were found. A Kruskal-Wallis rank sum test found a significant difference of task order on mean first-stage confidence on the DDTi, $\chi^2(1) = 7.41$, $p = .006$, $\varepsilon^2 = .06$, with the 70 participants who completed the DDTi first having higher confidence in first-stage decisions (M = 62.56, SD = 5.43) than the 65 participants who completed the ISTc first (M = 59.97, SD = 5.31). Participants who completed the DDTi first also took less time on the subsequent ISTc than participants who were assigned to do the ISTc first [$M_{DDTi\text{-}ISTc}$ = 23.1 s vs. $M_{ISTc\text{-}DDTi}$ = 26.1 s, $\chi^2(1)$= 3.86, p = .050, ε^2 = .03]. No significant effect of task order was found on the number of tiles seen before decision [$M_{DDTi\text{-}ISTc}$ = 12.12 vs. $M_{ISTc\text{-}DDTi}$ = 13.47, $\chi^2(1)$ = 2.04, p = .153] and RTs for the DDTi [$M_{DDTi\text{-}ISTc}$ = 2145 ms vs. $M_{ISTc\text{-}DDTi}$ = 1967 ms, $\chi^2(1)$ = 2.04, p = .154].

Trait anxiety as a predictor of confidence and information-seeking

Our first key hypothesis was that trait anxiety would predict lower subjective confidence and more information-seeking behaviour. To test our hypothesis, we conducted separate Bonferroni-corrected linear regressions and looked at the standardised regression coefficients of [1] trait anxiety score predicting confidence and [2] trait anxiety score predicting information-seeking, when adjusted for the effects of age and gender. We modelled this relationship with a linear regression *measure* ~ *trait anxiety* + *age* + *gender*, where *measure* is the *z*-scored continuous variable for our measure of interest (confidence or information-seeking), *trait anxiety* is the *z*-scored STAI score, *age* is the *z*-scored age in years, and *gender* is a *z*-scored dummy-coded Boolean variable indicating gender (0 for male, 1 for female).

As shown in Figure 3.2 and summarised in Table 3.2, we found non-significant negative associations between trait anxiety scores and confidence on the ISTc (β = -.07, p = .429) and DDTi (β = -.09, p = .300). The three-predictor model of trait anxiety, age, and gender explained a small but statistically significant amount of variance in the confidence measure for the ISTc [adj. R^2 = .04, $F(3, 131) = 2.70$, p = .048] but not the DDTi [adj. R^2 = .03, $F(3, 131) = 2.32$, p = .078]. Trait anxiety also failed to predict information-seeking on the ISTc (β = -.10, p = .237) and the DDTi (β = .01, p = .898). The three-predictor model explained a small but significant variance in the information-seeking measure for the ISTc [adj. R^2 = .04, F(3, 131) = 2.79, p = .043] and DDTi [adj. R^2 = .08, F(3, 131) = 5.09, p = .002]. Because no significant association was found between trait anxiety and confidence or information-seeking on either task, mediation models testing our second and third preregistered hypotheses — namely, uncertainty intolerance and confidence as mediators of the relationship between trait anxiety and confidence and between trait anxiety and information-seeking, respectively — were poor fits for the data and did not provide meaningful insights (see Appendix S3.3 for mediation model diagrams).

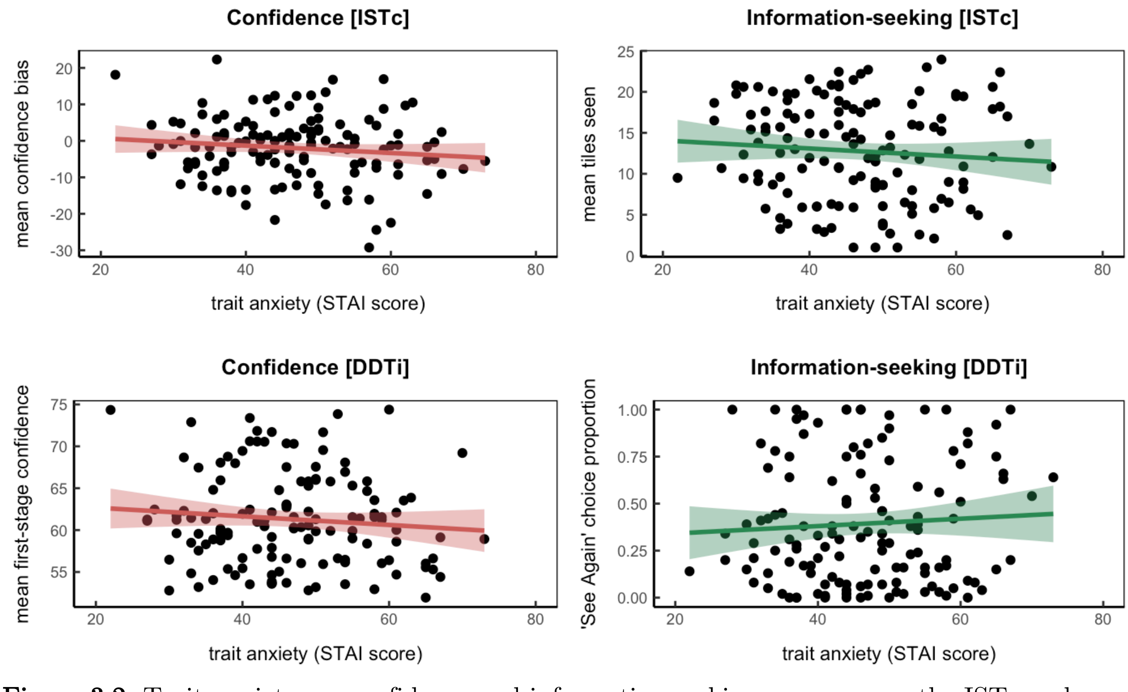


Figure 3.2. Trait anxiety vs. confidence and information-seeking measures on the ISTc and DDTi.

Table 3.2

Standardised regression coefficients for three-predictor regression model of confidence and information-seeking

	Confidence [ISTc]	Confidence [DDTi]	Information-seeking [ISTc]	Information-seeking [DDTi]
Predictor				
STAI	-.07	-.09	-.10	.01
age	.20*	-.03	-.01	-.15
gender	-.13	-.20*	.23**	.32***

Gender dummy-coded as 0 (male) and 1 (female).

Significance levels: . $p < .10$ * $p < .05$ ** $p < .01$ *** $p < .001$

Trait anxiety also failed to predict other preregistered measures of confidence and information-seeking on the two tasks. For instance, a secondary preregistered measure of information-seeking in the ISTc was the difference between the mean number of tiles seen before decision in the Fixed Win and Decreasing Win score conditions. We were interested in this possible difference in information-seeking tendency between scoring conditions because previous studies using the information sampling task (e.g., Clark et al., 2006; Hauser, Moutoussis, et al., 2017) had found the difference between tiles seen in Fixed Win and Decreasing Win to be greater in patients with OCD or substance abuse than controls. However, in Experiment 1, we found that trait anxiety did not predict this difference (β = -.15, p = .086) when adjusting for age and gender. Participants' trait anxiety scores also failed to predict the strength of the correlation between confidence and P(correct) at the point of decision in the ISTc ($\beta < .01$, $p = .991$) when adjusting for age and gender. Gender and age were not significant predictors in either case, and both model fits were poor, explaining less than 1% of variance. Overall, we did not observe any strong relationships between the trait anxiety score and primary or secondary measures of confidence and information-seeking from either task.

Other psychiatric predictors of confidence and information-seeking

To put our findings regarding trait anxiety into context, we looked at whether the other psychiatric dimensions measured would be stronger predictors of confidence and information-seeking on our tasks, using the same three-predictor regression model structure specified for trait anxiety.

Depressive severity, as measured by the PHQ-9, was not found to predict confidence on the ISTc (β = -.02, p = .834), confidence on the DDTi (β = .06, p = .514), or information-seeking on the DDTi (β = -.09, p = .319). Depressive severity did significantly predict information-seeking on the ISTc (β = -.20, p = .026), with the three-predictor model explaining a small but statistically significant amount of variance [adj. R^2 = .06, $F(3, 131)$ = 4.08, p = .008]. However, the direction of this effect was surprising: contrary to our intuition that more depressed people would seek more information before decision, participants with more severe self-rated depressive symptoms tended to seek less information (i.e., see fewer tiles) before making a decision.

Similarly, uncertainty intolerance, as measured by the IUS-27, was not found to predict confidence on the ISTc ($\beta < .01$, p = .998) or the DDTi (β = .04, p = .665), or information-seeking on the DDTi (β = .05, p = .566). But it significantly predicted information-seeking on the ISTc (β = -.18, p = .040), with the three-predictor model explaining a small but significant amount of variance [adj. R^2 = .06, F(3, 131) = 3.79, p = .012]. Again, the direction of the effect was in the opposite direction to our hypothesis: participants with greater self-rated uncertainty intolerance tended to see fewer tiles before decision on the ISTc.

Schizotypy, as measured by the SSMS, was not found to predict any of the four decision process measures: confidence on the ISTc (β = -.03, p = .725), confidence on the DDTi (β = .10, p = .257), information-seeking on the ISTc (β = -.06, p = .463), or information-seeking on the DDTi (β = -.06, p = .500). A summary of these findings, in standardised regression betas for comparability, is presented below in Figure 3.3.

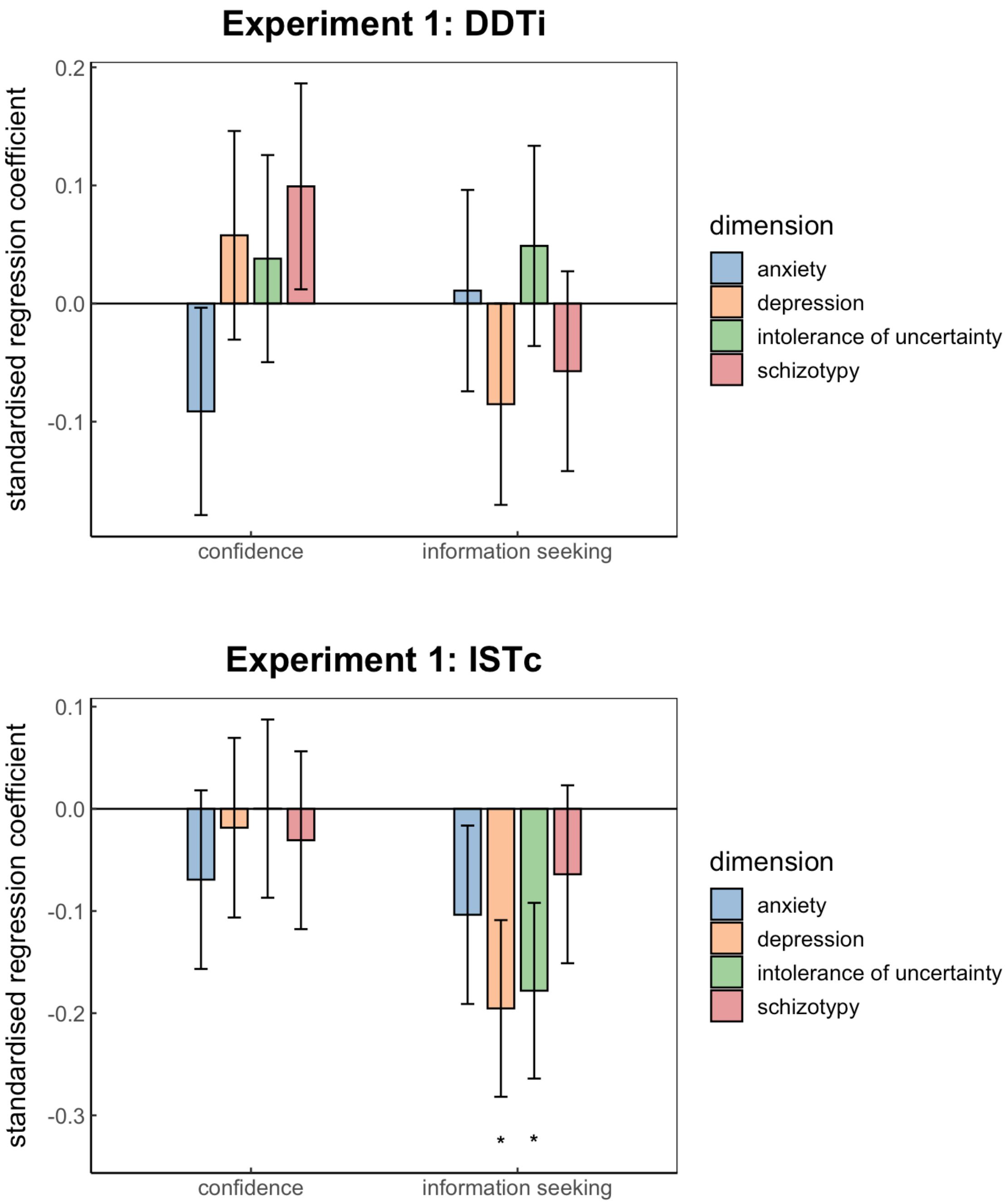


Figure 3.3. Associations between self-reported psychopathology and confidence and information-seeking measures. Error bars indicate ±1 standard error.

Relating psychiatric measures and first-order decision process measures

Although we observed weak and inconsistent associations between our psychiatric measures and first-order decision process measures, we found strong and consistent relationships within each subset of measures, as shown in the correlation matrix below [Figure 3.4]. There were strong and significant positive correlations between our four psychiatric measures, with all p-values $< .001$. Confidence measures across tasks were positively correlated [$r(133) = .35$, $p < .001$], as were information-seeking measures across tasks [$r(133) = .28$, $p < .001$]. Confidence and information-seeking measures were negatively correlated with each other within each task [ISTc: $r(133) = -.44$, $p < .001$; DDTi: $r(133) = -.29$, $p < .001$], but there were also significant cross-task correlations: confidence on the ISTc had a moderate negative correlation with information-seeking on the DDTi [$r(133) = -.22$, $p = .009$], and confidence on the DDTi had a moderate negative correlation with information-seeking on the ISTc [$r(133) = -.34$, $p < .001$]. The high degree of covariance within our psychiatric measures and within our task performance measures suggests, like Rouault et al. (2018), we may be able to observe stronger and more consistent relationships between latent variables than between specific task or questionnaire measures.

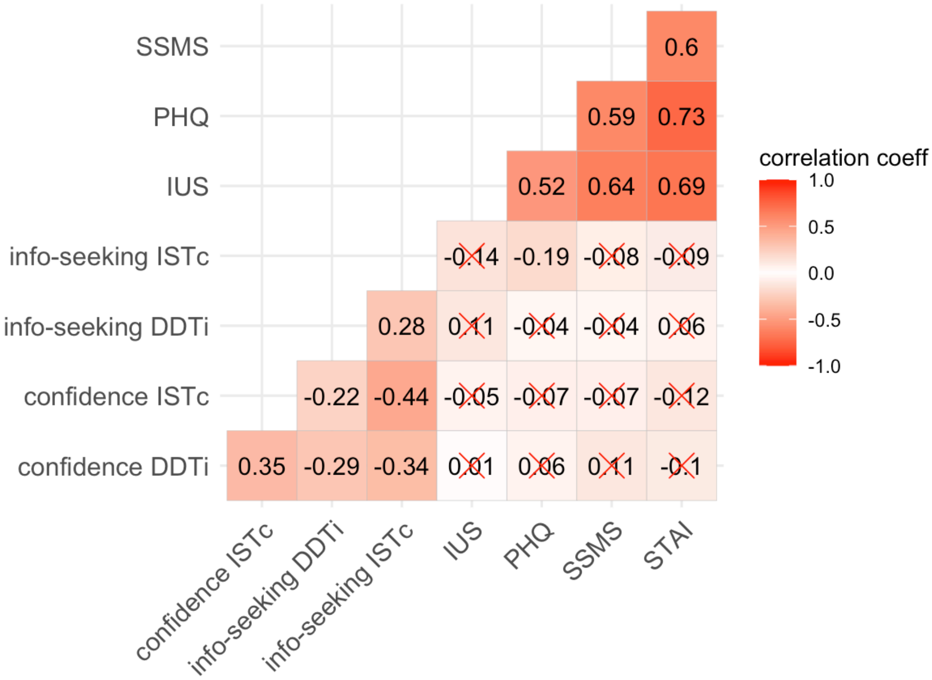


Figure 3.4. Correlation matrix of psychiatric, confidence, and information-seeking measures. Red crosses over correlation coefficients indicate failure to reach statistical significance.

Modelling the relationship between psychiatric and first-order decision process measures

Based on the high degree of covariance within our measures, we decided to assess the relationship between three latent variables reflecting self-reported psychiatric symptom severity, decision-point confidence bias, and information-seeking frequency before decision. This initial model structure reflects our initial query at the beginning of this chapter about the nature of the relationship between psychiatric symptomatology, subjective confidence, and information-seeking strategy. We assessed this initial structure using structural equation modelling (SEM) analysis with maximum likelihood parameter estimation in R's lavaan package (version 0.6-5). Our initial SEM is visualised graphically in Figure 3.5.

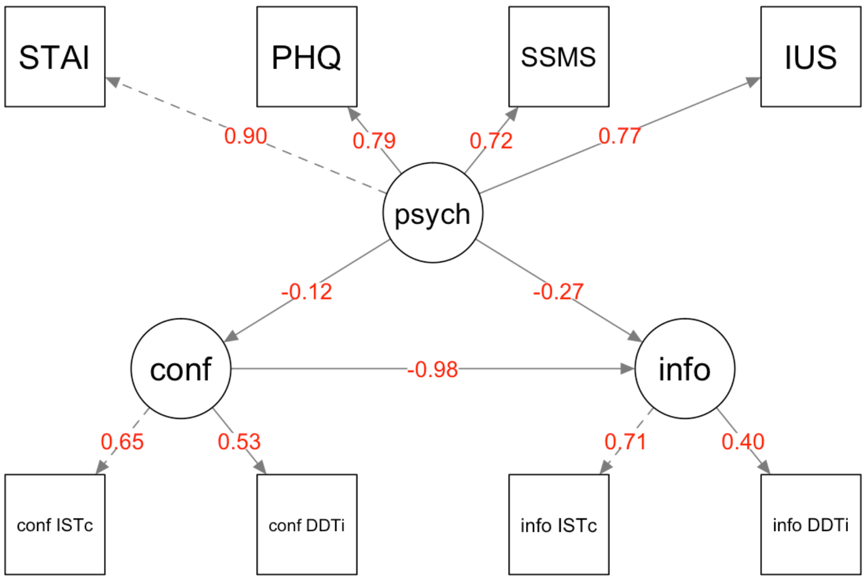


Figure 3.5. Structural equation model for Experiment 1 measures. The circles represent the latent variables: psychiatric symptomatology (psych), subjective confidence (conf), and information-seeking strategy (info). The rectangles represent our measured variables: four self-report questionnaire scores (STAI, PHQ, SSMS, and IUS) and four first-order decision process measures across our two tasks (ISTc confidence, DDTi confidence, ISTc information-seeking, DDTi information-seeking). Dotted lines represent fixed effects.

As expected, we found strong relationships between each of the questionnaire measures and our latent variable for psychiatric symptomatology. We also found a very strong relationship between the latent variables for confidence and information-seeking (β = -.98, p < .001), confirming the significant within- and cross-task associations between confidence and information-seeking seen in Figure 3.4. However, the latent variable structure found no significant association between psychiatric symptomatology and confidence (β = -.12, p = .332). The association between psychiatric symptomatology and information-seeking was stronger (β = -.27, p = .029), most likely driven by the significant negative relationships that depressive severity and uncertainty intolerance have with information-seeking. Overall, the model fit could be improved, TLI = .883, CFI = .929, RMSEA = .105, and post-hoc modification suggestions indicated that covariation between the questionnaire measures could be added to the model to improve fit. Incorporating this covariance into the revised model [Appendix S3.4] did significantly improve fit above the initial model [$\chi^2(4) = 19.89$, $p < .001$], but the revised model still exhibited a similarly weak relationship between individual differences measures and decision process measures.

Relating psychiatric dimensions and second-order decision process measures

The results from the previous analyses indicate weak relationships between first-order measures of individual differences and task performance, but possibly stronger relationships between latent or derived variables. Hence, as a final exploratory analysis, we looked at whether psychiatric dimensions would be more strongly associated with a second-order measure of task performance that we had preregistered.

In our pilot studies, we had observed rated confidence to be much higher at the point of decision than during the information sampling phase of the ISTc, even when taking into account the objective probability of being correct; this is the effect we termed the 'decision-evoked confidence boost' in Chapter 2. For each participant, we can model this effect with the linear regression *confidence ~ pcorrect + decision status*, where subjective confidence ratings are predicted by *pcorrect* (continuous numerical variable ranging from 50 to 100) and *decision status* (dummy-coded Boolean variable indicating whether the confidence rating was pre-decisional or decision-point). Each participant's pcorrect beta represents the degree of calibration between objective probability and their subjective confidence, while the decision status beta represents the magnitude of the decision-point confidence boost. We were interested in exploring whether there were associations between individual differences in these measures of the evidence-confidence relationship and individual differences along psychiatric dimensions.

To do this, we conducted separate linear regressions for each psychiatric dimension, where the z-scored questionnaire score (e.g., STAI) was entered as a predictor of z-scored decision status and pcorrect betas separately, adjusting for the effects of age and gender. A comparison of standardised regression coefficients for the four questionnaires is shown in Figure 3.6 below. The relationships between pcorrect betas and questionnaire scores were all below the significance threshold, with depressive severity coming the closest to significance ($\beta = .17$, $p = .064$). The analysis corroborates our earlier finding of no significant associations between trait anxiety scores and P(correct)-confidence correlation coefficients. However, certain questionnaire scores did more strongly predict decision status betas: namely, depressive severity (β = -.19, p = .047) and uncertainty intolerance (β = -.19, p = .033). The results suggest that participants who self-report more severe depressive symptoms and intolerance of uncertainty tend to exhibit a smaller decision-evoked confidence boost. There is a stronger relationship between these two dimensions and this second-order measure of decision confidence than with our first-order measure of confidence (confidence bias at the point of decision).

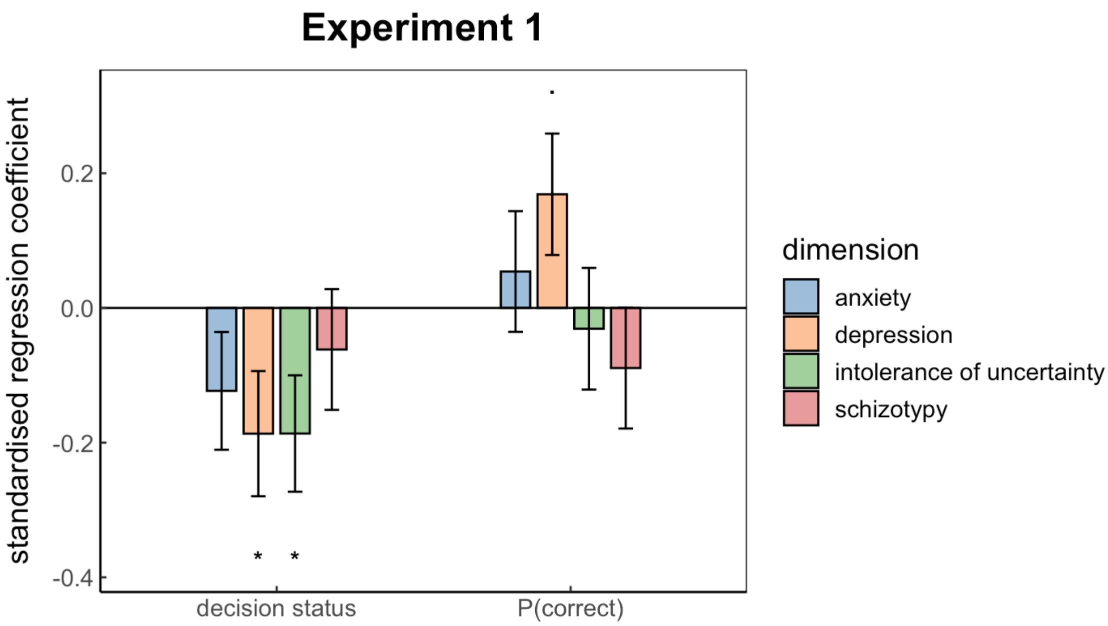


Figure 3.6. Associations between self-reported psychopathology and pcorrect and decision status betas in Experiment 1. Error bars indicate ±1 standard error.

3.2.4 Discussion

The central psychiatric dimension we were investigating in Experiment 1 was trait anxiety. We were interested in understanding whether individual differences along this psychiatric dimension of interest could reliably predict confidence and information-seeking tendencies across tasks in different decision-making domains, namely, perceptual discrimination and probabilistic judgment. Trait anxiety was of interest because a previous study by Rouault et al. (2018) using an analogous perceptual discrimination task had found anxiety measures to be the strongest and most consistent predictor of confidence bias. It is also interesting because of its inconsistent relationship with information-seeking behaviours. While a positive relationship between trait anxiety and information-seeking is intuitively consistent with confidence literature (e.g., Desender et al., 2018) and with observations of psychiatric patients (e.g., Kesby et al., 2017), there has been little direct empirical support for this relationship.

Our results failed to find support for the hypothesis that trait anxiety predicts lower decision confidence and greater information-seeking before decision. The association between trait anxiety and information-seeking was very weak and inconsistent in direction between tasks. On the other hand, the direction of the association between trait anxiety and confidence measures on both tasks was consistent with our preregistered predictions and with previous literature, but did not achieve statistical significance. However, when we compare our results with those of Rouault et al. (2018), we see that our standardised regression coefficients — in particular, trait anxiety predicting confidence on the DDTi — are comparable in size to what Rouault et al. (2018) found using a very similar dot-count discrimination task. Rouault et al.'s sample size was much larger, which suggests that the sample size in Experiment 1 may not have given us sufficient power to detect these effects with statistical significance. The reason we did not use a similar sample size to Rouault et al. was that, when we preregistered this experiment, we were unsure about the effect sizes we would see with the ISTc. In the interests of practicality for a first study, and because we anticipated larger effect sizes with the ISTc than with the dot-count discrimination task, we settled on a smaller sample size than Rouault et al., with sufficient power to observe a moderately sized effect. However, our results indicate that the effect sizes of the association between trait anxiety and confidence were similar across the two tasks, and to the effect sizes seen for the other psychiatric dimensions tested.

In contrast to confidence, information-seeking was found to be more strongly associated with depressive severity and uncertainty intolerance, but in the opposite direction to what we anticipated: greater depressive severity or uncertainty intolerance predicted less information-seeking on the ISTc. This finding is unusual because previous literature had seemed to suggest a trend in the other direction. Previous studies had found that uncertainty intolerance significantly predicts information-seeking even when more specific measures of psychiatric dimensions might not (Hauser, Moutoussis, et al., 2017; Rosen & Knauper, 2009). Because IUS scores are strongly correlated with scores on other psychiatric inventories, including anxiety and depression, in this experiment as well as others (Boswell et al., 2013; Dugas et al., 2004; Jensen et al., 2016), we expected both depression and uncertainty intolerance to be associated with greater information-seeking, not less. A possible rationale for this discrepancy could be the task design. In the ISTc, unlike the DDTi, participants received feedback after completing each trial. While participants can reduce uncertainty by seeking more information prior to committing to a decision on the ISTc, they may also be able to reduce uncertainty by ending the trial early, either to obtain information about the trial through the feedback provided or to escape from the uncertain situation.

Examining the relationship between these measures helps us place these varied results within the broader context and identify overarching trends. Correlation analyses helped us initially spot two clusters of variables — the psychiatric measures (trait anxiety, depressive severity, schizotypy, and uncertainty intolerance) and the decision process measures (confidence and information-seeking on the two tasks) — where measures were highly intercorrelated within each cluster but not very strongly correlated with measures in the other cluster. Based on these findings, we constructed a SEM that analysed this relationship between latent variables derived from the covariance structure. While our SEM confirmed findings in previous literature of strong associations between psychiatric symptomatology measures (e.g., Rouault et al., 2018) and between confidence and information-seeking (e.g., Desender et al., 2018), it also highlighted the extremely weak association between these two variable clusters. Taken together, the results seem to suggest that, while states of lower confidence may motivate greater information-seeking prior to decision commitment, individual differences along dimensions of psychiatric interest are not associated with large, systematic between-subjects differences in decision confidence or information-seeking strategy.

However, some interesting associations emerged in our exploratory analyses of second-order measures of decision-making. These second-order decision process measures, such as the P(correct)-confidence calibration and the magnitude of the decision-evoked confidence boost on the ISTc, quantify individual differences in how people translate evidence strength or objective likelihood into decision confidence. Only the decision-evoked confidence boost showed strong associations with psychiatric dimensions: we found that greater depressive severity and uncertainty intolerance significantly predicted smaller decision-evoked confidence boosts, even after adjusting for the objective likelihood of being correct. This suggests that the relationship we expected between confidence and individual differences of psychiatric significance might be more clearly seen in this second-order measure of confidence. The findings also raise the idea that the reduced confidence associated with psychiatric symptomatology in previous literature (e.g., Rouault et al., 2018) might be driven by a change in confidence around the point of decision commitment rather than a general underweighting or overweighting of evidence throughout the information sampling phase.

In summary, Experiment 1 did not find a significant overall relationship between psychiatric dimensions, confidence, and information-seeking in first-order measures or latent variable analysis. However, the direction and effect size of the relationship between trait anxiety and confidence we found were in line with previous findings in this field. The experiment also produced some interesting and unexpected results that warrant further exploration and replication: namely, a relationship between depressive severity and information seeking on the ISTc in the opposite direction to what we expected, and a strong association between depressive severity and uncertainty intolerance and the magnitude of the decision-evoked confidence boost in the same direction as what we expected. In the following study, Experiment 2, we seek to replicate these novel findings first before drawing further conclusions.

3.3 Experiment 2

3.3.1 Aims & hypotheses

Experiment 2 focused on replicating two key and surprising findings from Experiment 1, specifically that:

1. depressive severity and uncertainty intolerance would each predict a smaller decision-evoked confidence boost; and
2. depressive severity and uncertainty intolerance would each predict less information-seeking before decision on the ISTc.

Because the experiment focused on confidence and information-seeking measures from the ISTc only, we dropped the DDTi to shorten the experiment. Similarly, we dropped the schizotypy measure (SSMS) because it was not a key predictor of any first-order or second-order measures in Experiment 1.

3.3.2 Methods

Participants & exclusions

We decided to use a larger sample size in Experiment 2 to increase the power to observe associations of interest between trait anxiety and the decision status beta. Having observed in Experiment 1 effect sizes (standardised regression betas) of around -.20 for depressive severity and uncertainty intolerance and anticipating a slightly larger effect after eliminating task order effects associated with the DDTi, we determined from a power analysis that an intended sample size of 199 participants would be sufficient to detect a potential correlation of .225 or greater between our variables of interest with $\alpha = .05$ and 90% power.

A total of 223 people were recruited using the online recruitment platform Prolific. Participants provided informed consent in accordance with procedures approved by the University of Oxford Central University Research Ethics Committee (R56625/RE003). All participants were paid a base amount of around £5/hour plus a bonus of up to £1.02.

Participants were prescreened on Prolific by self-reported age (18 to 60 years) and visual acuity (normal or corrected-to-normal vision). After completing the PHQ-9, participants were precluded from going further if they indicated suicidal or self-harm thoughts at a frequency of “nearly every day” on the questionnaire, as we would be unable to provide adequate safety monitoring for these participants in our experimental design. All 223 participants in our collected dataset passed these prescreening measures.

Experiment 2 was not preregistered but used the same preregistered exclusion criteria as Experiment 1 [Appendix S3.2]. Based on these exclusion criteria, 25 people were further excluded from the final analysis, leaving a final dataset size of 198 participants ranging in age from 18 to 58 years (M = 31, SD = 10). 105 participants self-reported as female, 91 as male, and 2 as other. In the final dataset, 57 trials on the ISTc (0.74% of task total) were further removed from analysis, as they met the exclusion criteria for individual trials [see Appendix S3.2].

As the focus of the main experiment was on the continuous variation of psychopathology in the general population, no information about psychiatric diagnosis or medication was recorded, so there is a possibility that some participants may have qualified for a psychiatric diagnosis and been in treatment with psychotropic medication during the study.

Measures & procedure

Experiment 2 omitted the DDTi and the SSMS, but otherwise used the same tasks and questionnaires as Experiment 1. Participants took on average around 28 minutes to complete the experiment, as compared to 50 minutes for Experiment 1.

All participants were asked for their demographics (age, gender, handedness) before completing the three self-report questionnaires and the ISTc. Participants always completed the questionnaires in the specified order (PHQ, STAI, IUS) before commencing the ISTc, and were allocated by random assignment to one of two score rule conditions (Fixed Win then Decreasing Win, or Decreasing Win then Fixed Win). All other aspects of experimental procedure remained the same as in Experiment 1.

3.3.3 Results

Questionnaire scores

Distributions of scores in the three psychiatric survey measures are shown in Figure 3.7 and summarised in Table 3.3. Like Experiment 1, participants' trait anxiety scores were relatively normally distributed around a mean of 45.2. A slight right skew was seen in the depressive severity scores. Approximately 22% of our sample (43 participants) reported moderate to severe depression; the proportion was consistent with that observed in Experiment 1.

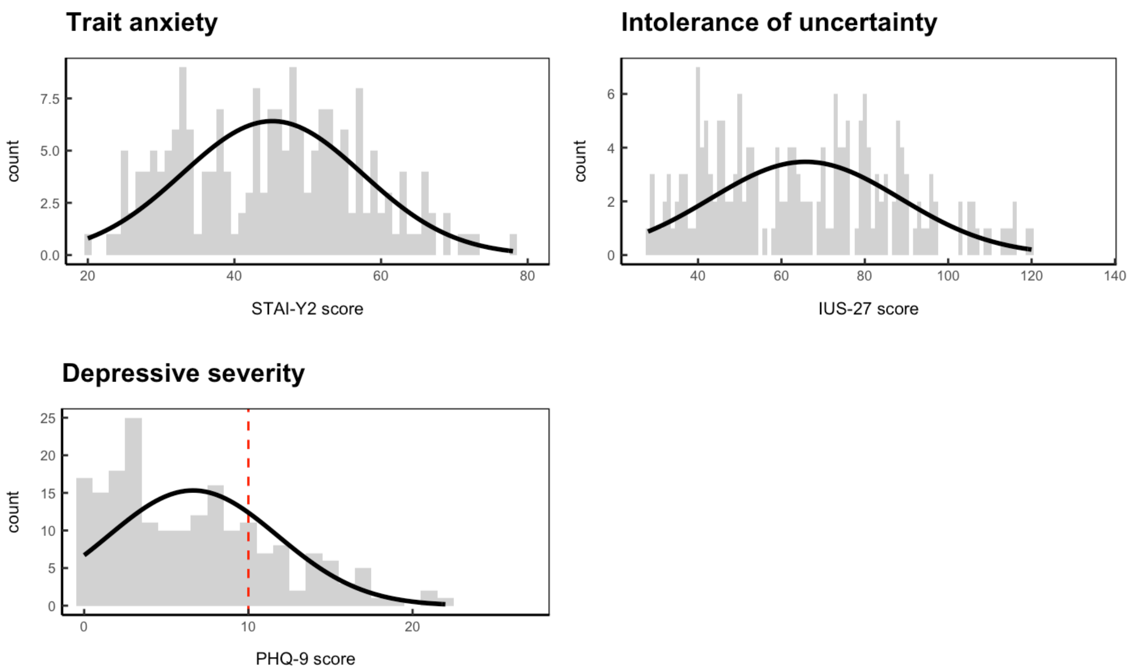


Figure 3.7. Self-report psychiatric questionnaire score distributions in Experiment 2. The red line on the PHQ-9/depressive severity distribution indicates the clinical cut-off score of 10.

Table 3.3

Questionnaire scores in Experiment 2

	Mean (SD)	Median	Range
Questionnaire			
Trait anxiety (STAI-Y2)	45.2 (12.3)	46.0	20 – 78
Depressive severity (PHQ-9)	6.6 (5.2)	6.0	0 – 22
Uncertainty intolerance (IUS-27)	65.8 (22.7)	64.0	28 – 120

General task performance

Participant performance on the ISTc was broadly consistent with findings from Experiment 1. ISTc scores were approximately normally distributed, and participants scored on average 1562 points across 20 trials. There was no significant difference between mean scores in the two experiments, $t < 1$. Like Experiment 1, trial accuracy was high (M = .86, SD = .12) and trial RTs were long (M = 29.4 s, SD = 16.2 s). Participants tended to see more tiles [M_{FW} = 16.12 vs. M_{DW} = 12.78, $t(197) = 8.71$, $p < .001$], take longer [M_{FW} = 31.4 s vs. M_{DW} = 27.4 s, $t(197) = 3.79$, $p < .001$], and be more accurate [M_{FW} = .89 vs. M_{DW} = .84, $t(197) = 4.51$, $p < .001$] in the Fixed Win than in the Decreasing Win score condition without having a significant difference in confidence bias at the point of decision [M_{FW} = -4.11 vs. M_{DW} = -4.31, $t(197) = .28$, $p = .783$]. Taken together, this suggests that participants were completing the task in a sensible and strategic manner, with comparable performance to the sample in Experiment 1. Given this broad comparability between the two experiments, we proceeded to assess whether confidence and information seeking measures in Experiment 2 would replicate the two novel findings of Experiment 1.

Psychiatric predictors of the decision-evoked confidence boost, revisited

The first hypothesis we tested in this experiment was whether we would find psychiatric questionnaire scores predicting a smaller decision-evoked confidence boost on the ISTc, as had been observed in Experiment 1. As in the previous experiment, we assessed this relationship with separate linear regressions for each psychiatric dimension, where the participant's *z*-scored questionnaire score (e.g., STAI) was a predictor of their decision status and pcorrect betas, adjusted for the effects of age and gender. The standardised regression coefficients for the three questionnaire scores are compared in Figure 3.8 below.

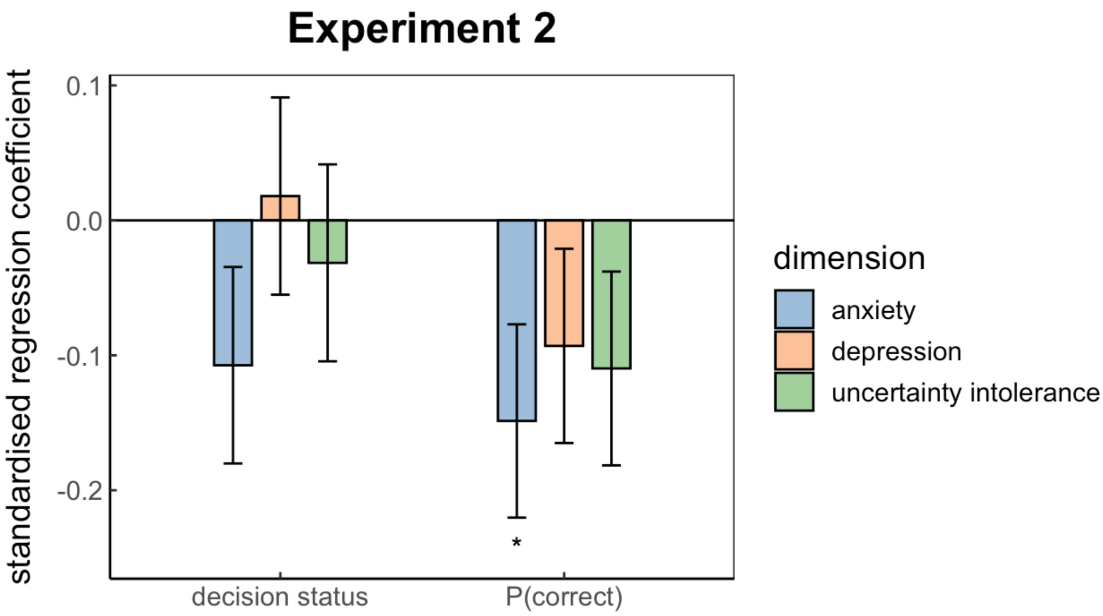


Figure 3.8. Associations between self-reported psychopathology and pcorrect and decision status beta in Experiment 2. Error bars indicate ±1 standard error.

Our results failed to replicate the finding in Experiment 1: none of our psychiatric measures significantly predicted the decision status beta. The effect size of the regression coefficients in the analyses was generally smaller; in the case of the dimensions of interest (depressive severity and uncertainty intolerance), the association with the decision status beta disappeared. In contrast to Experiment 1, trait anxiety proved to be generally the strongest predictive dimension. Higher trait anxiety scores were significantly associated smaller pcorrect betas (β = -.15, p = .039), suggesting that more trait-anxious participants were reporting lower confidence than less trait-anxious participants given the same level of evidence (i.e., objective probability of being correct). Moreover, while the association between trait anxiety and decision status betas was non-significant, it was in the same direction as we had seen in Experiment 1.

To assess the overall strength of the relationship between our psychiatric measures and the pcorrect and decision status betas, we pooled data from both Experiments 1 and 2 and reran these analyses. We found that, in the combined dataset, only trait anxiety remained significantly associated with the decision status beta (β = -.12, p = .039): higher trait anxiety appeared to predict a smaller decision-evoked confidence boost. A summary of these findings, in standardised regression betas for comparability, is presented below in Figure 3.9.

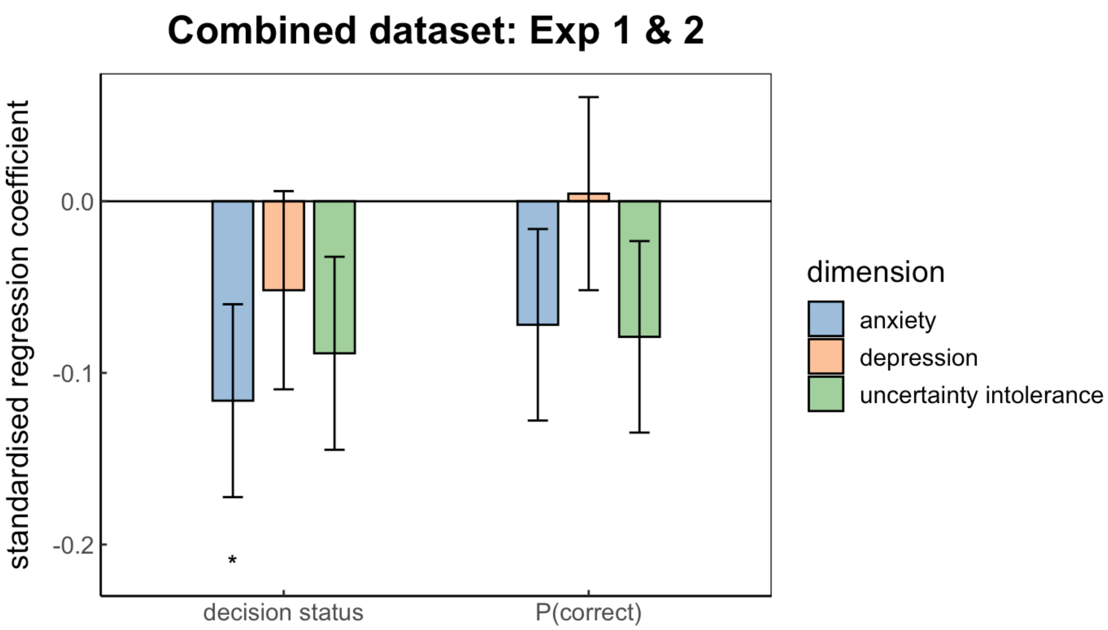


Figure 3.9. Associations between self-reported psychopathology and pcorrect and decision status betas (Experiments 1 and 2 combined dataset). Error bars indicate ±1 standard error.

Psychiatric predictors of confidence and information-seeking, revisited

The second hypothesis we tested in this experiment was whether we could replicate the surprising negative relationship between information-seeking and two questionnaire measures, depressive severity and uncertainty intolerance, found in Experiment 1. Previously, we had observed participants with greater depressive severity and uncertainty intolerance exhibiting reduced information-seeking on the ISTc: instead of turning over more tiles prior to decision, they turned over fewer relative to participants who scored lower on those dimensions. For a more complete comparison with Experiment 1, we conducted regression analyses for information-seeking and confidence, using the same three-predictor regression model structure as in the previous experiment. In addition to replicating the significant effects observed in Experiment 1, we were also interested in seeing if the previously observed non-significant trend towards reduced confidence as a function of trait anxiety would achieve significance with the larger sample size in the current experiment.

In contrast to the findings of but in line with our preregistered hypothesis for Experiment 1, we found a significant positive association between depressive severity and information-seeking on the ISTc (β = .17, p = .017), suggesting that participants with more severe depressive symptoms saw more tiles prior to decision in the current experiment. Looking at the other psychiatric predictors of information-seeking, we find a similar flip in direction of the results from the previous experiment. A positive but non-significant association was observed between trait anxiety and information-seeking (β = .13, p = .065). The association between uncertainty intolerance and information-seeking also flipped in direction, but it too failed to reach statistical significance in the current experiment (β = .12, p = .100).

In the current experiment, we also found a significant negative association between trait anxiety and confidence on the ISTc (β = -.23, p = .001), again in line with the preregistered hypothesis of the previous experiment. This appeared to be a strengthening of the negative effect previously observed in Experiment 1, which previously did not reach statistical significance. Depressive severity and uncertainty intolerance also had numerically smaller negative betas for confidence in Experiment 2 compared to Experiment 1, though neither reached significance levels [PHQ: $\beta = -.11$, $p = .113$; IUS: $\beta = -.10$, $p = .185$].

Model fit for all measures was similar to what was seen in Experiment 1: the three-predictor model explained on average less than 5% of the variance. A summary of these findings, in standardised regression betas for comparability, is presented below in Figure 3.10.

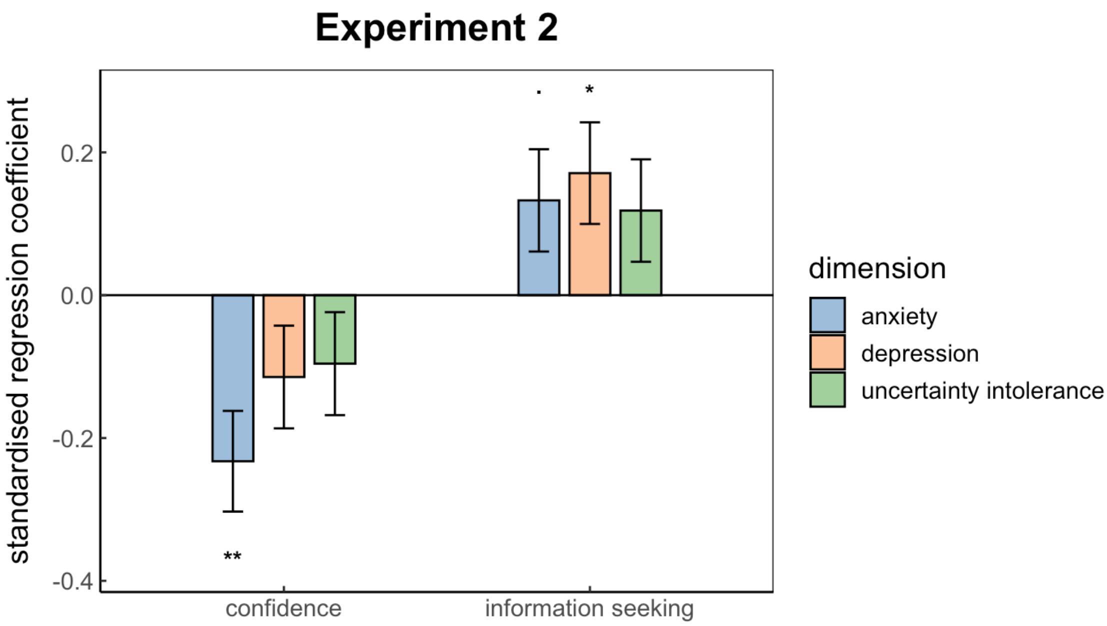


Figure 3.10. Associations between self-reported psychopathology and confidence and information-seeking in Experiment 2. Error bars indicate ±1 standard error.

Again, to assess the overall strength of the relationship between our psychiatric measures and confidence and information-seeking measures, we pooled data from both Experiments 1 and 2 and reran these analyses. The standardised regression coefficients for the three questionnaire scores are compared in Figure

3.11 below. In the combined dataset, we found that only trait anxiety was significantly associated with a greater decision-point confidence bias on the ISTc (β = -.17, p = .002).

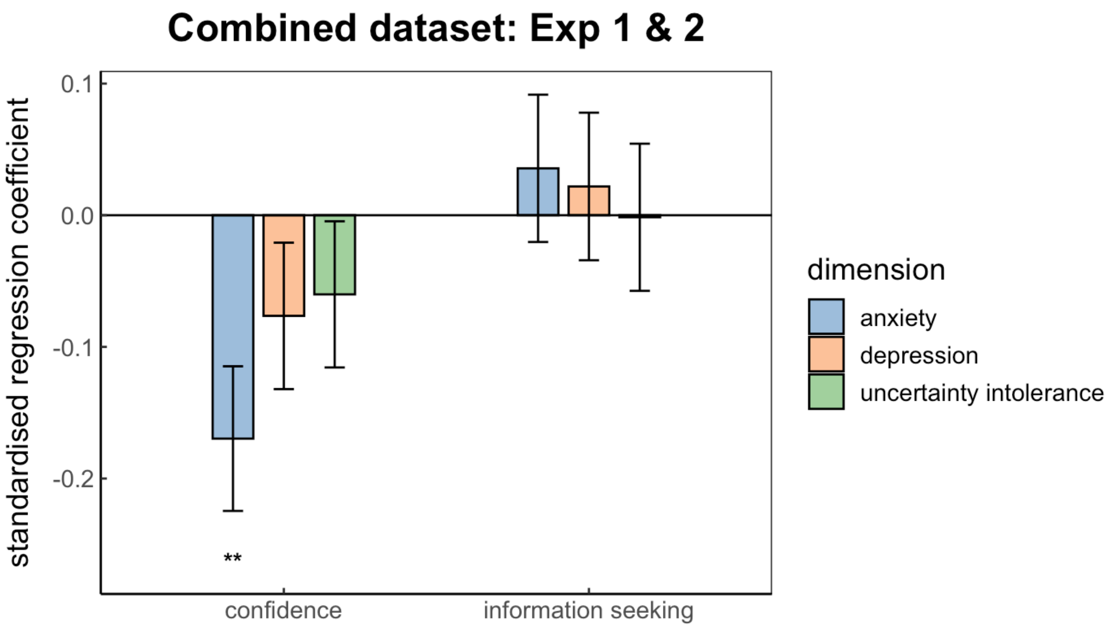


Figure 3.11. Associations between self-reported psychopathology and confidence and information-seeking (Experiments 1 and 2 combined dataset). Error bars indicate ±1 standard error.

3.3.4 Discussion

The aim of Experiment 2 was to replicate two surprising findings from Experiment 1: [1] that depressive severity and uncertainty intolerance would each significantly predict a smaller decision-evoked confidence boost, and [2] that they would also be associated with less information-seeking on the ISTc.

The results of Experiment 2 failed to replicate either finding. The three psychiatric dimensions we preserved in our battery were not reliably associated with differences in the magnitude of the decision-evoked confidence boost. In addition, while the experiment did find a significant association between depressive severity and information-seeking on the ISTc, the effect was in the opposite direction to that of Experiment 1, with depressive severity predicting greater information-seeking in Experiment 2. Like with depressive severity, the relationship between uncertainty intolerance and information-seeking also flipped in direction in Experiment 2 but failed to reach statistical significance.

When we combined the datasets of Experiments 1 and 2, we saw trait anxiety emerge as the only significant predictor of our first-order and our second-order confidence measures on the ISTc (decision-point confidence bias and decision-point confidence boost, respectively). Higher trait anxiety was associated with an underconfidence bias at the point of decision, and this may be causally linked to the smaller decision-evoked confidence boost in high-trait anxious individuals relative to low-trait anxious individuals. However, we found no significant associations in the combined dataset between trait anxiety, depressive severity, or uncertainty intolerance and information-seeking prior to decision on the ISTc.

3.4 Chapter discussion

This chapter of my thesis looks at whether individual differences along psychiatric dimensions of interest can reliably predict confidence and information-seeking tendencies. My research question was inspired by two ideas based on the previous literature. The idea, following from recent individual differences and psychopathology research, was to explore the relationships between different psychiatric dimensions and disturbances in confidence, and the possibility of a transdiagnostic or latent factor mediating the effects of these psychiatric dimensions on confidence. The second idea, following from cognitive science and decision-making research, was to explore the strong link between confidence and information-seeking in different decision-making contexts. In my first line of experimental work, these ideas came together in the hypothesis that psychiatric dimensions like trait anxiety and uncertainty intolerance might indirectly drive greater information-seeking through reduced confidence. In Experiment 1, I looked at whether confidence and information-seeking measures on the ISTc and DDTi correlate with four dimensions of psychiatric interest: self-reported trait anxiety, depressive severity, uncertainty intolerance, and schizotypy. Experiment 2 attempted to replicate the two key and surprising findings from Experiment 1 with a condensed version of the experimental procedure and a larger sample of participants.

Across the two experiments and two tasks, we saw an overall trend towards reduced confidence as a function of trait anxiety and, more weakly, related dimensions like depressive severity and uncertainty intolerance. This was reflected in various measures of confidence, including first-order confidence measures from both the ISTc (confidence bias at decision point) and DDTi (first-stage decision confidence) as well as second-order measures from the ISTc (decision-evoked confidence boost, evidence-confidence calibration). Our various confidence measures are sensible reflections of decision confidence: across participants, they scale systematically with decision evidence and are robustly predictive of information-seeking, in a manner consistent with previous literature on confidence and information-seeking (Desender et al., 2018).

The associations between our confidence measures and trait anxiety were small and comparable to effect sizes in previous studies. In particular, in the ISTc, a novel task paradigm for investigating confidence, we found a statistically significant association between trait anxiety and confidence in the combined dataset that was slightly larger in size than what Rouault et al. (2018) had previously found with the dot discrimination task. However, these effects were inconsistently found in the individual experiments, most likely due to limited sample sizes, and did not explain much variance in the data. Overall, our findings suggest that, while there is a relationship between trait anxiety and underconfidence in decision-making tasks like the ISTc and the DDTi, the relationship tends to be weak, and the differences in confidence seen on these tasks may not be reliable enough to be clinically significant or predictive.

For information-seeking measures, the relationship with psychiatric dimensions was even less consistent. Previous literature suggests that reduced confidence is strongly predictive of greater information-seeking prior to decision (e.g., Desender et al., 2018). Those with anxiety, depression, and related comorbid disorders, whom we might expect to be systematically less confident, were expected therefore to engage in more information-seeking behaviours; this is the direction of the relationship between psychiatric dimensions and information-seeking we had initially preregistered. However, what we observed was large variability in the direction of the relationship between psychiatric dimensions and information-seeking between the two experiments. More specifically, depressive severity and uncertainty intolerance predicted greater information-seeking on the ISTc in Experiment 2 but less information-seeking on the ISTc in Experiment 1. In the combined dataset, there were no significant relationships between our information-seeking measure and any of the three psychiatric dimensions we tested in both experiments.

It is not entirely clear why there is a lack of consistency in the relationships that depressive severity and uncertainty intolerance have with information-seeking in Experiments 1 and 2. One possibility is that, like what we saw with confidence, the effects could be weak and thus inconsistent at the sample sizes of the individual experiments. Another possibility is that the changes to the design of Experiment 2, which created a shorter experimental procedure, may have led to participants experiencing less subjective time pressure. In Experiment 1, we noticed that participants spent less time on the ISTc when it was completed after the DDTi. This suggests that participants in Experiment 1 might have experienced more time pressure on the ISTc when it followed the DDTi.

Previous research into human decision-making suggests that time pressure is a stressor and an important factor influencing decisions and task performance; however, there are large individual differences in how much time pressure people experience from the same external time constraints and how they regulate their behavioural response to time pressure (Rastegary & Landy, 1993). Some researchers have speculated that, in the face of an uncertain situation and especially under time pressure, decision-makers who are highly intolerant of uncertainty might adopt maladaptive information-seeking strategies — for instance, seeking less information or failing to search for information in an organised or coherent way — in order to reach a decision more quickly and reduce the perception of uncertainty (Janis, 1982; Keinan, 1987; Smock, 1955). It is possible that the design of Experiment 1 inadvertently led to heightened time pressure during the ISTc, and that participants who scored more highly for uncertainty intolerance and depressive severity might have reduced their information-seeking as a way of avoiding or escaping the feeling of uncertainty exacerbated by the heightened time pressure. The unintentional lessening of time pressure through the design changes in Experiment 2 might have allowed those participants with higher uncertainty intolerance and depressive severity to be more comfortable with remaining engaged in the task and seeking more information prior to decision.

In summary, the first experimental line of my thesis investigated the three-way relationship between psychiatric symptom dimensions, subjective confidence, and information-seeking strategy. Experiments 1 and 2 found a small but significant association between trait anxiety and confidence on the ISTc, as seen in confidence bias and the size of the decision-evoked confidence boost, but no significant association with information-seeking or with other psychiatric dimensions. Overall, we did not find individual differences along psychiatric dimensions capable of reliably predicting confidence and information-seeking tendencies or explaining much of the variance in decision process measures.

Thus, following this chapter, our focus shifted from relationships between psychiatric and task performance measures to relationships within task performance measures themselves — specifically, novel and interesting features of confidence and information-seeking on the ISTc. One subsequent line of work explored in more depth how people might reduce uncertainty in different ways, for instance, through feedback-seeking or escape; this line of research is reported in Chapter 4. Another line of experimentation follows up on the decision-evoked confidence boost effect, which was robustly observed in both experiments in this chapter as well as the pilot studies in Chapter 2, and aims to understand the mechanisms behind the differing relationship between P(correct) and confidence before versus at the point of decision. This series of experiments and simulations is reported in Chapter 5.

4
ASSESSING ALTERNATIVE STRATEGIES FOR MANAGING UNCERTAINTY

4.1 Introduction

In Chapter 3, we were surprised to find that psychiatric dimensions like uncertainty intolerance and trait anxiety, which are characterised by and contain items that probe people's fear and distress concerning uncertain situations, were not predictive of information-seeking in either of our tasks. This made us consider more generally what different individuals might do to reduce or manage uncertainty when making decisions. Our interest in these ideas evolved into an investigation in this chapter of different strategies for managing uncertainty in the decision process, as well as the extrinsic or intrinsic conditions that might push people towards favouring one strategy over another.

Uncertainty is an inherent property of our external environment. Features of the environment change over time, other autonomous agents in the environment behave unpredictably, and the information afforded to us about these changes is inconsistently available and frequently ambiguous (Gigerenzer & Gaissmaier, 2015; Kaye et al., 2020). As a decision agent in this uncertain environment, an adaptive approach would be to make choices or otherwise act strategically to reduce the degree of uncertainty and make the external environment more predictable (Mushtaq et al., 2011; Pasquini et al., 2019; Peters, McEwen, & Friston, 2017). This is a fundamental principle underlying a large class of modern learning and decision making theories in neuroscience (predictive coding, cf. Rao & Ballard, 1999), computational psychology (the free energy principle, cf. Friston, 2010), and artificial intelligence (informational gain, cf. Quinlan, 1986). Common to these theories is a Bayesian model of learning and decision making as a process of uncertainty reduction, whereby agents update prior beliefs with newly acquired information to gradually reach a more certain, or at least a more accurate and more precise, model of the external environment (van Es, 2020). Depending on the structure of the task environment or other constraints, this acquisition of information can take the form of actively sampling more information in the environment, seeking the advice of other agents, or even passively waiting for more information to be revealed before making a decision (Gottlieb, 2018; Pasquini et al., 2019).

However, depending on the decision situation, uncertainty reduction via acquiring information before decision may not be the optimal strategy (Gigerenzer & Gaissmaier, 2015). Uncertainty reduction may work better for some types of uncertainty than others: for instance, when the decision context is high in ambiguity (i.e., when the potential outcomes of the choice and the probabilities associated with each outcome are not explicitly known to the decision agent), then the decision agent might learn more about these outcomes and probabilities through feedback about previous choices than by acquiring more information before decision (Arend, 2020). In the ISTc, for example, information-seeking would appear to be an excellent strategy for reducing uncertainty, since the decision space is fixed (25 tiles, therefore a minimum majority threshold of 13 tiles) and the exact probability of a choice being correct is calculable at any point in the information sampling process if the underlying tile distribution is known (Axelsen et al., 2018; Clark et al., 2006). However, because the difficulty of the task (actual number of majority tiles) varies pseudo-randomly in the ISTc, participants are likely to perceive the task as more ambiguous than what the mathematical models of P(correct) would suggest. Thus, participants completing the ISTc may perceive information-seeking prior to decision as a less optimal strategy for reducing uncertainty than we might expect.

The attractiveness of information-seeking as a strategy also depends on the nature of the information that can be acquired. This might be the quality of that information (i.e., its accuracy and precision), the value of that information (in terms of the amount of uncertainty reduced and the increase in probability of gain), and the cost associated with acquiring the information (e.g., time, reward value, or missed opportunities). Relative to the possibility of further information acquisition, the agent may have access to more favourable alternatives. For instance, if an agent needs to make a series of decisions under ambiguity, the agent may choose to save time and learn more about the underlying contingencies of the situation directly through feedback instead of through information seeking. Alternatively, an agent may choose to accept a small but certain reward option in lieu of making a high-uncertainty decision, as observed in the opt-out paradigms used in studies of uncertainty monitoring in other species (Smith et al., 1995; 1997; Hampton, 2001).

So far we have assumed that agents are fundamentally rational, weighing up the expected rewards and costs of available strategies to select the one optimised for maximal probability of gain or minimised probability of loss. However, information processing, such as relative gain/loss probability estimates and expected gain/loss decision criteria, may be highly influenced by apparently non-rational factors like curiosity (Gottlieb, 2018) or emotion (Fessler, 2001; Hickson & Khemka, 2014). Markiewicz and Kubińska (2015), for instance, found that participants performed differently on two versions of the same decision-making task which promoted affective (hot) versus deliberative (cold) decision-making styles. The authors suggested that affective factors in the decision process may shift agents' focus to certain parameters of the decision situation (e.g., probability of loss/gain) at the expense of others (e.g., magnitudes of losses/gains).

More generally, the stress and discomfort people experience in an uncertain situation may lead them to adopt suboptimal coping strategies, such as denying or ignoring the situation, using heuristics to oversimplify the decision, or escaping or avoiding the situation altogether (Pasquini et al., 2019; Peters, McEwen, & Friston, 2017).

Taken together, these factors may explain why uncertainty intolerance and depressive severity were associated with reduced information seeking in Experiment 1. People who are more intolerant of uncertainty may be motivated to seek less information for two, mutually reinforcing reasons: the increased discomfort associated with and desire to escape from the uncertain situation, and the relative attractiveness of other sources of information that are more predictive (e.g., feedback). Previous studies have found that people who are more intolerant of uncertainty will engage in more feedback seeking behaviours and may regard feedback as a more attractive source of information in ambiguous situations (Bennett et al., 1990). There is also some evidence that people who are more intolerant of uncertainty may prefer to escape an uncertain situation, for instance, by coming to a decision more quickly or choosing a less rewarding option (Luhmann et al., 2011; Smock, 1955). However, the association between these alternative strategies of managing uncertainty and intolerance of uncertainty or related psychiatric dimensions (e.g., trait anxiety, OCD symptom severity) has not been consistently observed in information sampling paradigms like the Beads Task (Morein-Zamir et al., 2020).

In summary, existing literature in learning and decision making largely takes a rationalist approach to uncertainty management and assumes that information seeking prior to decision is the optimal strategy to adopt when making decisions under uncertainty. However, other strategies for managing uncertainty exist, and people may favour these alternative strategies because of the nature of the uncertainty, the task structure, or affect-related individual differences. Investigating these alternative strategies might help us develop measures that are more sensitive to individual differences and more reliably predicted by dimensions of psychiatric interest.

In this chapter, we decided to focus on assessing the plausibility of these alternative strategies. In particular, we investigated two possible uncertainty management strategies that might account for the reduced information-seeking we saw in Experiment 1: feedback-seeking and escaping uncertain situations. Experiment 3 investigates whether some participants might be using a feedback-seeking strategy on the ISTc; we assess whether information seeking behaviour shifts in response to changes in feedback availability. Experiment 4 explores whether escape behaviour can be seen in the ISTc when participants are presented with an explicit opt-out choice and what task factors might promote greater opt-out frequency.

4.2 Experiment 3

4.2.1 Aims & hypotheses

Experiment 3 investigates the plausibility that participants engage in feedback-seeking as an alternative to pre-decisional information-seeking on the ISTc. We speculated that, because feedback followed the decision on every trial, participants might be making earlier decisions to receive feedback instead of gathering more information prior to the decision; this might account for some of the individual differences in information-seeking behaviour observed in our previous experiments. We hypothesised that if some participants are using a feedback-seeking strategy, then the average tiles seen before decision would be higher for trials where post-decisional feedback was not available.

4.2.2 Methods

Pre-registration

The question and hypotheses, methods, data exclusion criteria, measures of interest, and anticipated analyses of Experiment 3 were preregistered with AsPredicted. The preregistration document is not yet publicly available, but a copy of the preregistration document is provided in Appendix S4.1.

Participants & exclusions

Our preregistration specified that we would recruit participants in batches of 10 until the Bayes Factor on the main contrast of interest fell above 3 or below ⅓, up to the maximum sample size of 55 (after exclusions). The upper limit for our sample size was determined from a power analysis intended to detect a two-tailed paired *t*-test result with Cohen's *d* of 0.5 at $\alpha = .05$ and 95% power.

A total of 35 people were recruited using the online recruitment platform Prolific. Participants provided informed consent in accordance with procedures approved by the University of Oxford Central University Research Ethics Committee (R56625/RE003). All participants were paid a base amount of around £7.50/hour plus a variable bonus of up to £0.60.

Participants were prescreened on Prolific by self-reported age (18 to 60 years) and visual acuity (normal or corrected-to-normal vision). All 35 participants in our collected dataset passed these prescreening measures. Based on the exclusion criteria listed in Appendix S4.2, 3 people were excluded from the final analysis, leaving a final dataset size of 32 participants ranging in age from 18 to 60 years ($M = 27$, $SD = 9$). 14 self-reported as female and 18 as male. 3 trials in the final dataset were further removed from analysis, as they met the exclusion criteria for individual trials [Appendix S4.2].

Design & measures

To assess the influence of feedback availability on information-seeking behaviour prior to decision, we modified the ISTc paradigm used in Experiments 1 and 2 by replacing the Decreasing Win (DW) block with a "delayed-feedback" block. The Fixed Win (FW) block from the original experiments was preserved and renamed the "standard-feedback" block.

In the standard-feedback block, immediately after making a decision about the majority colour, participants would see feedback about what the majority colour was, whether they had been correct in their estimate, and the number of points they had earned or lost on the trial. In the new delayed-feedback block, however, participants were not given any feedback after each trial, and were only informed about the total number of points they had earned or lost across all ten trials in the block after the block was completed. The main measure of interest in the task was the average number of tiles that people saw prior to making a final decision in each block.

Procedure

All participants were asked for their demographics (age, gender, handedness) before being given a walkthrough of the task aim, controls, and options available to them. They were then asked to demonstrate their understanding of the task by completing 3 practice trials before continuing onto the two blocks of the ISTc. After the first block, participants were presented with a click-through notice page reminding them of the changed feedback schedule before proceeding onto the second block.

Participants completed in total 20 trials, 10 in the standard-feedback block and 10 in the delayed-feedback block. The order of the two blocks was randomly assigned and counterbalanced between participants. At the end of the two blocks, participants were informed of the total number of points and the equivalent monetary bonus they had earned across the experiment. Participants were also given a chance to give feedback about the experimental design and to report any technical issues before being debriefed. Participants took on average 27 minutes to complete the experiment.

The entire experimental procedure was coded in JavaScript with jsPsych version 6.0.2 (de Leeuw, 2015) and run as a browser-based web application (Chrome and Firefox compatible).

4.2.3 Results

General task performance

Participants generally completed the task in a sensible and strategic manner. The mean score was 1650 points (median: 1800 points) across the 20 trials, but scores were highly left-skewed. Both the mean and median total score exceeded the at-chance mean score of 0 (midpoint of the minimum possible score of -2000 and the maximum possible score of 2000), suggesting that participants were performing the task strategically. This is corroborated by generally high trial accuracy (M = .91, SD = .12) and lengthy trial RTs (M = 42.2 s, SD = 24.1 s).

When comparing performance between the standard-feedback and delayed-feedback blocks, we found that they were generally similar. Participants did not differ in their accuracy [M_{standard} = .92 vs. M_{delayed} = .91, $t < 1$], RT [M_{standard} = 43.0 s vs. M_{delayed} = 41.4 s, $t < 1$], or confidence bias (i.e., reported confidence minus P(correct)) at decision-point [M_{standard} = -3.75 vs. M_{delayed} = -3.22, $t < 1$] between the two blocks. As a preregistered secondary analysis, we also explored whether there were differences in the magnitude of the decision-evoked confidence boost effect between the two blocks. We calculated the magnitude of the confidence boost effect by fitting each participants' data from standard-feedback and delayed-feedback trials separately to the regression model *confidence* ~ *pcorrect* + *decision status*, thereby obtaining standardised betas of pcorrect and decision status for the two types of trials (shown in Figure 4.1).

Similar to findings from our previous experiments, a decision-evoked confidence boost was seen after adjusting for P(correct) in both the standard-feedback block (M = 17.49, SD = 10.82) and the delayed-feedback block (M = 15.57, SD = 8.82). We ran a paired Bayes t-test to compare the magnitude of the decision status betas in the two conditions. The estimated BF_{01} of 3.23 indicated some evidence for the null hypothesis that the two decision status betas came from the same distribution.

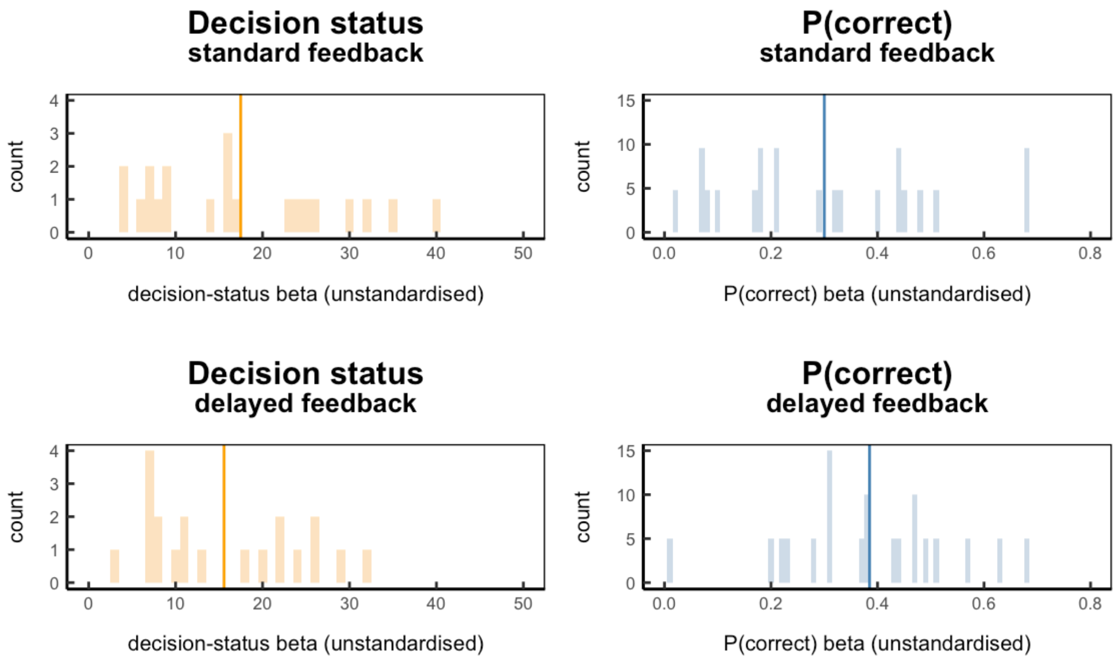


Figure 4.1. Distributions of pcorrect and decision status betas in Experiment 3.

Reducing uncertainty via information-seeking vs. feedback-seeking

Our hypothesis was that, if some participants use a feedback-seeking strategy as an alternative to information-seeking, then we would observe more tiles seen before decision in the delayed-feedback condition (where post-decisional feedback was not available) than in the standard-feedback condition. To test our hypothesis, we performed a paired Bayes t-test to compare the mean number of tiles seen in the standard-feedback block with the number seen in the delayed-feedback block.

As shown in Figure 4.2, there was no significant difference in the mean number of tiles seen before decision between the two blocks. The estimated BF_{01} of 4.95 indicated some evidence for the null hypothesis that the total number of tiles seen in standard-feedback (M = 18.12, SD = 6.26) versus delayed-feedback trials (M = 18.32, SD = 6.51) came from the same distribution.

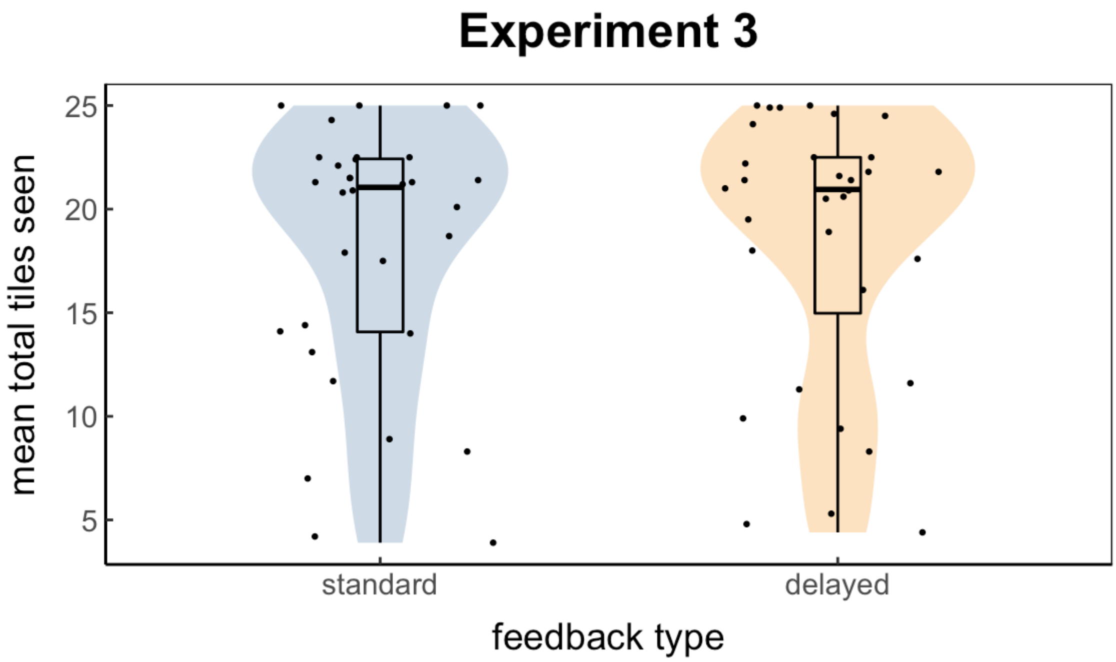


Figure 4.2. Total number of tiles seen before decision in delayed- vs. standard-feedback trials in Experiment 3.

As one of our preregistered manipulation checks, we also looked at whether there were block order effects on the total number of tiles seen. A 2 (feedback type: standard, delayed) × 2 (block order: standard - delayed, delayed - standard) mixed-model ANOVA found no significant main effects of feedback type [$F < 1$] or block order [$F < 1$], and no significant interaction between the two variables [$F(1, 30) = 2.03$, $p = .165$].

4.2.4 Discussion

The broader question that motivated Experiment 3 is whether people might manage uncertainty in different ways. Inspired by our finding in Experiment 1 that some participants were seeking less information prior to decision on the ISTc, we wanted to see if we could find evidence of participants using post-decisional feedback as a source of information about the ISTc instead of seeing tiles before decision. We assessed our primary hypothesis by manipulating the availability of that post-decisional feedback. We expected our participants to seek more information prior to decision when feedback about their performance was withheld compared to when they received immediate post-decisional feedback.

Our results did not support our primary hypothesis. Participants saw on average very similar numbers of tiles prior to decision in the standard-feedback and delayed-feedback trials. On delayed-feedback trials, participants performed sensibly and exhibited a similar relationship between confidence and information-seeking, as evidence by the decision-evoked confidence boost, to that on standard-feedback trials. Thus, the experiment did not find evidence that participants employ a feedback-seeking strategy on the ISTc (ending a trial early to see post-trial feedback).

Experiment 3 appears to discount one of our two overarching hypotheses about the drivers of individual differences in information-seeking. Given the unlikeliness that participants are motivated to end trials early to get feedback about the task, Experiment 4 assessed our other interpretation, that reduced information-seeking might be attributable to a desire to escape and avoid prolonging discomfort in an uncertain situation. The empirical question that follows is whether people would choose to escape an uncertain decision if they could opt out of it for a reduced reward instead of persevering with information sampling and decision. In the subsequent exploratory experiment, we investigated whether participants would make use of an explicit opt-out choice on the ISTc.

4.3 Experiment 4

4.3.1 Aims & hypotheses

Experiment 4 consisted of a series of four pilot studies (Studies 4A to 4D) that used variations of the ISTc to establish the frequency of explicit opt-out behaviour. The aim was to determine task factors, such as the cost of information sampling or the reward associated with the opt-out choice, that might influence opt-out frequency. This pilot series was intended as an initial proof-of-concept for our hypothesis that a desire to escape uncertainty on the ISTc might have partially contributed to the relationship between uncertainty intolerance and information seeking observed in Experiment 1. Due to the iterative nature of the task design changes between the studies, all four are presented and discussed together in this section for brevity.

The two key questions we addressed in Experiment 4 were:

1. how frequently participants in a simplified information sampling task paradigm would select an explicit opt-out choice that came with a reduced reward, rather than continuing with information-seeking, and;
2. whether differences in score rule (explicit information sampling cost), the amount of time needed for the task (increased implicit information sampling cost), or the value of the opt-out choice reward would systematically shift the opt-out frequency.

Studies 4A and 4B looked at whether participants would make use of an explicit opt-out choice on the information sampling task. Studies 4C and 4D explored the effectiveness of two different manipulations (increased opt-out reward and a tile onset delay) intended to increase opt-out frequency.

4.3.2 Methods

Participants & exclusions

A total of 39 participants were recruited using the online recruitment platform Prolific. Participants were recruited in batches of 10 for each pilot study (Study 4A, 4B, and 4D); the exception is Study 4B, where only 9 people were recruited due to a recruitment bug on the Prolific portal. All participants provided informed consent in accordance with procedures approved by the University of Oxford Central University Research Ethics Committee (R56625/RE003). All participants were paid a base amount of around £7.50/hour plus a variable bonus of up to £1.44.

Participants were prescreened on Prolific by self-reported age (18 to 60 years) and visual acuity (normal or corrected-to-normal vision). Some participants unfortunately experienced a bug that randomly mixed up the order of the trials between the calibration and experimental blocks; 2 participants (Study 4B: 1; Study 4D: 1) were excluded from final analysis due to these technical issues with our online task. The final dataset thus consisted of 37 participants, ranging in age from 18 to 52 years ($M = 27$, $SD = 9$). 10 self-reported as female and 27 as male. A more detailed breakdown of gender and age distributions for each pilot study is reported in Table 4.1.

Table 4.1
Demographics in Experiment 4, Studies 4A to 4D

	N	Gender (F/M)	Mean age in years (SD)	Age range in years
Study				
4A	10	1/9	30 (9)	19 – 47
4B	8	3/5	25 (12)	18 – 52
4C	10	4/6	27 (7)	19 – 40
4D	9	2/7	26 (9)	19 – 49

Design & measures

Because Experiment 4 was a proof of concept that focused on eliciting opt-out choices, we decided to omit confidence reporting for all studies in Experiment 4 to keep the task designs shorter and simpler for the participants. Thus, all modifications were made to original information sampling task (IST) rather than the ISTc.

In trials where the opt-out choice was available, the IST design was modified by the inclusion of a third button on the bottom row: in these trials, the scoreboard in previous IST/ISTc versions was replaced with a button labelled 'skip', which was enabled after participants saw one tile [Figure 4.3]. Pressing the 'skip' button allowed participants to bypass having to make a choice about the majority colour, rewarded them with a fixed number of points, and moved them on to the next trial.

Study 4A and 4B differed only in the score rule implemented: Study 4A used the Fixed Win score rule, while Study 4B used the Decreasing Win score rule. In both studies, the opt-out choice (selecting the 'skip' button after at least one tile seen) was rewarded with 10 points. Study 4C had the same design as Study 4A (Fixed Win) but increased the value of the opt-out option to 50 points (half the value of a correct answer). Study 4D had the same design as Study 4B (Decreasing Win) but increased the implicit time cost of information sampling by adding a tile onset delay of 1000 ms. In Study 4D, tiles, after being selected, would first turn grey during this delay period before changing to one of the two colours.

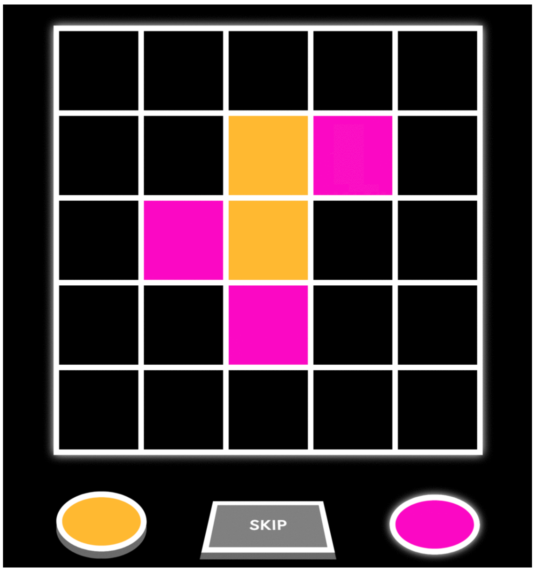


Figure 4.3. IST design in Experiment 4, showing the third "skip" button alongside the coloured response buttons at the bottom of the task window.

Procedure

In each of the four designs, participants were asked for their demographics (age, gender, handedness) before being given a walkthrough of the task aim, controls, and options available to them. They were then asked to demonstrate their understanding of the task by completing a practice trial, before continuing onto a calibration block of standard IST trials without an explicit opt-out choice. After the calibration block, participants were presented with a walkthrough tutorial of the new opt-out ("skip") choice, the associated opt-out reward, and the tile onset delay if applicable, and were again asked to demonstrate their understanding of the changes to the task by completing another practice trial with the skip button enabled. Afterwards, they completed the experimental block. The order of the calibration and experimental blocks was fixed.

Altogether, participants in each group completed 20 trials, 10 in the calibration block and 10 in the experimental block. At the end of the two blocks, participants were informed of the total number of points and the equivalent monetary bonus they had earned across the experiment. Participants were also given a chance to give feedback about the experimental design and to report any technical issues before being debriefed. Across the four designs, participants took on average around 13 minutes to complete the experiment.

The entire experimental procedure was coded in JavaScript with jsPsych version 6.0.2 (de Leeuw, 2015) and run as a browser-based web application (Chrome and Firefox compatible).

4.3.3 Results

General task performance

Participants generally completed the task in a sensible and strategic manner. We observed high trial accuracy (M = .88, SD = .13) and lengthy trial RTs (M = 13.1 s, SD = 7.9 s). Like what we saw in Experiments 1 and 2, the average total number of tiles seen per trial was significantly lower, and mean accuracy was correspondingly lower, in the DW-scored studies (4B and 4D) than the FW-scored studies (4A and 4C).

Table 4.2
Task performance means across participants for Studies 4A to 4D

	Score rule	Total duration (min)	Trial RT (s)	Accuracy	Total tiles seen	Opt-out proportion	Total score	Max score
Study								
4A	FW	17	17.6	.98	21.4	.02	1862	2000
4B	DW	12	9.6	.75	9.5	.11	1778	4800
4C	FW	12	13.4	.96	17.5	.10	1470	2000
4D	DW	12	11.1	.79	9.3	0	1946	4800
mean		*13*	*13.1*	*.88*	*14.9*	*.06*	*1758*	

Opt-out frequency

For each of the four studies, we calculated the proportion of trials skipped in the experimental block of 10 trials for each participant to derive that participant's opt-out proportion. The opt-out proportion across participants for each study is less than .12, as seen in Table 4.2. Figure 4.4 illustrates how the number of trials skipped in the study varies widely between participants. 35 of the 37 participants (95%) across the four studies skipped once or less and 25 participants (68%) never chose to skip a trial. Only 2 participants across the studies (5%) chose to skip half or more of the trials.

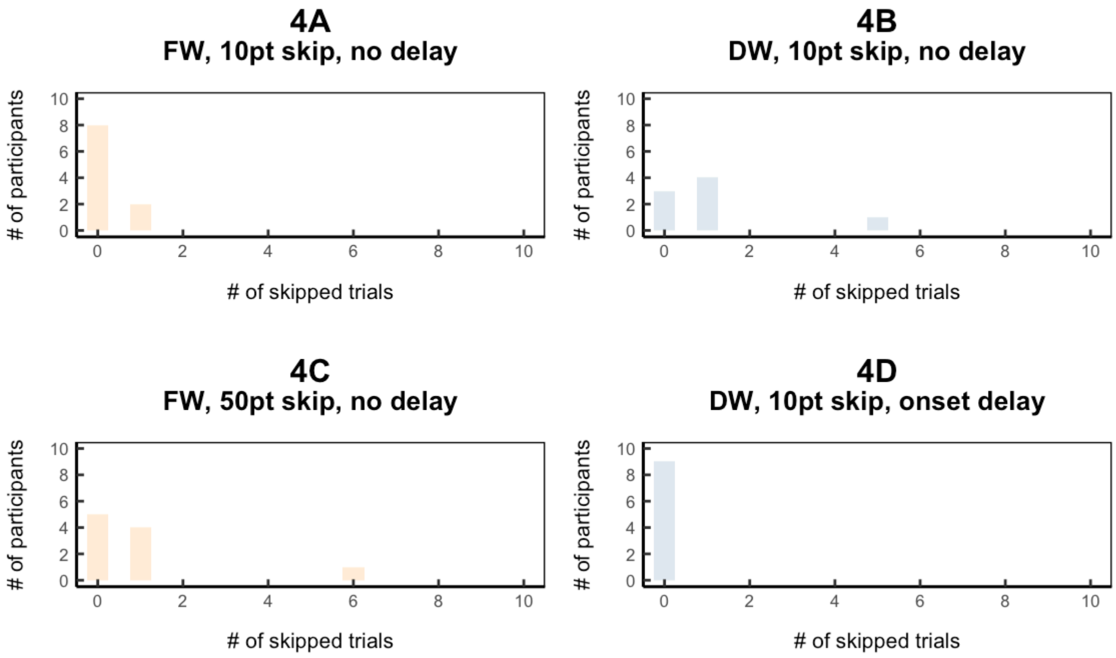


Figure 4.4. Number of trials skipped per participant in Studies 4A to 4D

The initial studies (4A and 4B) were not particularly successful at inducing a normal-like distribution of opt-out frequency, and neither of the manipulations used in studies 4C and 4D were particularly effective at upshifting the opt-out frequency distributions. The addition of the tile onset delay in Study 4D actually seemed to downshift the opt-out frequency as compared to the similarly DW-scored Study 4B. The opt-out proportion for DW-scored studies (M = .05, SD = .12) was very similar to that of the FW-scored studies (M = .06, SD = .14), suggesting that score rule does not affect the opt-out frequency. Altogether, these results give little indication that participants in the IST are inclined to make an explicit opt-out choice.

Patterns in opt-out selection

We were also interested in seeing if there were any trends in when participants chose to opt-out. Because participants across the four studies only chose to opt out of a small proportion of the trials, the following results are presented mainly as observations because of the difficulty of assessing statistical significance in many cases.

Our first hypothesis was that participants might be opting out predominantly on more difficult trials. The IST (and ISTc) features five difficulty levels based on how many tiles of the majority colour are present, from least difficult (18-tile majority / 7-tile minority) to most difficult (13-tile majority, 12-tile minority). As seen in Figure 4.5, we did observe slightly more skipped trials at higher difficulty levels (13- and 14-tile majorities) than at lower difficulty levels (16- and 17-tile majorities).

Another hypothesis was that opt-out choices might tend to be made later in the information sampling phase, perhaps as a means of escaping prolonged uncertainty in trials lacking clear evidence early on in favour of one colour over the other. To assess this, we wanted to look at the number of tiles seen and the P(correct) at the point of choice for the skipped trials versus the decided trials in the experimental block. As in previous experiments, the total number of tiles seen and, by association, the P(correct) differs significantly between Fixed Win and Decreasing Win modes, so we sought to examine studies with different score rules separately.

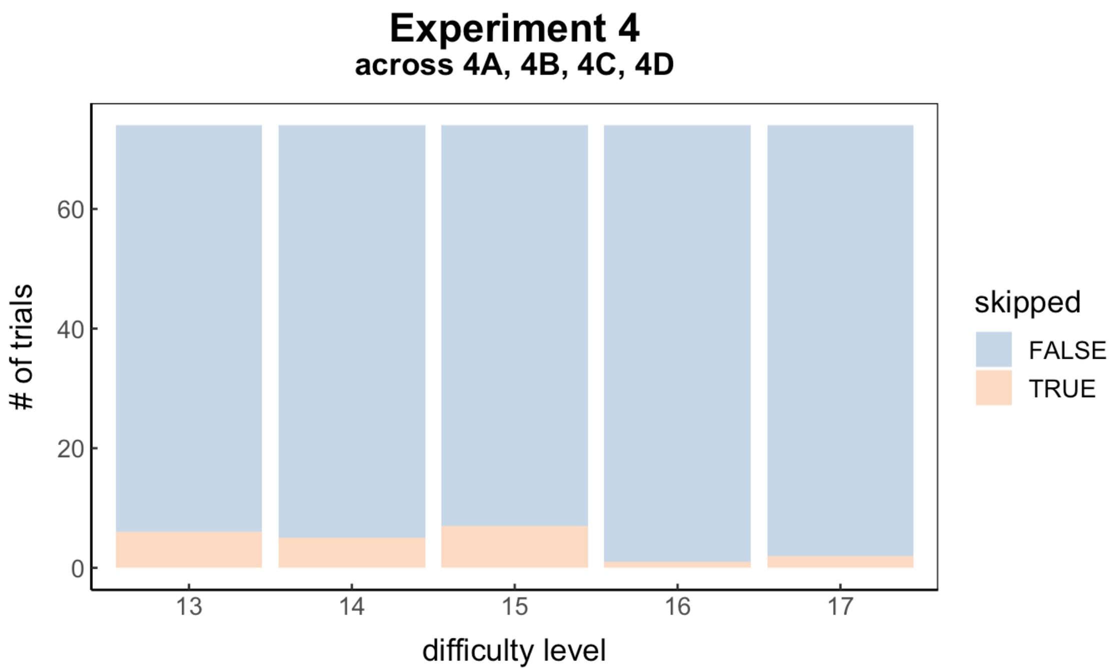


Figure 4.5. Skipped versus non-skipped trials by difficulty level, aggregated across participants in Studies 4A to 4D.

Figure 4.6 compares the total number of tiles seen per trial in each study, grouped according to the three trial types: calibration trials, skipped experimental trials, and decided (unskipped) experimental trials. In the two FW-scored studies (4A and 4C), of the 7 participants who chose to skip the trial at least once during the experimental block, we saw that participants' mean total tiles seen before decision tended to be slightly higher for skipped trials (M = 16.86, SD = 8.32) than for decided trials (M = 14.40, SD = 7.61). For 5 participants who skipped the trial at least once in the DW-scored Study 4B, their average number of tiles seen in skipped trials (M = 13.24, SD = 5.88) was also higher than their average number of tiles seen in decided trials ($M = 8.92$, $SD = 3.17$). These results suggest that participants who chose the skip option tended to do so later in the information sampling phase, and not earlier than when they might otherwise make a decision about the colour.

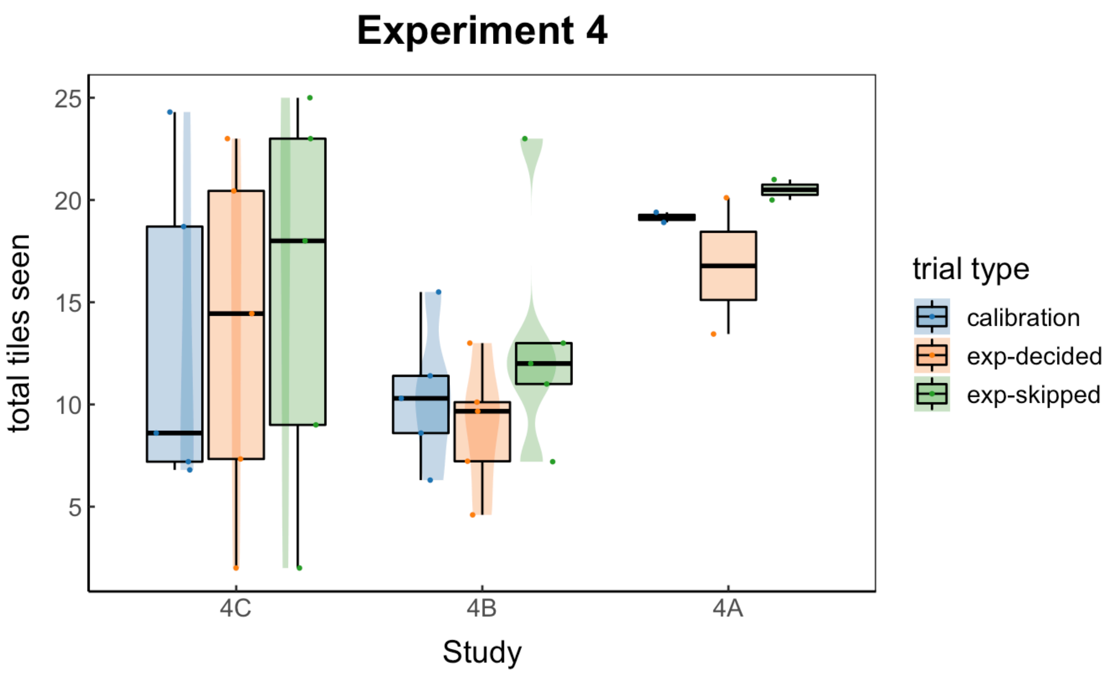


Figure 4.6. Average number of tiles seen per participant by study and trial type. Each dot represents a single participant's average total tiles seen within a particular trial type.

Finally, we looked at the P(correct) of skipped versus non-skipped trials in FW- versus DW-scored studies [Figure 4.7]. In the two FW-scored studies (4A and 4C), of the 7 participants who chose to skip the trial at least once during the experimental block, participants' mean P(correct) at the point of decision tended to be similar for skipped trials (M = 71.50%, SD = 22.33%) and decided trials (M = 64.86%, SD = 16.99%). However, for the 5 participants who skipped the trial at least once in the DW-scored Study 4B, however, their average P(correct) at the point of decision in skipped trials (M = 58.56%, SD = 11.30%) tended to be a little lower than for decided trials (M = 68.75%, SD = 14.60%). Overall, these results suggest that there were no consistent differences in the objective probability of being correct, or evidence strength, when people chose to skip a trial rather than commit to a decision about the majority colour.

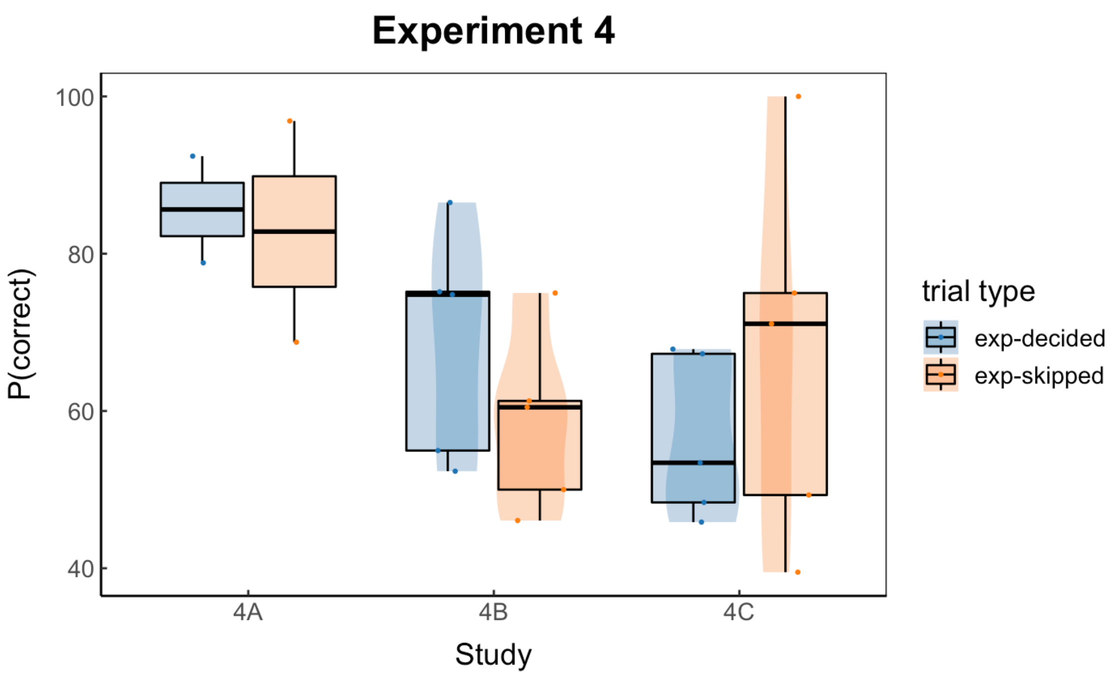


Figure 4.7. Average P(correct) at the point of decision per participant by study and trial type. Each dot represents a single participant's average decision-point P(correct) within a particular trial type.

4.3.4 Discussion

Experiment 4 addressed our alternative hypothesis about reduced information-seeking: that it reflects a tendency among people who are intolerant of uncertainty to prefer avoiding or escaping uncertain situations. To test this overarching hypothesis, we looked at whether people choose to escape an uncertain decision if offered an explicit and reasonable opt-out choice instead of persevering with information sampling and decision. The intention of this experiment was to afterwards relate opt-out frequency back to psychiatric dimensions like uncertainty intolerance. Through a series of four pilot studies, we attempted to establish a baseline frequency of opt-out choice and to test the effectiveness of several task design manipulations (e.g., score rule, tile onset delay, and reward associated with the opt-out choice) in changing opt-out frequency. We expected that both increased information sampling cost and increased rewardingness of the opt-out choice would upshift the opt-out frequency.

In reality, we did not observe much opt-out choices at all across our four studies; the average percentage of opt-out responses was less than 15%, and the majority of participants never opted out of a trial. Some of the opt-out behaviour in the studies was very much counter-strategic — for instance, choosing the skip button despite seeing all the tiles onscreen in a FW-scored block — and suggests that some of the opt-out choices we observed were due to mechanical error or task misunderstanding (cf. Morein-Zamir et al., 2020). Our manipulations of information-seeking cost and opt-out reward also generally failed to systematically upshift opt-out frequency. The addition of an implicit information sampling cost (i.e., tile onset delay) actually seemed to reduce opt-out frequency. While we were unable to induce systematic changes in opt-out frequency, our results indicate, on a more positive note, that participants are generally engaged by the IST paradigm of this length and do not appear to be quitting or escaping trials early out of fatigue or boredom.

This conclusion is further corroborated by our analysis for the point at which participants choose to opt out. If participants were choosing to skip trials to reduce the amount of time spent on the task, then we would expect much fewer tiles being selected before the skip option was chosen compared to before a decision was made. In reality, we observed similar numbers of total tiles seen in skipped and decided trials across the three studies where people had chosen the skip option, even when there was increased financial incentive for people to opt out earlier and more frequently (4C). There was also no evidence that participants were skipping trials at a consistent level of evidence strength (P(correct)).

Taken together, our results from Experiment 4 provide little evidence that reduced information-seeking can be attributed to the desire to escape an uncertain or ambiguous situation. Individuals' discomfort with uncertainty may not tend to manifest in explicit opt-out choice, at least in the IST paradigm. This may be because the uncertainty in the task does not evoke sufficiently intense feelings of discomfort; participants in our studies generally rate the task paradigm as interesting and engaging. It is likely that moderating factors, such as perceived time pressure, are important in determining whether participants experience higher levels of uncertainty-related distress.

4.4 Chapter discussion

This chapter investigated strategies other than information seeking that people might use to deal with uncertainty when making decisions. This line of work was inspired by the finding in Experiment 1 of a potential association between higher intolerance of uncertainty and reduced information seeking. Previous literature points to two possible mechanisms for reduced information-seeking: people may be engaging in feedback-seeking instead of information-seeking, and they may be preferring to escape making decisions under uncertainty. In this chapter, we tested the plausibility of both explanations in the context of the IST paradigm. In Experiment 3, we manipulated the availability of feedback to see if it would change information-seeking behaviour. In Experiment 4, we added an explicit opt-out choice and manipulated the information sampling costs and opt-out choice reward to explore possible effects of these task factors on opt-out frequency.

Across the two experiments, we did not observe people engaging in alternatives to information-seeking. There was no compelling evidence in Experiment 3 that feedback-seeking was being employed as a competing strategy to information-seeking. In Experiment 4, we observed neither a high baseline opt-out frequency nor a systematic shift in opt-out frequency in response to our manipulations of the information sampling cost or the reward associated with the opt-out choice. Participants did not seem to be opting out as a way of finishing the task more quickly. Overall, we observed high participant engagement with the IST paradigm, with most participants engaging in lots of information-seeking prior to decision, and not much evidence of participants engaging in alternative, time-effective strategies (e.g., feedback-seeking, opting out) even when they are available. With the dearth of evidence for alternative strategies for managing uncertainty in the ISTc, we decided not to pursue this line of work further.

Going back to Experiment 1, then, our initial findings of reduced information-seeking appear more surprising, as they were not replicated in Experiment 2 and failed to be accounted for by the alternative explanations tested in Experiments 3 and 4. As it stands, the most likely explanation for the findings of Experiment 1 — assuming that the observed association between intolerance of uncertainty and depressive symptoms with reduced information seeking was not a false positive — is that the longer experimental design, in which participants had to complete two tasks and four questionnaires, may have led to heightened perception of time pressure and subsequent rushing behaviour (e.g., reduced information seeking) in participants with higher depressive severity and intolerance of uncertainty. This is an idea that is supported by previous decision-making literature, but in subsequent experiments, we have not been able to replicate or manipulate that degree of subjective time pressure because our procedures are shorter and omit key sections of Experiment 1 (i.e., the DDTi and some or all of the psychiatric questionnaires). However, with previous studies like Morein-Zamir et al. (2020) and Markiewicz and Kubińska (2015) reporting inconsistent relationships between psychiatric dimensions of interest and measures from “cold” decision-making task designs, it is possible that the relationship between these psychiatric dimensions and information-seeking may not be found consistently in paradigms like the ISTc.

If information seeking is indeed the primary method that people use to manage uncertainty in the ISTc, one of the questions that follows is what the relationship between information seen (i.e., decision evidence) and subjective confidence is like. Although we intuitively expect a monotonic positive relationship between evidence strength and confidence throughout the decision process, our previous findings with the ISTc in Chapters 2 and 3 indicate that the relationship may be less linear than expected. Chapter 5 investigates the decision-evoked confidence boost effect first reported in the pilot studies and subsequently observed in Experiments 1 to 3, with the aim of understanding why the P(correct)-confidence relationship is different before and at the point of decision.

5
EXAMINING THE DECISION-EVOKED CONFIDENCE BOOST EFFECT

5.1 Introduction

In our previous experiments with the ISTc, something we consistently noticed was that participants reported much higher subjective confidence at the point of decision than during information sampling, even when accounting for differences in the objective likelihood of being correct given the evidence. Moreover, the magnitude of this decision-point confidence boost was quite similar across our pilot studies [Chapter 2] and Experiments 1, 2 [Chapter 3], and 3 [Chapter 4]. The large and consistent nature of this effect prompted us, in the final line of experimental work for the thesis, to think more deeply about the underlying relationship between acquired information, choice, and confidence during the decision process.

One of the fundamental principles of standard theories of decision-making is that evidence drives choice: when the accumulation of evidence causes the internal decision variable to exceed the threshold level needed for a decision in favour of one choice over alternatives, a decision is precipitated and that choice is selected (Pleskac & Busemeyer, 2010; Fleming & Daw, 2017). Confidence, in this class of theories, reflects the strength of the evidence for the chosen option relative to the other choices, and is formalised as either as the decision variable itself (Balsdon et al., 2020; Lee & Pezzulo, 2021) or a continuous variable integrating quantity of evidence and the rate of evidence accumulated at decision-point (Fleming & Daw, 2017; Kiani & Shadlen, 2009; Kiani et al., 2014; Yeung & Summerfield, 2012). A recent human neurophysiological study by Murphy et al. (2015) found a build-up of centroparietal positivity during the evidence accumulation phase, in a manner that is consistent with evidence accumulation models of decision-making.

However, what is less obvious from these standard theories of decision-making is that the act of committing to a choice can itself have an effect on the way that a current choice is evaluated or future decisions are made. Choice commitment can lead to biases in favour of the chosen option in subsequent confidence judgments (positive evidence bias, cf. Navajas et al., 2016; Peters, Thesen, et al., 2017; Samaha & Denison, 2020) and value-based preference judgments (endowment effect and choice-induced preference change, cf. Bem, 1967; Brehm, 1956; Jiwa et al., 2021; Kahneman et al., 1991). Choice commitment can also lead to biases in action — for instance, a confirmation bias in post-decisional information-seeking and evidence accumulation — that can further reinforce the positive evidence bias for confidence (Navajas et al., 2016; Rollwage et al., 2020; Rollwage & Fleming, 2021).

Biases in reported confidence can also come from their explicit and communicative nature or intent. While many papers over the years (cf. Overgaard & Sandberg, 2012) have suggested a dissociation between the internal confidence people use for reaching a decision and the external confidence that they report, only recently have researchers attributed this to the social and communicative role of external confidence, in particular its role in supporting group decision-making by allowing individuals to make their private mental states and information accessible to other decision agents (Shea et al., 2014; Heyes et al., 2020). This makes external confidence more susceptible to socially motivated biases – for instance, reporting greater confidence than one feels in order to seem more credible, competent, or attractive to others (Buunk et al., 2002; Schlenker & Leary, 1982; Whitley & Greenberg, 1986).

Up to now, there has been limited evidence from human behavioural studies to support this dissociation between internal and external confidence. Human behavioural studies usually do not provide insight into how confidence evolves over the evidence accumulation phase, as decision confidence is almost always explicitly reported only at the end of the trial after the decision has been made in a standard binary-choice decision paradigm. The confidence reported by participants in most behavioural studies therefore tends to reflect a judgment about the choice that was made. The ISTc, on the other hand, can give us insight into how subjective confidence changes in response to the evidence accumulated, and resulting changes to the objective P(correct), during the decision process. Using the ISTc, we can investigate differences in confidence before and after decision commitment.

In this chapter, we explore the relationship between evidence strength, choice, and confidence in the decision process. In particular, we will examine the relationship between P(correct) and confidence in the ISTc, and investigate the mechanisms underlying the decision-evoked confidence boost effect. Our primary question is whether the act of making a decision is responsible for evoking the boost in confidence at decision point on the ISTc. We start with models that test (and provide evidence against) the hypothesis that the confidence boost effect is merely the result of a selection bias in confidence reporting. Then, in Experiment 5, we look at whether making a decision, even when forced to do so, evokes a confidence boost. In Experiment 6, we take that idea further and investigate whether forcing a decision at a much lower level of evidence strength than what participants would normally need to commit to a choice can still evoke a confidence boost and have lasting effects beyond the initial decision.

5.2 Simulations

5.2.1 Aims & hypotheses

One of the hypotheses that we had about the decision-point confidence boost in Experiments 1 and 2 was that the decision-point confidence boost could reflect a selection bias. Assuming noise in the relationship between accumulated evidence and corresponding decision confidence, decisions would more likely be precipitated when noise increased confidence from accumulated evidence in the direction of the decision threshold. The higher confidence we observed for decision-point confidence might therefore simply reflect instances where noise boosted confidence above the threshold for decision.

In this section, we assessed the selection bias hypothesis for the decision-point confidence boost effect by constructing two models of the relationship between P(correct) and confidence. We then used the models to simulate the expected size of the confidence boost resulting from noise alone. In the first simulation, we used a model of individuals' confidence as a function of P(correct) with a stopping criterion based on individuals' reported decision-point confidence. In the second simulation, we used the same model of individuals' confidence, but with a stopping criterion based on the number of tiles individuals actually saw prior to decision. We then compared the size of the decision-point confidence boost from each of these simulations to that from Experiments 1 and 2.

5.2.2 Simulation 1: basic model

Model specifications

Our basic model of participants' confidence on the ISTc assumes that their internal decision confidence scales linearly but noisily with P(correct). After each tile selection, the evidence accumulated (number of tiles for one colour versus the other) is used to assess subjective confidence that one colour or the other is the majority colour. If that confidence exceeds a threshold value, a decision is made, otherwise further information (tiles) will be sought and the balance of evidence reassessed. We expect a selection bias effect where decisions tend to be precipitated when noise pushes confidence higher. Because of the boost from noise, confidence at the decision point would be expected to be generally higher than confidence prior to decision.

The first simulation used a model with parameters based on empirically observed behaviour in Experiments 1 and 2. Based on the regression *confidence ~ pcorrect* for pre-decision confidence reports described in Chapter 2, we derived two parameters for each participant: [1] the slope of the relationship between P(correct) and confidence, reflecting the individual scaling factor of P(correct) to confidence, and [2] the intercept of this relationship, reflecting the individual confidence bias. Additionally, we calculated the mean decision-point confidence for each participant and used this as the stopping rule (i.e., confidence threshold for decision) for their model.

The model prediction for an individual's confidence C in choice ϑ can thus be given by:

[Eq. 5.1]

$$\hat{C}(\vartheta) = k\ P(\vartheta\ |E) +\ C_0 +\ e$$

where $P(\vartheta\ |E)$ is the probability of the choice ϑ given the evidence E available, k is the individual's scaling factor, C_0 is the individual's confidence reporting bias, and e is the noise in the relationship between confidence and P(correct). The noise parameter e was sampled from a Gaussian distribution with a mean of 0 and a standard deviation equal to the standard deviation of residuals from each participant's *confidence ~ pcorrect + decision status* regression for pre-decisional confidence reports.

We assumed that this internal decision confidence variable $\hat{C}(\vartheta)$ would build until a threshold value, at which point the model would output a decision about the majority colour. The stopping criterion for this model was assumed to be an internal confidence threshold that varied between individuals. For Simulation 1, we modelled the threshold value as:

[Eq. 5.2]

$$C_{threshold} = \mu_{C'(\vartheta)} +\ e_{C'(\vartheta)}$$

where $\mu_{C'(\vartheta)}$ is the mean of the individual's actual reported confidence at the point of decision $C'(\vartheta)$, and $e_{C'(\vartheta)}$ is a noise parameter sampled from a Gaussian distribution with a mean of 0 and a standard deviation equal to the standard deviation of the individual's reported confidence at the point of decision.

Model assessment

We tested our model on a simulated version of the ISTc in Experiments 1 and 2. The structure of the ISTc in the Experiment 1 and 2 datasets was 20 trials divided into 2 blocks of 10 trials with different scoring rules (Fixed Win or Decreasing Win). Trials varied in difficulty from a 13 majority/12 minority tile split to a 17 majority/8 minority tile split, with 2 trials within each block at each level of difficulty. Replicating this task structure, we generated 20 new sequences of 25 tiles drawn from binomial distributions that matched the distribution of difficulty levels in the original task, and subsequently derived P(correct) sequences for these simulated trials, which we used as the input for our decision confidence model.

To increase the robustness of our model, we ran the simulation on participants in the combined dataset of Experiments 1 and 2 ($N = 333$). However, the decision confidence model was unable to simulate the performance of 20 participants because they did not see enough tiles before decision to yield pre-decision confidence data, leaving 313 participants to be simulated in our final dataset.

Confidence was reported by the model prior to decision, with the same pseudo-random frequency as in the original task structure, as well as at the point of decision. To assess how well the basic model reflected the observed data, we compared the simulated data and previous experimental data on two key metrics: the total number of tiles seen prior to decision and the decision status beta from the regression model *confidence ~ pcorrect + decision status.*

Results

When we simulated the performance of participants, we found a decision-point confidence boost that was significantly greater than zero for both Experiment 1 [$M = 5.23$, $t(124) = 9.36$, $p < .001$] and 2 [$M = 4.32$, $t(187) = 10.69$, $p < .001$]. As seen in Figure 5.1, the distributions of the empirical and simulated decision status betas are distinct, with the distribution of simulated decision status betas strongly right-skewed and narrowly peaked.

However, participants' simulated decision status beta was significantly smaller than their matching empirically observed decision status beta in Experiments 1 ($M = 14.48$, $t(124) = 22.25$, $p < .001$] and 2 [$M = 15.40$, $t(187) = 25.81$, $p < .001$]. Figure 5.2 provides examples of empirical versus simulated P(correct)-confidence regressions that illustrate the variability in the sizes of the decision-point confidence boost effect. The analyses suggest that the current model is underestimating the size of the real decision-evoked confidence boost, indicating that the selection bias hypothesis — as captured by this model — does not adequately explain the confidence boost effect observed in Experiments 1 and 2.

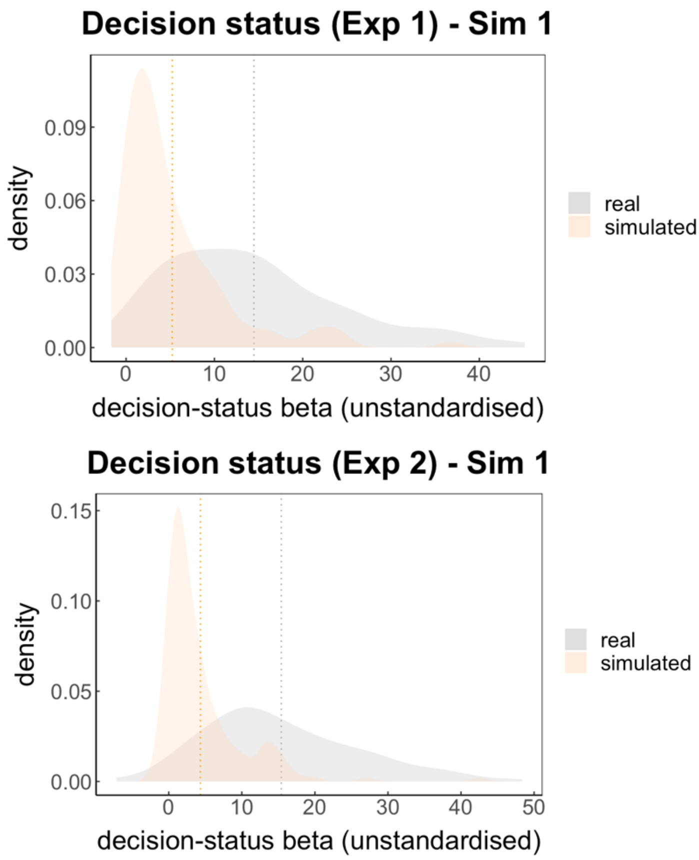


Figure 5.1. Distributions of real versus simulated (Simulation 1) decision status betas in Experiments 1 and 2. Dotted lines indicate the means of the colour-matched distribution.

Furthermore, when we look at the other model metric, total number of tiles seen, we see that the mean number of tiles before decision reported by the simulation is significantly greater than that in both Experiment 1 [$M_{real} = 13.61$ vs. $M_{sim} = 18.52$, $t(124) = 13.80$, $p < .001$] and Experiment 2 [$M_{real} = 15.09$ vs.

$M_{sim} = 19.88$, $t(187) = 16.05$, $p < .001$]. Many participant simulations failed to stop before the last tile, indicating that the model required much more evidence to reach the same level of confidence reported at the point of decision in our original experiments, and that the decision-point confidence in the original experiments was much higher than what could be predicted from noisy variability in the data.

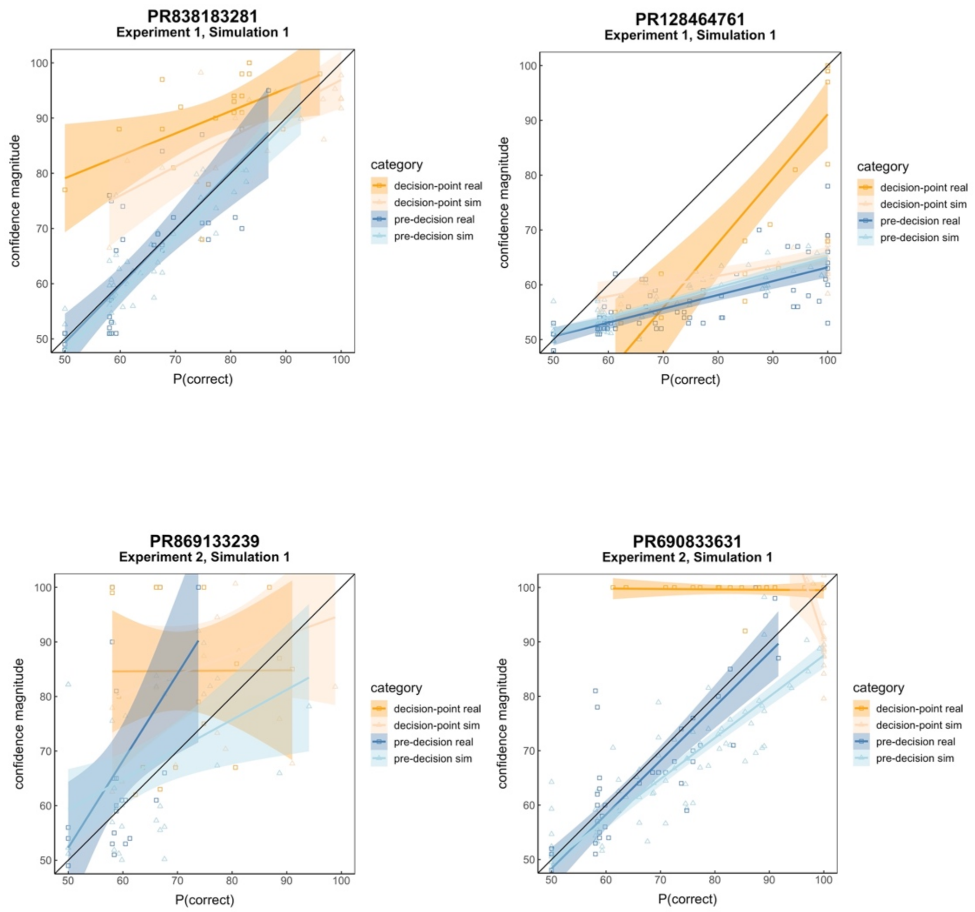


Figure 5.2. Real versus simulated P(correct)-confidence correlations of example participants in Simulation 1. Solid diagonal line represents perfect correspondence between P(correct) and confidence.

Taken together, these results suggest that the basic model fails to informatively simulate how participants in Experiments 1 and 2 might be inferring confidence from P(correct) during the decision process. We hypothesise that the model might be failing because the confidence threshold for decision in this model, which we set as reported confidence at the point of decision in the experimental data, might not be the threshold value of confidence that participants are using internally to decide when to commit to a choice — perhaps reflecting the dissociation between publicly reported confidence and private internal confidence that has been previously proposed (cf. Bang et al., 2020; Massoni et al., 2014; Overgaard & Sandberg, 2012). In our second simulation, therefore, we revised the model to infer an individual's possible confidence threshold for decision from empirical data about the total number of tiles they saw prior to decision. We then use this inferred decision threshold value as the stopping criteria for our confidence model when simulating the decision-point confidence boost.

5.2.3 Simulation 2: inferred criterion model

Model specifications

In the revised simulation, we retained the P(correct) to confidence model [Eq. 5.1] from Simulation 1. However, because Simulation 1 frequently failed to reach the confidence threshold for decision — essentially running out of tiles before confidence reaches the empirically determined decision threshold [Eq. 5.2]

— we decided to revise the stopping criterion so that the model would come to a decision at a similar point to what we had observed empirically.

For each possible integer value of the stopping criterion $\hat{C}_{thresh}$ (51 to 100 inclusive), we ran the confidence model described in Simulation 1 [Eq. 5.1] on the real P(correct) sequences from trials that the participant had completed in Experiment 1 or 2. For each trial (i.e., P(correct) sequence), the confidence model was run 10 times, and a simulated number of tiles seen for that trial at that stopping criterion $\hat{C}_{thresh}$ was derived by averaging across the 10 runs of the 10 trials. By comparing this simulated mean number of tiles seen to the real mean number of tiles seen in the experimental dataset, we were able to infer which $\hat{C}_{thresh}$ produced values that were the closest match to the empirical data. We then set the stopping criterion for that participant to that $\hat{C}_{thresh}$ value.

Model assessment

Using these new inferred decision threshold values, we reran the same decision confidence model with the same simulated task structure and with the same metrics of interest as described in Simulation 1.

Results

Simulation 2 was generally more successful in modelling when participants made decisions as compared to Simulation 1, but was not a perfect fit for the empirical data: the total number of tiles seen by the model was still greater than in the original datasets of both Experiment 1 [M_{sim} = 14.61 vs. M_{real} = 13.82, $t(119) = 5.33$, $p < .001$] and Experiment 2 [M_{sim} = 16.32 vs. M_{real} = 15.41, $t(177) = 7.20$, $p < .001$]. The difference between the simulated and real data likely reflects the way in which the stopping criterion in our model works: since confidence needs to meet or exceed the threshold value to reach a decision in our model, evidence strength would necessarily be higher than the participant's mean in the original dataset.

When we simulated the performance of participants, we found a decision-point confidence boost that was significantly greater than zero for both Experiment 1 [$M = 7.74$, $t(119) = 15.13$, $p < .001$] and 2 [$M = 6.93$, $t(177) = 18.67$, $p < .001$]. As seen in Figure 5.3, the distributions of the simulated decision status betas from both experiments are still very strongly right-skewed, narrowly peaked, and distinct from the empirical decision status distributions.

Although larger than what we saw in Simulation 1, participants' simulated decision status beta in Simulation 2 was still significantly smaller than (approximately half the size of) their empirically observed decision status beta in Experiments 1 ($M = 14.81$, $t(119) = 15.48$, $p < .001$] and 2 [$M = 16.08$, $t(177) = 22.42$, $p < .001$]. Figure 5.4 provides examples of empirical versus simulated P(correct)-confidence regressions that illustrate the variability in the sizes of the decision-point confidence boost effect.

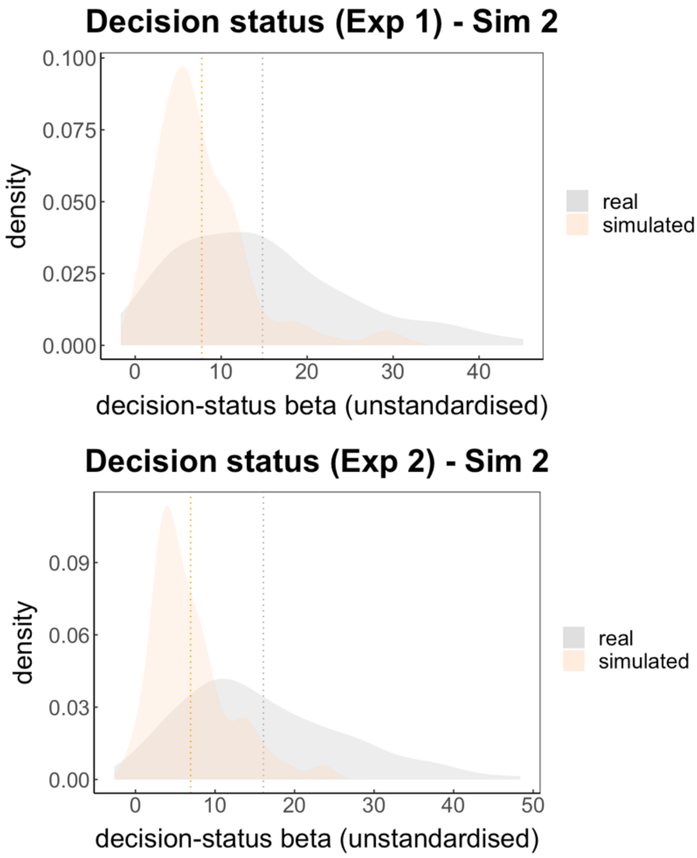


Figure 5.3. Distributions of real versus simulated (Simulation 2) decision status betas in Experiments 1 and 2. Dotted lines indicate the means of the colour-matched distribution.

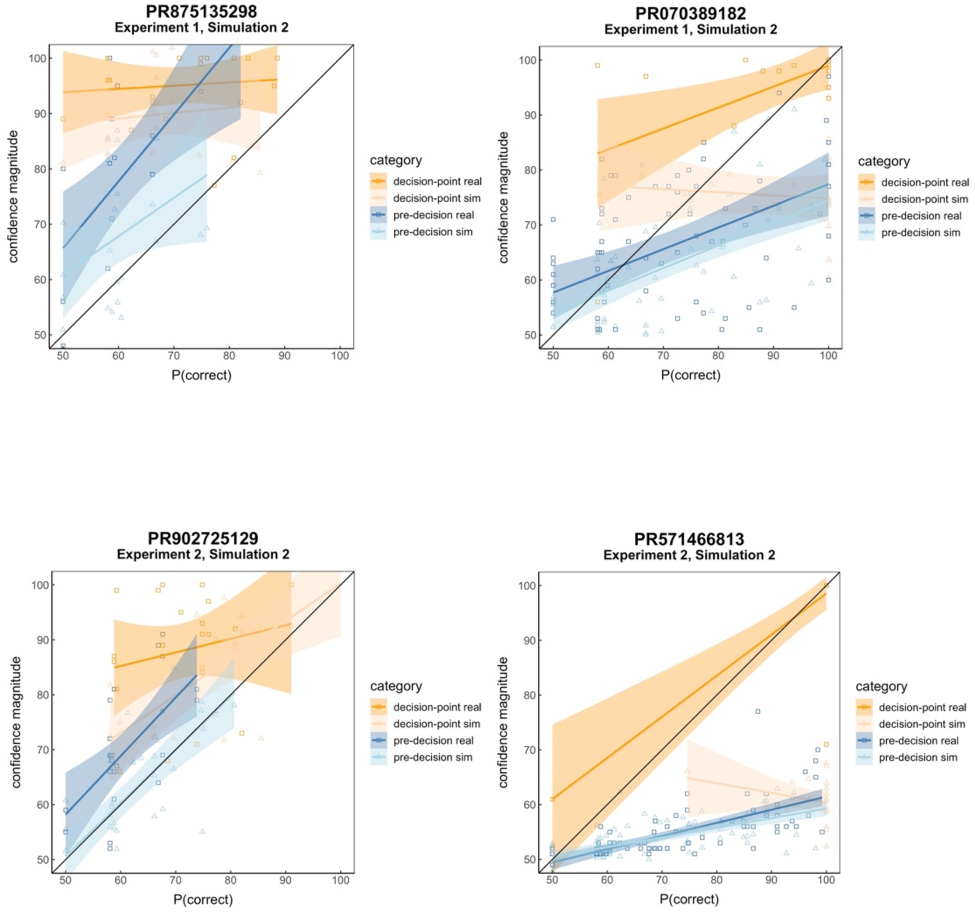


Figure 5.4. Real versus simulated P(correct)-confidence correlations for Simulation 2. Solid diagonal line represents perfect correspondence between P(correct) and confidence.

5.2.4 Discussion

The aim of the two simulations described in this section was to test the simplest explanation for the decision-point confidence boost effect: that decision-point confidence reports overwhelmingly captured instances where random noise was in the direction of the confidence threshold and had "boosted" the underlying confidence signal over the confidence threshold for decision. Using a basic decision confidence model that was fitted to each participant's P(correct)-confidence relationship parameters and a stopping criterion that was based on either their decision-point confidence reports (Simulation 1) or inferred from their total number of tiles seen (Simulation 2) in the empirical data, we were able to simulate the confidence boost that would arise from stochastic noise in the relationship between P(correct) and confidence.

These models of the P(correct)-confidence relationship underlying the simulations were not intended to be detailed and comprehensive models of how the objective probability of being correct was being 'translated' into reported confidence; instead, they were intended to illustrate to what extent the experimentally observed decision-point confidence boost could be explained by noise within a basic linear model of P(correct) to confidence, i.e., the selection bias hypothesis.

In fact, both simulations were poor fits for the actual data; neither was able to closely replicate the shape of the distribution or the average magnitude of the decision-point confidence boost observed in the empirical data for Experiments 1 and 2. The results of our simulations suggest that the selection bias hypothesis is not a sufficient explanation for the confidence boost effect we observed in the previous experiments. If the magnitude of the empirically observed confidence boost at decision-point cannot be entirely accounted for by noise in the P(correct)-confidence relationship, then the confidence boost may be driven by something that is specific to making a decision, for instance, choice commitment. A basic prediction that might follow from this is that a confidence boost would be seen whenever an individual commits to a choice, voluntarily or otherwise. In Experiment 5, therefore, we investigated whether the confidence boost effect would remain and how the effect size would change if we forced people to commit to a choice on the ISTc.

5.3 Experiment 5

5.3.1 Aims & hypotheses

In Experiment 5, we wanted to get some insight into whether the act of committing to a choice was partly responsible for the large decision-point confidence boost effect observed in our previous experiments. Our preregistered hypothesis was that we would observe the confidence boost both when the decision was endogenously motivated (i.e., when participants were free to commit to a choice whenever they felt ready) and when it was exogenously motivated (i.e., when participants were forced to make a choice at a point in time decided by the experimental programme). We were also interested in exploring the effect of this decision agency on the magnitude of the boost, but did not have a specified hypothesis as to the direction of this relationship.

5.3.2 Methods

Pre-registration

The question and hypotheses, methods, data exclusion criteria, measures of interest, and anticipated analyses of Experiment 5 were preregistered with the Open Science Framework (OSF). The preregistration document is not yet publicly available, but a copy of the preregistration document is provided in Appendix S5.1.

Participants & exclusions

Our intended sample size, as specified in our pre-registration, was 30 participants after exclusions. The sample size was specified based on a power analysis to detect a confidence boost (i.e., regression coefficient of *decision status*) with an anticipated Cohen's d of 0.73 — half the effect size of our original results — at $\alpha = .05$ and 95% power.

A total of 40 people were recruited, in batches of 10 until 30 or more participants had passed the exclusion criteria listed in Appendix S5.2, using the online recruitment platform Prolific. Participants provided informed consent in accordance with procedures approved by the University of Oxford Central University Research Ethics Committee (R56625/RE003). All participants were paid a base amount of around £6/hour plus a variable bonus of up to £0.60.

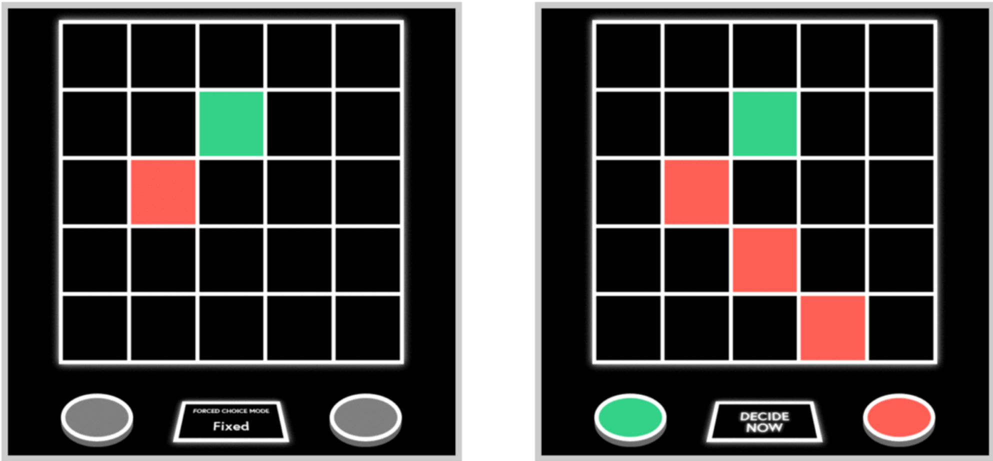
Participants were prescreened on Prolific by self-reported age (18 to 60 years) and visual acuity (normal or corrected-to-normal vision). All 40 participants in our collected dataset passed these prescreening measures. Based on the exclusion criteria listed in Appendix S5.2, 7 people were excluded from the final analysis, leaving a final dataset size of 33 participants ranging in age from 18 to 41 years (M = 26, SD = 6). 16 self-reported as female and 17 as male. 23 trials in the final dataset (0.38% of task total) were further removed from analysis, as they met the exclusion criteria for individual trials.

Design & measures

To assess the influence of decision agency on the presence and magnitude of the decision-evoked confidence boost, we used a within-participants design, comparing ISTc performance — specifically, decision status betas derived from P(correct) and confidence ratings before and at the point of decision — in two types of trials: free-choice versus forced-choice.

Free-choice trials followed the same procedure as in previous experiments: participants were allowed to turn over as many tiles as they liked before committing to a choice about the majority colour. Using on their performance over the block, we were able to calculate the mean and standard deviation of P(correct) at the point of decision, as well as the minimum number of tiles they saw on a trial during the free-choice block. Because the free-choice trials were used as calibration for the subsequent forced-choice trials, we required participants to complete two blocks of free-choice trials: one with a Fixed Win scoring rule and the other with a Decreasing Win scoring rule.

During forced-choice trials, participants were forced to commit to a choice at a certain point in the trial (stopping point). The stopping point was determined separately for each participant according to their performance on the free-choice block: it was drawn from a random normal distribution with mean and standard deviation equal to the participant's decision-point P(correct) values in the previous free-choice block with the matching score rule (Fixed or Decreasing Win). After each tile selection, the running P(correct) was calculated. As seen in Figure 5.5, participants would be stopped from sampling further tiles and told to commit to a decision about the majority colour (see Figure 5.5) if both [1] the running P(correct) reached or exceeded the stopping point P(correct) and [2] the number of tiles seen at that point met or exceeded the minimum number of tiles seen in the matching free-choice block.



A B

Figure 5.5. ISTc design in Experiment 5 forced-choice trials. (A) Response buttons are greyed out and unselectable during the information sampling phase. Bottom-row message informs the participant of the scoring rule. (B) At the calculated decision-point, response buttons become enabled and further tiles are unselectable. Bottom-row message board informs participants that they are required to decide now.

As in the ISTc task designs used in Experiments 1 and 2, confidence was sampled pseudo-randomly during the pre-decision phase, and was asked again at the point of decision.

Procedure

All participants were asked for their demographics (age, gender, handedness) before being given a walkthrough of the task aim, controls, and options available to them. They were then asked to demonstrate their understanding of the task by completing a practice trial before continuing onto the first free-choice block. After the first block, participants were presented with a click-through notice page reminding them of the changed score rule and reward schedule before proceeding onto the second free-choice block.

Following completion of both free-choice blocks, participants were given a walkthrough of the new forced-choice task format, highlighting changes to the task from the previous blocks. As in the first phase, they were forced to complete a practice trial before continuing onto the forced-choice block, and were presented with a notice page reminding them of the score rule change after the first forced-choice block.

Participants completed 40 trials in total: 20 free-choice trials (10 Fixed Win, 10 Decreasing Win) and 20 forced-choice trials (10 Fixed Win, 10 Decreasing Win). The two free-choice blocks always came before the two forced-choice blocks, but the score rule order (Fixed Win, then Decreasing Win vs. Decreasing Win, then Fixed Win) was randomly assigned, counterbalanced among participants, and consistent between free-choice and forced-choice sections of the experiment. At the end of the experiment, participants were informed of the total number of points and the equivalent monetary bonus they had earned across the experiment. Participants were also given a chance to give feedback about the experimental design and to report any technical issues before being debriefed. Participants took on average around 31 minutes to complete the experiment.

The entire experimental procedure was coded in JavaScript with jsPsych version 6.0.2 (de Leeuw, 2015) and run as a browser-based web application (Chrome and Firefox compatible).

5.3.3 Results

General task performance

Participants generally completed the task in a sensible and strategic manner. The mean score was 3537 points (median: 3520 points) across the 40 trials, with scores roughly normally distributed. Both the mean and median total score significantly exceeded the at-chance mean score of 1400 (midpoint of the minimum possible score of -4000 and the maximum possible score of 6800), suggesting that participants were performing the task strategically.

As seen in previous versions of the ISTc, score rule had a significant effect on several task performance measures. Participants saw more tiles prior to decision [M_{FW} = 17.83 vs. M_{DW} = 13.83, $t(32) = 7.24$, $p < .001$] and took longer [M_{FW} = 33.5 s vs. M_{DW} = 26.1 s, $t(32) = 4.05$, $p < .001$] on Fixed Win trials than on Decreasing Win trials. However, there were no significant differences between Fixed Win and Decreasing Win trials on other measures of task performance, including task accuracy [M_{FW} = .93 vs. M_{DW} = .92, $t(32) = 1.11$, $p = .276$] and confidence bias (i.e., difference between reported confidence and P(correct)) at decision point [M_{FW} = -6.88 vs. M_{DW} = -5.61, $t(32) < 1$].

We also observed some differences between free-choice and forced-choice trials in these general task performance measures. Participants spent similar amounts of time on free-choice and forced-choice trials [$M_{free} = 29.4$ s vs. $M_{forced} = 30.2$ s, $t(32) < 1$], but generally saw slightly more tiles before decision [$M_{forced} = 16.24$ vs. $M_{free} = 15.42$, $t(32) = 3.05$, $p = .005$] and were more accurate [$M_{forced} = .96$ vs. $M_{free} = .89$, $t(32) = 4.39$, $p < .001$] on forced-choice trials than on free-choice trials. It is interesting to note, however, that participants exhibited more of an underconfidence bias in their decision-point confidence reports in forced-choice trials [$M_{forced} = -9.39$ vs. $M_{free} = -3.17$, $t(32) = 4.75$, $p < .001$], despite the greater accuracy and number of tiles seen prior to decision. There were no significant interactions between condition (free- vs. forced-choice) and score rule (FW vs. DW) for any of these task performance measures.

Decision-evoked confidence boost effect

As a preregistered manipulation check, we first replicated our previous finding of a decision-evoked confidence boost in free-choice trials, by fitting each participant's data from free-choice trials to the regression model *confidence ~ pcorrect + decision status*. We found a mean free-choice decision status beta of 14.83, which was statistically greater than the null hypothesis mean of 0 in a one-tailed t-test, $t(32) = 8.20$, $p < .001$. This free-choice decision status beta was similar in magnitude to decision status betas we had observed in free-choice ISTc trials in Experiments 1 and 2.

Our primary hypothesis for this experiment, however, was that this decision-evoked confidence boost would persist in forced-choice trials. After fitting participants' data from forced-choice trials to the same regression model, we found a mean forced-choice decision status beta of 8.19, which was also statistically greater than the null hypothesis mean of 0 in a one-tailed t-test, $t(32) = 6.12$, $p < .001$. As seen in Figure 5.6, participants generally reported a smaller confidence boost in forced-choice[†], $t(32) = 5.20$, $p < .001$. The paired Bayes t-test gave an estimated BF_{10} of 1882 in favour of the alternative hypothesis. This suggests there is very strong evidence that the decision-point regression betas of free-choice and forced-choice trials come from different distributions, and that the forced-choice confidence boost is indeed smaller than the free-choice confidence boost.

[†] This and subsequent analyses in Chapter 5 compare unstandardised betas across regression models. We felt it was appropriate to compare regression betas derived from these experiments without further standardisation, as all regression analyses in Chapter 5 compare *P(correct) magnitude* against *confidence magnitude* (see Chapter 2 for further details of the regression model and how these variables are derived). Both variables share the same scale (50-100) and are derived from raw scales of P(correct) and reported confidence that remain consistent across experiments, and thus effectively standardise the scale for these analyses.

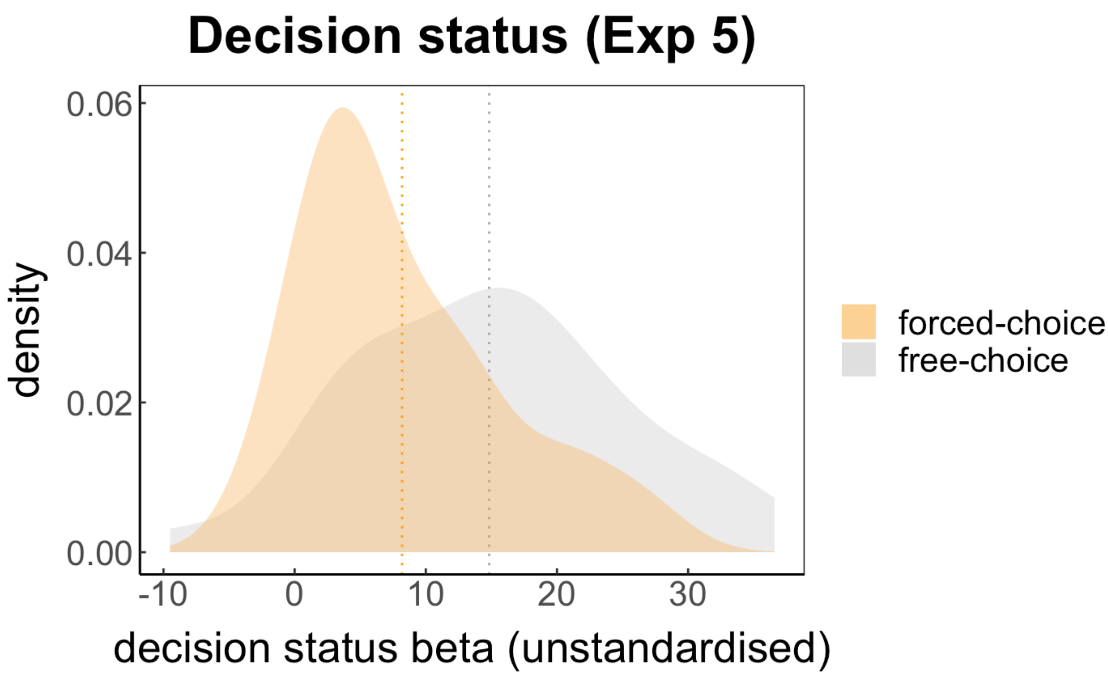


Figure 5.6. Distributions of decision status betas for free-choice vs. forced-choice trials in Experiment 5. Dotted lines indicate the means of the colour-matched distribution.

5.3.4 Discussion

Experiment 5 sought to evaluate how the act of committing to a choice might contribute to the decision-evoked confidence boost effect observed. Specifically, we wanted to test whether the effect would persist when the decision was exogenously motivated and whether the magnitude of the boost in these cases would be equivalent to that seen when the decision was endogenously motivated. We assessed these hypotheses by modifying the ISTc to include trials where participants were forced to commit to a decision at an equivalent P(correct) to when they freely chose to commit to a decision.

We found a strong and statistically significant decision-evoked confidence boost effect for forced-choice trials. However, the magnitude of this effect was smaller than that seen for free-choice trials. In fact, the confidence boost we observed in forced-choice trials was just over half the size of that observed in the free-choice trials, and similar in magnitude to that observed in our simulations.

There are three main implications of these findings. Firstly, the fact that we see a confidence boost effect in forced-choice trials, where participants had no control over when they could commit to a decision, further confirms our conclusion that the empirically observed confidence boost effect is not solely attributable to a selection bias. Secondly, removing participants' agency over when they commit to a choice seems to reduce the size of the confidence boost effect, suggesting that the voluntary nature of choice commitment and the decision agency associated with it may contribute to the confidence boost effect. Thirdly, given the relative effect sizes of the confidence boost in free-choice, forced-choice, and simulated trials, it is possible that the free-choice confidence boost may consist of two separate additive effects: the effect of noise (as reflected in our simulations) and the effect of committing to a choice (as reflected in forced-choice trials).

The questions that remain are whether the effect of choice commitment on confidence is independent of related factors such as evidence strength, and whether it produces a lasting change in decision confidence. Because forced-choice trials in Experiment 5 asked for a choice commitment at an equivalent evidence strength (i.e., P(correct)) to that in free-choice trials, we were unable to determine whether the confidence boost effect in forced-choice trials was being driven by confidence boosts on trials where people would have been endogenously motivated to commit to a choice anyway. In addition, the paradigm fails to clarify whether the confidence boost is limited to the reporting of decision confidence (cf. Bang et al., 2020) or whether it would influence subsequent confidence reports and information-seeking behaviour like previously observed positive evidence and confirmation biases (cf. Navajas et al., 2016). Experiment 6 addressed these questions by testing whether a decision-evoked confidence boost would still be seen when people were forced to commit to an initial choice at a much lower level of evidence strength than they would ordinarily need, and whether we would observe changes in confidence and information-seeking for a subsequent decision.

5.4 Experiment 6

5.4.1 Aims & hypotheses

In Experiment 6, we wanted to understand whether committing to a choice would lead to a confidence boost effect, independent of evidence strength, and whether this increase in confidence reflects a lasting change in decision processes or a more temporary uplift in confidence. Our pre-registered hypotheses were that:

1. even if participants were forced to make a binary choice at an "interim point" with a much lower evidence strength (i.e., P(correct)) than what

they would voluntarily require for making a decision, we would still observe a decision-evoked confidence boost effect for that interim-point decision; and

2. if participants were given the opportunity after the interim point to seek more information before committing to a second, final decision, participants who had been asked to make a binary choice at the interim point will seek less information and express greater subjective confidence than if they had not been asked to commit to a choice at the interim point.

5.4.2 Methods

Pre-registration

The question and hypotheses, methods, data exclusion criteria, measures of interest, and anticipated analyses of Experiment 6 were preregistered with AsPredicted. The preregistration document is not yet publicly available, but a copy of the preregistration document is provided in Appendix S5.3.

Participants & exclusions

Our intended sample size, as specified in our pre-registration, was 50 participants after exclusions. The sample size was specified based on a power analysis to detect a confidence boost (i.e., regression coefficient of *decision status*) with an anticipated Cohen's d of 0.53 — half the effect size of our original results — at $\alpha = .05$ and 95% power.

A total of 68 people were recruited, in batches of 10 until 50 or more participants had passed the exclusion criteria listed in Appendix S5.4, using the online recruitment platform Prolific. Participants provided informed consent in accordance with procedures approved by the University of Oxford Central University Research Ethics Committee (R56625/RE003). All participants were paid a base amount of around £7/hour plus a variable bonus of up to £1.50.

Participants were prescreened on Prolific by self-reported age (18 to 60 years) and visual acuity (normal or corrected-to-normal vision). All 68 participants in our collected dataset passed these prescreening measures. Based on the exclusion criteria listed in Appendix S5.4, 16 people were excluded from the final analysis, leaving a final dataset size of 52 participants ranging in age from 18 to 55 years ($M = 27$, $SD = 8$). 18 self-reported as female and 33 as male, with 1 participant opting out of the gender self-report. 12 trials in the final dataset were further removed from analysis, as they met the exclusion criteria for individual trials.

Design & measures

We used a within-participants design featuring matched tile sequences between blocks. In the initial calibration block, participants were asked to complete a number of free-choice trials. For each of those trials, we extracted the tile sequence (e.g., [1] majority colour, [2] minority colour, [3] majority colour, [4] majority colour, ...) and found the "interim point" (i.e., the halfway point of the total number of tiles seen).

In the subsequent experimental block, each trial consisted of one of these specific tile sequences. We randomised the order of the trials and used new randomised combinations of majority and minority colours so that the same sequences would appear novel to the participants. During each of these trials, if their tile sampling reached the interim point, participants were asked one of two things:

1. In some trials ("interim-decision trials"), participants were asked to make a binary choice at the interim point (Figure 5.7B-C). After committing to this initial decision, they were then allowed to see as many tiles as they liked before committing to a final decision.
2. In other trials ("interim-rating trials"), participants were asked to make a confidence rating at the interim point (Figure 5.7G). For the participant, there was nothing special or different about this rating to distinguish it from other confidence ratings during the trial, but critically for the experimenter, this confidence rating was taken at a point in the tile sequence matched to the interim decision.

Additionally:

3. to incentivise participants to take the interim decision seriously, we interspersed some “catch trials” where the binary choice at the interim point was the final decision and participants were not allowed to see more tiles afterwards (Figure 5.7J-L).

As in previous ISTc designs, confidence was sampled pseudo-randomly during the pre-decision phase. Participants were also asked to make a confidence rating immediately after the final decision, unless the decision (i.e., clicking the response button) had been immediately preceded by a preset confidence rating. The last confidence rating participants made, via either of these scenarios, was deemed the decision-point confidence rating.

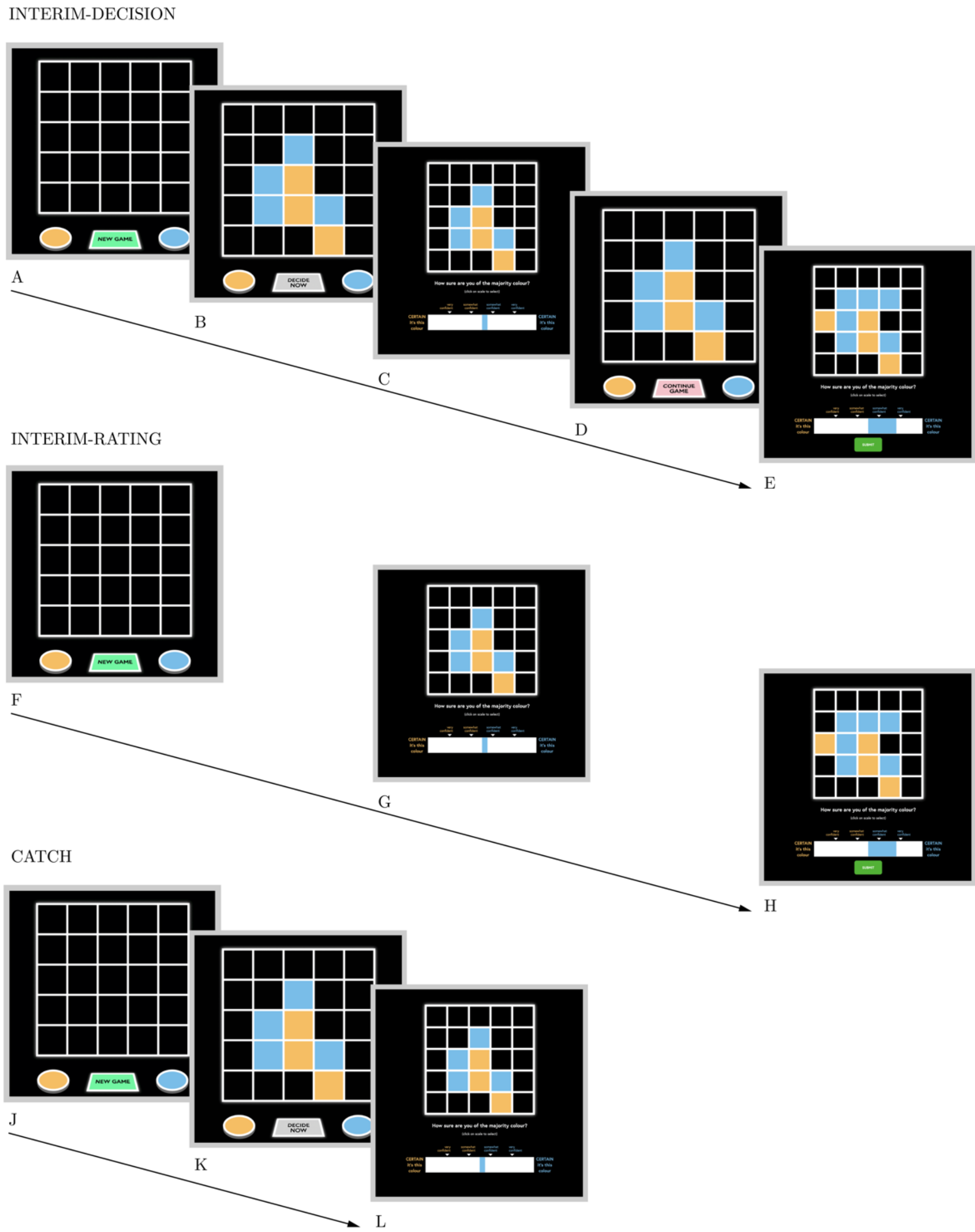


Figure 5.7. Structure of the three types of trials in Experiment 6. In interim-decision trials, participants selected tiles until they reached the interim point (B), where they were asked to select one of the two response buttons and give a confidence rating for this decision (C). Participants were then allowed to continue information sampling (D) before submitting a final decision and associated confidence rating (E). The message board at the bottom of the window informs the participant of what their primary action is ("New Game" in green, "Decide Now" in grey, and "Continue Game" in pink). In interim-rating trials, participants are asked to give confidence ratings during the information sampling phase, including, critically, at the interim point (G) and the point of decision (H). In catch trials, participants are forced to make a decision (K) and give a confidence rating (L) at the interim point, after which the trial ends.

Procedure

At the start of the experiment, participants were asked for their demographics (age, gender, handedness) before being given a walkthrough of the task aim, controls, and options available to them. They were then asked to demonstrate their understanding of the task by completing a practice trial, before continuing onto the initial calibration block, which consisted of 25 free-choice trials.

After completing the calibration block, participants were given a walkthrough of the interim-choice task format, specifically the interim-decision and catch trial types. Participants completed a practice trial before continuing onto the experimental block, which consisted of 10 interim-decision trials, 10 interim-rating trials, and 5 catch trials. Participants thus completed 50 trials in total, all with a Fixed Win scoring rule.

At the end of the experiment, participants were informed of the total number of points and the equivalent monetary bonus they had earned across the experiment. Participants were also given a chance to give feedback about the experimental design and to report any technical issues before being debriefed. Participants took on average around 45 minutes to complete the experiment. The entire experimental procedure was coded in JavaScript with jsPsych version 6.0.2 (de Leeuw, 2015) and run as a browser-based web application (Chrome and Firefox compatible).

5.4.3 Results

General task performance

Participants generally completed the task in a sensible and strategic manner. The mean score was 4077 points (median: 4400 points) across the 50 trials, with the scores distribution significantly left-skewed. Both the mean and median total score significantly exceeded the at-chance mean score of 0 (midpoint of the minimum possible score of -5000 and the maximum possible score of 5000), suggesting that participants were performing the task strategically.

As seen in Table 5.1, there were more similarities than differences between free-choice, interim-decision, and interim-rating trials. Participants were similarly accurate, saw similar numbers of tiles, and had a similar trial RT. Despite an apparently smaller mean underconfidence bias for free-choice trials, a four-way ANOVA found no significant differences in decision-point confidence bias between the trial types, $F < 1$.

Table 5.1
Task performance metrics by trial type in Experiment 5

	Accuracy	Trial RT (s)	Total tiles seen	Decision-point confidence bias	Trial score
Trial type					
interim-catch	.76 (.20)	19.8 (12.0)	9.6 (2.6)	-5.82 (9.15)	52.69 (39.36)
interim-decision	.92 (.14)	35.8 (15.7)	18.3 (5.4)	-5.74 (10.75)	84.23 (28.10)
interim-rating	.92 (.13)	33.7 (15.8)	19.0 (4.9)	-5.36 (8.95)	83.38 (25.67)
free-choice	.93 (.09)	40.0 (17.2)	19.3 (4.2)	-4.22 (7.42)	86.29 (18.23)

Data presented as mean (SD).

Looking for the interim-point confidence boost

As in Experiment 5, we preregistered a replication of our previous finding of a decision-evoked confidence boost in free-choice trials as a manipulation check. After fitting participants' free-choice trial data to the regression model *confidence ~ pcorrect + decision status*, we found a mean free-choice decision status beta of 19.17, which was statistically greater than the null hypothesis mean of 0 in a one-tailed t-test, $t(51) = 13.39$, $p < .001$. The free-choice decision status beta was greater in magnitude to the decision status betas we had observed in similarly structured ISTc trials in Experiments 1, 2, and 5; we suspect this might be due to the increased number of free-choice trials in this experiment and the use of the same Fixed Win scoring rule throughout the experiment, which gave people more opportunity to learn the optimal strategy for maximising reward for this score rule.

Our first hypothesis for Experiment 6 was that we would continue to see this decision-evoked confidence boost effect, even if that decision was prompted at the midpoint of their previous information sampling phase. To test our hypothesis, we focused on the P(correct)-confidence relationship up to and including the interim point for interim-catch, interim-decision, and interim-rating trials. We split these data into two groups: one group contained trials where a decision was made at the interim point (i.e., interim-catch and interim-decision trials), and the other contained trials where a decision was not made at the interim point (i.e., interim-rating trials only). We then fitted the groups (decision vs. rating) separately to the regression model *confidence ~ pcorrect + interim status*, where interim status was a Boolean variable indicating whether the confidence rating was made prior to or at the interim point (the equivalent of decision status in the confidence boost regression models of previous experiments), and compared the interim status betas for the two groups of trials.

In trials where participants had been forced to make a decision at the interim point, we found a mean interim status beta of 3.94 (SD = 6.57), which was statistically greater than zero in a one-tailed t-test, $t(51) = 4.32$, $p < .001$. In trials where participants only made a confidence rating at the interim point, the mean interim status beta (M = 1.25, SD = 3.41) was also statistically greater than zero in a one-tailed t-test, $t(51) = 2.65$, $p = .011$. However, as seen in Figure 5.8, interim status betas for trials where a decision was made at the interim point were larger than for trials where a rating was made at the interim point, $t(51) = 2.94$, $p = .005$. The paired Bayes t-test gave an estimated BF_{10} of 6.90 in favour of the alternative hypothesis. This suggests moderate evidence that the interim-point regression betas of trials where a decision was made and trials where a rating was made come from different distributions.

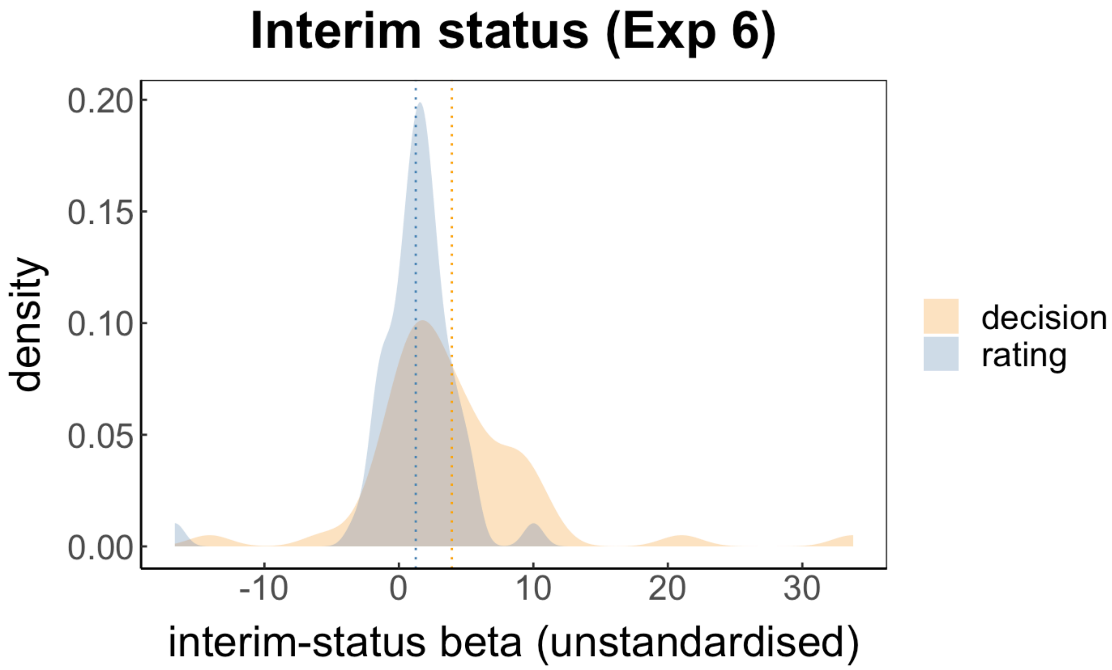


Figure 5.8. Distributions of interim status betas for decision-made vs. rating-made trials in Experiment 6. Dotted lines indicate the means of the colour-matched distributions.

Investigating the post-decisional phase of interim decisions

Our second hypothesis concerned the effects of decision commitment at the interim point on the subsequent information sampling phase for the second, final decision. Because we hypothesised that decision commitment at the interim point would lead to a lasting boost in confidence, we predicted that we would also see greater confidence during information sampling after the interim point and fewer total tiles seen at the final point of decision following an interim-point decision compared to an interim-point rating.

However, as shown in Figure 5.9, there was no clear difference between the total number of tiles, with numerically fewer tiles seen in interim-decision trials (M = 18.33, SD = 5.38) than interim-rating trials (M = 19.00, SD = 4.86), t(51)

= 1.71, p = .094. The paired Bayes t-test of the total number of tiles seen in interim-decision and interim-rating trials gave an estimated BF_{01} of 1.71, indicating weak evidence for the null hypothesis that the mean total tiles seen for interim-decision and interim-rating trials came from the same distribution. A one-way ANOVA found no statistically significant difference in the total number of tiles seen between free-choice, interim-decision, and interim-rating trials, $F(2, 27) < 1$.

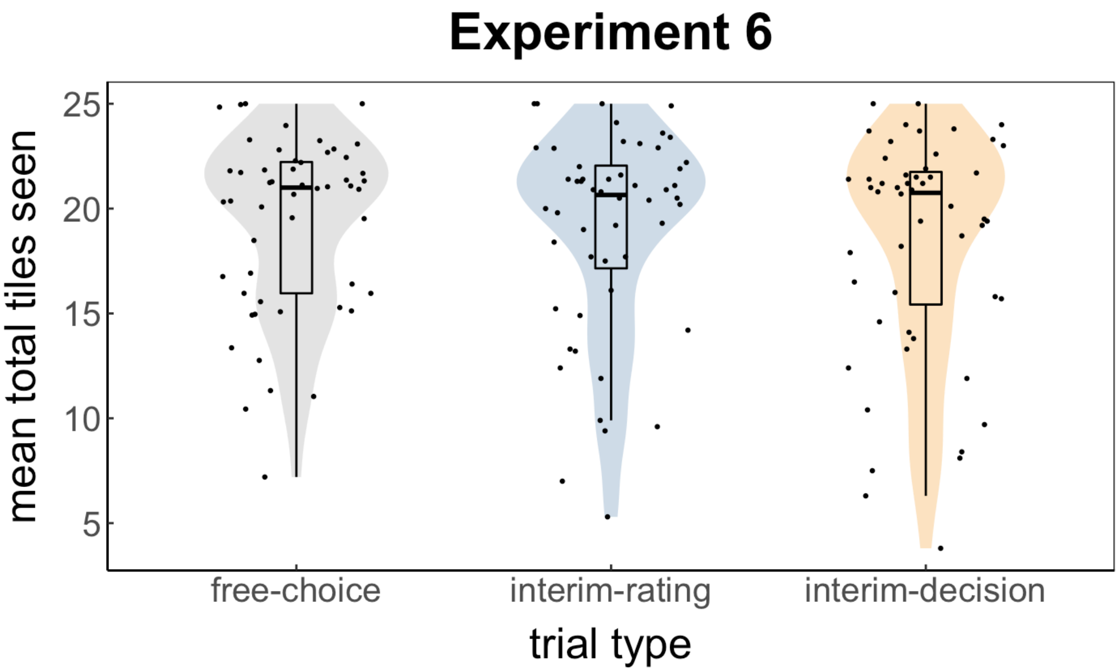


Figure 5.9. Total tiles seen in free-choice, interim-decision, and interim-rating trials in Experiment 6.

To assess the difference in post-decisional confidence between interim-decision and interim-rating trials, we fitted each participant's confidence and P(correct) data, separately for interim-decision and interim-rating trials and excluding any confidence ratings from the interim point or final decision point, to the regression model *confidence ~ pcorrect + timepoint*, where timepoint was a Boolean variable indicating whether the confidence rating was made before or after the interim point. On interim-decision trials, there was a small effect of timepoint (before vs. after the interim point) on confidence ($M = 4.17$, $SD = 5.60$) when accounting for P(correct) ($M = .61$, $SD = .28$), and this was statistically greater than zero, $t(46) = 5.11$, $p < .001$. On interim-rating trials, there was a similarly small effect of timepoint on confidence ($M = 4.04$, $SD = 5.85$) when accounting for P(correct) ($M = .64$, $SD = .26$), and this was also greater than zero, $t(51) = 4.97$, $p < .001$. However, as seen in Figure 5.10, the size of the timepoint beta did not significantly differ between interim-decision and interim-rating trials, $t(46) < 1$. In a paired Bayes t-test of the timepoint betas from interim-decision and interim-rating trials, the estimated BF_{01} of 4.31 indicated some evidence for the null hypothesis that the timepoint betas of interim-decision and interim-rating trials came from the same distribution.

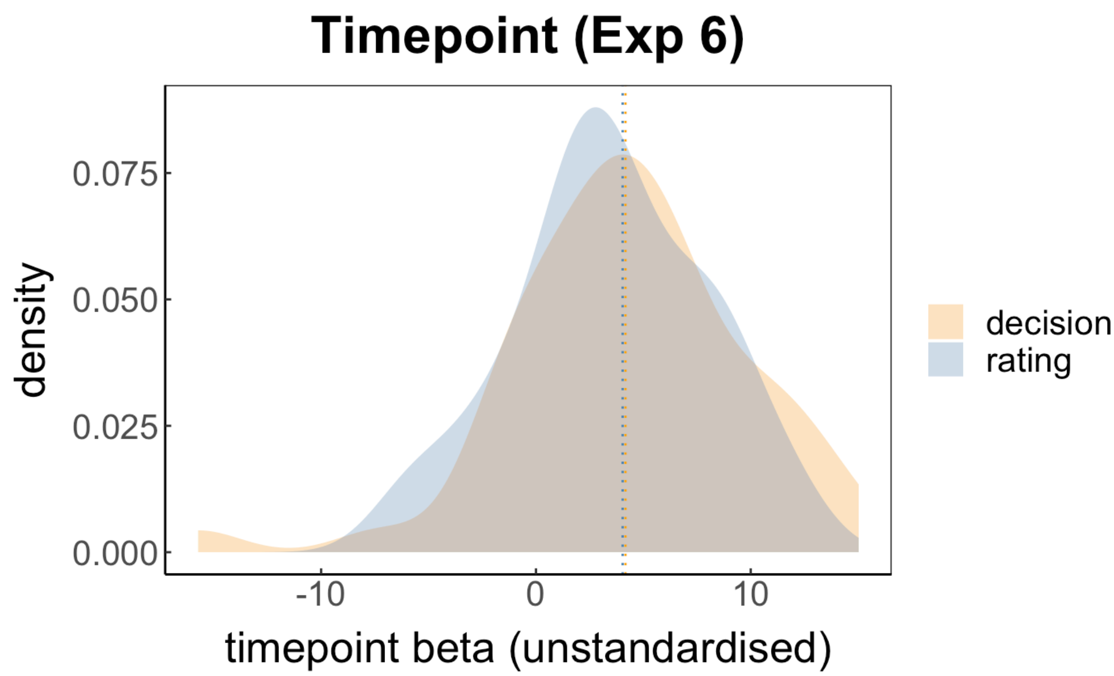


Figure 5.10. Distributions of timepoint betas for decision-made vs. rating-made trials in Experiment 6. Dotted lines indicate the means of the colour-matched distributions.

The relationship between interim choice and final choice

The results from the previous analyses suggest that the interim decision did not have noticeable effects on subsequent confidence and information seeking. In our final set of exploratory analyses, we wanted to explore the relationship between the interim decision and the final decision, which has been excluded from analysis so far.

There were two questions we wanted to address. The first question is whether the decision-evoked confidence boost only happens once in the decision process. Specifically, for interim-decision trials where two decisions were being made per trial, we wanted to see if confidence boost effects would be observed at both the interim point and the final decision point. We found the decision status betas (for the final decision) to be on average much larger (M = 16.99, SD = 13.80) and similar in magnitude to that seen for free-choice trials in the calibration block of this experiment [$t(49) = 1.44$, $p = .158$] as well as for other free-choice trials in previous experiments. In a paired Bayes t-test of the interim status and decision status betas from these interim-decision trials, the estimated BF_{10} was 104820 in favour of the alternative hypothesis, indicating very strong evidence that the interim status and decision status betas came from different distributions.

The second question is whether an explicit choice made at the interim point (i.e., an interim-point decision) would be more predictive of the final decision than an implicit choice at the interim point (i.e., an interim-point confidence rating). To assess this, we compared the mean proportion of matches between interim and final choices in interim-decision trials (M = .77, SD = .13) and interim-rating trials (M = .77, SD = .14) in a paired Bayes t-test. The estimated BF_{01} of 6.37 indicates moderate evidence for the null hypothesis that the mean proportions of matching interim and final choices for interim-decision and interim-rating trials come from the same distribution.

5.4.4 Discussion

This experiment further probed the relationship between decision commitment and confidence. Earlier in Experiment 5, we saw that the decision-point confidence boost persisted even when that decision was externally prompted rather than endogenously motivated. In Experiment 6, we extended this line of thought by testing whether a decision-point confidence boost would still be seen when that decision was both externally prompted and at a much earlier point in the information sampling phase (well below the participant's confidence threshold for decision).

Our results give support to our first hypothesis: we found a small but statistically significant confidence boost effect for decisions made at the interim point. This interim-point effect (the interim status beta) is much smaller in magnitude than the confidence boost effect (decision status beta) previously found in free-choice and forced-choice trials. On trials in which people were able to make both an interim and final decision, the final decision-point confidence boost was larger than the interim-point confidence boost and similar in magnitude to the confidence boost effects found for free-choice trials in this and previous experiments. Thus, although the act of committing to a decision does seem to boost reported confidence, the effect appears to be significantly smaller when it is far away from the typical threshold for decision.

On the other hand, the evidence tended to go against our second hypothesis that the interim decision would lead to changes in subsequent confidence and information seeking. We predicted that participants who had made an interim-point decision would seek less information afterwards. In reality, most people tended to continue seeking information after the interim point when given the option to do so, and the final decision was made after seeing similar numbers of tiles in all trial types (free-choice, interim-rating, and interim-decision). This suggests that participants' natural stopping points for information-seeking seem to reflect relatively stable internal decision thresholds. We also expected to see higher confidence ratings following an interim-point decision than an interim-point confidence rating. However, there were no significant differences in confidence following an interim-point decision versus rating after accounting for P(correct). This suggests that the interim decision and the confidence boost it evokes do not seem to have lasting effects on the continued decision process. The interim decision is also generally not more predictive of final choice than an interim confidence rating. The interim decision does not seem to be given a special weight in subsequent decisions, and when they are free to choose their own decision strategy, people seek information and compute confidence in a way that appears to be independent of the outcome of any previous decision calculations.

Overall, the findings of Experiment 6 support the idea that committing to a decision leads to a boost in confidence, but the size of this confidence boost effect is reduced when a decision is induced below the internal decision threshold. Furthermore, the confidence boost from these early decisions does not seem to influence information-seeking strategy and reported confidence in the subsequent information sampling phase. Interim decisions also fail to be more predictive of the final decision than a confidence rating at the equivalent point. Taken together, these findings suggest that the decision-evoked confidence boost does not have a lasting effect on internal decision confidence. Rather than a positive evidence bias or confirmation bias, which involve changes in the internal confidence governing information-seeking and choice commitment, the decision-evoked confidence boost may reflect a more transient change in externally reported confidence.

5.5 Chapter discussion

In this chapter, we looked more closely into the decision-evoked confidence boost effect observed in previous experiments with the ISTc. Previous literature on decision confidence promotes the idea of a steady graded relationship between evidence and confidence. However, our findings from the ISTc — a behavioural task which enables us to collect confidence judgments during the evidence accumulation phase, in contrast to the standard one-shot perceptual judgment tasks used to study decision confidence — suggest that the act of making a decision strongly biases confidence in favour of the choice made, above and beyond the evidence available. This chapter tested several hypotheses around the idea that choice commitment is responsible for the decision-point confidence boost that we observed. In our simulations, we tested the null hypothesis that the confidence boost was due to a selection bias in the decision-point confidence reports. In Experiment 5, we investigated whether forcing people to commit to a choice, instead of allowing them to make a decision at a point of their own choosing, would evoke a decision-point confidence boost. In Experiment 6, we took this idea further, testing whether a forced-choice commitment and a reduced level of evidence strength could still evoke a decision-point confidence boost, and whether an early boost would influence subsequent confidence and information seeking.

Our simulations ruled out the idea of a noise in the P(correct)-confidence bias sufficiently accounting for the confidence boost, as neither of our models was able to capture the magnitude of or the variability in the confidence boost effect that we had observed in our empirical data. In our experiments, therefore, we looked at other factors that are specific to the final decision in the ISTc, such as explicit choice commitment or the self-determined nature of this decision point. Across Experiments 5 and 6, we found that we were able to evoke a confidence boost even when people were forced to commit to a decision at a much lower level of evidence strength or objective likelihood. However, the magnitude of the confidence boost did seem to be sensitive to the level of evidence strength and the decision agency participants had. In Experiment 5, where choice commitment was exogenously motivated (forced-choice) but at a similar level of evidence strength to endogenously motivated (free-choice) decisions, the confidence boost in the forced-choice condition was observed to be about half the size of the boost in the free-choice condition. In Experiment 6, where choice commitment was both exogenously motivated and at a much lower level of evidence than for endogenously motivated choices, the confidence boost in this interim-point decision was then about half the size of that observed for forced-choice decisions in Experiment 5.

The task design of Experiment 6 also allowed us to address the causal relationship between choice commitment and the confidence boost effect more directly. Because we allowed participants to continue information sampling after making that initial interim-point decision, we were able to see if this initial confidence boost influenced confidence and information seeking in the second information sampling phase, and if it might affect the choice or confidence boost associated with the final decision. We found that the interim-point decision had no significant effects on the subsequent information sampling phase: it did not lead to a noticeable increase in subsequently reported confidence or a decrease in the amount of information sought before the final decision. Most participants in Experiment 6 elected to see very similar numbers of tiles before making a final decision, regardless of whether they made an interim decision or not, suggesting that individuals have distinct but stable decision threshold values. This is consistent with previous findings of individual thresholds for decisions (Ais et al., 2016; Hausmann-Thürig & Läge, 2008). The initial decision did not seem to have any particular bearing on the final decision either: the initial decision was not significantly more predictive of the final decision than a confidence rating at that same interim point, and the magnitude of the final confidence boost remained large and similar in magnitude to the decision-point confidence boost effects observed in previous free-choice conditions. This seems to suggest that the confidence boost is a phenomenon exclusively associated with the act of making a choice.

Taken together, the results of the simulations and experiments in this chapter support the idea that the act of committing to a decision evokes a boost in confidence at the point of decision, but that this boost in confidence fails to have lasting effects on subsequent confidence, information-seeking, or choices. This is in contrast to evidence supporting more long-lasting effects of choice commitment, such as a positive evidence bias in confidence (cf. Navajas et al., 2016; Peters et al., 2017b) or a post-decisional confirmation bias in information-seeking (cf. Schulz et al., 2020; Rollwage & Fleming, 2021). Mechanistically, the lack of a lasting effect of the boost after the interim-point decision in Experiment 6 bears some similarity to the "baseline resetting" effect observed by Parés-Pujolràs et al. (2020) for the P3 signal in human EEG. The P3 signal is thought to encode a build-to-threshold variable that tracks evidence accumulation in real-time during the decision process (Kelly & O'Connell, 2013; O'Connell et al., 2012; Twomey et al., 2015). In a two-decision task structure, Parés-Pujolràs et al. found that the P3 signal decreases dramatically after the initial decision and resumes the second evidence accumulation phase from a starting point close to the first, suggesting that evidence does not simply continue to accumulate after the initial decision for a subsequent related decision.

If the confidence boost does not reflect a lasting change in internal decision confidence, then it is possible that the confidence boost might reflect some sort of reporting bias or difference in communicative intent between pre-decisional and decision-point confidence reports. When reporting confidence in the ISTc, participants could be indicating an estimation of evidence strength related to a choice and/or a more global self-belief or performance estimate regarding this decision (cf. Rouault et al., 2019). There has been some discussion in the confidence literature about the social communicative nature of explicitly reported confidence, as distinct from the private internal confidence used for decision-making and cognitive control (Bang et al., 2020; Heyes et al., 2020). People may be motivated to report higher confidence in their decision, possibly both to reinforce the choice to themselves and to signal to others that they are credible, especially if they are seeking to influence others in a social setting (Pulford et al., 2018; Whitley & Greenberg, 1986; Schlenker & Leary, 1982).

Further research into the decision-evoked confidence boost and the reporting bias hypothesis might directly investigate how a social context — for instance, running the ISTc within a group decision-making or advice paradigm — might enhance the magnitude of the decision-evoked confidence boost. Alternatively, further research may also look at neural mechanisms of this decision-evoked confidence effect to see if the boost in confidence is reflected in measures other than explicit report. This would be complemented by work on the modelling side– for instance, building a more general model of how objective probability of being correct gets ‘translated’ into reported confidence and fitting behavioural and neural data from participants to model parameters – to clarify the nature of the P(correct)-confidence relationship leading up to the decision-point confidence boost.

6
GENERAL DISCUSSION

This chapter concludes the presented work on confidence and information-seeking during decision-making under uncertainty. In the following sections, I will summarise the methods used and the results found, and highlight two common themes that are woven through our experimental motivations and findings. I will then talk about how the present work fits into the broader context of existing literature, as well as remaining questions and future directions that research in this field could take.

6.1 Characterising psychopathologies with metacognitive phenotypes

6.1.1 Motivations for the work

One major theme in the research described in this thesis was an exploration of individual differences in measures of confidence and information-seeking, and their relationship to and potential relevance for understanding psychiatric disorders. Over the past two decades, there has been increasing discontent with the existing nosological systems for psychopathology, such as the DSM-5 and ICD-10, because of their conceptualisation of psychiatric disorders as categorical, distinct, and independent constructs, when in practice psychopathologies tend to be continuous, overlapping, and comorbid (Haslam, 2003; Kessler et al., 2012; Krueger & Eaton, 2015). In that time, there has also been increasing interest among researchers in applying computational approaches, such as network analyses of symptoms or Bayesian modelling of decision processes, to refine the cognitive profiles or phenotypes associated with psychiatric disorders (Borsboom, 2017; Montague et al., 2012). Some of the overarching aims of this research are to better understand the underlying structure and mechanisms of common symptomatology between disorders, to develop more precise and effective treatments for complex disorder cases, and to work towards a reconceptualisation of the psychiatric nosology that is more reflective of the clinical reality (Borsboom, 2017; Montague et al., 2012; Petzschner et al., 2017).

Part of this interest extends to looking at changes in behaviour and cognition that are more domain-general, or at least not directly relevant to the disorder — for instance, cognitive changes in non-social contexts for an individual with a social anxiety disorder — and that are not typically assessed in psychiatric diagnosis. Measures of metacognitive monitoring and control, such as confidence and information-seeking, may fit the bill as domain-general measures that are not typically assessed in relation to psychiatric disorders. There are typically large individual differences in reported confidence and information-seeking frequency between individuals, even at the level of very basic perceptual judgment or choice preference, and these differences appear to be stable within and idiosyncratic between individuals (Ais et al., 2016; Call & Carpenter, 2001; Desender et al., 2018; Navajas et al., 2017). There have also been suggestions that, in uncertain decision contexts, individuals might employ alternatives to information-seeking, such as seeking feedback or opting out of uncertain decisions, and that there may be individual differences in preferences for managing uncertainty that reflect differing levels of uncertainty tolerance (Jensen et al., 2014; Pasquini et al., 2019; Salvador et al., 2020). Unsurprisingly, then, there is a growing list of research exploring on how these measures might be linked with, and might be used to characterise, various disorder-specific and transdiagnostic dimensions of psychiatric interest (Hausmann-Thürig & Läge, 2008; Rouault et al., 2018; Salvador et al., 2020; Schulz et al., 2020).

In the work described in this thesis, we built on the previous research in two ways. Firstly, with work on metacognitive phenotypes so far tending to focus on confidence or information-seeking separately, we explored whether relationships between metacognition and psychiatric dimensions might be better captured by dual measures of confidence and information-seeking, given the close interplay between these two as metacognitive monitoring and control processes.

Secondly, we wanted to explore whether any differences in confidence and information-seeking would be domain-general, and thus used tasks that involved complementary decision styles and different types of uncertainty.

6.1.2 Discussion of findings

As described in Chapter 2, we developed two tasks — the ISTc and the DDTi — to measure confidence and information-seeking in different domains of decision-making. Whereas the precursor of the DDTi, the dot discrimination task, is a perceptual judgment task that has been used extensively in previous literature to help understand and formalise decision-making and confidence, the information sampling task (and the related Beads Task) on which the ISTc is based has been used more frequently as a diagnostic paradigm for information-seeking and probabilistic reasoning in psychiatric disorders. By adapting these paradigms (adding a confidence measure to the information sampling paradigm and an information-seeking measure to the dot discrimination paradigm), we were able to derive parallel measures of confidence and information-seeking for probabilistic reasoning and perceptual judgment, which we then used as the basis for our subsequent experiments.

Anxiety symptoms predict a domain-general underconfidence bias

Of the psychiatric dimensions we tested in general population samples in Experiments 1 and 2, we found the most consistent association between confidence and trait anxiety, with higher trait anxiety predicting an underconfidence bias in decision-point confidence reports and a smaller decision-evoked confidence boost in the ISTc. This corroborates Rouault et al.'s (2018) findings of anxiety inventory scores (trait anxiety, social anxiety, and generalised anxiety) being the most predictive of an underconfidence bias in perceptual judgment within general population samples. In combination with the Rouault et al. study, our results support the idea that this underconfidence bias is domain-general and can be seen in both first-order and second-order measures of confidence.

As found by previous psychiatric literature (Jensen et al., 2016; Lauriola et al., 2018; Rouault et al., 2018), the psychiatric dimensions we assessed were all strongly correlated with trait anxiety — most clearly, depressive severity and uncertainty intolerance, and to a lesser extent, schizotypy — but none of the other dimensions remained significant predictors of confidence bias or decision-evoked confidence boost across the two experiments. Somewhat surprisingly for us, people's levels of uncertainty intolerance, as measured by the Intolerance of Uncertainty Scale (Buhr & Dugas, 2002; Freeston et al., 1994), did not have any significant associations with their reported confidence and failed to mediate the relationship between trait anxiety and decision-point confidence. In contrast, Rouault et al.'s (2018) study found confidence bias to be more strongly predicted by transdiagnostic latent factors than scores on disorder-specific questionnaires or clinical inventories. This suggests that bottom-up factor analytic approaches might produce transdiagnostic dimensions that better predict confidence than top-down conceptualised ones. However, this factor-analytic approach comes with the risk of creating symptom dimensions that might be strongly predictive of confidence but are in practice quite heterogeneous in composition and difficult to apply coherent and clinically useful labels to.

No reliable evidence of differences in managing uncertainty

Across Experiments 1 and 2, psychiatric dimensions did not consistently predict our measures of information-seeking from the two tasks. However, in both experiments, we found a strong inverse relationship between confidence and information-seeking, as had been described in previous studies like Desender et al. (2018) and Hauperich et al. (2017): this was true for measured variables on the DDTi and ISTc, as well as for the latent factors of decision confidence and information-seeking derived from the structural equation model in Chapter 3. There was no evidence for our initial hypothesis of confidence mediating the relationship between psychiatric dimensions and information-seeking. Instead, our structural equation model suggests that confidence and information-seeking may have independent relationships with a latent psychopathology factor.

One of the alternative hypotheses we explored in Chapter 4 was whether individual differences in uncertainty intolerance would be seen in preferences for alternatives to information-seeking, such as feedback or opt-out choices. However, Experiments 3 and 4 did not find evidence of individual preferences for feedback or opt-out over information-seeking: the vast majority of individuals in the vast majority of instances chose to seek information prior to decision rather than engage in alternatives that are associated with reduced information-seeking. This suggests that the surprising relationship we saw in Experiment 1, of uncertainty intolerance and depressive severity being associated with reduced information-seeking on the ISTc, is most likely explained as variability in the data due to sample size, rather than a consistent relationship between the factors. It also suggests that uncertainty intolerance, in practice, does not appear to have a pronounced effect on information-seeking in the decision contexts we investigated, and does not seem to motivate alternative strategies of reducing or managing uncertainty.

The question is why information-seeking fails to have a discernible relationship with psychiatric dimensions, even though disturbances in information-seeking are seen in psychiatric disorders like obsessive-compulsive disorder and schizophrenia (Garety & Freeman, 1999; Hoven et al., 2019; Strauss et al., 2020). One possibility is that with the task designs we were using, participants found the information-seeking process more valuable than not: in

Experiment 4, we observed participants generally engaging in high levels of information-seeking prior to decision. This suggests that the hedonic value of the process was high (i.e., participants found it fun or entertaining, as suggested by our experiment feedback), the consequences of poor choice were not aversive or immediate enough, or the alternatives offered (e.g., escape for a much smaller reward or non-social task feedback) were not attractive enough relative to the net value of information-seeking. Another possibility is information-seeking is simply not a sensitive metacognitive measure at subclinical levels. For instance, Morein-Zamir et al. (2020) were unable to find a significant association between information-seeking on the Beads Task and uncertainty intolerance or obsessive-compulsive symptoms in a general population sample either. In contrast, Hauser, Moutoussis, et al. (2017) found greater information-seeking in a high-compulsive group than a low-compulsive group when comparing high- and low-compulsive subsampled groups derived from the extremes of the normal population distribution. This suggests that we might see stronger effects when using a clinical sample or when testing samples from extremes of these psychopathological dimensions.

Current limitations of metacognitive phenotyping

Our key takeaway from this line of experimental work is that there is some evidence for metacognitive phenotyping being possible. Even within the limited number of measures we used, we found that certain psychiatric dimensions (e.g., trait anxiety) are much more strongly associated with certain metacognitive measures (e.g., confidence) than other dimensions or measures. This suggests that, if we were to test widely enough (with enough psychiatric dimensions and more precise measures that are more suited to certain dimensions), we might be able to find specific profiles of metacognitive differences for certain disorders or groups of disorders.

However, the effect sizes we observed in our experiments and in the previous literature are small (cf. Rouault et al., 2018), to the point of being inconsistent in direction or magnitude in smaller samples. Psychopathologies do not explain much of variance in the measures, but this is unsurprising given the degree of heterogeneity and comorbidity within disorder diagnoses. However, this suggests that future work attempting to find domain-general metacognitive phenotypes may need to use larger samples, and that this metacognitive fingerprinting for psychopathological dimensions may not be reliable enough to be clinically significant for nosology or diagnosis in the foreseeable future.

6.1.3 Future directions

There are several ways in which future research in metacognitive phenotyping can build on the work reported in this thesis. One follow-up investigation we would like to see is whether metacognitive differences might be more clearly identified in clinical or patient populations than a general population sample. Stronger evidence of relationships between psychopathology and some of measures, such as information-seeking, have been found when comparing samples from dimension extremes (e.g., Hauser, Moutoussis, et al., 2017). Most of our participants, on the other hand, were subclinical and skewed towards the lower end of the dimensions measured. Even participants in Experiment 1 and 2 who scored at higher, possibly clinical, levels on the dimensions may still be quite unrepresentative of those with similar levels of psychopathology, as participants registered on Prolific might be assumed to be more technologically advanced and high-functioning than members of the general population.

Another research direction following from our work might be to look into whether affective states and decision-making contexts might induce stronger relationships between metacognitive measures and psychopathology. Previous studies have found that emotion influences perception, for example, biasing attention and pre-attentive perceptual processes in favour of affective stimuli (Phelps et al., 2006; Zadra & Clore, 2011). The relationships between our psychiatric dimensions, confidence, and information-seeking might become more evident under negative affective states. This could be studied by adding an affect induction procedure to our tasks, such as presenting affective cues (e.g., fearful versus neutral facial expressions) prior to the primary visual stimuli in the DDTi (cf. Joseph et al., 2020; Phelps et al., 2006). Alternatively, we might want to assess how individuals differ in their performance on tasks that require affective ("hot") versus deliberative ("cold") decision styles. Previous literature (cf. Markiewicz & Kubińska, 2015) has found stronger individual differences in the relationship between decision parameters and psychopathology in hot than cold versions of the same task. In their current versions, the DDTi is a cold task, while the ISTc has elements of both (i.e., it can be played in a risky or deliberate way, and the degree of ambiguity changes over the course of the game non-linearly). For an initial assessment of this hypothesis, we might choose to use existing task paradigms with clearly delineated hot and cold versions, like the Columbia Card Task (cf. Figner et al., 2009); if successful, we might then look into how our tasks might be developed into parallel hot and cold versions.

Finally, we might want to add or assess additional measures in the ISTc and the DDTi. We might look into whether psychopathology is more consistently associated with second-order computational parameters of confidence and information-seeking, such as metacognitive efficiency. Alternatively, we may want to contrast our existing local (trial-by-trial) measures of confidence with global post-test measures of confidence, to see if global confidence, as a measure sitting between long-term metacognitive beliefs and short-term decision confidence, might have stronger relationships with psychiatric dimensions.

6.2 Metacognition and the illusion of control

6.2.1 Motivations for the work

The second major theme in the thesis is about understanding how confidence evolves during the decision process in response to incoming information and what reported confidence reflects beyond the information acquired. While prevalent theories and computational models of decision-making generally describe confidence as a decision variable, or the output of a decision variable, that tracks evidence accumulation (cf. Balsdon et al., 2020; Lee & Pezzulo, 2021; Pleskac & Busemeyer, 2010; Moran et al., 2015), there is abundant evidence suggesting that confidence is affected by factors other than evidence strength (de Gardelle & Mamassian, 2015; Kiani et al., 2014; Navajas et al., 2016; Bang et al., 2017). In addition, there is debate over whether additional motivations and biases differentially affect the confidence that underlies decision processes and the confidence that is explicitly reported (Massoni et al., 2014; Overgaard & Sandberg, 2012). The dissociation of internal and external confidence has increasingly gained traction over the years as researchers have recognised the significance of confidence in facilitating social communication and group decision-making (Bang et al., 2017; Bang et al., 2020; Heyes et al., 2020).

Confidence may also be biased by personal history and preferences. In cognitive and applied consumer decision-making literature, there is significant evidence of a choice-induced preference, where people like and value things that they have chosen more following choice commitment. We see these effects extended as well to confidence about their choice as well, in what is known as a positive evidence bias (Navajas et al., 2016; Peters, Thesen, et al., 2017; Samaha & Denison, 2020; Jiwa et al., 2021). These biases may be reinforced by confirmation bias, where people seek and overweigh information that supports their choice (Navajas et al., 2016; Rollwage et al., 2020). The combination of overinflated confidence and biased information-seeking can lead to a kind of maladaptive reinforcement of false or unrealistic beliefs, which might underlie the development and maintenance of certain psychiatric conditions like psychosis (Salvador et al., 2020).

In contrast to the neurophysiological research on changes in confidence and information-seeking over the course of the decision (cf. Boldt & Yeung, 2015; Boldt et al., 2019; Parés-Pujolràs et al., 2020), there is comparatively little behavioural research on the evolution of confidence during the decision process and after decision commitment. This is partly because confidence and information-seeking have predominantly been investigated in single-shot decision tasks, where participants only have one opportunity to request further information and one chance post-decision to register their confidence about the decision.

The ISTc paradigm that we developed earlier in this work is ideally placed to build on the previous behavioural research for two reasons. Firstly, it features related measures of confidence and information-seeking throughout the decision process up until the decision point. Secondly, there is an objective measure of evidence strength for the chosen option that can be used to benchmark reported confidence. With the ISTc paradigm, we had the opportunity to extend the previous literature by conducting behavioural and computational investigations of how confidence and information-seeking evolve over the decision process.

6.2.2 Discussion of findings

A decision-evoked confidence boost

From the initial pilots of the ISTc paradigm, we noticed a surge in reported confidence relative to the objective probability of being correct at the point of decision versus before pre-decisionally. Because people were significantly underconfident in the pre-decisional P(correct)-confidence relationship, this meant that the boost actually reflected an increase in calibration at the point of decision. The effect size of this boost was quite large and consistent, usually about 15 confidence points in size, and was observed in all subsequent experiments with the ISTc.

In Chapter 5, we ran two simulations to test whether the decision-point confidence boost was being caused by a selection bias, namely, a bias for instances where noise boosted confidence over the confidence threshold for decision. The results of the two simulations suggested that noise alone could not sufficiently account for the confidence boost effect we were seeing. Experiment 5 subsequently found that this decision-point confidence boost occurs even when the decision is forced, although the effect size was approximately halved. Forcing a decision at an earlier point in the information sampling phase, we still observed a decision-point boost in Experiment 6, but at half the size of the forced choice effect in Experiment 5.

The consistency and significance of the decision-evoked confidence boost indicate that confidence reflects more than just the objective evidence strength. This corroborates previously findings of confidence being affected by several other factors, including previous choices (Navajas et al., 2016; Peters, Thesen, et al., 2017; Samaha & Denison, 2020). The question that remains is what else confidence reports are communicating, beyond the amount of evidence we have accumulated for a choice.

Reported confidence may be dissociable from internal confidence

One of our hypotheses about the decision-evoked confidence boost is that it represents a reporting bias. This reflects a tendency, conscious or not, to inflate confidence estimates when reporting them in relation to a decision. In Experiment 6, we assessed whether the decision-evoked confidence boost from the initial decision at the interim point would persist and have an influence on subsequent confidence and information-seeking if participants were allowed to sample more information before committing to a final decision.

We found that confidence did not continue to be higher after the initial decision-evoked confidence boost, and that the total number of tiles seen prior to final decision was relatively consistent within individuals regardless of whether participants had made an initial decision or not, as had been suggested by Huasmann-Thürig and Läge (2008). The final choice, total amount of information sought, and overall level of confidence at final decision were all not predicted by the initial decision or associated confidence boost.

Our findings in Experiment 6 suggest that the decision-evoked confidence boost is a phenomenon exclusively associated with reported confidence, and does not appear to mark a sustained shift in the internal confidence value that tracks the accumulated evidence. We have a few hypotheses as to what might be driving this phenomenon in reported confidence. The first is that the differences we observe between pre-decisional and decision-point confidence reports might indicate differing interpretations of the confidence report prompt at different times: participants may use confidence reports to say something about the evidence strength for the choice pre-decisionally, and then use confidence reports at decision-point to say something about their own decision choice accuracy or ability (Navajas et al., 2017).

Another hypothesis is that reported confidence may be inherently biased because of its communicative intent (Massoni et al., 2014; Overgaard & Sandberg, 2012). Some researchers have proposed a dual theory of confidence, suggesting that while internal confidence plays an important role in intrapersonal metacognition, external confidence plays more of a social and communicative role supporting group decision-making (Shea et al., 2014; Heyes et al., 2020). Because of this, external confidence may be more susceptible to self-serving biases. For instance, one might express an external or public-facing confidence that is higher than what one used to make a decision because higher expressed confidence increases one's perceived credibility, competence, or attractiveness among others (Buunk et al., 2002; Pulford et al., 2018; Schlenker & Leary, 1982; Whitley & Greenberg, 1986). Such a reporting bias may be adaptive in helping us achieve our goals in decision contexts with multiple decision agents.

The desire for agency in decision-making under uncertainty

The reporting bias hypothesis of the decision-evoked confidence boost highlights a common idea in our work with confidence and information-seeking: the extent of our control over an uncertain reality.

It has been suggested that people prefer the "illusion of control", expressing higher confidence about and subjectively valuing information more in decision contexts that are self-determined rather than assigned (Jiwa et al., 2021). Across our experiments in Chapters 4 and 5, we saw evidence of people's preferences for agency in their information-seeking behaviour and associated confidence boost. In Experiment 5, we saw that the size of the decision-evoked confidence boost was halved when participants were forced to make a decision than when they voluntarily made a decision, even when the objective evidence strength was matched. In Experiment 6, we saw a further halving in the size of the confidence boost relative to the forced-choice decision when active information sampling phase was curtailed. In Experiment 3 and 4, we found that people overwhelmingly preferred to seek information on the ISTc than to skip to feedback or opt out of the information sampling process. Taken together, the evidence strongly suggests that people prefer to feel like they have agency over the uncertainty reduction process during decision-making.

This preference for agency may speak to a principle underlying human decision-making under uncertainty: the desire, and in some cases need, to feel like we have some degree of control in an uncertain decision situation. This can come from either being able to influence the outcome or through being able to anticipate and respond to it. The preference for agency and control is generally adaptive, as it allows us to better situate our behavioural response to changing external circumstances and motivates productive strategies like information-seeking that reduce decision uncertainty. However, this desire has a darker side, where individuals may become obsessed with retaining a feeling of control over an uncertain or changing situation, to the point where they may engage in more maladaptive strategies to exert control. For instance, to minimise the difference between the external reality and their cognitive models of the world, some people may choose to filter the reality of which they are aware, and thereby persist in false beliefs, instead of updating their cognitive models with new information about the external world that contradicts their worldview (Salvador et al., 2020).

This gets to the fundamental dilemma that metacognition presents: to some extent, metacognition is all about having control over our reality, but at what point does this need for control become problematic?

6.2.3 Future directions

A natural follow-up to the current experiments is to further understand the dissociation between internal and external confidence. One way to do this is through neurophysiological studies that could give us insight into how signals of internal confidence evolve over the decision process. Previous studies like Boldt and Yeung (2015) have looked at how neural signals like error positivity vary with confidence and error detection likelihood in one-shot perceptual decision-making tasks. Parés-Pujolràs et al. (2020) recently completed a study looking at how the P3 signal might track a decision variable and predict choice in a two-step sequential sampling paradigm. Combining neurophysiological approaches with the sequential sampling and confidence reporting of the ISTc might help us better understand, beyond our initial observations of a dissociation between pre-decisional and decision-point confidence reports, where exactly external reported confidence and internal decision confidence might be most differentiated in the decision process.

Another line of investigation would be to experimentally investigate agency in the ISTc paradigm by manipulating the perceived degree of control people have during information-seeking. For instance, we might present a version of the ISTc where participants are not engaging in active information-seeking; instead, the computer flips over a certain number of tiles that is matched with the mean number of tiles that people actively sample before decision. Given earlier findings in the literature of people's preference for and greater confidence in sources of information they select (cf. Jiwa et al., 2021) and our own findings in Chapter 5 of smaller decision-evoked confidence boosts in forced-choice conditions, we would expect, in a version of the ISTc where people are effectively stripped of their ability to actively dictate the information-seeking process, to see significantly reduced effect sizes for the decision-evoked confidence boost. This version of the ISTc could help us get a better sense of the extent to which the active nature of information-seeking in the ISTc plays a role in the decision-evoked confidence boost effect, and brings us closer to understanding the role that decision agency plays in confidence.

This line of investigation is linked to our questions about individual differences in preferences for perceived agency or sense of control. While there are many psychometrically sound questionnaires measuring need for control, such as the desirability of control scale (Burger & Cooper, 1979) or the Intolerance of Uncertainty B subscale (Gosselin et al., 2008; Carleton et al., 2010), it would be interesting to see if these established transdiagnostic dimensions about control and agency have implications at a basic decisional level as reflected in confidence and information-seeking measures. We would also like to see how universal this preference for agency and control is, or whether it might be influenced by differing sociocultural contexts (e.g., individualistic vs. collectivistic communities).

6.3 Afterword

In Chapter 1, we introduced two hypothetical Oxford students, Jane and John, whose daily struggle with anticipating the weather illustrated the close association between confidence and information-seeking, as well as how individual differences influence this relationship. But, looking back now, we also see how their struggle points to a bigger idea about the role of metacognition and how we make decisions in an uncertain world.

The world we live in is dynamic and volatile, and it is difficult to accurately and consistently control or predict future events. In this last year and a half of the pandemic, we have all perhaps experienced how little actual control we have over the events that shape our lives. When faced with external uncertainty, it is natural to seek some degree of perceived control over the situation. Confidence and information-seeking are tools that metacognition equips us with to make adaptive decisions under uncertainty, but they are not perfect, and the degree of perceived control they can afford us is sometimes sadly little. In some ways, to remain healthy, we must be prepared to, or at least acknowledge the probability that we will need to, relinquish some degree of control in our lives.

For some people though, when confronted with the great uncertainty — political, economic, cultural, and epistemic — that has defined the postmodern zeitgeist, they have responded by employing confidence and information-seeking in maladaptive ways, for instance, engaging in extensive information-seeking about conspiracy theories, or expressing heightened confidence for choices that are not supported by the scientific evidence. Unsurprisingly, we have seen a concurrent mental health crisis unfolding around the world during this time (Mind, 2021; The Health Foundation, 2021). This suggests that our metacognitive abilities are a two-edged sword: on one hand, they can help us anticipate and adapt to our reality around us, and on the other, they can be used to isolate us from and alter the reality in which we operate.

Returning to the case of Jane and John then, perhaps the most adaptive principle we can advise them to adopt is simply that of acceptance: accepting the inherent uncertainty of the weather and accepting that, despite their best-laid plans, they still might get caught in the rain on the rare occasion, but that this hiccup in their day would not be the end of the world.

Carpe diem quam minimum credula postero
[Seize the day, trusting as little as possible in the next one]

- Horace, *Odes*

REFERENCES

Ahrends, C., Bravo, F., Kringelbach, M. L., Vuust, P., & Rohrmeier, M. A. (2019). Pessimistic outcome expectancy does not explain ambiguity aversion in decision-making under uncertainty. Scientific Reports, 9(1), 12177. https://doi.org/10.1038/s41598-019-48707-y

Ais, J., Zylberberg, A., Barttfeld, P., & Sigman, M. (2016). Individual consistency in the accuracy and distribution of confidence judgments. Cognition, 146, 377–386. https://doi.org/10.1016/j.cognition.2015.10.006

Arend, R. J. (2020). Strategic decision-making under ambiguity: a new problem space and a proposed optimization approach. Business Research, 13(3), 1231–1251. https://doi.org/10.1007/s40685-020-00129-7

Ashford, S. J. (1986). Feedback-Seeking in Individual Adaptation: A Resource Perspective. Academy of Management Journal. Academy of Management, 29(3), 465–487. https://doi.org/10.2307/256219

Averbeck, B. B., Evans, S., Chouhan, V., Bristow, E., & Shergill, S. S. (2011). Probabilistic learning and inference in schizophrenia. Schizophrenia Research, 127(1-3), 115–122. https://doi.org/10.1016/j.schres.2010.08.009

Axelsen, M. C., Jepsen, J. R. M., & Bak, N. (2018). The Choice of Prior in Bayesian Modeling of the Information Sampling Task [Review of The Choice of Prior in Bayesian Modeling of the Information Sampling Task]. Biological Psychiatry, 83(12), e59–e60. https://doi.org/10.1016/j.biopsych.2017.04.021

Azzopardi, P., & Cowey, A. (1998). Blindsight and visual awareness. Consciousness and Cognition, 7(3), 292–311. https://doi.org/10.1006/ccog.1998.0358

Bahrami, B., Olsen, K., Latham, P. E., Roepstorff, A., Rees, G., & Frith, C. D. (2010). Optimally interacting minds. Science, 329(5995), 1081–1085. https://doi.org/10.1126/science.1185718

Balsdon, T., Wyart, V., & Mamassian, P. (2020). Confidence controls perceptual evidence accumulation. Nature Communications, 11(1), 1753. https://doi.org/10.1038/s41467-020-15561-w

Bang, D., Aitchison, L., Moran, R., Herce Castanon, S., Rafiee, B., Mahmoodi, A., Lau, J. Y. F., Latham, P. E., Bahrami, B., & Summerfield, C. (2017). Confidence matching in group decision-making. Nature Human Behaviour, 1(6), 0117. https://doi.org/10.1038/s41562-017-0117

Bang, D., Ershadmanesh, S., Nili, H., & Fleming, S. M. (2020). Private-public mappings in human prefrontal cortex. eLife, 9. https://doi.org/10.7554/eLife.56477

Bechara, A., Damasio, A. R., Damasio, H., & Anderson, S. W. (1994). Insensitivity to future consequences following damage to human prefrontal cortex. Cognition, 50(1-3), 7–15. https://doi.org/10.1016/0010-0277(94)90018-3

Bem, D. J. (1967). Self-perception: An alternative interpretation of cognitive dissonance phenomena. Psychological Review, 74(3), 183–200. https://doi.org/10.1037/h0024835

Bennett, D., Bode, S., Brydevall, M., Warren, H., & Murawski, C. (2016). Intrinsic Valuation of Information in Decision Making under Uncertainty. PLoS Computational Biology, 12(7), e1005020. https://doi.org/10.1371/journal.pcbi.1005020

Bennett, D., Sutcliffe, K., Tan, N. P.-J., Smillie, L. D., & Bode, S. (2021). Anxious and obsessive-compulsive traits are independently associated with valuation of noninstrumental information. Journal of Experimental Psychology. General, 150(4), 739–755. https://doi.org/10.1037/xge0000966

Bennett, N., Herold, D. M., & Ashford, S. J. (1990). The effects of tolerance for ambiguity on feedback-seeking behaviour. Journal of Occupational Psychology, 63(4), 343–348. https://doi.org/10.1111/j.2044-8325.1990.tb00535.x

Berlyne, D. E. (1954). A theory of human curiosity. British Journal of Psychology , 45(3), 180–191. https://doi.org/10.1111/j.2044-8295.1954.tb01243.x

Bjork, R. A., Dunlosky, J., & Kornell, N. (2013). Self-regulated learning: beliefs, techniques, and illusions. Annual Review of Psychology, 64, 417–444. https://doi.org/10.1146/annurev-psych-113011-143823

Blais, A.-R., Thompson, M. M., & Baranski, J. V. (2005). Individual differences in decision processing and confidence judgments in comparative judgment tasks: The role of cognitive styles. Personality and Individual Differences, 38(7), 1701–1713. https://doi.org/10.1016/j.paid.2004.11.004

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. https://doi.org/10.1037/xhp0000404

Boldt, A., Schiffer, A.-M., Waszak, F., & Yeung, N. (2019). Confidence Predictions Affect Performance Confidence and Neural Preparation in Perceptual Decision Making. Scientific Reports, 9(1), 4031. https://doi.org/10.1038/s41598-019-40681-9

Boldt, A., & Yeung, N. (2015). Shared neural markers of decision confidence and error detection. The Journal of Neuroscience: The Official Journal of the Society for Neuroscience, 35(8), 3478–3484. https://doi.org/10.1523/JNEUROSCI.0797-14.2015

Borsboom, D. (2017). A network theory of mental disorders. World Psychiatry: Official Journal of the World Psychiatric Association , 16(1), 5–13. https://doi.org/10.1002/wps.20375

Boswell, J. F., Thompson-Hollands, J., Farchione, T. J., & Barlow, D. H. (2013). Intolerance of uncertainty: a common factor in the treatment of emotional disorders. Journal of Clinical Psychology, 69(6), 630–645. https://doi.org/10.1002/jclp.21965

Brehm, J. W. (1956). Postdecision changes in the desirability of alternatives. Journal of Abnormal Psychology, 52(3), 384–389. https://doi.org/10.1037/h0041006

Brenner, L. A. (2003). A random support model of the calibration of subjective probabilities. Organizational Behavior and Human Decision Processes, 90(1), 87–110. https://doi.org/10.1016/S0749-5978(03)00004-9

Brevers, D., Cleeremans, A., Bechara, A., Greisen, M., Kornreich, C., Verbanck, P., & Noël, X. (2014). Impaired metacognitive capacities in individuals with problem gambling. Journal of Gambling Studies / Co-Sponsored by the National Council on Problem Gambling and Institute for the Study of Gambling and Commercial Gaming, 30(1), 141–152. https://doi.org/10.1007/s10899-012-9348-3

Browning, M., Behrens, T. E., Jocham, G., O'Reilly, J. X., & Bishop, S. J. (2015). Anxious individuals have difficulty learning the causal statistics of aversive environments. Nature Neuroscience, 18(4), 590–596. https://doi.org/10.1038/nn.3961

Brownlee, J. (2019, October 15). Information Gain and Mutual Information for Machine Learning. Machine Learning Mastery. https://machinelearningmastery.com/information-gain-and-mutual-information/

Bruno, N., Sachs, N., Demily, C., Franck, N., & Pacherie, E. (2012). Delusions and metacognition in patients with schizophrenia. Cognitive Neuropsychiatry, 17(1), 1–18. https://doi.org/10.1080/13546805.2011.562071

Brydevall, M., Bennett, D., Murawski, C., & Bode, S. (2018). The neural encoding of information prediction errors during non-instrumental information seeking. Scientific Reports, 8(1), 6134. https://doi.org/10.1038/s41598-018-24566-x

Buhr, K., & Dugas, M. J. (2002). The Intolerance of Uncertainty Scale: psychometric properties of the English version. Behaviour Research and Therapy, 40(8), 931–945. https://doi.org/10.1016/s0005-7967(01)00092-4

Burger, J. M., & Cooper, H. M. (1979). The desirability of control. Motivation and Emotion, 3(4), 381–393. https://doi.org/10.1007/BF00994052

Buunk, B. P., Dijkstra, P., Fetchenhauer, D., & Kenrick, D. T. (2002). Age and gender differences in mate selection criteria for various involvement levels. Personal Relationships, 9(3), 271–278. https://doi.org/10.1111/1475-6811.00018

Call, J., & Carpenter, M. (2001). Do apes and children know what they have seen? Animal Cognition, 3(4), 207–220. https://doi.org/10.1007/s100710100078

Camp, C. J. (1986). I am curious-grey: Information seeking and depression across the adult lifespan. Educational Gerontology, 12(4), 375–384. https://doi.org/10.1080/0380127860120412

Campbell, J. D. (1990). Self-esteem and clarity of the self-concept. Journal of Personality and Social Psychology, 59(3), 538–549. https://doi.org/10.1037/0022-3514.59.3.538

Carlebach, N., & Yeung, N. (2020). Subjective confidence acts as an internal cost-benefit factor when choosing between tasks. Journal of Experimental Psychology. Human Perception and Performance, 46(7), 729–748. https://doi.org/10.1037/xhp0000747

Carleton, R. N., Gosselin, P., & Asmundson, G. J. G. (2010). The intolerance of uncertainty index: replication and extension with an English sample. Psychological Assessment, 22(2), 396–406. https://doi.org/10.1037/a0019230

Cartwright-Hatton, S., & Wells, A. (1997). Beliefs about worry and intrusions: the Meta-Cognitions Questionnaire and its correlates. Journal of Anxiety Disorders, 11(3), 279–296. https://doi.org/10.1016/s0887-6185(97)00011-x

Chambers, J. A., Power, K. G., & Durham, R. C. (2004). The relationship between trait vulnerability and anxiety and depressive diagnoses at long-term follow-up of Generalized Anxiety Disorder. Journal of Anxiety Disorders, 18(5), 587–607. https://doi.org/10.1016/j.janxdis.2003.09.001

Charles, L., & Yeung, N. (2019). Dynamic sources of evidence supporting confidence judgments and error detection. Journal of Experimental Psychology. Human Perception and Performance, 45(1), 39–52. https://doi.org/10.1037/xhp0000583

Chou, C.-Y., & Zou, N.-B. (2020). An analysis of internal and external feedback in self-regulated learning activities mediated by self-regulated learning tools and open learner models. International Journal of Educational Technology in Higher Education, 17(1), 1–27. https://doi.org/10.1186/s41239-020-00233-y

Clark, L., Robbins, T. W., Ersche, K. D., & Sahakian, B. J. (2006). Reflection impulsivity in current and former substance users. Biological Psychiatry, 60(5), 515–522. https://doi.org/10.1016/j.biopsych.2005.11.007

Clarke, F. R., Birdsall, T. G., & Tanner, W. P., Jr. (1959). Two types of ROC curves and definitions of parameters. The Journal of the Acoustical Society of America, 31(5), 629–630. https://doi.org/10.1121/1.1907764

Cogliati Dezza, I., Maher, C., & Sharot, T. (2021). People adaptively use information to improve their internal and external states. https://doi.org/10.31234/osf.io/f5vyq

Dashiell, J. F. (1925). A quantitative demonstration of animal drive. Journal of Comparative Psychology, 5(3), 205–208. https://doi.org/10.1037/h0071833

Davies, G., Rae, C. L., Garfinkel, S. N., Seth, A. K., Medford, N., Critchley, H. D., & Greenwood, K. (2018). Impairment of perceptual metacognitive accuracy and reduced prefrontal grey matter volume in first-episode psychosis. Cognitive Neuropsychiatry, 23(3), 165–179. https://doi.org/10.1080/13546805.2018.1444597

de Gardelle, V., & Mamassian, P. (2015). Weighting mean and variability during confidence judgments. PloS One, 10(3), e0120870. https://doi.org/10.1371/journal.pone.0120870

de Leeuw, J. R. (2015). jsPsych: a JavaScript library for creating behavioral experiments in a Web browser. Behavior Research Methods, 47(1), 1–12. https://doi.org/10.3758/s13428-014-0458-y

Desender, K., Boldt, A., & Yeung, N. (2018). Subjective Confidence Predicts Information Seeking in Decision Making. Psychological Science, 29(5), 761–778. https://doi.org/10.1177/0956797617744771

Desender, K., Donner, T. H., & Verguts, T. (2021). Dynamic expressions of confidence within an evidence accumulation framework. Cognition, 207, 104522. https://doi.org/10.1016/j.cognition.2020.104522

Diamond, A. (2013). Executive functions. Annual Review of Psychology, 64, 135–168. https://doi.org/10.1146/annurev-psych-113011-143750

Dienes, Z., & Seth, A. K. (2010). Measuring any conscious content versus measuring the relevant conscious content: comment on Sandberg et al [Review of Measuring any conscious content versus measuring the relevant conscious content: comment on Sandberg et al]. Consciousness and Cognition, 19(4), 1079–1080; discussion 1081–1083. https://doi.org/10.1016/j.concog.2010.03.009

Doron, G., Derby, D. S., Szepsenwol, O., & Talmor, D. (2012). Flaws and all: Exploring partner-focused obsessive-compulsive symptoms. Journal of Obsessive-Compulsive and Related Disorders, 1(4), 234–243. https://doi.org/10.1016/j.jocrd.2012.05.004

Dugas, M. J., Schwartz, A., & Francis, K. (2004). Brief Report: Intolerance of Uncertainty, Worry, and Depression. Cognitive Therapy and Research, 28(6), 835–842. https://doi.org/10.1007/s10608-004-0669-0

Ellsberg, D. (1961). Risk, Ambiguity, and the Savage Axioms. The Quarterly Journal of Economics, 75(4), 643–669. https://doi.org/10.2307/1884324

Eysenck, H. J. (1983). Stress, disease, and personality: The inoculation effect. In C. L. Cooper (Ed.), Stress and Research (pp. 121–146). John Wiley & Sons.

Fernandez-Duque, D., Baird, J. A., & Posner, M. I. (2000). Executive attention and metacognitive regulation. Consciousness and Cognition, 9(2 Pt 1), 288–307. https://doi.org/10.1006/ccog.2000.0447

Fessler, D. M. T. (2001). Emotions and cost–benefit assessment: The role of shame and self-esteem in risk taking. In G. Gigerenzer & R. Selten (Eds.), Bounded rationality: The adaptive toolbox (Vol. 377, pp. 191–214). The MIT Press. https://psycnet.apa.org/fulltext/2001-00702-011.pdf

Figner, B., Mackinlay, R. J., Wilkening, F., & Weber, E. U. (2009). Affective and deliberative processes in risky choice: age differences in risk taking in the Columbia Card Task. Journal of Experimental Psychology. Learning, Memory, and Cognition, 35(3), 709–730. https://doi.org/10.1037/a0014983

FitzGibbon, L., Lau, J. K. L., & Murayama, K. (2020). The seductive lure of curiosity: information as a motivationally salient reward. Current Opinion in Behavioral Sciences, 35, 21–27. https://doi.org/10.1016/j.cobeha.2020.05.014

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

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. https://doi.org/10.1037/rev0000045

Fleming, S. M., & Lau, H. C. (2014). How to measure metacognition. Frontiers in Human Neuroscience, 8, 443. https://doi.org/10.3389/fnhum.2014.00443

Folke, T., Jacobsen, C., Fleming, S. M., & De Martino, B. (2016). Explicit representation of confidence informs future value-based decisions. Nature Human Behaviour, 1(1), 1–8. https://doi.org/10.1038/s41562-016-0002

Foote, A. L., & Crystal, J. D. (2007). Metacognition in the rat. Current Biology: CB, 17(6), 551–555. https://doi.org/10.1016/j.cub.2007.01.061

Frank, R. A., van der Klaauw, N. J., & Schifferstein, H. N. (1993). Both perceptual and conceptual factors influence taste-odor and taste-taste interactions. Perception & Psychophysics, 54(3), 343–354. https://doi.org/10.3758/bf03205269

Freeston, M. H., Rhéaume, J., Letarte, H., Dugas, M. J., & Ladouceur, R. (1994). Why do people worry? Personality and Individual Differences, 17(6), 791–802. https://doi.org/10.1016/0191-8869(94)90048-5

Friston, K. (2010). The free-energy principle: a unified brain theory? Nature Reviews. Neuroscience, 11(2), 127–138. https://doi.org/10.1038/nrn2787

Fu, T. S.-T., Koutstaal, W., Poon, L., & Cleare, A. J. (2012). Confidence judgment in depression and dysphoria: the depressive realism vs. negativity hypotheses. Journal of Behavior Therapy and Experimental Psychiatry, 43(2), 699–704. https://doi.org/10.1016/j.jbtep.2011.09.014

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. https://doi.org/10.3758/bf03196546

Garety, P. A., & Freeman, D. (1999). Cognitive approaches to delusions: a critical review of theories and evidence. The British Journal of Clinical Psychology / the British Psychological Society, 38(2), 113–154. https://doi.org/10.1348/014466599162700

Garety, P. A., Hemsley, D. R., & Wessely, S. (1991). Reasoning in deluded schizophrenic and paranoid patients. Biases in performance on a probabilistic inference task. The Journal of Nervous and Mental Disease, 179(4), 194–201. https://doi.org/10.1097/00005053-199104000-00003

Gigerenzer, G., & Gaissmaier, W. (2015). Decision Making: Nonrational Theories. In J. D. Wright (Ed.), International Encyclopedia of the Social & Behavioral Sciences, 2nd Edition (Vol. 5, pp. 911–916). Elsevier. https://doi.org/10.1016/B978-0-08-097086-8.26017-0

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning (Adaptive Computation and Machine Learning series) (Illustrated edition). The MIT Press. https://www.amazon.com/Deep-Learning-Adaptive-Computation-Machine/dp/0262035618/ref=as_li_ss_tl?keywords=Deep+Learning&qid=1563148923&s=books&sr=1-1&linkCode=sl1&tag=inspiredalgor-20&linkId=6e39ca521cd3c37b4a192f0ad8cfcf8c&language=en_US

Goodie, A. S. (2005). The Role of Perceived Control and Overconfidence in Pathological Gambling. Journal of Gambling Studies / Co-Sponsored by the National Council on Problem Gambling and Institute for the Study of Gambling and Commercial Gaming, 21(4), 481–502. https://doi.org/10.1007/s10899-005-5559-1

Goodman, J. S. (1998). The Interactive Effects of Task and External Feedback on Practice Performance and Learning. Organizational Behavior and Human Decision Processes, 76(3), 223–252. https://doi.org/10.1006/obhd.1998.2805

Goodman, L. A., & Kruskal, W. H. (1954). Measures of association for cross classifications. Journal of the American Statistical Association, 49, 732–769. https://doi.org/10.2307/2281536

Gosselin, P., Ladouceur, R., Evers, A., Laverdière, A., Routhier, S., & Tremblay-Picard, M. (2008). Evaluation of intolerance of uncertainty: development and validation of a new self-report measure. Journal of Anxiety Disorders, 22(8), 1427–1439. https://doi.org/10.1016/j.janxdis.2008.02.005

Gottlieb, J. (2018). Understanding active sampling strategies: Empirical approaches and implications for attention and decision research. Cortex; a Journal Devoted to the Study of the Nervous System and Behavior, 102, 150–160. https://doi.org/10.1016/j.cortex.2017.08.019

Green, D. M., & Swets, J. A. (1966). Signal detection theory and psychophysics. John Wiley. https://psycnet.apa.org/fulltext/1967-02286-000.pdf

Grupe, D. W., & Nitschke, J. B. (2013). Uncertainty and anticipation in anxiety: an integrated neurobiological and psychological perspective. Nature Reviews. Neuroscience, 14(7), 488–501. https://doi.org/10.1038/nrn3524

Hampton, R. R. (2001). Rhesus monkeys know when they remember. Proceedings of the National Academy of Sciences of the United States of America, 98(9), 5359–5362. https://doi.org/10.1073/pnas.071600998

Hancock, J. A. (1996). "Depressive Realism" assessed via Confidence in Decision-making. Cognitive Neuropsychiatry, 1(3), 213–220. https://doi.org/10.1080/135468096396514

Handelzalts, J. E., Levy, S., Peled, Y., Binyamin, L., Wiznitzer, A., Goldzweig, G., & Krissi, H. (2016). Information seeking and perceptions of anxiety and pain among women undergoing hysterosalpingography. European Journal of Obstetrics, Gynecology, and Reproductive Biology, 202, 41–44. https://doi.org/10.1016/j.ejogrb.2016.04.037

Harmon-Jones, E., & Harmon-Jones, C. (2007). Cognitive Dissonance Theory After 50 Years of Development. Zeitschrift Für Sozialpsychologie, 38(1), 7–16. https://doi.org/10.1024/0044-3514.38.1.7

Hart, J. T. (1965). Memory and the feeling-of-knowing experience. Journal of Educational Psychology, 56(4), 208–216. https://doi.org/10.1037/h0022263

Haslam, N. (2003). Categorical versus dimensional models of mental disorder: the taxometric evidence. The Australian and New Zealand Journal of Psychiatry, 37(6), 696–704. https://doi.org/10.1111/j.1440-1614.2003.01258.x

Hauperich, A., Pescetelli, N., & Yeung, N. (2017). Advice-seeking is Driven by Internal Confidence rather than Perceptual Difficulty (N. Yeung (ed.)) [MSc Psychological Research]. University of Oxford.

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. https://doi.org/10.1038/s41598-017-06116-z

Hauser, T. U., Moutoussis, M., NSPN Consortium, Dayan, P., & Dolan, R. J. (2017). Increased decision thresholds trigger extended information gathering across the compulsivity spectrum. Translational Psychiatry, 7(12), 1296. https://doi.org/10.1038/s41398-017-0040-3

Hausmann-Thürig, D., & Läge, D. (2008). Sequential evidence accumulation in decision making: the individual desired level of confidence can explain the extent of information acquisition. Judgment and Decision Making, 3(3), 229–243. https://doi.org/10.5167/uzh-6226

Hayden, B. Y., Heilbronner, S. R., & Platt, M. L. (2010). Ambiguity aversion in rhesus macaques. Frontiers in Neuroscience, 4. https://doi.org/10.3389/fnins.2010.00166

Hebb, D. O. (1955). Drives and the C. N. S. (conceptual nervous system). Psychological Review, 62(4), 243–254. https://doi.org/10.1037/h0041823

Heyes, C., Bang, D., Shea, N., Frith, C. D., & Fleming, S. M. (2020). Knowing Ourselves Together: The Cultural Origins of Metacognition. Trends in Cognitive Sciences, 24(5), 349–362. https://doi.org/10.1016/j.tics.2020.02.007

Hickson, L., & Khemka, I. (2014). The Psychology of Decision Making. In R. M. Hodapp (Ed.), International Review of Research in Developmental Disabilities (Vol. 47, pp. 185–229). Academic Press. https://doi.org/10.1016/B978-0-12-800278-0.00006-3

Higham, P. A., & Higham, D. P. (2019). New improved gamma: Enhancing the accuracy of Goodman-Kruskal's gamma using ROC curves. Behavior Research Methods, 51(1), 108–125. https://doi.org/10.3758/s13428-018-1125-5

Hoven, M., Lebreton, M., Engelmann, J. B., Denys, D., Luigjes, J., & van Holst, R. J. (2019). Abnormalities of confidence in psychiatry: an overview and future perspectives. Translational Psychiatry, 9(1), 268. https://doi.org/10.1038/s41398-019-0602-7

Howell, D. C. (2010). Statistical Methods for Psychology, 7th Edition. Wadsworth Cengage Learning. https://play.google.com/store/books/details?id=ymAJzgEACAAJ

Hunt, J. M. (1965). Intrinsic Motivation and Its Role in Psychological Development. In D. Levine (Ed.), Nebraska Symposium on Motivation (Vol. 13, pp. 189–282). University of Nebraska Press.

Hunt, L. T., Rutledge, R. B., Malalasekera, W. M. N., Kennerley, S. W., & Dolan, R. J. (2016). Approach-Induced Biases in Human Information Sampling. PLoS Biology, 14(11), e2000638. https://doi.org/10.1371/journal.pbio.2000638

Huq, S. F., Garety, P. A., & Hemsley, D. R. (1988). Probabilistic judgements in deluded and non-deluded subjects. The Quarterly Journal of Experimental Psychology. A, Human Experimental Psychology, 40(4), 801–812. https://doi.org/10.1080/14640748808402300

Jackson, S. A., Kleitman, S., Stankov, L., & Howie, P. (2016). Decision pattern analysis as a general framework for studying individual differences in decision making. Journal of Behavioral Decision Making, 29(4), 392–408. https://doi.org/10.1002/bdm.1887

Jackson, S. A., Kleitman, S., Stankov, L., & Howie, P. (2017). Individual differences in decision making depend on cognitive abilities, monitoring and control. Journal of Behavioral Decision Making, 30(2), 209–223. https://doi.org/10.1002/bdm.1939

Janis, I. L. (1982). Stress, attitudes, and decisions. Praeger Publishers.

Jensen, D., Cohen, J. N., Mennin, D. S., Fresco, D. M., & Heimberg, R. G. (2016). Clarifying the unique associations among intolerance of uncertainty, anxiety, and depression. Cognitive Behaviour Therapy, 45(6), 431–444. https://doi.org/10.1080/16506073.2016.1197308

Jensen, D., Kind, A. J., Morrison, A. S., & Heimberg, R. G. (2014). Intolerance of Uncertainty and Immediate Decision-Making in High-Risk Situations. Journal of Experimental Psychopathology, 5(2), 178–190. https://doi.org/10.5127/jep.035113

Jezzini, A., Bromberg-Martin, E. S., Trambaiolli, L. R., Haber, S. N., & Monosov, I. E. (2021). A prefrontal network integrates preferences for advance information about uncertain rewards and punishments. Neuron, 109(14), 2339–2352.e5. https://doi.org/10.1016/j.neuron.2021.05.013

Jiwa, M., Cooper, P. S., Chong, T. T.-J., & Bode, S. (2021). Choosing increases the value of non-instrumental information. Scientific Reports, 11(1), 8780. https://doi.org/10.1038/s41598-021-88031-y

Johnson, S. U., & Hoffart, A. (2016). Metacognitive Therapy for Comorbid Anxiety Disorders: A Case Study. Frontiers in Psychology, 7, 1515. https://doi.org/10.3389/fpsyg.2016.01515

Johnson, S. U., Hoffart, A., Nordahl, H. M., & Wampold, B. E. (2017). Metacognitive therapy versus disorder-specific CBT for comorbid anxiety disorders: A randomized controlled trial. Journal of Anxiety Disorders, 50, 103–112. https://doi.org/10.1016/j.janxdis.2017.06.004

Joseph, D. L., Chan, M. Y., Heintzelman, S. J., Tay, L., Diener, E., & Scotney, V. S. (2020). The manipulation of affect: A meta-analysis of affect induction procedures. Psychological Bulletin, 146(4), 355–375. https://doi.org/10.1037/bul0000224

Kaanders, P., Nili, H., O'Reilly, J. X., & Hunt, L. (2021). Medial Frontal Cortex Activity Predicts Information Sampling in Economic Choice. The Journal of Neuroscience: The Official Journal of the Society for Neuroscience, 41(40), 8403–8413. https://doi.org/10.1523/JNEUROSCI.0392-21.2021

Kaczmarek, Ł. D., Bączkowski, B., Enko, J., Baran, B., & Theuns, P. (2014). Subjective well-being as a mediator for curiosity and depression. Polish Psychological Bulletin, 45(2), 200–204. https://doi.org/10.2478/ppb-2014-0025

Kagan, J. (1966). Reflection--impulsivity: the generality and dynamics of conceptual tempo. Journal of Abnormal Psychology, 71(1), 17–24. https://doi.org/10.1037/h0022886

Kagan, J. (1972). Motives and development. Journal of Personality and Social Psychology, 22(1), 51–66. https://doi.org/10.1037/h0032356

Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1991). Anomalies: The Endowment Effect, Loss Aversion, and Status Quo Bias. The Journal of Economic Perspectives: A Journal of the American Economic Association, 5(1), 193–206. https://doi.org/10.1257/jep.5.1.193

Kang, M. J., Hsu, M., Krajbich, I. M., Loewenstein, G., McClure, S. M., Wang, J. T.-Y., & Camerer, C. F. (2009). The wick in the candle of learning: epistemic curiosity activates reward circuitry and enhances memory. Psychological Science, 20(8), 963–973. https://doi.org/10.1111/j.1467-9280.2009.02402.x

Kaye, P. H., Smith, A. D., Strudwick, M. J., White, M., Bird, C. E. L., Aggarwal, G., Durkin, T., Marcuson, T. A. G., Masters, T. R., Regan, N., Restrepo, S., Toller, J. R., White, S., & Wilkinson, R. (2020). Managing uncertainty: Principles for improved decision making. British Actuarial Journal, 25. https://doi.org/10.1017/S135732172000015X

Keinan, G. (1987). Decision making under stress: Scanning of alternatives under controllable and uncontrollable threats. Journal of Personality and Social Psychology, 52(3), 639–644. https://doi.org/10.1037/0022-3514.52.3.639

Kelly, S. P., & O'Connell, R. G. (2013). Internal and External Influences on the Rate of Sensory Evidence Accumulation in the Human Brain. The Journal of Neuroscience: The Official Journal of the Society for Neuroscience, 33(50), 19434–19441. https://doi.org/10.1523/JNEUROSCI.3355-13.2013

Kelly, C. A., & Sharot, T. (2021). Individual differences in information-seeking. Nature Communications, 12(1), 7062. https://doi.org/10.1038/s41467-021-27046-5

Kepecs, A., & Mainen, Z. F. (2012). A computational framework for the study of confidence in humans and animals. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 367(1594), 1322–1337. https://doi.org/10.1098/rstb.2012.0037

Kepecs, A., Uchida, N., Zariwala, H. A., & Mainen, Z. F. (2008). Neural correlates, computation and behavioural impact of decision confidence. Nature, 455(7210), 227–231. https://doi.org/10.1038/nature07200

Kesby, A., Maguire, S., Brownlow, R., & Grisham, J. R. (2017). Intolerance of Uncertainty in eating disorders: An update on the field. Clinical Psychology Review, 56, 94–105. https://doi.org/10.1016/j.cpr.2017.07.002

Kessler, R. C., Petukhova, M., Sampson, N. A., Zaslavsky, A. M., & Wittchen, H.-U. (2012). Twelve-month and lifetime prevalence and lifetime morbid risk of anxiety and mood disorders in the United States. International Journal of Methods in Psychiatric Research, 21(3), 169–184.

Kiani, R., Corthell, L., & Shadlen, M. N. (2014). Choice certainty is informed by both evidence and decision time. Neuron, 84(6), 1329–1342. https://doi.org/10.1016/j.neuron.2014.12.015

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

Kleitman, S., & Stankov, L. (2007). Self-confidence and metacognitive processes. Learning and Individual Differences, 17(2), 161–173. https://doi.org/10.1016/j.lindif.2007.03.004

Kobayashi, K., & Hsu, M. (2019). Common neural code for reward and information value. Proceedings of the National Academy of Sciences of the United States of America, 116(26), 13061–13066. https://doi.org/10.1073/pnas.1820145116

Komura, Y., Nikkuni, A., Hirashima, N., Uetake, T., & Miyamoto, A. (2013). Responses of pulvinar neurons reflect a subject's confidence in visual categorization. Nature Neuroscience, 16(6), 749–755. https://doi.org/10.1038/nn.3393

Koriat, A., & Levy-Sadot, R. (2001). The combined contributions of the cue-familiarity and accessibility heuristics to feelings of knowing. Journal of Experimental Psychology. Learning, Memory, and Cognition, 27(1), 34–53. https://doi.org/10.1037/0278-7393.27.1.34

Kornell, N., Son, L. K., & Terrace, H. S. (2007). Transfer of metacognitive skills and hint seeking in monkeys. Psychological Science, 18(1), 64–71. https://doi.org/10.1111/j.1467-9280.2007.01850.x

Köther, U., Veckenstedt, R., Vitzthum, F., Roesch-Ely, D., Pfueller, U., Scheu, F., & Moritz, S. (2012). "Don't give me that look" - overconfidence in false mental state perception in schizophrenia. Psychiatry Research, 196(1), 1–8. https://doi.org/10.1016/j.psychres.2012.03.004

Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The PHQ-9: validity of a brief depression severity measure. Journal of General Internal Medicine, 16(9), 606–613. https://doi.org/10.1046/j.1525-1497.2001.016009606.x

Krueger, R. F., & Eaton, N. R. (2015). Transdiagnostic factors of mental disorders. World Psychiatry: Official Journal of the World Psychiatric Association , 14(1), 27–29. https://doi.org/10.1002/wps.20175

Kunimoto, C., Miller, J., & Pashler, H. (2001). Confidence and Accuracy of Near-Threshold Discrimination Responses. Consciousness and Cognition, 10(3), 294–340. https://doi.org/10.1006/ccog.2000.0494

Ladouceur, R., Talbot, F., & Dugas, M. J. (1997). Behavioral expressions of intolerance of uncertainty in worry. Experimental findings. Behavior Modification, 21(3), 355–371. https://doi.org/10.1177/01454455970213006

Lakey, C. E., Campbell, W. K., Brown, K. W., & Goodie, A. S. (2007). Dispositional Mindfulness as a Predictor of the Severity of Gambling Outcomes. Personality and Individual Differences, 43(7), 1698–1710. https://doi.org/10.1016/j.paid.2007.05.007

Lanzetta, J. T., & Kanareff, V. T. (1962). Information cost, amount of payoff, and level of aspiration as determinants of information seeking in decision making. Behavioral Science, 7(4), 459–473. https://doi.org/10.1002/bs.3830070406

Lauriola, M., Mosca, O., Trentini, C., Foschi, R., Tambelli, R., & Carleton, R. N. (2018). The Intolerance of Uncertainty Inventory: Validity and Comparison of Scoring Methods to Assess Individuals Screening Positive for Anxiety and Depression. Frontiers in Psychology, 9, 388. https://doi.org/10.3389/fpsyg.2018.00388

Lazarov, A., Dar, R., Liberman, N., & Oded, Y. (2012). Obsessive-compulsive tendencies and undermined confidence are related to reliance on proxies for internal states in a false feedback paradigm. Journal of Behavior Therapy and Experimental Psychiatry, 43(1), 556–564. https://doi.org/10.1016/j.jbtep.2011.07.007

Le Berre, A.-P., Pinon, K., Vabret, F., Pitel, A.-L., Allain, P., Eustache, F., & Beaunieux, H. (2010). Study of metamemory in patients with chronic alcoholism using a feeling-of-knowing episodic memory task. Alcoholism, Clinical and Experimental Research, 34(11), 1888–1898. https://doi.org/10.1111/j.1530-0277.2010.01277.x

Lee, A. L. F., de Gardelle, V., & Mamassian, P. (2021). Global visual confidence. Psychonomic Bulletin & Review, 28(4), 1233–1242. https://doi.org/10.3758/s13423-020-01869-7

Lee, D., & Daunizeau, J. (2020). Choosing what we like vs liking what we choose: How choice-induced preference change might actually be instrumental to decision-making. PloS One, 15(5), e0231081. https://doi.org/10.1371/journal.pone.0231081

Lee, D. G., & Pezzulo, G. (2021). Evidence or Confidence: What Really Accumulates During a Decision? In bioRxiv (p. 2021.04.02.438275). https://doi.org/10.1101/2021.04.02.438275

Lichtenstein, S., & Fischhoff, B. (1977). Do those who know more also know more about how much they know? Organizational Behavior and Human Performance, 20(2), 159–183. https://doi.org/10.1016/0030-5073(77)90001-0

Lichtenstein, S., Fischhoff, B., & Phillips, L. D. (1982). Calibration of probabilities: The state of the art to 1980. In Judgment under Uncertainty: Heuristics and Biases (pp. 306–334). Cambridge University Press. https://doi.org/10.1017/CBO9780511809477.023

Locander, W. B., & Hermann, P. W. (1979). The Effect of Self-Confidence and Anxiety on Information Seeking in Consumer Risk Reduction. JMR, Journal of Marketing Research, 16(2), 268–274. https://doi.org/10.1177/002224377901600211

Loewenstein, G. (1994). The psychology of curiosity: A review and reinterpretation. Psychological Bulletin, 116(1), 75–98. https://doi.org/10.1037/0033-2909.116.1.75

Luhmann, C. C., Ishida, K., & Hajcak, G. (2011). Intolerance of uncertainty and decisions about delayed, probabilistic rewards. Behavior Therapy, 42(3), 378–386. https://doi.org/10.1016/j.beth.2010.09.002

Manea, L., Gilbody, S., & McMillan, D. (2012). Optimal cut-off score for diagnosing depression with the Patient Health Questionnaire (PHQ-9): a meta-analysis. CMAJ: Canadian Medical Association Journal = Journal de l'Association Medicale Canadienne, 184(3), E191–E196. https://doi.org/10.1503/cmaj.110829

Maniscalco, B., & Lau, H. (2012). A signal detection theoretic approach for estimating metacognitive sensitivity from confidence ratings. Consciousness and Cognition, 21(1), 422–430. https://doi.org/10.1016/j.concog.2011.09.021

Maniscalco, B., & Lau, H. (2014). Signal detection theory analysis of type 1 and type 2 data: Meta-d', response-specific meta-d', and the unequal variance SDT model. In S. M. Fleming & C. D. Frith (Eds.), The cognitive neuroscience of metacognition (Vol. 407, pp. 25–66). Springer-Verlag Publishing. https://doi.org/10.1007/978-3-642-45190-4_3

Mann, M., Hosman, C. M. H., Schaalma, H. P., & de Vries, N. K. (2004). Self-esteem in a broad-spectrum approach for mental health promotion. Health Education Research, 19(4), 357–372. https://doi.org/10.1093/her/cyg041

Markiewicz, Ł., & Kubińska, E. (2015). Information Use Differences in Hot and Cold Risk Processing: When Does Information About Probability Count in the Columbia Card Task? Frontiers in Psychology, 6, 1727. https://doi.org/10.3389/fpsyg.2015.01727

Martí, L., Mollica, F., Piantadosi, S., & Kidd, C. (2018). Certainty is primarily determined by past performance during concept learning. Open Mind, 2(2), 47–60. https://doi.org/10.1162/opmi_a_00017

Mason, O., Linney, Y., & Claridge, G. (2005). Short scales for measuring schizotypy. Schizophrenia Research, 78(2-3), 293–296. https://doi.org/10.1016/j.schres.2005.06.020

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), 509–527. https://doi.org/10.1037/a0014876

Massoni, S., Gajdos, T., & Vergnaud, J.-C. (2014). Confidence measurement in the light of signal detection theory. Frontiers in Psychology, 5, 1455. https://doi.org/10.3389/fpsyg.2014.01455

McKay, R., Langdon, R., & Coltheart, M. (2006). Need for closure, jumping to conclusions, and decisiveness in delusion-prone individuals. The Journal of Nervous and Mental Disease, 194(6), 422–426. https://doi.org/10.1097/01.nmd.0000221353.44132.25

Metcalfe, J., & Finn, B. (2008). Evidence that judgments of learning are causally related to study choice. Psychonomic Bulletin & Review, 15(1), 174–179. https://doi.org/10.3758/pbr.15.1.174

Mind. (2021). Coronavirus: the consequences for mental health. Mind. https://www.mind.org.uk/media/8962/the-consequences-of-coronavirus-for-mental-health-final-report.pdf

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "Frontal Lobe" tasks: a latent variable analysis. Cognitive Psychology, 41(1), 49–100. https://doi.org/10.1006/cogp.1999.0734

Montague, P. R., Dolan, R. J., Friston, K. J., & Dayan, P. (2012). Computational psychiatry. Trends in Cognitive Sciences, 16(1), 72–80. https://doi.org/10.1016/j.tics.2011.11.018

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. https://doi.org/10.1016/j.cogpsych.2015.01.002

Morein-Zamir, S., Shapher, S., Gasull-Camos, J., Fineberg, N. A., & Robbins, T. W. (2020). Avoid jumping to conclusions under uncertainty in Obsessive Compulsive Disorder. PloS One, 15(1), e0225970. https://doi.org/10.1371/journal.pone.0225970

Moritz, S., & Lysaker, P. H. (2018). Metacognition - What did James H. Flavell really say and the implications for the conceptualization and design of metacognitive interventions. Schizophrenia Research, 201, 20–26. https://doi.org/10.1016/j.schres.2018.06.001

Moritz, S., Ramdani, N., Klass, H., Andreou, C., Jungclaussen, D., Eifler, S., Englisch, S., Schirmbeck, F., & Zink, M. (2014). Overconfidence in incorrect perceptual judgments in patients with schizophrenia. Schizophrenia Research. Cognition, 1(4), 165–170. https://doi.org/10.1016/j.scog.2014.09.003

Moritz, S., & Woodward, T. S. (2006). The contribution of metamemory deficits to schizophrenia. Journal of Abnormal Psychology, 115(1), 15–25. https://doi.org/10.1037/0021-843X.15.1.15

Moritz, S., Woodward, T. S., & Ruff, C. C. (2003). Source monitoring and memory confidence in schizophrenia. Psychological Medicine, 33(1), 131–139. https://doi.org/10.1017/s0033291702006852

Moutoussis, M., Bentall, R. P., El-Deredy, W., & Dayan, P. (2011). Bayesian modelling of Jumping-to-Conclusions bias in delusional patients. Cognitive Neuropsychiatry, 16(5), 422–447. https://doi.org/10.1080/13546805.2010.548678

Murphy, P. R., Robertson, I. H., Harty, S., & O'Connell, R. G. (2015). Neural evidence accumulation persists after choice to inform metacognitive judgments. eLife, 4. https://doi.org/10.7554/eLife.11946

Mushtaq, F., Bland, A. R., & Schaefer, A. (2011). Uncertainty and cognitive control. Frontiers in Psychology, 2, 249. https://doi.org/10.3389/fpsyg.2011.00249

Navajas, J., Bahrami, B., & Latham, P. E. (2016). Post-decisional accounts of biases in confidence. Current Opinion in Behavioral Sciences, 11, 55–60. https://doi.org/10.1016/j.cobeha.2016.05.005

Navajas, J., Hindocha, C., Foda, H., Keramati, M., Latham, P. E., & Bahrami, B. (2017). The idiosyncratic nature of confidence. Nature Human Behaviour, 1(11), 810–818. https://doi.org/10.1038/s41562-017-0215-1

Nelson, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. Psychological Bulletin, 95(1), 109–133. https://doi.org/10.1037/0033-2909.95.1.109

Nelson, T. O., Gerler, D., & Narens, L. (n.d.). Accuracy of feeling-of-knowing judgments for predicting perceptual identification and relearning. Journal of Experimental Psychology. General, 113(2), 282–300. https://doi.org/10.1037/0096-3445.113.2.282

Nelson, T. O., & Narens, L. (1990). Metamemory: A Theoretical Framework and New Findings. In G. H. Bower (Ed.), Psychology of Learning and Motivation (Vol. 26, pp. 125–173). Academic Press. https://doi.org/10.1016/S0079-7421(08)60053-5

Nordahl, H., & Wells, A. (2017). Testing the metacognitive model against the benchmark CBT model of social anxiety disorder: Is it time to move beyond cognition? PloS One, 12(5), e0177109.

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

OED Online. (2021). confidence, n. Oxford English Dictionary Online. https://www.oed.com/view/Entry/38806?rskey=d5ixQM&result=1

Overgaard, M., & Sandberg, K. (2012). Kinds of access: different methods for report reveal different kinds of metacognitive access. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 367(1594), 1287–1296. https://doi.org/10.1098/rstb.2011.0425

Pallier, G., Wilkinson, R., Danthiir, V., Kleitman, S., Knezevic, G., Stankov, L., & Roberts, R. D. (2002). The role of individual differences in the accuracy of confidence judgments. The Journal of General Psychology, 129(3), 257–299. https://doi.org/10.1080/00221300209602099

Parés-Pujolràs, E., Hatchuel, J., & Haggard, P. (2020). A neural turning point - the EEG P3 component tracks unfolding changes of mind. In bioRxiv (p. 2020.11.18.388363). https://doi.org/10.1101/2020.11.18.388363

Pasquini, L., Steynor, A., Waagsaether, K., & CSAG. (2019). The Psychology of Decision-Making Under Uncertainty: A Literature Review (No. AID-OAA-I-14-00013). United States Agency for International Development. https://www.climatelinks.org/sites/default/files/asset/document/2019_USAID_ATLAS_The%20Psychology%20of%20Decision-Making_Literature%20Review.pdf

Perry, C. J., & Barron, A. B. (2013). Honey bees selectively avoid difficult choices. Proceedings of the National Academy of Sciences of the United States of America, 110(47), 19155–19159. https://doi.org/10.1073/pnas.1314571110

Persaud, N., McLeod, P., & Cowey, A. (2007). Post-decision wagering objectively measures awareness. Nature Neuroscience, 10(2), 257–261. https://doi.org/10.1038/nn1840

Pescetelli, N., Hauperich, A.-K., & Yeung, N. (2021). Confidence, advice seeking and changes of mind in decision making. Cognition, 215, 104810. https://doi.org/10.1016/j.cognition.2021.104810

Peters, A., McEwen, B. S., & Friston, K. (2017). Uncertainty and stress: Why it causes diseases and how it is mastered by the brain. Progress in Neurobiology, 156, 164–188. https://doi.org/10.1016/j.pneurobio.2017.05.004

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

Petrusic, W. M., & Baranski, J. V. (2003). Judging confidence influences decision processing in comparative judgments. Psychonomic Bulletin & Review, 10(1), 177–183. https://doi.org/10.3758/bf03196482

Petzschner, F. H., Weber, L. A. E., Gard, T., & Stephan, K. E. (2017). Computational Psychosomatics and Computational Psychiatry: Toward a Joint Framework for Differential Diagnosis. Biological Psychiatry, 82(6), 421–430. https://doi.org/10.1016/j.biopsych.2017.05.012

Phelps, E. A., Ling, S., & Carrasco, M. (2006). Emotion facilitates perception and potentiates the perceptual benefits of attention. Psychological Science, 17(4), 292–299. https://doi.org/10.1111/j.1467-9280.2006.01701.x

Phillips, B., R. Prybutok, V., & A. Peak, D. (2014). Decision confidence, information usefulness, and information seeking intention in the presence of disconfirming information. Information Sciences, 17, 001–024. https://doi.org/10.28945/1932

Phillips, L. D., & Edwards, W. (1966). Conservatism in a simple probability inference task. Journal of Experimental Psychology, 72(3), 346–354. https://doi.org/10.1037/h0023653

Piaget, J. (1952). The Origins of Intelligence in Children. International Universities Press. https://sites.pitt.edu/~strauss/origins_r.pdf

Pleskac, T. J., & Busemeyer, J. R. (2010). Two-stage dynamic signal detection: a theory of choice, decision time, and confidence. Psychological Review, 117(3), 864–901. https://doi.org/10.1037/a0019737

Pouget, A., Drugowitsch, J., & Kepecs, A. (2016). Confidence and certainty: distinct probabilistic quantities for different goals. Nature Neuroscience, 19(3), 366–374. https://doi.org/10.1038/nn.4240

Pulford, B. D., Colman, A. M., Buabang, E. K., & Krockow, E. M. (2018). The persuasive power of knowledge: Testing the confidence heuristic. Journal of Experimental Psychology. General, 147(10), 1431–1444. https://doi.org/10.1037/xge0000471

Quinlan, J. R. (1986). Induction of decision trees. Machine Learning, 1(1), 81–106. https://doi.org/10.1007/BF00116251

Rachman, S. (2002). A cognitive theory of compulsive checking. Behaviour Research and Therapy, 40(6), 625–639. https://doi.org/10.1016/s0005-7967(01)00028-6

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. https://doi.org/10.1038/4580

Rastegary, H., & Landy, F. J. (1993). The Interactions among Time Urgency, Uncertainty, and Time Pressure. In O. Svenson & A. J. Maule (Eds.), Time Pressure and Stress in Human Judgment and Decision Making (pp. 217–239). Plenum Press.

Rausch, M., Hellmann, S., & Zehetleitner, M. (2018). Confidence in masked orientation judgments is informed by both evidence and visibility. Attention, Perception & Psychophysics, 80(1), 134–154. https://doi.org/10.3758/s13414-017-1431-5

Rausch, M., & Zehetleitner, M. (2019). The folded X-pattern is not necessarily a statistical signature of decision confidence. PLoS Computational Biology, 15(10), e1007456. https://doi.org/10.1371/journal.pcbi.1007456

Robinson, G., & Freeston, M. (2015). Intolerance of Uncertainty as a Predictor of Performance Anxiety and Robustness of Sport Confidence in University Student-Athletes. Journal of Clinical Sport Psychology, 9(4), 335–344. https://doi.org/10.1123/jcsp.2015-0008

Rodrigue, J. R., Olson, K. R., & Markley, R. P. (1987). Induced mood and curiosity. Cognitive Therapy and Research, 11(1), 101–106. https://doi.org/10.1007/BF01183135

Rodriguez Cabrero, J. A. M., Zhu, J.-Q., & Ludvig, E. A. (2019). Costly curiosity: People pay a price to resolve an uncertain gamble early. Behavioural Processes, 160, 20–25. https://doi.org/10.1016/j.beproc.2018.12.015

Roebers, C. M. (2017). Executive function and metacognition: Towards a unifying framework of cognitive self-regulation. Developmental Review: DR, 45, 31–51. https://doi.org/10.1016/j.dr.2017.04.001

Rollwage, M., & Fleming, S. M. (2021). Confirmation bias is adaptive when coupled with efficient metacognition. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 376(1822), 20200131. https://doi.org/10.1098/rstb.2020.0131

Rollwage, M., Loosen, A., Hauser, T. U., Moran, R., Dolan, R. J., & Fleming, S. M. (2020). Confidence drives a neural confirmation bias. Nature Communications, 11(1), 2634. https://doi.org/10.1038/s41467-020-16278-6

Rosati, A. G., & Hare, B. (2011). Chimpanzees and bonobos distinguish between risk and ambiguity. Biology Letters, 7(1), 15–18. https://doi.org/10.1098/rsbl.2010.0927

Rosati, A. G., & Santos, L. R. (2016). Spontaneous Metacognition in Rhesus Monkeys. Psychological Science, 27(9), 1181–1191. https://doi.org/10.1177/0956797616653737

Rosen, N. O., & Knäuper, B. (2009). A little uncertainty goes a long way: state and trait differences in uncertainty interact to increase information seeking but also increase worry. Health Communication, 24(3), 228–238. https://doi.org/10.1080/10410230902804125

Rouault, M., Dayan, P., & Fleming, S. M. (2019). Forming global estimates of self-performance from local confidence. Nature Communications, 10(1), 1141. https://doi.org/10.1038/s41467-019-09075-3

Rouault, M., Seow, T., Gillan, C. M., & Fleming, S. M. (2018). Psychiatric Symptom Dimensions 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

Salvador, A., Arnal, L. H., Vinckier, F., Domenech, P., Gaillard, R., & Wyart, V. (2020). Premature commitment to uncertain beliefs during human NMDA receptor hypofunction. In bioRxiv (p. 2020.06.17.156539). https://doi.org/10.1101/2020.06.17.156539

Samaha, J., & Denison, R. (2020). The positive evidence bias in perceptual confidence is not post-decisional. In bioRxiv (p. 2020.03.15.991513). https://doi.org/10.1101/2020.03.15.991513

Sanders, J. I., Hangya, B., & Kepecs, A. (2016). Signatures of a Statistical Computation in the Human Sense of Confidence. Neuron, 90(3), 499–506. https://doi.org/10.1016/j.neuron.2016.03.025

Schlenker, B. R., & Leary, M. R. (1982). Audiences' reactions to self-enhancing, self-denigrating, and accurate self-presentations. Journal of Experimental Social Psychology, 18(1), 89–104. https://doi.org/10.1016/0022-1031(82)90083-X

Schulz, L., Rollwage, M., Dolan, R. J., & Fleming, S. M. (2020). Dogmatism manifests in lowered information search under uncertainty. Proceedings of the National Academy of Sciences of the United States of America, 117(49), 31527–31534. https://doi.org/10.1073/pnas.2009641117

Shannon, C. E. (1948). A mathematical theory of communication. The Bell System Technical Journal, 27(3), 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Sharot, T., & Sunstein, C. R. (2020). How people decide what they want to know. Nature Human Behaviour, 4(1), 14–19. https://doi.org/10.1038/s41562-019-0793-1

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

Shimamura, A. P. (2000). Toward a cognitive neuroscience of metacognition [Review of Toward a cognitive neuroscience of metacognition]. Consciousness and Cognition, 9(2 Pt 1), 313–323; discussion 324–326. https://doi.org/10.1006/ccog.2000.0450

Shimamura, A. P. (2008). A neurocognitive approach to metacognitive monitoring and control. In Handbook of Metamemory and Memory. Routledge. https://doi.org/10.4324/9780203805503.ch19

Smith, J. D., Schull, J., Strote, J., McGee, K., Egnor, R., & Erb, L. (1995). The uncertain response in the bottlenosed dolphin (Tursiops truncatus). Journal of Experimental Psychology. General, 124(4), 391–408. https://doi.org/10.1037//0096-3445.124.4.391

Smith, J. D., Shields, W. E., Schull, J., & Washburn, D. A. (1997). The uncertain response in humans and animals. Cognition, 62(1), 75–97. https://doi.org/10.1016/s0010-0277(96)00726-3

Smith, J. D., & Washburn, D. A. (2005). Uncertainty Monitoring and Metacognition by Animals. Current Directions in Psychological Science, 14(1), 19–24. https://doi.org/10.1111/j.0963-7214.2005.00327.x

Smock, C. D. (1955). The influence of psychological stress on the “intolerance of ambiguity.” Journal of Abnormal and Social Psychology, 50(2), 177–182. https://doi.org/10.1037/h0047951

Spielberger, C. D., Gorsuch, R. L., Lushene, R., Vagg, P. R., & Jacobs, G. A. (1983). State-Trait Anxiety Inventory for Adults Manual. Mind Garden.

Stankov, L., Morony, S., & Lee, Y. P. (2014). Confidence: the best non-cognitive predictor of academic achievement? Educational Psychology Review, 34(1), 9–28. https://doi.org/10.1080/01443410.2013.814194

Steinhauser, M., & Yeung, N. (2010). Decision processes in human performance monitoring. The Journal of Neuroscience: The Official Journal of the Society for Neuroscience, 30(46), 15643–15653. https://doi.org/10.1523/JNEUROSCI.1899-10.2010

Stigler, G. J. (1961). The Economics of Information. The Journal of Political Economy, 69(3), 213–225. http://www.jstor.org/stable/1829263

Strauss, A. Y., Fradkin, I., McNally, R. J., Linkovski, O., Anholt, G. E., & Huppert, J. D. (2020). Why check? A meta-analysis of checking in obsessive-compulsive disorder: Threat vs. distrust of senses. Clinical Psychology Review, 75, 101807. https://doi.org/10.1016/j.cpr.2019.101807

The Health Foundation. (2021, May 5). Latest data highlights a growing mental health crisis in the UK. The Health Foundation. https://www.health.org.uk/news-and-comment/news/latest-data-highlights-a-growing-mental-health-crisis-in-the-uk

Thura, D., Beauregard-Racine, J., Fradet, C.-W., & Cisek, P. (2012). Decision making by urgency gating: theory and experimental support. Journal of Neurophysiology, 108(11), 2912–2930. https://doi.org/10.1152/jn.01071.2011

Trudel, N., Scholl, J., Klein-Flügge, M. C., Fouragnan, E., Tankelevitch, L., Wittmann, M. K., & Rushworth, M. F. S. (2021). Polarity of uncertainty representation during exploration and exploitation in ventromedial prefrontal cortex. Nature Human Behaviour, 5(1), 83–98. https://doi.org/10.1038/s41562-020-0929-3

Twomey, D. M., Murphy, P. R., Kelly, S. P., & O'Connell, R. G. (2015). The classic P300 encodes a build-to-threshold decision variable. The European Journal of Neuroscience, 42(1), 1636–1643. https://doi.org/10.1111/ejn.12936

van den Berg, R., Zylberberg, A., Kiani, R., Shadlen, M. N., & Wolpert, D. M. (2016). Confidence Is the Bridge between Multi-stage Decisions. Current Biology: CB, 26(23), 3157–3168. https://doi.org/10.1016/j.cub.2016.10.021

van den Hout, M., & Kindt, M. (2003). Repeated checking causes memory distrust. Behaviour Research and Therapy, 41(3), 301–316. https://doi.org/10.1016/s0005-7967(02)00012-8

van Es, T. (2020). Minimizing prediction errors in predictive processing: from inconsistency to non-representationalism. Phenomenology and the Cognitive Sciences, 19(5), 997–1017. https://doi.org/10.1007/s11097-019-09649-y

van Lieshout, L. L. F., de Lange, F. P., & Cools, R. (2021). Uncertainty increases curiosity, but decreases happiness. Scientific Reports, 11(1), 14014. https://doi.org/10.1038/s41598-021-93464-6

Warman, D. M. (2008). Reasoning and delusion proneness: confidence in decisions. The Journal of Nervous and Mental Disease, 196(1), 9–15. https://doi.org/10.1097/NMD.0b013e3181601141

Wells, A. (2000). Emotional disorders and metacognition: Innovative cognitive therapy. John Wiley & Sons. https://psycnet.apa.org/fulltext/2001-18281-000.pdf

Wells, A., Fisher, P., Myers, S., Wheatley, J., Patel, T., & Brewin, C. R. (2012). Metacognitive therapy in treatment-resistant depression: a platform trial. Behaviour Research and Therapy, 50(6), 367–373. https://doi.org/10.1016/j.brat.2012.02.004

White, J. K., Bromberg-Martin, E. S., Heilbronner, S. R., Zhang, K., Pai, J., Haber, S. N., & Monosov, I. E. (2019). A neural network for information seeking. Nature Communications, 10(1), 5168. https://doi.org/10.1038/s41467-019-13135-z

Whitley, B. E., & Greenberg, M. S. (1986). The role of eyewitness confidence in juror perceptions of credibility. Journal of Applied Social Psychology, 16(5), 387–409. https://doi.org/10.1111/j.1559-1816.1986.tb01148.x

Yeung, N., & Summerfield, C. (2012). Metacognition in human decision-making: confidence and error monitoring. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 367(1594), 1310–1321. https://doi.org/10.1098/rstb.2011.0416

Yu, A. J. (2015). Decision-Making Tasks. In D. Jaeger & R. Jung (Eds.), Encyclopedia of Computational Neuroscience (pp. 931–937). Springer New York. https://doi.org/10.1007/978-1-4614-6675-8_314

Yu, S., Pleskac, T. J., & Zeigenfuse, M. D. (2015). Dynamics of postdecisional processing of confidence. Journal of Experimental Psychology. General, 144(2), 489–510. https://doi.org/10.1037/xge0000062

Zadra, J. R., & Clore, G. L. (2011). Emotion and perception: the role of affective information. Wiley Interdisciplinary Reviews. Cognitive Science, 2(6), 676–685. https://doi.org/10.1002/wcs.147

Zhang, L., Wang, K., Zhu, C., Yu, F., & Chen, X. (2015). Trait Anxiety Has Effect on Decision Making under Ambiguity but Not Decision Making under Risk. PloS One, 10(5), e0127189. https://doi.org/10.1371/journal.pone.0127189

Zhang, Z., Wang, M., Miao, X., Li, Y., Hitchman, G., & Yuan, Z. (2017). Individuals with high obsessive-compulsive tendencies or undermined confidence rely more on external proxies to access their internal states. Journal of Behavior Therapy and Experimental Psychiatry, 54, 263–269. https://doi.org/10.1016/j.jbtep.2016.09.003

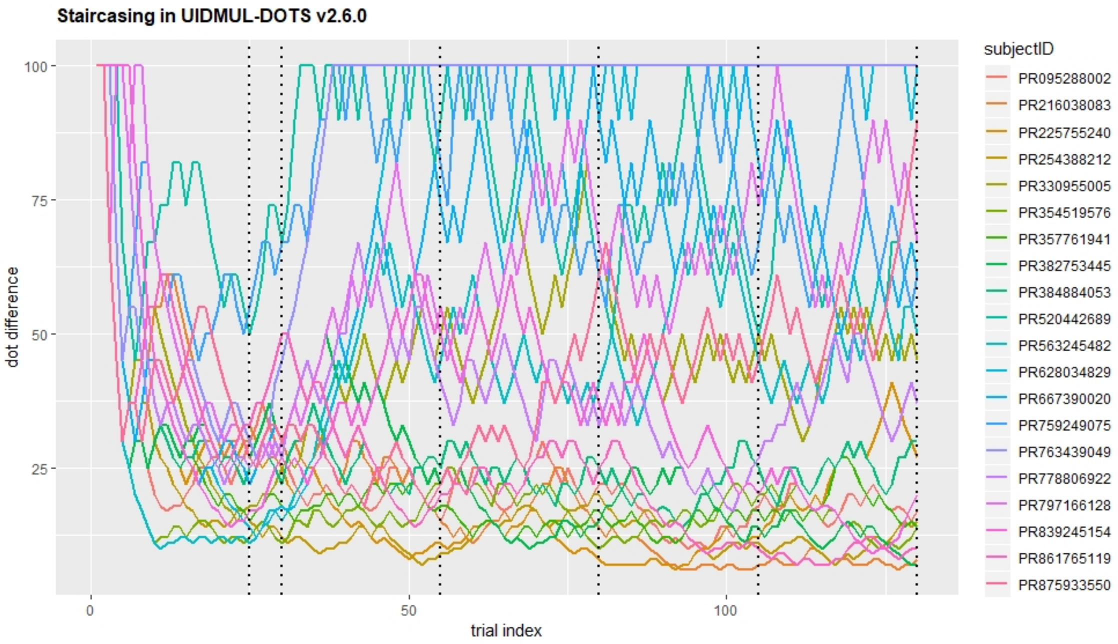
Zuckerman, D. M. (1989). Stress, self-esteem, and mental health: How does gender make a difference? Sex Roles: A Journal of Research, 20(7), 429–444. https://doi.org/10.1007/BF00288001

Zylberberg, A., Fetsch, C. R., & Shadlen, M. N. (2016). The influence of evidence volatility on choice, reaction time and confidence in a perceptual decision. eLife, 5. https://doi.org/10.7554/eLife.17688

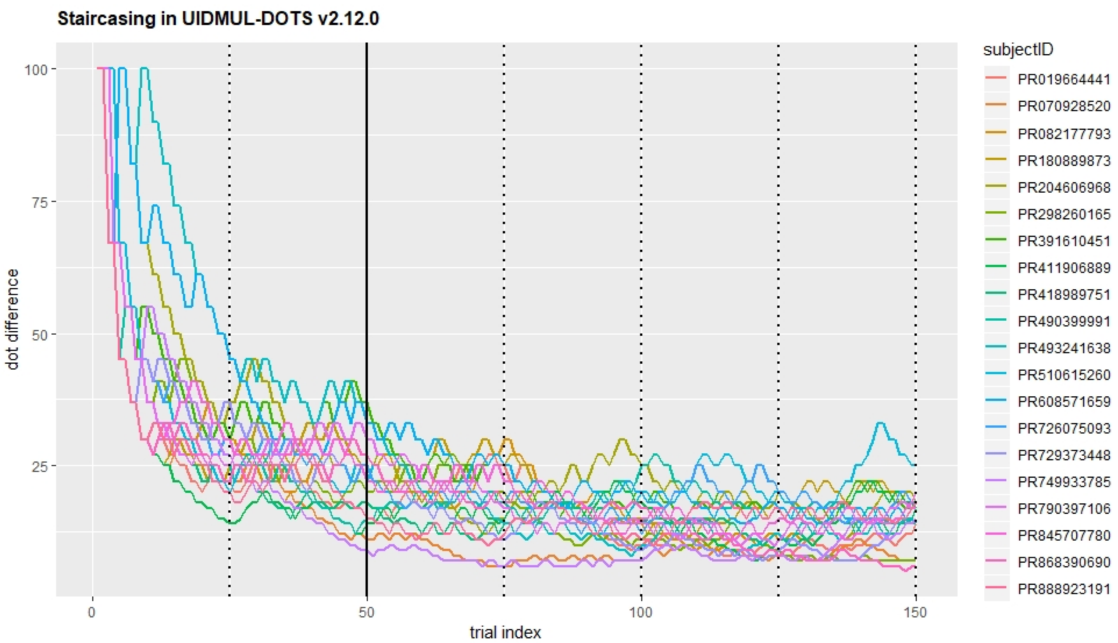
Appendix
SUPPLEMENTARY MATERIAL

S2. Chapter 2 supplementary material

S2.1 Staircasing in DDTi Pilots 1 and 2



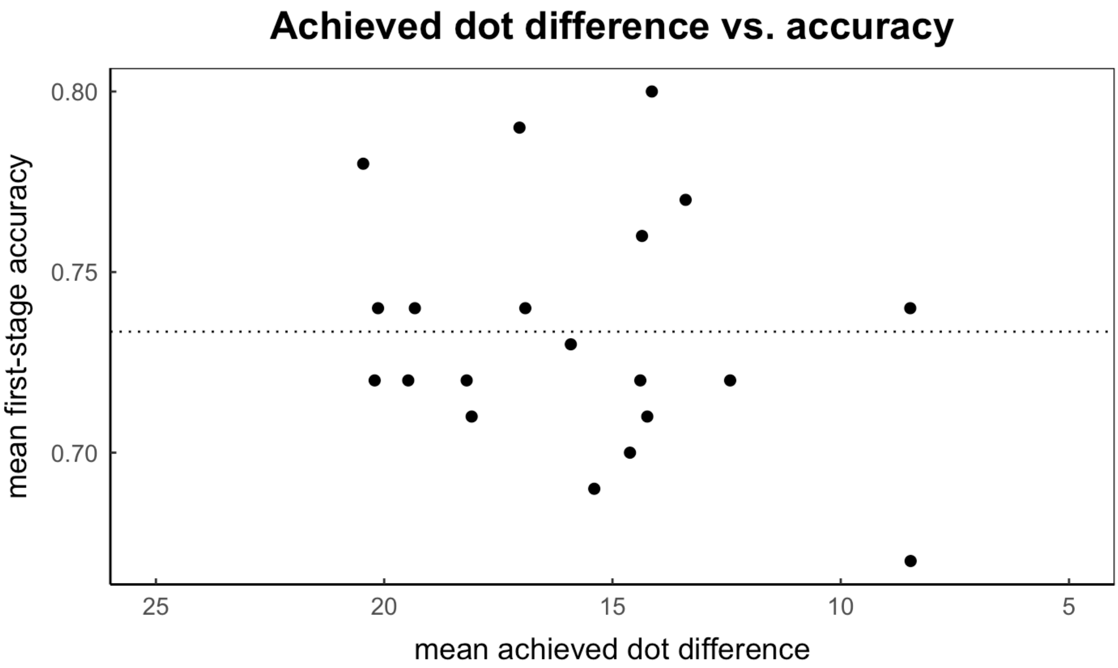
A



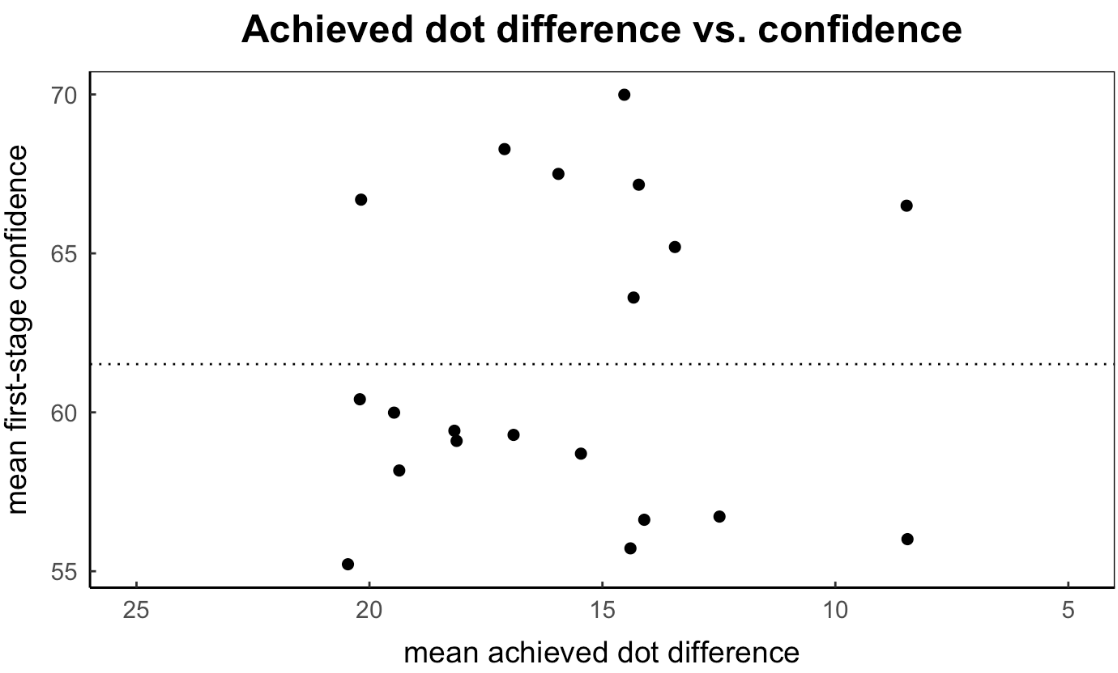
B

Figure S2.1. Dot difference selected by the staircasing procedure by trial index in DDTi Pilots 1 (A) and 2 (B). Black vertical line represents the end of the practice/calibration block. Each line is an individual participant in the study.

Difficulty was not varied systematically within or between participants during the DDTi. The staircasing procedure controlling dot difference was designed to ensure that task difficulty, and thus first-stage accuracy, would be similar between participants. Participants were similarly accurate [$r(18) = .22$, $p = .352$] and confident [$r(18) = -.08$, $p = .734$] on the first-stage decision regardless of the absolute value of the dot difference that they settled on after the burn-in period [Figure S2.2].



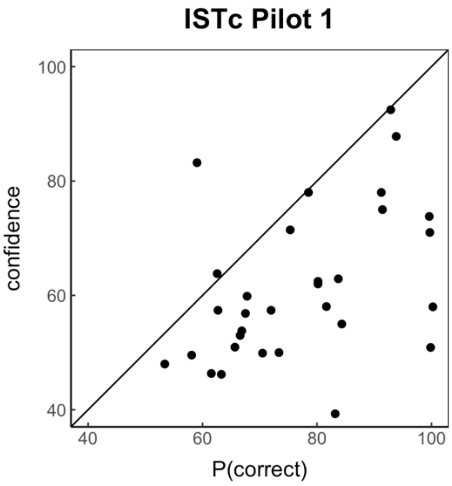
A



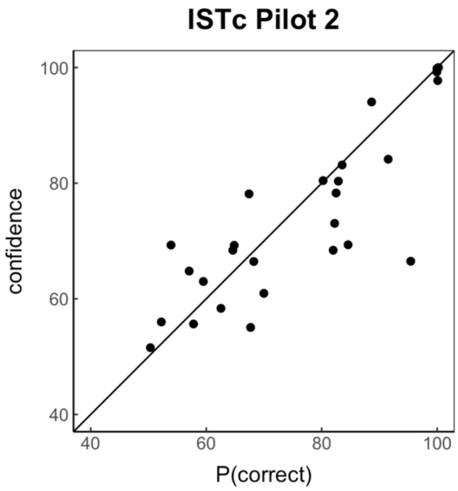
B

Figure S2.2. Accuracy (A) and confidence (B) in the first-stage decision by achieved dot difference (i.e., dot difference after the calibration phase) in DDTi Pilot 2. Dotted horizontal lines represent mean first-stage accuracy and confidence across participants, respectively.

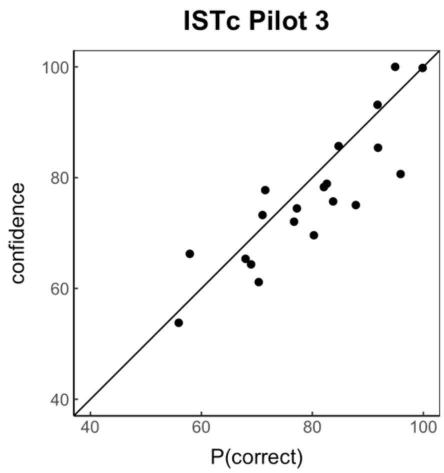
S2.2 Calibration plots for ISTc Pilots 1-3



A



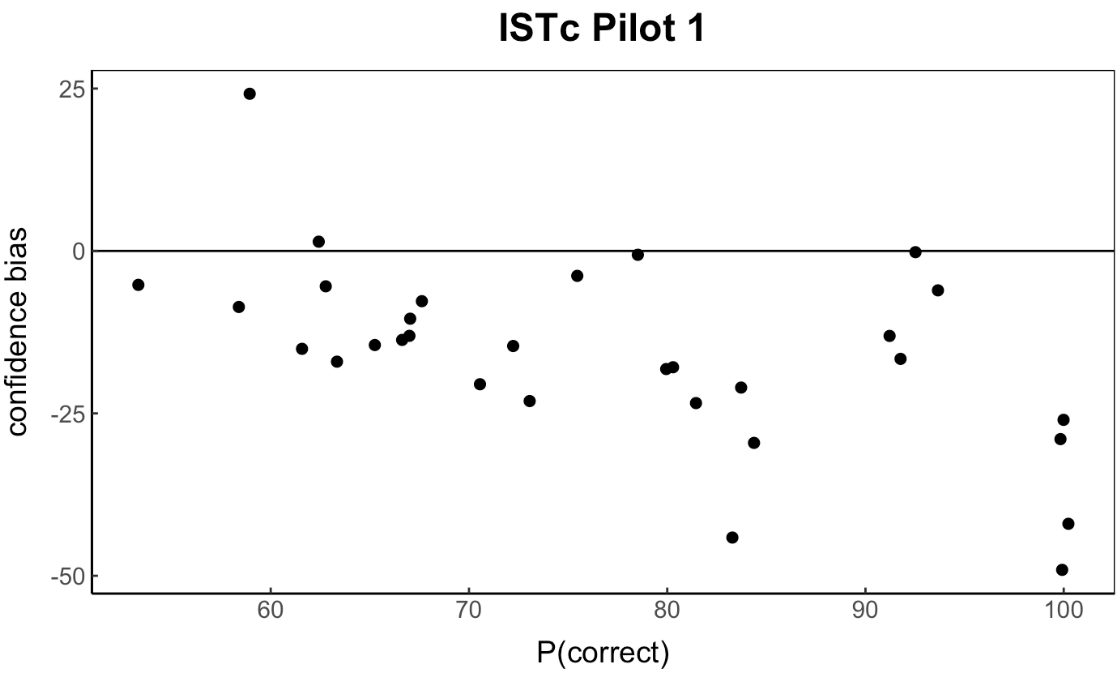
B



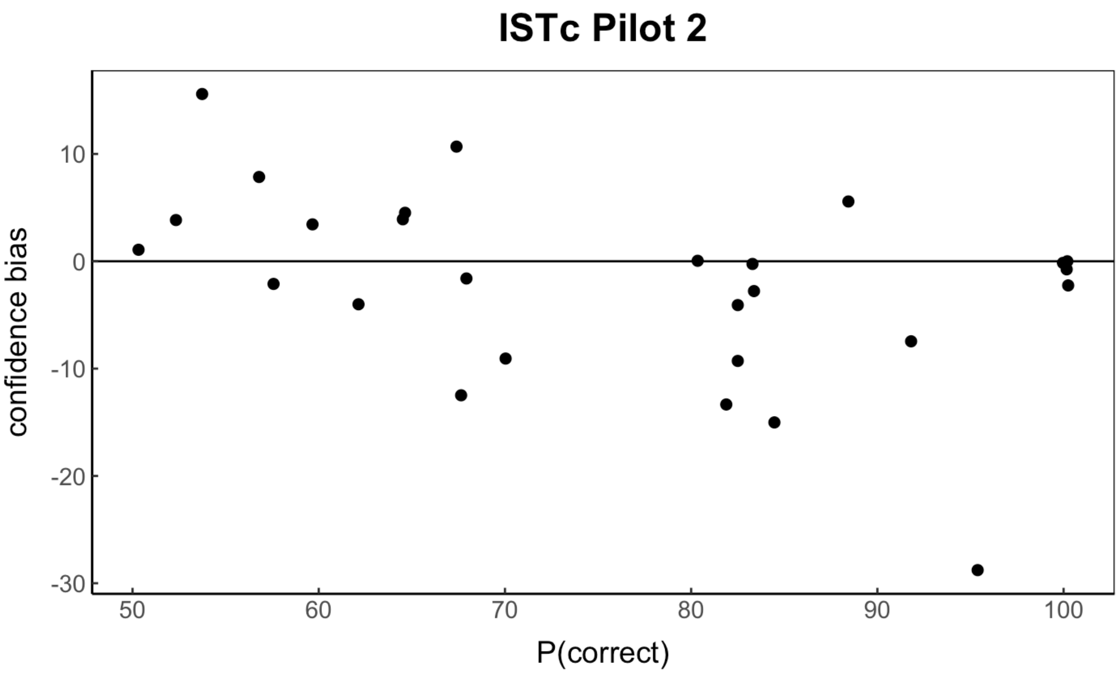
C

Figure S2.3. Plot of participants' average reported confidence at the point of decision against their average P(correct) at the point of decision (scaled by a factor of 100 to match confidence) for trials in ISTc Pilots 1 (A), 2 (B), and 3 (C). Each dot in the plots represents a single participant. Diagonal line represents perfect correspondence (calibration) between P(correct) and confidence.

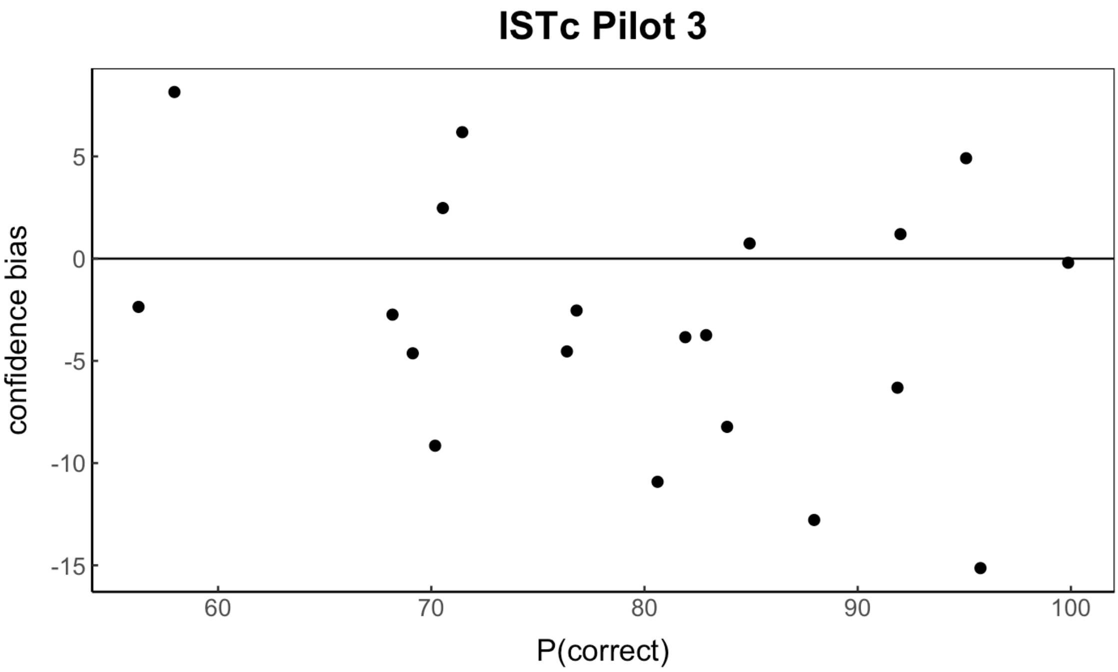
S2.3 Confidence bias plots for ISTc Pilots 1-3



A



B



C

Figure S2.4. Plot of participants' average decision-point confidence bias (reported decision-point confidence minus decision-point P(correct)) against their average P(correct) at the point of decision for trials in ISTc Pilot 1 (A), 2 (B), and 3 (C). Each dot in the plots represents a single participant. Negative bias values indicate underconfidence; positive bias values indicate overconfidence.

S3. Chapter 3 supplementary material

S3.1 Preregistration document for Experiment 1

Experiment 1 was preregistered with an AsPredicted.org template on OSF on 7 November 2018. It is now publicly available at https://osf.io/pr6nv.

S3.2 Exclusion criteria summary for Experiments 1 and 2

Table S3.1

Participant exclusion criteria and excluded counts in Experiments 1 and 2

	Number of participants excluded	
	Experiment 1	Experiment 2
Overall		
Reported technical difficulties	4	4
Missing data from one or more psychiatric questionnaires	13	19
ISTc		
Failed to complete all 20 trials	0	1
All decision-point confidence ratings in the range of 49 to 51	0	0
More than 2 trials in either score rule condition (FW / DW) where the confidence rating value did not correspond to the accuracy of the response on that trial (i.e., confidence greater than 50 if incorrect and less than 50 if correct)	3	5
DDTi		
Task accuracy fell outside the range of 60-85% correct	0	n/a
Selected the same confidence rating on every trial	0	n/a
Total	18	25
Original dataset size	153	223
Final dataset size	135	198

Participant excluded if they meet one or more of the criteria listed above.

Table S3.2

Trial exclusion criteria and excluded counts in Experiments 1 and 2

	Number of trials excluded	
	Experiment 1	Experiment 2
ISTc		
At decision-point but confidence rating value does not correspond to the accuracy of the response on that trial (i.e., confidence greater than 50 if incorrect and less than 50 if correct)	51 (0.49%)	71 (0.43%)
DDTi		
Trial RT exceeds 10 seconds or 3 standard deviations from mean RT	621 (2.06%)	n/a

Data reported as count (percentage of total trials for the task).

S3.3 Mediation analyses for Experiment 1

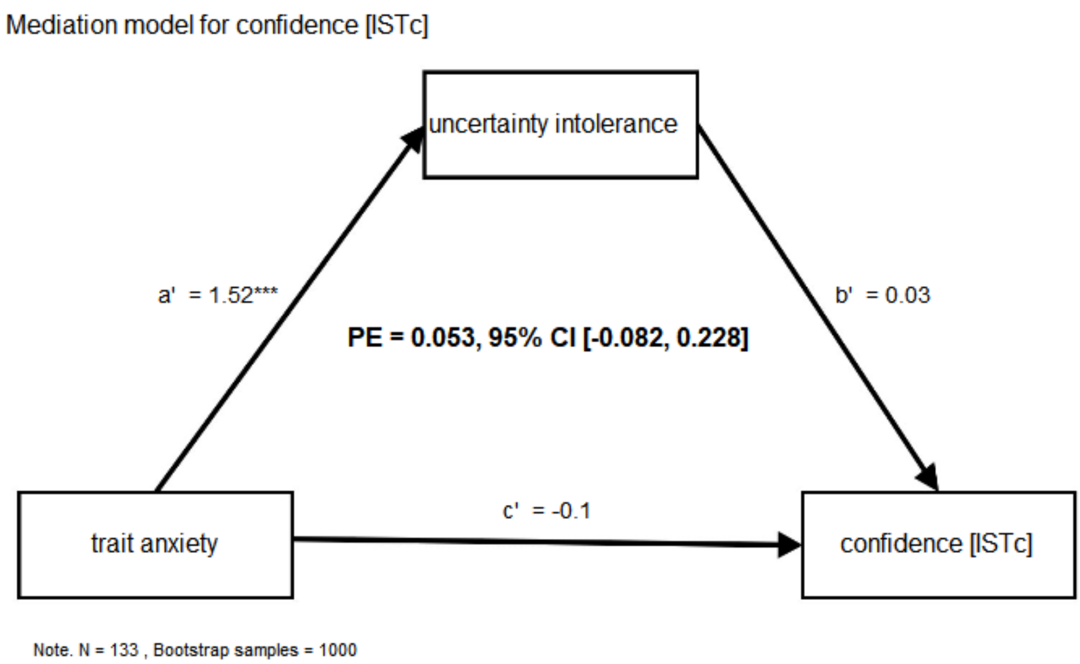


Figure S3.1. Mediation model with uncertainty intolerance (IUS-27 score) as a mediator for the relationship between trait anxiety (STAI score) and confidence (decision-point confidence minus P(correct)) on the ISTc.

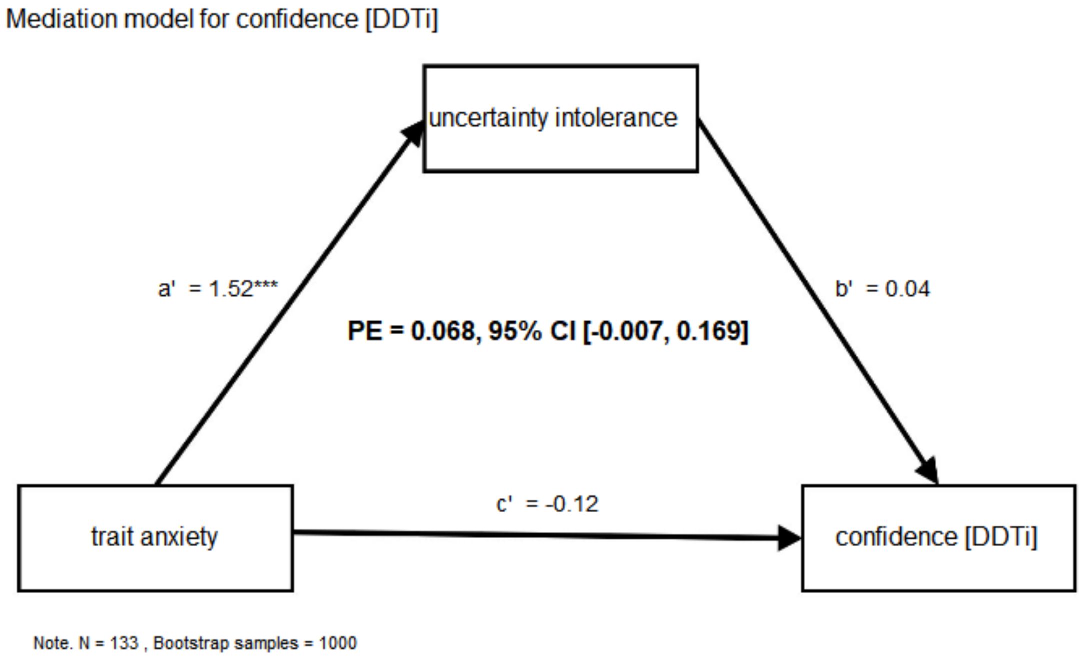


Figure S3.2. Mediation model with uncertainty intolerance (IUS-27 score) as a mediator for the relationship between trait anxiety (STAI score) and confidence (first-stage decision confidence) on the DDTi.

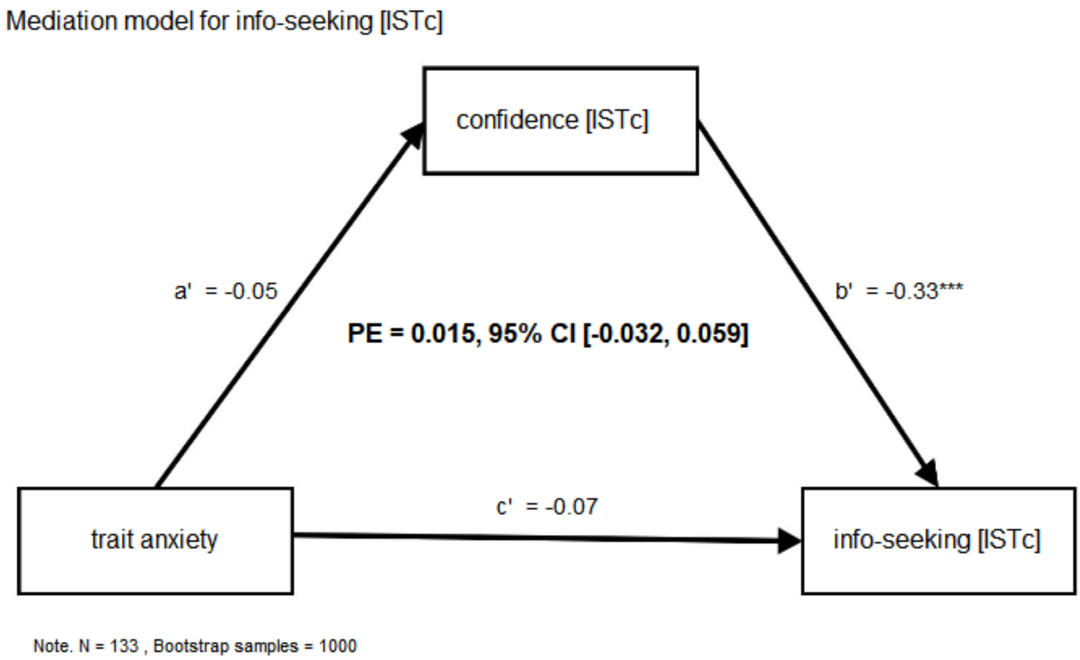


Figure S3.3. Mediation model with confidence (decision-point confidence minus P(correct)) as a mediator for the relationship between trait anxiety (STAI score) and confidence (number of tiles seen before decision) on the ISTc.

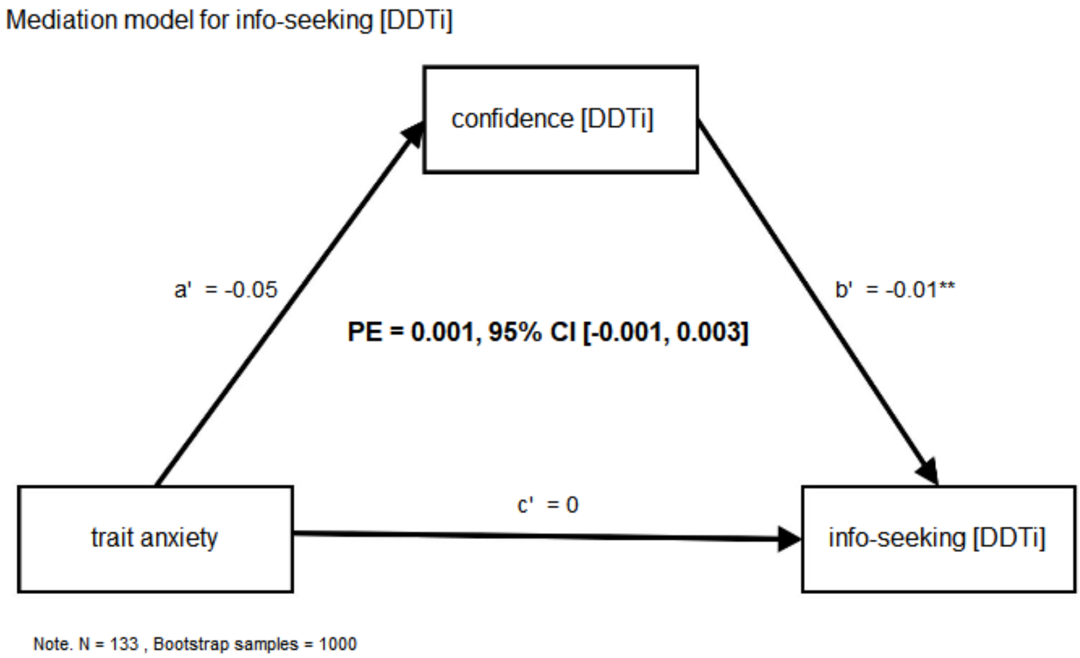


Figure S3.4. Mediation model with confidence (first-stage decision confidence) as a mediator for the relationship between trait anxiety (STAI score) and confidence ('See Again' choice proportion) on the DDTi.

S3.4 Structural equation model for Experiment 1

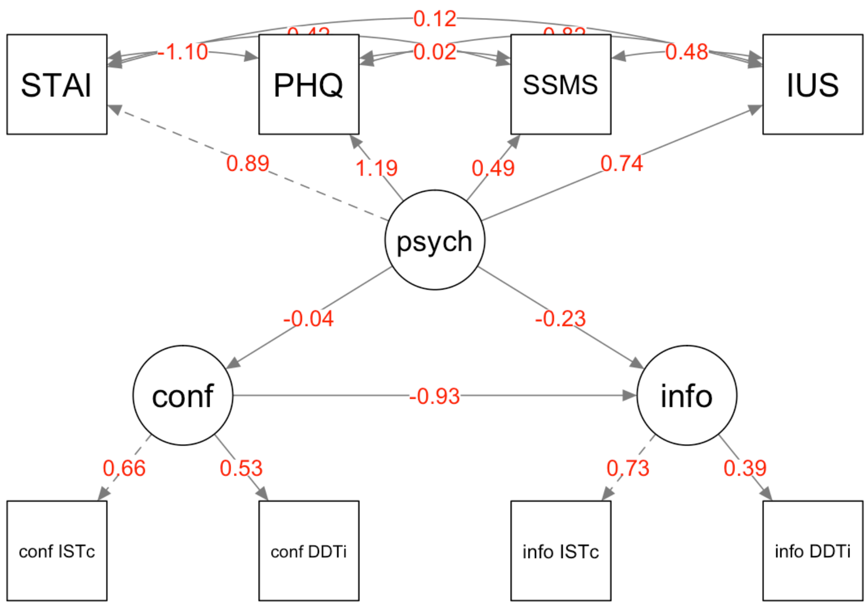


Figure S3.5. Revised structural equation model for Experiment 1 measures, with added covariation between the questionnaire measures. The circles represent the latent variables: psychiatric symptomatology (psych), subjective confidence (conf), and information-seeking strategy (info). The rectangles represent our measured variables: four self-report questionnaire scores (STAI, PHQ, SSMS, and IUS) and four first-order decision process measures across our two tasks (ISTc confidence, DDTi confidence, ISTc information-seeking, DDTi information-seeking). Dotted lines indicate fixed effects.

S4. Chapter 4 supplementary material

S4.1 Preregistration document for Experiment 3

Experiment 3 was preregistered on AsPredicted.org on 30 May 2020 (reg. no. 42046). The document is currently a private link on AsPredicted (https://aspredicted.org/CBB_VVJ). A copy of the document is provided below:

1) Have any data been collected for this study already?

No, no data have been collected for this study yet.

2) What's the main question being asked or hypothesis being tested in this study?

In a previous experiment using the Information Sampling Task with Confidence (ISTc — adapted from the IST introduced by Clark et al., 2006), we found that people who report greater intolerance of uncertainty (IUS-27) saw fewer tiles on average before making a decision about the majority colour on the ISTc (as did people reporting greater depressive severity via the PHQ-9). This surprising result — intuitively, one would expect more information seeking, rather than less, in people who express greater intolerance of uncertainty — led us to hypothesise that there are two methods of reducing uncertainty on the ISTc:

through increased information-seeking before decision (i.e., choosing to see more tiles before decision) and through increased feedback-seeking after decision (i.e., choosing to end a trial earlier to get feedback on whether the response was correct or not). People who are more intolerant of uncertainty and/or have stronger symptoms of negative affect may be more motivated to reduce uncertainty quickly, and thus favour the second method of increased feedback-seeking.

We are testing whether people's behavioural performance on the ISTc reflects these two distinct methods of reducing uncertainty: information-seeking versus feedback-seeking. We hypothesise that, on ISTc trials for which post-decision feedback is delayed, participants will choose to see more tiles on average than on ISTc trials for which post-decision feedback is immediately available.

3) Describe the key dependent variable(s) specifying how they will be measured.

The Information Sampling Task with Confidence (ISTc) is a probabilistic reasoning task adapted from the well-characterised information sampling task described by Clark et al. (2006). The ISTc additionally has participants rate their subjective confidence during the sampling phase as well as at the point of decision. Participants are allowed to turn over as many tiles as they like before committing to a binary choice about the majority colour.

In the trial-by-trial feedback block, participants are told whether how many points they have earned on each trial immediately after response. In the delayed feedback block, participants are not given trial-by-trial feedback and instead are told the total number of points they have earned on the block only at the end of the block.

To assess our hypothesis, we will look at the average number of tiles seen across trials in trial-by-trial feedback and delayed feedback trials separately.

4) How many and which conditions will participants be assigned to?

The main experimental manipulation is within-participants, relating to whether feedback is provided trial-by-trial (trial-by-trial feedback condition) or only at the end of a block of trials (delayed feedback condition). The order of these two conditions is counterbalanced in a between-participants design: the block order (delayed - trial-by-trial, trial-by-trial - delayed) is randomly assigned and counterbalanced between participants. Both blocks consist of 10 trials and use a "fixed win" (FW) scoring rule (as defined in previous research).

5) Specify exactly which analyses you will conduct to examine the main question/hypothesis.

To assess our hypothesis, we will run a paired Bayes t-test to compare the mean number of tiles seen in trial-by-trial feedback versus delayed feedback trials.

6) Describe exactly how outliers will be defined and handled, and your precise rule(s) for excluding observations.

Participants are pre-screened on Prolific based on age (excluded if outside our required age bracket of 18 to 60 years) and vision (excluded if they do not have normal or corrected-to-normal vision) requirements.

Participants will be excluded from analysis if:

- all their confidence ratings are in the range of 49 to 51;
- data are not available for all trials (total: 20) of the experiment;
- they show 2 or more instances across all trials of the experiment (total: 20) of discrepancy between majority colour choice as indicated by their rated confidence at decision-point (where confidence greater than or equal to 50 indicates confidence rated in favour of the correct choice and confidence less than 50 indicates confidence rated in favour of the incorrect choice) and their actual choice behaviour.

7) How many observations will be collected or what will determine sample size? No need to justify decision, but be precise about exactly how the number will be determined.

We will recruit up to 55 participants for this study, in batches of 10 participants. We will continue running batches until the BF on the main contrast of interest falls above 3 or below 1/3, or until the maximum number of participants is reached. The maximum sample size is based on a G*Power estimate of 54 participants to observe, in a two-tailed paired t-test, an effect size (Cohen's d) of 0.5 at alpha = .05 and 95% power.

8) Anything else you would like to pre-register? (e.g., secondary analyses, variables collected for exploratory purposes, unusual analyses planned?)

To check for block order effects (trial-by-trial - delayed vs. delayed - trial-by-trial), we will run a one-way ANOVA on the mean number of tiles seen. If the interaction between condition (trial-by-trial vs. delayed) and condition order is significant, we will subsequently run separate paired Bayes t-tests comparing the mean number of tiles seen in trial-by-trial versus delayed feedback trials for the different order subgroups.

In addition, we will check the magnitude of the decision-evoked confidence boost effect seen in previous experiments by fitting each participant's data from trial-by-trial feedback and delayed feedback trials separately to the regression model `confidence ~ pcorrect + decision-status`, thereby obtaining standardised betas of P(correct) and decision-status for trial-by-trial feedback trials and delayed feedback trials. We will then run (1) a paired Bayes t-test to assess the likelihood that the decision-status regression betas from the trial-by-trial feedback block and the delayed feedback block come from the same distribution and (2) a paired Bayes t-test to assess the likelihood that the P(correct) betas from the trial-by-trial feedback block and the delayed feedback block come from the same distribution.

S4.2 Exclusion criteria summary for Experiment 3

Table S4.1

Participant exclusion criteria and excluded counts in Experiment 3

	Number of participants excluded
Overall	
Reported technical difficulties	1
ISTc	
Failed to complete all trials	0
All decision-point confidence ratings in the range of 49 to 51	0
2 or more instances across the 20 experimental trials of discrepancy between majority colour choice as indicated by their rated confidence at decision-point (where confidence greater than or equal to 50 indicates confidence rated in favour of the correct choice and confidence less than 50 indicates confidence rated in favour of the incorrect choice) and their actual choice behaviour	3
Total	3
Original dataset size	35
Final dataset size	32

Participant excluded if they meet one or more of the criteria listed above.

Table S4.2

Trial exclusion criteria and excluded counts in Experiment 3

	Number of trials excluded
At decision-point but confidence rating value does not correspond to the accuracy of the response on that trial (i.e., confidence greater than 50 if incorrect and less than 50 if correct)	3

S5. Chapter 5 supplementary material

S5.1 Pregistration document for Experiment 5

Experiment 5 was preregistered with an AsPredicted.org template on OSF on 10 January 2020. It is currently embargoed at https://osf.io/ym4za until 1 January 2022. A copy of the document is provided below:

Data collection: Have any data been collected for this study already? Note: 'Yes' is a discouraged answer for this preregistration form.

No, no data have been collected for this study yet.

Hypothesis

In previous experiments using the Information Sampling Task with Confidence (ISTc — adapted from the IST introduced by Clark et al., 2006), we have observed that at the same level of objective certainty ("P(correct)"), participants register greater subjective confidence if their rating is made after they have committed to a choice about the majority colour ("decision-point") as opposed to before they have committed ("pre-decision"). We are testing whether this commitment-related boost in confidence relative to P(correct) is observed regardless of whether that commitment is endogenously motivated (i.e., participants are free to commit to a choice whenever they feel ready) or exogenously provoked (i.e., participants are forced to make a choice at a time decided by the experimental program). We hypothesise that a decision-point confidence boost will be observed even when participants are forced to commit to this decision instead of having freely chosen to make it. We will also assess whether this commitment-related confidence boost in forced-choice trials on the ISTc will be of a comparable magnitude to that previously observed in free-choice trials. Such a finding would indicate that the act of committing to a choice alters confidence above and beyond the strength of evidence leading to that choice, and help us to rule out alternative explanations of the confidence boost we have seen in the free-choice version of the task (e.g., in terms of a sampling bias, if participants committed to a choice upon reaching a criterion level of confidence).

Dependent variable

The Information Sampling Task with Confidence (ISTc) is a probabilistic reasoning task adapted from the well-characterised information sampling task described by Clark et al. (2006). The ISTc additionally has participants rate their subjective confidence during the sampling phase as well as at the point of decision. In free-choice trials, participants are allowed to turn over as many tiles as they like before committing to a choice about the majority colour. In forced-choice trials, participants are forced to commit to a choice after turning over a certain number of tiles determined by the computer; the number is based on the previous P(correct) at which they made their previous choices.

We will be assessing behaviour in free-choice and forced-choice trials separately, using two measures: firstly, the standardised beta of decision status (binary factor variable: pre-decision vs. decision-point) in the regression model `confidence ~ pcorrect + decision-status`, and secondly, the difference in intercept between the separate regression lines for `confidence ~ pcorrect` pre-decision and `confidence ~ pcorrect` at decision-point.

Conditions: How many and which conditions will participants be assigned to?

The experiment uses a within-participants design, comparing ISTc performance in free vs. forced-choice conditions. All participants will perform 20 free-choice trials followed by 20 forced-choice trials. Each set of 20 trials will comprise a block of 10 trials with a "fixed win" (FW) scoring rule and a block of 10 trials with a "decreasing win" (DW) scoring rule (as defined in previous research). The order in which the blocks (FW and DW) are presented will be randomly assigned and counterbalanced between participants.

Analyses

For the primary measure, we will fit each participant's data from free-choice and forced-choice trials separately to the same regression model `confidence ~ pcorrect + decision-status`, thereby obtaining a standardised beta of decision-status for free-choice trials and a standardised beta of decision-status for free-choice trials for each participant. As a manipulation check, we will first replicate our previous finding of a decision-point confidence boost in free-choice trials, running a t-test to compare the mean of the free-choice regression beta across participants against the null hypothesis mean of 0.

Key analyses:

(1) To test our key hypothesis that a decision point confidence boost will also be observed in forced-choice trials, we will run a t-test to compare the mean of the forced-choice regression beta across participants against the null hypothesis mean of 0.

(2) To compare the decision-point confidence boost observed across the two conditions, we will run a paired t-test to compare the free- vs. forced-choice regression betas, using Bayes factor calculation to evaluate evidence in favour of the null (i.e., that the betas for the two conditions come from the same distribution).

For the secondary measure, which allows for the possibility that the slope of the P(correct) vs. confidence relationship will vary for the pre-decision vs. decision-point data, we will fit each participant's data from free-choice and forced-choice trials separately to two regression models: `confidence ~ pcorrect` for pre-decision ratings, and `confidence ~ pcorrect` for ratings at the point of decision, and calculate the difference between the intercepts of these two regression lines. We will then run a paired t-test to compare the means of the intercept difference for free-choice trials versus forced-choice trials across participants.

Outliers and Exclusions

Participants are pre-screened on Prolific based on age (excluded if outside our required age bracket of 18 to 60 years) and vision (excluded if they do not have normal or corrected-to-normal vision) requirements.

Participants will be excluded from analysis if: [1] all their confidence ratings are in the range of 49 to 51, [2] data are not available for all trials (total: 40) of the experiment, [3] they show 3 or more instances across the 40 experimental trials of discrepancy between majority colour choice as indicated by their rated confidence at decision-point (where confidence greater than or equal to 50 indicates confidence rated in favour of the correct choice and confidence less than 50 indicates confidence rated in favour of the incorrect choice) and their actual choice behaviour, or [4] if they have not made at least 3 pre-decision confidence ratings in each block of 10 trials (FW-free, DW-free, FW-forced, DW-forced).

Sample Size

We will recruit 30 participants for this study. The sample size is based on a G*Power estimate of 27 participants to observe, in a two-tailed t-test against zero, a regression coefficient of decision-status with half the effect size of our original results (originally observed Cohen's d = 1.47, anticipated Cohen's d = 0.73) at alpha = .05 and 95% power.

S5.2 Exclusion criteria summary for Experiment 5

Table S5.1

Participant exclusion criteria and excluded counts in Experiment 5

	Number of participants excluded
Overall	
Reported technical difficulties	0
ISTc	
Failed to complete all trials	0
All decision-point confidence ratings in the range of 49 to 51	0
3 or more instances across the 40 experimental trials of discrepancy between majority colour choice as indicated by their rated confidence at decision-point (where confidence greater than or equal to 50 indicates confidence rated in favour of the correct choice and confidence less than 50 indicates confidence rated in favour of the incorrect choice) and their actual choice behaviour	7
Less than 3 pre-decision confidence ratings in each block of 10 trials (FW-free, DW-free, FW-forced, DW-forced)	5
Total	7
Original dataset size	40
Final dataset size	33

Participant excluded if they meet one or more of the criteria listed above.

Table S5.2

Trial exclusion criteria and excluded counts in Experiment 5

	Number of trials excluded
At decision-point but confidence rating value does not correspond to the accuracy of the response on that trial (i.e., confidence greater than 50 if incorrect and less than 50 if correct)	60 (1.00%)

Data reported as count (percentage of total trials for the task).

S5.3 Preregistration document for Experiment 6

Experiment 6 was preregistered on AsPredicted.org on 25 April 2020 (reg. no. 39915). The document is currently a private link on AsPredicted (https://aspredicted.org/XRK_OVS). A copy of the document is provided below:

1) Have any data been collected for this study already?

No, no data have been collected for this study yet.

2) What's the main question being asked or hypothesis being tested in this study?

In previous experiments using the Information Sampling Task with Confidence (ISTc — adapted from the IST introduced by Clark et al., 2006), we have observed that at the same level of objective certainty ("P(correct)"), participants register greater subjective confidence if their rating is made after they have committed to a choice about the majority colour ("decision-point") as opposed to before they have committed ("pre-decision"). A commitment-related boost in confidence relative to P(correct) was seen regardless of whether the commitment was endogenously motivated (i.e., participants are free to commit to a choice whenever they feel ready) or exogenously provoked (i.e., participants are forced to make a choice at a time decided by the experimental program).

We are testing whether the act of initially committing to a binary choice has a lasting and measurable impact on subsequent choice behaviour and subjective confidence:

HYPOTHESIS 1 — on ISTc trials in which participants are forced to make an interim binary choice, a decision-evoked boost in confidence will be observed at this initial decision point.

HYPOTHESIS 2 — if further information is sought before committing to a final decision, participants will subsequently [a] seek less information and [b] express greater confidence than if they had not been asked to make a binary choice at the interim point (and had merely been asked to give a confidence rating).

3) Describe the key dependent variable(s) specifying how they will be measured.

The Information Sampling Task with Confidence (ISTc) is a probabilistic reasoning task adapted from the well-characterised information sampling task described by Clark et al. (2006). The ISTc additionally has participants rate their subjective confidence during the sampling phase as well as at the point of decision. In the free-choice (calibration) block, participants are allowed to turn over as many tiles as they like before committing to a binary choice about the majority colour. In the interim-choice (experimental) block, there are two types of trials: interim-decision and interim-rating. On interim-decision trials, participants are forced to commit to an initial decision at the "interim point" — approximately half of the number of tiles they turned over for the same P(correct) sequence during the calibration block — before they are allowed to freely turn over as many more tiles as they like and/or make a final decision. On interim-rating trials, participants are simply asked to make a confidence rating at the interim point. In addition, we included a number of catch trials where participants are not allowed to turn over more tiles after their interim decision.

To assess Hypothesis 1, we will look at the standardised beta of interim status (before interim point vs. at interim point) in the regression model `confidence ~ pcorrect + interim-status` for confidence ratings up to and including the interim point.

To assess Hypothesis 2, we will look at [a] the total number of tiles seen and [b] the standardised beta of timepoint (before the interim point vs. after the interim point) in the regression model `confidence ~ pcorrect + timepoint` for confidence ratings excluding those at the interim point and the final decision point.

4) How many and which conditions will participants be assigned to?

The experiment uses a within-participants design, comparing performance (confidence and total tiles seen) between the free-choice block and the interim-choice block, and between interim-decision and interim-rating trials. All participants will perform a calibration block of 25 free-choice trials followed by an experimental block of 25 trials consisting of 10 interim-decision trials (initial choice at the interim point), 10 interim-rating trials (confidence rating at the interim point), and 5 catch trials (forced/final choice at the interim point). All trials in the calibration and experimental blocks will use a “fixed win” (FW) scoring rule (as defined in previous research), and the P(correct) sequences (i.e., the order of majority vs. minority tiles seen as the participant samples the gameboard) of the calibration block will be matched in the experimental block.

5) Specify exactly which analyses you will conduct to examine the main question/hypothesis.

To assess Hypothesis 1, we will fit each participant’s confidence data, separately for [Type A] interim-decision and catch trials (which involve a decision being made at the interim point) and [Type B] interim-rating trials (which involve only a confidence rating at the interim point), to the regression model `confidence ~ pcorrect + interim-status`, thereby obtaining standardised betas of interim status from Type A trials and Type B trials for each participant. We will then run a t-test to compare the mean of the Type A interim status regression betas across participants against the null hypothesis mean of 0. We will also run a paired Bayes t-test to assess the likelihood that the interim status regression betas from Type A and Type B trials come from the same distribution.

To assess Hypothesis 2A, we will run a paired Bayes t-test to compare the mean number of tiles seen at the final decision point in interim-decision versus interim-rating trials.

To assess Hypothesis 2B, we will fit each participants' confidence data, separately for interim-decision trials and interim-rating trials, to the regression model `confidence ~ pcorrect + timepoint`, thereby obtaining standardised betas of timepoint from interim-decision and interim-rating trials for each participant. We will then run a paired Bayes t-test to assess the likelihood that the timepoint regression betas from the two types of experimental trials come from the same distribution.

6) Describe exactly how outliers will be defined and handled, and your precise rule(s) for excluding observations.

Participants are pre-screened on Prolific based on age (excluded if outside our required age bracket of 18 to 60 years) and vision (excluded if they do not have normal or corrected-to-normal vision) requirements.

Participants will be excluded from analysis if:

- all their confidence ratings are in the range of 49 to 51;
- data are not available for all trials (total: 50) of the experiment;
- they show 3 or more instances across all trials of the experiment (total: 50) of discrepancy between majority colour choice as indicated by their rated confidence at decision-point (where confidence greater than or equal to 50 indicates confidence rated in favour of the correct choice and confidence less than 50 indicates confidence rated in favour of the incorrect choice) and their actual choice behaviour;
- if they have not made at least 3 pre-interim-point confidence ratings in each condition (interim-decision, interim-rating) in the experimental block.

7) How many observations will be collected or what will determine sample size? No need to justify decision, but be precise about exactly how the number will be determined.

We will recruit 50 participants for this study. The sample size is based on a G*Power estimate of 49 participants to observe, in a two-tailed t-test against zero, a regression coefficient of decision-status with half the effect size of our original results (originally observed Cohen's d = 1.07, anticipated Cohen's d = 0.53) at alpha = .05 and 95% power. Sample size is based on a halved effect size because the interim forced choice occurs earlier during the trial than in our previous forced-choice study, such that choice-related changes in confidence may be of smaller magnitude.

8) Anything else you would like to pre-register? (e.g., secondary analyses, variables collected for exploratory purposes, unusual analyses planned?)

We will also run a number of secondary analyses:

1. We will replicate our previous finding of a decision-point confidence boost in free-choice trials, running a t-test to compare the mean of the free-choice decision-status regression beta across participants against the null hypothesis mean of 0.
2. To examine whether the decision-point confidence boost is a non-cumulative effect in the decision process, we will fit each participants' confidence data from interim-decision trials to the regression model `confidence ~ pcorrect + decision type`, where 'decision type' is a categorical variable indicating whether the decision was an interim or a final decision. We will then run a t-test to compare the mean of the decision-type regression beta across participants against the null hypothesis mean of 0.
3. To assess whether an explicit interim choice is more predictive of final choice than an interim choice implied by a subjective confidence rating, we

will run a paired Bayes t-test to compare the mean proportion of matches between choices in interim and final decisions in interim-decision trials with the mean proportion of matches in interim-rating trials.

S5.4 Exclusion criteria summary for Experiment 6

Table S5.3

Participant exclusion criteria and excluded counts in Experiment 6

	Number of participants excluded
Overall	
Reported technical difficulties	0
ISTc	
Failed to complete all trials	0
All decision-point confidence ratings in the range of 49 to 51	0
3 or more instances across the 50 experimental trials of discrepancy between majority colour choice as indicated by their rated confidence at decision-point (where confidence greater than or equal to 50 indicates confidence rated in favour of the correct choice and confidence less than 50 indicates confidence rated in favour of the incorrect choice) and their actual choice behaviour	6
Less than 3 pre-interim-point confidence ratings in each condition (interim-decision, interim-rating) in the experimental block	14
Total	16
Original dataset size	68
Final dataset size	52

Participant excluded if they meet one or more of the criteria listed above.

Table S5.4

Trial exclusion criteria and excluded counts in Experiment 6

	Number of trials excluded
At decision-point but confidence rating value does not correspond to the accuracy of the response on that trial (i.e., confidence greater than 50 if incorrect and less than 50 if correct)	12