On the uses of confidence judgements to guide behaviour

Nomi Carlebach

Wolfson College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity 2020

In loving memory of my dad,
Shimmy Carlebach

Acknowledgements

First and foremost, I would like to thank my supervisor, Nick Yeung, who has guided me through a new field with endless patience and encouragement. Your support, generosity, and willingness to always answer questions have made this work possible. Your humour and calmness have made it all the more enjoyable.

The ACClab has changed and evolved over these years, but always been a pleasure to be part of. Aaron, Niccolo, Raluca, Chia, and Karen have somehow managed to make the Tinbergen a beautiful memory. Linda, Sarah, Raj and Maja have brought life to the NRH. A special thanks goes to Matt D. for sharing his infinite EEG wisdom, and Matt J. for the help with Javascripting; To Michael for all the whispers, and to Joshua for putting up with them.

I am grateful also for the friends outside the lab. Thank you Arianna for all our talks and good times at Wolfson, Marlene for the shared breaks and Alon for your extraordinary talent for making everything fun. I'd also like to thank Roi, for providing some stability in a world filled with uncertainty: It is good to be confident that, no matter where you or I are, you will always be a pain.

In moving to Oxford I have been lucky enough to gain not one, but two new families. Marta, Tom, Brad and Pierre, thank you for being my home. The times shared at Ikhaya have shaped these last few years and made them special. I hope that some of the joy, energy and optimism that surround you have rubbed off on me. To Marta especially, thank you for your friendship and companionship during these years. I cannot imagine a better partner to go through the highs and lows of the DPhil with (except for maybe a lazier one, who would stick around for longer). To Nir, Michal, Smadar, and Daniel (+ David and Adam), thank you for accepting me into your family, and being such a loving and supportive one. Never before have I looked forward as much to Friday Kiddush. And of course, a special thanks to my family back home, mum and Daniel, who have somehow managed to provide so much support and love from far away. Finally, thank you Will for the much needed walks, the continuous encouragement and for providing my tomato fix during lockdown. I could not have finished this thesis without it.

I am grateful to the Department of Experimental Psychology, the British Friends of the Hebrew University, Funds for Women Graduates, B'nai B'rith Leo Baeck (London) Lodge, and the Anglo-Israel Association for their generous support that has made this research possible.

Abstract

The ability to reflect upon our own thoughts and actions is a defining feature of human cognition. People can evaluate the quality of their own performance and report levels of confidence that are correlated with objective accuracy. As such, the subjective sense of confidence that accompanies our decisions and beliefs could potentially be a useful tool that can guide adaptive behaviour. The research presented in this thesis explores this idea to examine the functional role of confidence and the scope of its influence on our behaviour.

Across three lines of work, the role of confidence is studied in the context of carefully-controlled perceptual decision making tasks. In the first part of this work, I examine the role of confidence in guiding task selections, acting as a valuable indicator of likely success. I present data showing that people exhibit a preference for tasks in which they report higher confidence. Furthermore, I show that confidence guides task choices in a more nuanced way than binary error detection, and that this preference is evident even when external feedback is available.

In the second part, I examine the proposal that people can use their confidence flexibly, depending on the specific context. Using an advice seeking task, I show that when advice quality is unknown, people select advice more often when their confidence is high, presumably using their confidence as a feedback proxy to learn about advisor quality. A contrasting pattern is observed when advice quality is known, with people requesting more advice when confidence is low, consistent with the idea that confidence is used as a self-monitoring tool that signals that external help is needed. Together, the results indicate that people use confidence in a way that is directed towards achieving their current goals.

The third part explores how confidence also modulates attention and, through this, influences lower-level processing of information. I present behavioural and electrophysiological data indicating that high confidence in the informativeness of task-relevant information is associated with greater attention towards it.

Overall, the three lines of work provide converging support for the view that confidence plays an important role in guiding behaviour in a flexible, goal-directed way, and suggest that it influences a wide range of both high-level and low-level processes.

Publications

Chapter 2 has been adapted from the following publication:

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.*

Contents

1 Introduction

1.1 A brief outline

We all know what confidence is intuitively, and can readily respond to the question *'how sure are you?'* in respect to our knowledge and beliefs. In a world filled with uncertainties, our beliefs and sense of confidence in them guide our decisions. These decisions range from brief and inconsequential, such as what meal to order at a restaurant, to those involving some deliberation, such as where to go on vacation, to those with lifelong impact, such as what career path to pursue. Regardless of its importance, upon reaching a decision we typically have a sense of the likelihood that the decision we reached was a good one. Thus, confidence is a ubiquitous part of human experience. Recent years have shown a surge of interest in studying confidence, with research conducted on the neural and computational basis of confidence judgements, and their function. The present work focusses on the latter, examining how our subjective sense of confidence is used to guide adaptive behaviour.

In this thesis, I consider three main questions relating to the functional role of confidence judgements: (1) does confidence help guide our choices of which situations and tasks to engage with, as a valuable indicator of likely success? (2) is confidence used strategically, according to the immediate context and goal at hand? and (3) does confidence in the relevance of information modulate the allocation of attention to it? Each of these questions is explored in a separate chapter, with the specific motivation and theoretical background driving each line of research reviewed at the beginning of the specific chapter. Hence, in the general introduction, I explore only those themes which are shared by the three lines of research, i.e., all things confidence. I begin by defining metacognition, and provide an overview of the theoretical and empirical work investigating metacognitive monitoring to date. Following this, I briefly discuss theories of how confidence judgements are generated, and describe several frameworks for metacognition. I then review previous work investigating the ways in which confidence regulates behaviour, and end with a brief description of the present work.

1.2 From introspection to metacognition

Our ability to self-reflect and evaluate our own knowledge and judgements has been a topic of interest in the disciplines of psychology and philosophy for many years. Wilhelm Wundt, widely regarded as the father of experimental psychology, conceived of psychology as the study of the mind and consciousness, and treated introspection as a scientific method to observe sensory experiences (Asthana, 2015). Following Wundt, other psychologists applied the use of introspection to study more complex mental processes, such as learning, memory and emotion (Boring, 1953). However, with emerging dissociations between subjective judgements and objective performance, concerns grew that judgements about our own cognitions cannot serve as accurate reflections of the processes they are examining (Nisbett & Wilson, 1977). Thus, the use of introspection was abandoned by cognitive psychologists for many years and only recently has interest in such self-evaluations resurfaced. This new interest is likely the result of an emerging view of people as active agents, who can use their subjective evaluations to control and regulate their cognitive processes, in a way that is consistent with achieving their goals (Koriat, 2007). Under such a view, self-evaluations have an importance in and of themselves — as tools that can be used to guide behaviour — and not just as a window into understanding the object of their examination. Hence, data from introspective reports are valuable despite, and perhaps even because, of their distortions.

The study of *metacognition*, broadly defined as cognitions about cognitions (Flavell, 1979) concerns metacognitive knowledge (what we know about cognition in general), and the online monitoring and control of our own cognitive processes (Nelson & Narens, 1990). Many different types of judgements about our cognitive processes fall within this definition of metacognitive evaluations. Confidence, defined as a belief in the accuracy of our thoughts, knowledge or performance (Grimaldi et al., 2015), is a form of metacognitive evaluation. Other examples of metacognitive reports include detecting errors in our performance, judging the difficulty of a task, and evaluating the likelihood that we will remember a certain piece of information. In the following section, I review some of the main experimental work into metacognitive monitoring, describing the different lines of research, their motives, and the main paradigms used. While the discussion will, in general, touch upon various forms of metacognition, the centre of attention will be confidence, as that is the focus of the present work.

1.3 Experimental paradigms studying metacognitive monitoring

Early research into metacognition focussed on the evaluation of memory processes, termed metamemory. Developmental psychologists, such as Flavell, were interested in how children's learning is influenced by their understanding of memory processes in general, and of their own memory capacities in particular. For example, when studying for an exam, we might ask ourselves if we have learned the information well enough to be able to remember it in a test. If the answer is no, we are likely to invest more time in studying before moving on to the next item. This reflects monitoring of our own knowledge of the studied topic, and the existence of general metacognitive knowledge that investing more time in studying can lead to better performance.

In a typical task investigating metamemory, participants are presented with material to learn, such as cue-target pairs of words, and asked to make judgements about their learning. Several types of reports have been used to examine whether people form accurate judgements of their learning (see Dunlosky and Metcalfe, 2008 for a review). These reports tap into various stages of the learning process, with prospective judgements elicited prior to task performance, and retrospective judgements after the task is completed. For example, ease of learning (EOL) judgments are made before items are learned, and refer to people's estimates of how difficult they think learning each item will be, based on prior information such as the meaningfulness of items (Richardson & Erlebacher, 1958; Underwood, 1966). Judgements of learning (JOL) are typically elicited after learning took place, with participants asked to report how likely they believe they will be to remember a learned item if prompted with the associated cue (Arbuckle & Cuddy, 1969). Feeling of knowing (FOK) judgements, first introduced by Hart (1965) are made during an initial test, with participants reporting how likely they believe they are to later recognise the correct answer even if they are currently unable to retrieve it from memory. Finally, at the time of test, participants can be asked to report their confidence that their response (i.e. the retrieved memory) is correct. Studies examining prospective judgements (FOK, JOL and EOL), have found that the predictive accuracy of these judgements is intermediate — they are above chance, but far from perfect (Leonesio & Nelson, 1990). Retrospective judgements of confidence, while also not perfect, have been shown to be better predictors of performance (Siedlecka et al., 2016).

In the field of cognitive psychology, researches have also studied metacognition, but for different motives. Specifically, they were interested in understanding how people know what they know. Studies in this area have mainly used decision-making tasks, allowing for more tractable neural and computational accounts of the basis of metacognition. In perceptual decision-making studies, participants are typically asked to make a binary choice, such as judging which of two lines is longer (Henmon, 1911), or whether the direction of motion of ambiguous dot arrays is towards the right or the left (Hauser et al., 2017). Then, they make a metacognitive judgement on the accuracy of their first-order performance (see Figure 1.1 for a

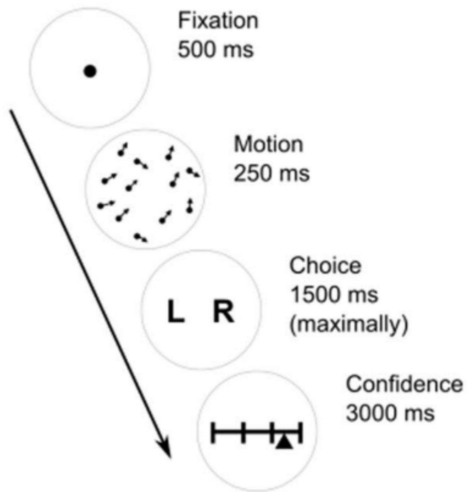

Figure 1.1: A sample perceptual decision task. Participants view a cloud of dots moving with a defined mean motion direction and added random movement noise. They have to make a categorical judgement of the main direction of motion (left or right) and then rate their confidence using a sliding scale. Figure adapted from Hauser et al. (2017).

sample task). Two main types of metacognitive judgements have been investigated in this context — error detection and confidence judgements (for a review, see Yeung & Summerfield, 2012). In error detection studies, participants are asked to make a binary metacognitive judgement, such as to judge each response as correct or not, or signal when they think they have made an error. In contrast, graded confidence judgements are rated on a scale. To elicit confidence judgements, studies have used both discrete scales with verbal labels such as 'perfectly confident', 'fairly confident', 'with little confidence' and 'doubtful' (Boldt & Yeung, 2015; Henmon, 1911), as well as continuous numerical or sliding scales, such as seen in Figure 1.1 (Hauser et al., 2017). Importantly, studies have demonstrated that people can reliably detect errors in their performance (Rabbitt, 1966), and that subjective confidence strongly reflects objective accuracy (Baranski & Petrusic, 1994). The critical underlying assumption of these studies, and the present work, is that simple perceptual decisions share common principles with the more complex choices that we face in our daily lives (i.e., tapping into our confidence regarding a perceptual decision should be similar to tapping into our confidence regarding a memory when deciding whether to continue studying). Thus, by examining how people behave under these controlled settings, we can learn about the mechanisms driving decisions and confidence, also in more complex situations (Gold & Shadlen, 2007).

Experimental work in the field of psychophysics has also focussed on people's subjective experiences of their mental processes since its early days. For example, Pierce and Jastrow (1884) asked participants to judge changes in tactile pressure, and then rate their confidence on a scale of 0 (absence of any preference) to 3 (as strong a confidence as one would have about such sensations), and studied the relationship between confidence and accuracy. They observed that confidence strongly predicts accuracy, but that people also tend to be correct in their first-order judgements even when they reported they had guessed, suggesting the existence of unconscious processing. Building upon these methods, researchers interested in consciousness have used the relationship between metacognitive reports and task accuracy on psychophysical discrimination tasks as a measure of awareness. Here, three main types of metacognitive reports have been elicited: retrospective confidence ratings, where people report their confidence that their first-order judgement was correct; the perceptual awareness scale (PAS), in which people report the quality of their subjective experience, rather than the quality of their response (e.g., the degree of visibility in a visual task; Ramsøy and Overgaard, 2004); and post-decision wagering (PDW), where participants make a decision and are then asked to place wagers on the outcomes of their decisions, with wagers acting as a measure of confidence (Persaud et al., 2007). There is a clear difference in approach between the first two measures, and the latter. Confidence reports and the PAS rely on participants' introspections and accurate reports of their experience. Some researchers have argued that these measures are not indexes of awareness, but, rather, of awareness of awareness (Persaud et al., 2007). Furthermore, the same researchers suggest that when using such reported measures, participants might provide inaccurate reports, as they are not given any motivation to reveal their actual knowledge. The PDW paradigm was introduced to overcome these limitations. In this paradigm, if the decision was correct, the amount gambled is kept, whereas if an error was made the amount is lost. Hence, participants are incentivised to report their real confidence by placing larger bets when they are confident and smaller bets when they are unconfident. It has additionally been argued that wagering is a more intuitive measure of reporting confidence than using an arbitrary scale (Persaud et al., 2007). However, various criticisms of the PDW paradigm have also been made. For example, it has been suggested that wagering behaviour is influenced by loss aversion, and thus is not a direct index of subjective awareness (Fleming and Dolan, 2010; see Sandberg et al., 2010 for a comparison of confidence, PAS and PDW).

The introduction of non-verbal reports of confidence, such as post-decision wagering, has allowed researchers to extend the study of metacognition from humans to various species of animals. These investigations are of importance for two main reasons. First, they are of interest to comparative psychologists, since establishing whether animals other than humans have metacognitive capacities can bear on their consciousness and self-awareness. Second, they allow for measuring confidence at a single-cell level, which would not normally be possible when studying humans. Typically, in studies of metacognition in animals, subjects encounter a perceptual task with varying levels of difficulty, and metacognition is measured using a behavioural proxy for confidence (see Kepecs and Mainen, 2012 for a review). For example, in *uncertain-option* and *opt-out* tasks, subjects can choose a response that allows them to get a small, but guaranteed reward, or to avoid a penalty by skipping a response altogether, instead of committing to one of the primary response options. Other paradigms include PDW and a similar *decision-restart* paradigm. Here, reward is delayed, and animals can either wait or abort the trial and immediately begin the next, with the time spent waiting acting as a measure of confidence. Overall, these studies have provided evidence that animals such as apes, monkeys, dolphins and rats make judgements that are driven by uncertainty (Smith, 2009). However, there is still much controversy regarding the ability of these tasks to reflect on the metacognitive capabilities of these species, with some researchers arguing that the results are mediated by associative learning (see Le Pelley, 2012 and Smith et al., 2016 for both sides of the argument).

While this review is by no means exhaustive, it demonstrates the richness of research touching upon metacognitive monitoring. The discussion thus far has centred around the different types of metacognitive judgements elicited in various lines of research. Given the wide range of paradigms, and the variety of experiences and self-reports studied under the umbrella of metacognitive monitoring, it is perhaps not surprising that differences in the predictive accuracy of these measures exist. For example, even when examining metamemory judgements relating to the same set of items, Leonesio and Nelson (1990) found that the predictive accuracy of FOK, JOL and EOL judgements differs. Moreover, they suggest that these judgements do not to rely on the same underlying processes. It has also been shown that confidence judgements are more reliable than FOK (Costermans et al., 1992). However, despite such differences, several observations can be made, that are supported by research from the wide variety presented above.

Some hints have been made already at the fact that metacognitive judgements are accurate to some extent, but that people are also subject to errors in their evaluations (Baranski & Petrusic, 1994; Nelson & Narens, 1990). This might come as no surprise. If you reflect upon your own experience, you will probably conclude that your confidence is often related to accuracy: when you are highly confident in a decision, you are more likely to be correct than when your confidence is lower. The degree to which metacognitive judgements reliably predict accuracy is termed metacognitive calibration. Calibration is assessed by measuring the correlation between the observed probability of a correct response given a level of subjective confidence (i.e. the subjective probability of being correct) and the subjective probability. A perfectly calibrated judge will rate their confidence so that, for all decisions assigned the same probability, the proportion correct is equal to the probability assigned (e.g. 60% of the responses given a .6 probability of being accurate are indeed correct). It is important to note however, that calibration is affected by biases in confidence, such as a tendency to be over- or underconfident. Furthermore, calibration depends on task difficulty: people tend to be overconfident when tasks are difficult, and underconfident when they are easy (Baranski & Petrusic, 1994; Lichtenstein & Fischhoff, 1977). In other words, people's ability to judge task performance is dependent on the task performance itself. Thus, measuring metacognitive accuracy independently from first-order performance proposes a challenge. Various approaches to measuring metacognition have been applied in order to produce a measure of metacognitive ability that is not connected to task performance, or to biases in confidence. Current approaches to this challenge rely on the principles of signal detection theory (Fleming & Lau, 2014), but are beyond the scope of the present discussion. Critically however, it has been shown that confidence — both in memory (Koriat & Goldsmith, 1996) and in perceptual decisions (Baranski & Petrusic, 1994) — often exhibits quite impressive calibration, thus providing a reliable internal signal about the quality of our knowledge and decisions.

1.4 How do we know what we know?

Various models have been proposed to account for our metacognitive abilities. Within the metamemory research tradition, there are two conflicting views. According to the *direct access* view, confidence is based on partial retrieval of the same information that is used for the first-order judgement. For example, to explain FOK judgements, Hart (1965) suggested that individuals can access their memory, and monitor whether a trace of the item exists or not. In contrast, the *inferential* view proposes that confidence is based on different cues and heuristics, rather than the information that guides the decision itself (see Koriat, 2007 for a review). Several cues that have been shown to influence FOK and JOL reports are the fluency or ease with which an item is processed (Begg et al., 1989), the ease with which information is retrieved, regardless of its correctness (Koriat, 1993), and the familiarity of the cue (Metcalfe et al., 1993; Reder & Ritter, 1992). For example, you might feel more confident that you remember who composed 'From the new world' symphony if you've recently heard it, regardless of whether you know answer (Dvorak).

In recent years, cognitive neuroscience approaches have started to examine the neural and computational basis of metacognition, showing significant progress in the development of formal accounts for confidence judgements and error detection

(Yeung & Summerfield, 2012). This work has built upon previous models that were developed to explain first-order choice behaviour, and extended these models to account for the relationship between evidence, decisions and second-order judgments of confidence. Typically, behaviour is examined in the context of simple perceptual judgement tasks such as those described above. Several types of models have been used to explain how confidence arises, including signal detection theory (Kepecs et al., 2008), evidence accumulation (Yeung & Summerfield, 2012), and Bayesian inference based models (Meyniel et al., 2015). Across the variety of existing computational models, it is generally agreed that the degree of confidence in a decision is informed by the decision variable (DV), which represents an internal state influenced by the incoming evidence (Gold & Shadlen, 2007). For example, according to SDT-based models, a decision is made by comparing the DV to a criterion, and confidence determined by the distance between them (Kepecs and Mainen, 2012; see Figure 1.2A). According to Bayesian models, confidence is the posterior probability of being correct, and beliefs can be modelled as a probability distribution, that evolves over time with new incoming evidence being weighed in a Bayesian way. Thus, the precision of the evidence (i.e. the inverse of its variance), is a key determinant of confidence in Bayesian frameworks (Meyniel et al., 2015).

Akin to the direct access model of metamemory, most of these computational models have assumed that confidence is based on the same information as the decision. This suggestion is supported by evidence from neural recordings in Rhesus monkeys, demonstrating that confidence may be reflected in activity of the same neurons that represent the formation of a decision in the lateral intraparietal (LIP) area (Kiani & Shadlen, 2009). However, there is a growing body of evidence demonstrating dissociations between confidence and performance, at a neural and behavioural level (for a review, see Fleming & Dolan, 2012). An intuitive demonstration of such a dissociation is our ability to detect errors, and change our minds after making a decision (Resulaj et al., 2009; Yeung & Summerfield, 2012). This implies that at the very least evidence continues to accumulate and inform confidence after the choice has been made. Therefore, models in which confidence is based solely on the same information as the decision cannot account for these dissociations.

Consider, for example, a perceptual decision where the observer must choose between two options. Evidence accumulation models, such as the drift diffusion and race models, postulate that the DV is accumulated sequentially until it reaches a decision threshold, at which point a choice is made. In the race model, evidence is accumulated for each option by competing accumulators, and the balance of evidence between them at the time of choice determines confidence (Vickers, 1979; see Figure 1.2B). In drift diffusion models, evidence for and against an option is accumulated in the same DV (see Figure 1.2C). By allowing for evidence integration over time, these models can account for both choice and reaction times (as opposed to static models such as SDT). However, they fail to explain changes of mind (Yeung & Summerfield, 2012). To account for such dissociations between first-order task performance and confidence, it has been suggested that confidence judgements depend on continued processing of information after a choice has been made. For example, the 2-stage dynamic signal detection model (2DSD) proposes that the DV continues to accumulate after a choice has been made, until a confidence report is elicited (Pleskac & Busemeyer, 2010), with confidence depending on the state of the DV at this later, post-decisional time. Thus, evidence that is conflicting to our choice can lead to a change of mind. Recent empirical evidence has provided behavioural (Charles & Yeung, 2019; Moran et al., 2015; Van Den Berg et al., 2016) and neural (Murphy et al., 2015) support for post-decisional evidence accumulation that influences confidence reports. As we will see later, this feature is exploited in Chapter 2, to manipulate confidence independently of accuracy. Further support for the proposal that confidence judgements are not based solely on the information integrated into the decision, comes from recent studies demonstrating the influence of various cues, such our beliefs about our abilities to perform a task (Boldt, Schiffer, et al., 2019; Rouault et al., 2019), and evidence variance (Boldt et al., 2017).

To conclude, various models of confidence judgements exist. While some models suggest that confidence is a direct readout of the information on which the

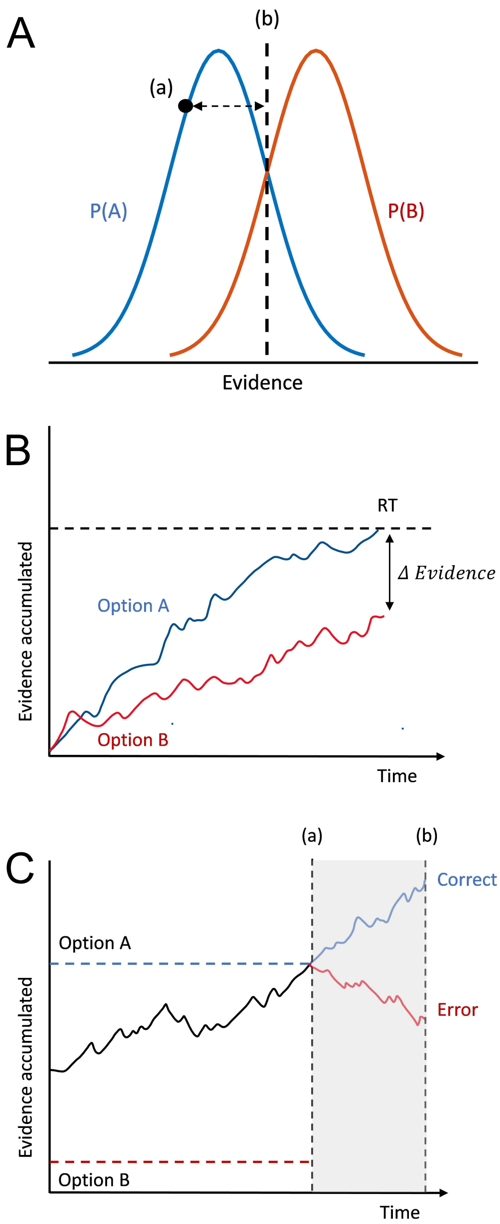

Figure 1.2: Computational models for decisions and confidence. (A) Signal-detection theory based model. The two stimulus distributions for option A and option B: *(a)* is the evidence on a single trial, *(b)* is the decision criterion, and the distance between them determines confidence. (B) Race model. Evidence for each of the options is accumulated in separate decision variables, until one DV reaches the decision threshold, represented by the dashed line. Confidence in a decision can be captured by the 'balance of evidence' at the time of decision. (C) Drift diffusion model, with evidence for both options accumulated in a single decision variable. Dashed horizontal lines represent the decision thresholds. A response is made at time *(a)*, and post-decision processing continues until time *(b)*, when confidence is reported. If option A was correct, evidence continues to accumulate in favour of the decision. If option B was correct (i.e. an error made), evidence in favour of option B will accumulate. Post decisional evidence accumulation accounts for varying confidence levels and can lead to changes of mind (e.g., if the DV crosses the threshold of the non-chosen option).

decision was made, it seems likely that confidence judgements are influenced by additional information. For example, confidence that a memory can be retrieved might be influenced by various heuristics, such as cue familiarity. Decisional confidence models traditionally focussed on the computation of confidence based on accumulated evidence, but recent advances have extended these models to include information that was not part of the decision-making process, notably continued evidence accumulation following an initial decision. Furthermore, behavioural and neurological dissociations between performance and confidence suggest that confidence is generated by a separate system to the decision itself. Several studies have suggested that this might take place in the pre-frontal cortex (e.g. Rounis et al., 2010, Fleming and Dolan, 2012 for a review). Critically, these findings suggest a domain-general system that represents confidence in an abstract way, consistent with the view that confidence can be used as a generalisable signal that supports flexible behaviours across different cognitive systems. This suggestion is discussed further in the following section.

1.5 Why do we know what we know?

In the previous sections, I have discussed various types of metacognitive judgements, the paradigms used to elicit them, and how they are generated. We have seen that people are able to provide metacognitive judgements that are correlated with performance across different domains and tasks such as memory and perception. Furthermore, there is some evidence that not only humans, but also other species, have access to the uncertainty of their representations. Given the demonstrated ubiquity of monitoring abilities, we might ask *why* we have this ability to explicitly represent our confidence.

Several theoretical frameworks for metacognition have been proposed that agree on an intuitive point: we must monitor in order to control. The highly influential framework suggested by Nelson and Narens (1990), described metacognition in terms of the interactions between two interrelated levels of cognitive processes: an *object-level* and a *meta-level*. The object-level is the level at which basic information processing, such as encoding, retention and retrieval, occur. The meta-level contains an imperfect model of the object-level, and monitors and modifies (i.e. controls) the operations occurring at the object-level, through a flow of information between them. Importantly, the flow of information is directional, so that the meta-level receives information about the object-level, and regulates it in a top-down way, but the object-level does not have any access to the meta-level.

Under this broad definition, any cognitive process that receives information from another cognitive process, and exerts control over it, would be deemed metacognitive. Indeed, the similarity between metacognitive and executive control functions has been noted before (Fernandez-Duque et al., 2000). This has led researchers to differentiate between explicit and implicit metacognitive monitoring (Fleming & Dolan, 2012; Shea et al., 2014; Shea & Frith, 2019). Shea et al. (2014) note that many cognitive control processes can occur at an automatic and implicit level. For example, motor plans can be automatically inhibited as a result of subliminal visual stimuli (Sumner et al., 2007), typists have been shown to slow down after errors they are not consciously aware of making (Logan & Crump, 2010), and neurological patients demonstrate control of executive attention as a result of conflict in the Stroop task, without conscious feeling of mental effort (Naccache et al., 2005). The authors suggest that these behaviours are driven by unconscious representations, such as decision uncertainty, and that metacognitive behaviours exhibited by animals are likely the result of such implicit metacognition. Thus, their framework divides metacognition into two separate systems: 'system 1' metacognition, which includes automatic, effortless, and typically associative processes, and 'system 2' metacognition, which is conscious, slower, and is most likely distinctive of human cognition. To the inevitable question of why we have explicit metacognitive representations, given the widespread influence of an implicit system, they offer an elegant solution: Explicit representations are necessary for social interaction. While learning and control can occur within an individual at an automatic and implicit level, metacognitive information has to be shared in order to allow coordination in social situations, where more than one agent is involved.

Indeed, there is a large body of evidence demonstrating the influence of confidence on decisions in social settings (see section on *'Guiding social decisions'* below).

In a recent paper, Shea and Frith (2019) extend this framework, and suggest that system 2 metacognition serves not only to make information available *between* people, but also *within* an individual. They highlight the link between metacognition and the global workspace — a mechanism that allows for representations to be made accessible to a variety of cognitive processes (Dehaene & Naccache, 2001). Specifically, they claim that all globally available representations are accompanied by a degree of confidence that they are likely to be correct, and the representation of confidence is needed for cognitive work to be carried out in the workspace. Since the workspace contains representations from various cognitive systems (e.g. perceptual, mnemonic and motor), there is a need for some form of rating that is scale-free and allows for weighing and integrating these different representations — confidence.

Two recent studies have demonstrated that confidence can indeed act as a 'common currency', allowing different types of decisions to be compared and prioritized (de Gardelle et al., 2016; de Gardelle & Mamassian, 2014). In an initial study, participants viewed Gabor stimuli and had to perform one of two tasks, comparing either their orientation, or their spatial frequency. After every pair of trials, they reported which trial they felt more confident that they succeeded on. Critically, results showed that participants' ability to discriminate between the quality of their performance was equally good for within-task and between-task confidence comparisons. Thus, the authors conclude that observers could access and use their confidence at an abstract level, that was not task-dependent (de Gardelle & Mamassian, 2014). In a second study, the authors extended these findings by demonstrating that confidence can further be compared across visual and auditory decisions, while still maintaining the same precision as within-task comparisons. They propose that, given these findings, it is plausible that subjective confidence is combined with utility, to form expected utilities of different choices and guide behaviour (de Gardelle et al., 2016). In addition to allowing for the comparison of different tasks, such an abstract representation of confidence could also support flexible, generalisable behaviour: Any situation that leads to the same explicit representation of confidence can then be associated with the same adaptive response, without having to learn the association for each situation separately. For example, low confidence might lead to gathering more information before committing to a choice (Desender et al., 2018; Desender, Murphy, et al., 2019), regardless of the source of the uncertainty, such as a noisy environment, lack of attention, or weak evidence.

In summary, initial frameworks of metacognition were defined broadly to include any type of cognitive process that monitors and controls another cognitive process. Recent frameworks have differentiated between an implicit metacognitive system, which is automatic and fast, and an explicit one, which is slower and conscious. Explicit confidence, as an important instance of the latter type of information, appears to act as a common scale that allows the comparison of different representations, both between and within individuals. When information is shared, there is need for an evaluation of its reliability to be broadcasted as well, in order to enable its effective integration and use in guiding behaviour. Thus, under this framework, an explicit representation of our confidence in our beliefs and decisions is critical for cognitive functions that rely on the communication and integration of various types of information. The following section examines evidence that, consistent with this framework, confidence is involved in regulating several aspects of adaptive behaviour such as learning, decision making and information seeking.

1.6 The use of confidence to regulate behaviour

Research in the area of metamemory has demonstrated the important role of metacognitive judgements, such as feeling of knowing and judgments of learning, in guiding successful self-regulated learning (see Bjork et al., 2013 for a review). In contrast, research on confidence within the field of cognitive psychology has traditionally focussed on its cognitive and neural basis, rather than its use to guide behaviour. However, recent years have shown an increase in studies examining the influence of confidence on a range of behaviours, as reviewed below.

Guiding decisions

An important aspect of making adaptive decisions is reaching the right balance between collecting enough evidence to make an accurate choice, and not expending unnecessary time and effort. Since confidence reliably predicts accuracy, it has been suggested that confidence can be used to enhance performance by guiding the amount of evidence required before reaching a decision (the decisional threshold). Specifically, when confidence is low, the probability of making a correct choice is low compared with when confidence is high, and people should gather more evidence before committing to a decision (Yeung & Summerfield, 2012). Consistent with this logic, it has been demonstrated that confidence in one decision modulates the amount of evidence needed before committing to a choice in a subsequent, similar decision: more evidence is accumulated following low-confidence choices, leading to slower but more accurate decisions (Desender, Boldt, Verguts, et al., 2019). Similarly, the decision whether or not to sample more information before committing to a choice on a given trial is driven by subjective confidence that a correct decision can be reached, rather than objective accuracy (Desender et al., 2018).

Whereas decision making studies in the lab typically examine isolated choices, with reward contingent on performance in a single trial, achieving goals in real-life often involves a sequence of multiple decisions. Critically, the success of each stage might not be known until the goal is reached. In such sequential tasks, it is useful to represent confidence in the previous stages, and adjust the time invested in subsequent stages accordingly. Consider, for example, arriving at your car in the morning, to discover the keys are not in your pocket as you had thought. The amount of time you spend searching will depend on how sure you are that you have the keys with you. If you are relatively certain, you will search for longer, while if unsure, you might only look briefly before returning home. A recent study tested whether, as illustrated, confidence in a previous choice is carried over and used in multiple stage decisions. Participants were asked to judge the direction of motion of a random-dot display. Critically, trials were paired, and participants had to respond correctly on both decisions in order to receive reward. The results were consistent with the prediction, and demonstrated that the decisional threshold of subsequent stages was adjusted according to confidence on the previous choice (Van Den Berg et al., 2016). Studies have also found that confidence modulates serial dependence in perceptual decisions, so that people judge a stimulus as more similar to previous stimuli following past decisions that were made with higher confidence (Samaha et al., 2019). Furthermore, in value-based decision making, people are less likely to change their mind if they previously experienced high confidence on the same choice (Folke et al., 2017). Taken together, these results demonstrate that confidence plays an important role in decision making, and confidence in one choice can influence the policy that is adopted in subsequent choices.

Guiding learning

To make correct decisions, we must learn from our past experiences and adjust behaviour according to experienced outcomes. Normative theories of learning consider how behaviour is adapted based on mismatches between expectations and outcomes (Rescorla, 1972; Sutton & Barto, 2018). While these models have been successful in describing learning when external feedback, such as reward, is available, they are unable to account for successful adjustments of behaviour when feedback is lacking. In contrast, in real-day life, feedback is often not available after every action, making the question of how people are able to learn under such conditions an important one. Since confidence is widely defined as the subjective probability of being correct, it provides a plausible candidate for an internal signal that can guide learning when external feedback is lacking.

This proposal was tested by two studies that recorded brain activity using fMRI while participants performed a categorisation task (Daniel & Pollmann, 2012) and a perceptual learning task (Guggenmos et al., 2016), in the absence of external feedback. Both studies found evidence that signals in the mesolimbic dopamine system are modulated by confidence. Specifically, mesolimbic activity reflected the difference between experienced confidence and predicted confidence (based on previous trials). The authors concluded that 'confidence prediction errors' act as learning signals in a similar way to reward prediction errors (Daniel & Pollmann, 2012; Guggenmos et al., 2016). Crucially, Guggenmos et al. (2016) further demonstrated that participants' performance in the task improved over time, showing that they were indeed able to learn without feedback. Extending these findings, a recent study demonstrated that people can learn about the reliability of others by comparing opinions to their own beliefs when they are confident they know the correct response, even when feedback is lacking (Pescetelli & Yeung, 2020). If, as suggested by these results, people rely on confidence in order to learn, then the ability to learn when feedback is absent should correlate with the quality of people's metacognitive judgements. This prediction was empirically tested in a recent study, and results showed that participants with better metacognitive efficiency (i.e., participants who evaluated their own performance more accurately) could also more successfully learn about predictive cues in a feedback-free environment (Hainguerlot et al., 2018).

The above studies have demonstrated that confidence can be used to guide learning, acting as a replacement for external feedback. Another suggested role of confidence in learning is as a weighing factor, balancing the influence of new information and prior knowledge according to their reliability. In other words, confidence can act as an adjustable learning rate (or the prior probability in Bayesian based models), so that new evidence has a smaller influence on our beliefs the stronger our confidence in previous knowledge is (Meyniel & Dehaene, 2017; Meyniel et al., 2015). Coherent with this proposal, recent work from our lab has shown that subjective confidence modulates feedback processing, so that higher confidence in learned associations leads to smaller updates in beliefs following negative feedback, and vice versa (Ben Yehuda, 2019).

Guiding strategic behaviour

The studies discussed thus far have highlighted the role of confidence in learning and decision making. There is growing evidence that confidence is further involved in higher-order strategic behaviour, for example in reaching the right balance between exploration and exploitation. When choosing between options in complex environments, people must decide whether to choose the option which they currently believe has the best outcome, or explore new alternatives in the hope of finding an even better option (e.g. should you listen to your favourite album, or try a band you've never heard before). It is well established that people take the level of uncertainty in the environment into account when making such choices (Sutton & Barto, 2018). Boldt, Blundell, et al. (2019) extended previous findings to show that people use their subjective degree of certainty in the beliefs they hold about the value of different options (measured by explicit confidence reports, rather than modelled based on objective certainty), to arbitrate between exploration and exploitation strategies. This finding was reflected in a tendency to explore more when certainty in the beliefs about the value of different options was low, a result that makes intuitive sense because low certainty implies that there is more to learn about the environment.

A strategic way that people often optimise and extend their abilities in everyday life is through the use of external tools that decrease the requirements on their cognitive system (Clark & Chalmers, 1998). For example, a common use of external tools might be writing down important information, such as a shopping list, to minimise demands on our memory system. Such use of physical actions and external devices to alter the information processing demands of a task on our cognitive system is termed *cognitive offloading* (Risko & Gilbert, 2016). According to a metacognitive framework of cognitive offloading, selecting whether to rely on internal processes or offload cognition on to external tools is determined by metacognitive evaluations (Risko & Dunn, 2015; Risko & Gilbert, 2016). These evaluations are based on our beliefs regarding the cost and benefits involved in each line of action. For example, a decision whether to set a reminder of your mother's next birthday on your phone will be informed by beliefs about how likely you are to forget (did you just remember she had a birthday last week?), how angry you think she will be if you forget again (did she hang up on you when you called just now?), and how much effort is involved in setting a reminder (you are now lying on the sofa and the phone is out of reach). Consistent with a metacognitive framework of cognitive offloading, recent empirical evidence has shown that people with lower confidence in their memory capabilities are more likely to set reminders than rely on their internal memory (Boldt & Gilbert, 2019; Gilbert, 2015). Importantly, these findings are present even when controlling for differences in objective memory abilities, demonstrating that the effects are driven by beliefs, and not objective capacities.

Guiding social decisions

People make use not only of external tools, but also of communication with other people, in order to achieve better performance. The role of confidence has been studied extensively in the fields of social and organisational psychology, with studies examining group decisions, as well as the influence of advice on a decision maker. It has been shown that more confident group members tend to be more accurate and have a stronger influence on the collective decision of a freely interacting group (Zarnoth & Sniezek, 1997). Bahrami et al. (2010) investigated whether, as the phrase goes, two heads are better than one. They found that the performance of paired observers in a perceptual task improved compared to their individual accuracies when they were allowed to communicate the reliability of their choice (i.e., their confidence judgements). The importance of communication for improving task performance via joint decision making has been supported by several other studies (Bang et al., 2014; Fusaroli et al., 2012).

Often in life, even when the final choice is made by an individual, people solicit information from others before reaching their decision. For example, people might seek advice from doctors, lawyers and other experts. Studies have demonstrated that people tend to prefer the advice of more confident advisors, and judge them as having better knowledge (Price & Stone, 2004). Similarly, people use confidence as a cue to advisors' expertise and therefore trust confident advisors more, and follow their advice more compared to less confident advisors (J. A. Sniezek & Van Swol, 2001). In eyewitness identification, the reliance placed upon confidence as a cue to eyewitness accuracy has been highlighted by courts, lawyers and jurors (see

Penrod and Cutler, 1995 for a review). Experimental studies with mock jurors have consistently shown that the confidence expressed by an eyewitness is a key factor in determining whether people believe the identification was accurate (Brewer & Hupfeld, 2004; Tenney et al., 2007). Furthermore, it has been demonstrated that the calibration of metacognitive judgements is also used as an important cue for juror judgements: If a witness expresses high confidence in his judgement but is shown to be in error, he will lose more credibility than a witness who made the same error but was originally unconfident (Tenney et al., 2007).

In summary, the importance of confidence judgements in social situations is two-fold. A person's confidence determines both the extent to which they rely on their own first-order cognitions (rather than seek or use advice), as well as the extent to which other people will be influenced by these judgements. The extensive influence of confidence in social settings is consistent with the suggestion that a key reason that we explicitly represent our metacognitive judgements is so that we can communicate them to others (Shea et al., 2014).

1.7 The role of confidence, an ongoing investigation

The literature reviewed above demonstrates that people are able to form and report judgements about their performance and knowledge that reflect objective accuracy, even if not perfectly. Educational and social psychologists have studied how confidence influences behaviour for many years, with research showing that confidence can guide several aspects of behaviour, such as learning strategies and group decision making. Within the field of cognitive psychology, studies have traditionally focussed on the neural and computational basis of confidence, or used confidence judgements as a means to investigate other constructs, such as awareness. However, as reviewed in the previous section, recent studies have begun to explore how internal representations of confidence may guide an individual decision maker. The research reported in this thesis builds on these ideas and approaches to explore the functional role of confidence, the specific ways in which it is used, and the scope of its influence. A theme of the research is emphasis on the question of why people have an explicit sense of confidence.

To study the role of confidence, I make use of carefully-controlled perceptual decision making paradigms, such as those described above. As noted, these tasks allows for precise control over task information, as well as precise measurement of performance. By manipulating the amount of information provided, it is possible to control also for performance levels. These features are particularly important to the present work, since they make it possible to disentangle the effects of confidence from those of objective performance.

In this thesis, I examine three ways in which confidence is used to guide behaviour. Chapter 2 examines the role of confidence in task selection. Specifically, it tests the hypothesis that confidence acts as an intrinsic cost-benefit factor when making decisions, biasing people towards situations in which they are more confident. Chapter 3 studies whether confidence can be used to guide behaviour flexibly, depending on the current context. This is studied by examining how confidence influences advice seeking, testing the hypothesis that the seemingly natural relationship between low confidence and advice requests is not a fixed, but can instead vary according to whether people know, or do not know, the quality of the advice. Finally, Chapter 4 looks at the relationship between confidence and attention, testing the hypothesis that the allocation of attention to information is modulated by confidence in its relevance. Together, these experiments investigate different levels at which confidence influences behaviour: from high-order strategic behaviour, to lower level effects on attention. Across this wide scope, all the experiments are driven by the underlying theory that explicit representations of confidence serve to guide behaviour in a flexible, goal-directed manner, as an abstract signal that is shared across multiple cognitive processes.

2

Confidence guides task choices

2.1 Introduction

In our daily lives, we often have to choose and prioritise between different tasks. For example, when arriving at work this morning, it is likely that you had to choose between reading and replying to emails, writing a paper, or any other work that has piled up. The present chapter considers the role of confidence in guiding such choices, thereby exploring its role beyond the scope of an immediate decision (‘Am I confident enough to make this decision now?’) to include strategic choices about which situations to pursue (‘Am I confident enough in this task to engage with it?’).

In this way, the current chapter draws research on decision confidence together with a hitherto separate line of work exploring how people choose between different tasks or lines of action. In this other work, a core hypothesis is that adaptive selection is guided by the learnt potential costs and benefits of each line of action. These costs and benefits include the outcomes of our actions, such as monetary loss or gain (Kahnemann, 1984), as well as costs and benefits intrinsic to the action itself. For example, a well-established intrinsic cost of an action is the amount of physical work or effort that it entails (Hull, 1943; Solomon, 1948; Walton et al., 2006). This concept has been extended to cognitive as well as physical effort (Apps et al., 2015; Kool et al., 2010; Westbrook et al., 2019). In particular, in a set of experiments that provided the methodological basis for the present experiments, Kool and colleagues employed a demand selection task in which participants chose between cues associated with tasks requiring different degrees of cognitive effort. The participants were not instructed about these associations, but could learn them through experience of the outcomes of their cue selections. Kool et al.'s results demonstrate that people tend to avoid selecting the line of action requiring more mental effort (Kool et al., 2010). Interestingly, although in this study they report that participants were unaware of the experimental manipulation, a later study found bias towards the less demanding task only when participants noticed the differences in effort demands (Gold et al., 2015). This finding raises questions regarding the role that perceived as opposed to objective costs and benefits play in the development of task preferences.

Two recent studies addressing this issue have found evidence for the involvement of metacognitive awareness in task selection. Dunn and colleagues found that even when objective performance is equal, participants avoid situations which they believed would be more effortful, time demanding and less accurate (Dunn et al., 2016). Similarly, in another study that manipulated cognitive demand using subliminal priming, bias towards a low-demand context was only apparent in participants who reported differences in task difficulty (Desender et al., 2017). Based on the results described above, it seems plausible that confidence could affect task selection in a fashion similar to related metacognitive judgements of effort and difficulty. There would be theoretical value in showing this relationship: Whereas there is ongoing debate about what exactly constitutes cognitive effort and what factors lead to its avoidance (Shenhav et al., 2017), decision confidence is well defined — as the internal belief that a response made was correct — and computationally well specified — as a transformation of the evidence accumulated in favor of a decision (even if there is debate about the precise detail of this transformation; Pleskac and Busemeyer, 2010; Pouget et al., 2016; Sanders et al., 2016). Thus, linking confidence to task choice would have meaningful theoretical implications through establishing a clear computational basis of task preferences.

Following this rationale, we hypothesized that people's confidence in their decisions acts as a cost or benefit factor intrinsic to the decision itself (cf. Lebreton et al., 2015). If this hypothesis is true, confidence should affect task selection even when controlling for other factors such as objective performance or external rewards. Importantly, we expected confidence to influence behavior in a graded fashion, beyond the binary classification of a response as being correct or incorrect. Specifically, we expected people to demonstrate a bias towards tasks in which they were more confident, even if they detected similar numbers of errors across the different tasks. As such, our predictions differ from those expected based solely on error aversion or classic reinforcement learning theories, whereby behavior is adjusted based on the outcome (i.e. accuracy or reward) of an action (Holroyd & Coles, 2002; Holroyd et al., 2005). Extending this idea, we asked whether confidence will influence behavior even when external feedback is present, so that participants can know objectively whether or not their responses were correct without having to rely on their own internal evaluations. If effects of confidence are apparent even in the presence of feedback, this result would support the view that the fine-grained confidence signal is integrated into decisions automatically (Lebreton et al., 2015), and not only used as an internal monitoring signal when external feedback is lacking.

We tested these predictions in a series of five experiments, using a task-selection paradigm in which participants chose freely between two cues that (initially unbeknownst to them) were associated with decisions of differing levels of confidence. Specifically, participants performed a perceptual decision task in which they reported which of two boxes contained more dots, with the number of dots in each box changing dynamically during stimulus presentation, both before and after the decision (Charles & Yeung, 2019). Regardless of the cue chosen, participants encountered a decision of equivalent difficulty (i.e., the mean dot difference between the two boxes was identical for both cues). However, after participants registered their decision, the stimulus continued to dynamically evolve, with the strength of post-decisional evidence varying according to the selected cue. This manipulation allowed us to break the typical link between subjective confidence and the initial difficulty and accuracy of decisions (Kepecs et al., 2008; Kiani & Shadlen, 2009), in line with recent theoretical models (Moran et al., 2015; Pleskac & Busemeyer, 2010; Yu et al., 2015) and neurophysiological evidence (Boldt & Yeung, 2015; Fleming et al., 2018; Murphy et al., 2015) suggesting that confidence incorporates information that is accumulated after the initial decision has been made. We predicted that differences in post-decisional evidence strength—either in favour of or against the initial choice—would lead to differences in the confidence levels associated with each cue, and that this in turn would lead to the development of a bias towards tasks associated with higher confidence, even across tasks matched for objective decision difficulty (as reflected in accuracy and reaction time). To anticipate, our results support this hypothesis, with people selecting more often the task in which they experience and report higher confidence.

2.2 General Methods

Task and Procedure

All five experiments shared the same general method, comprising 4 blocks of a task selection paradigm with 80 trials per block. Each trial was divided into three sections: Task selection, perceptual judgement task, and confidence rating (see Figure 2.1). During task selection, participants chose one of two cues associated with perceptual judgement tasks that were identical except for the strength of post-decisional evidence presented. Participants were instructed to choose between the cues freely, and told that if one began to feel preferable they could choose it more often. They were not told the nature of the differences between the tasks associated with the cues. Trials started with a grey fixation cross presented in the center of the screen. After 300 ms, the cues (different coloured circles) appeared symmetrically above and below the fixation cross. Eight circles in different colors (red, green, blue, yellow, orange, purple, pink and cyan) were used throughout the experiment, with two colors randomly selected for each block, and assigned randomly to one of two locations in which they appeared throughout the entire block. Participants had unlimited time to select a cue using the mouse cursor. 200 ms after cue selection, the

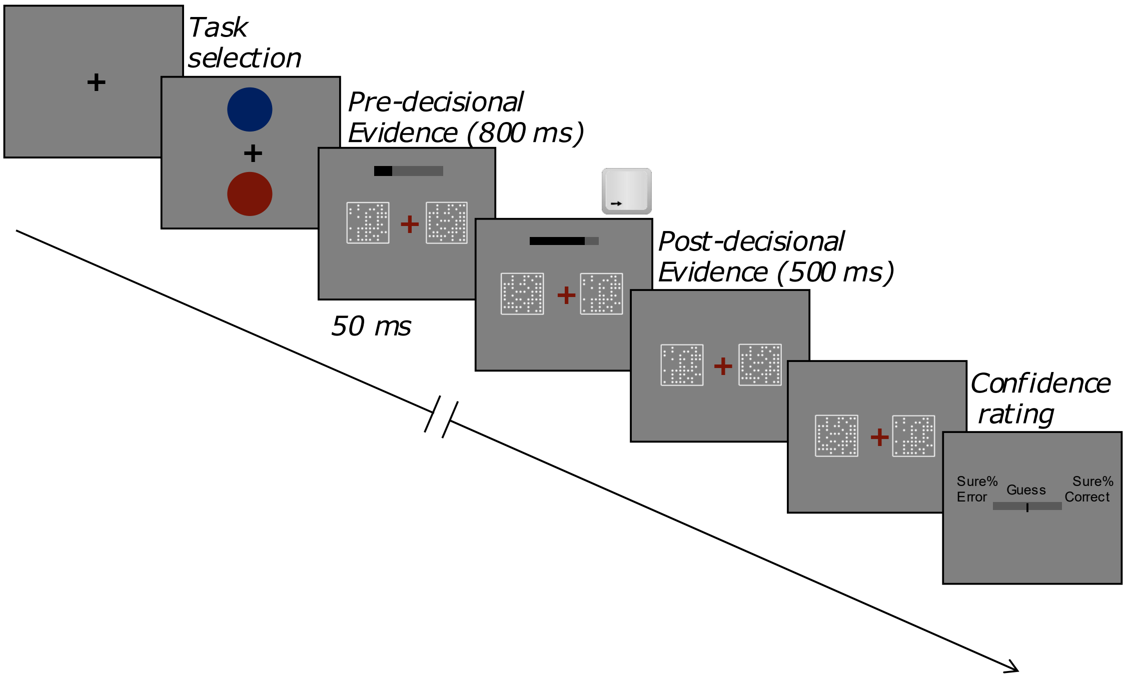

Figure 2.1: Task Design. Trials began with a choice between task cues associated with different post-decisional evidence. Following task choice, participants made a speeded judgement about which of two squares contained more dots on average. The number of dots was sampled from two normal distributions centred around different means, and was updated every 50 ms. Dot stimuli remained on the screen and continued to be updated every 50 ms after participants' responses. After viewing post-decisional evidence, participants rated their confidence in their response.

fixation cross changed color to that of the selected cue, and after another 500 ms the fixation cross flashed, indicating the perceptual task would start shortly. Following a varying delay of 101-500 ms, two empty grey squares appeared on either side of the fixation for 300 ms, after which dots appeared within the squares and participants were asked to judge which square contained more dots on average. The number of dots and their locations within the squares were updated every 50 ms, with dot numbers sampled randomly from two distributions with different means (see specific experiments) and an equal standard deviation of 20 dots. Thus, although the number of dots within each square changed during the trial, the correct square contained more dots on average. The correct square (left vs. right) was randomized across trials, with each square being correct on half of the trials in each block. Participants had up to 800 ms to respond by pressing the right or left mouse button, with a bar at the top of the screen indicating time progression. After participants responded, the stimulus remained on-screen and continued to be updated every 50 ms for an additional 500 ms, allowing the accumulation of additional post-decisional evidence. This specific period was chosen based on previous indications that evidence presented during this time influenced confidence ratings (Charles & Yeung, 2019). Importantly, the distances between the dot number distributions were changed in this post-decisional period, bringing the mean difference between boxes either further apart or closer together. Tasks associated with each cue differed in the strength of post-decisional evidence presented (i.e. the mean difference in dot numbers after the participant's response), allowing the manipulation of participants' confidence. The exact manipulation of post decisional evidence varied across experiments, and is detailed in the corresponding methods sections below.

In the final part of each trial, participants were asked to rate their confidence in their response on a 50-point scale ranging from 'sure incorrect' (low confidence) to 'sure correct' (high confidence), with 'guess' in the center. The starting location of the cursor on this scale varied randomly across trials, and participants responded by moving the cursor to the appropriate location and clicking the left mouse button. There was no time limit for this part. If participants did not respond to the dot task within the 800 ms time limit, the trial ended and the next trial began, with the post-decisional evidence and confidence sections skipped.

Before the main experiment began, participants were given written instructions and trained on the dot task. Participants were required to answer a minimum of 12 out of 20 trials (60%) correctly in the training block, and training was repeated until they reached this requirement. During the training block, visual feedback on the first-order judgement was given after every trial ('correct', 'incorrect', 'too slow') to assist participants in learning the task. Except for Experiment 4, feedback was not given during the main experimental blocks. However, participants were shown the percentage of correct first-order judgements they made at the end of every block to help keep them engaged. After completing the computer task participants answered the "need for cognition" scale (Cacioppo et al., 1984) and a debriefing questionnaire assessing their awareness of the experimental manipulation (adapted from Kool et al., 2010). Findings relating to these questionnaire measures are presented in combination for all 5 experiments below in the final section of results. All procedures were approved by the Oxford University Medical Sciences Interdivisional Research Ethics Committee.

Stimuli and Apparatus

All stimuli were created and presented in Matlab using Psychtoolbox3 extension (Brainard, 1997; Kleiner et al., 2007; Pelli, 1997). Stimuli were presented on a 20-inch CRT screen with grey background (128, 128, 128), a refresh rate of 60 Hz, and 1600 x 1200 resolution. The task cues were circles in 8 different colors (red, green, blue, yellow, orange, purple, pink and cyan). Dot stimuli were comprised of two grey squares (150, 150, 150), in which white dots were presented (200, 200, 200). The experiments took place in a dimly lit room, with participants sitting at a comfortable viewing distance from the screen.

2.3 Experiment 1

Experiment 1 aimed to test our main hypothesis: that confidence holds an intrinsic value in the decision-making process, so that people should prefer situations in which they are more confident (even when controlling for the difficulty of the first-order decision). Participants chose between two cues and performed a task associated with the chosen cue. To manipulate participants' confidence independent of task difficulty, cues were associated with tasks that were identical in the pre-decision stage, but which differed in terms of post-decisional evidence. One cue was associated with a task in which the evidence supporting the participant's decision was boosted after their decision, which should lead to relatively high confidence (Charles & Yeung, 2019; Fleming et al., 2018). The other cue was associated with a task in which the evidence supporting the decision was reduced, which should lead to lower confidence. We predicted that participants would select the cue associated with a boost in evidence supporting their choice (and hence with higher confidence) more often than the alternative cue.

Methods

Participants

Twenty participants took part in the experiment (14 females, ages 19-35 years, mean age 24.5, SD = 3.49). All participants had normal or corrected to normal vision. They provided written informed consent and were compensated with course credit or payment for their participation. To select an appropriate sample size and ensure that the current task selection paradigm would be able to capture task preferences, we ran a pilot study. The sample size for the pilot (n = 20) was chosen on the basis of previous experiments that have found effects of subjective confidence on decision making (specifically on information sampling) using this number of participants (Desender et al., 2018). In the pilot, each cue was associated with a different objective difficulty of the dot task (i.e., also varied in dot difference prior to participants' decisions). One cue was associated with strong evidence, with the distributions of dot numbers centered around $M_C = 222$, $M_I = 178$ (mean dot numbers in the correct and incorrect boxes, respectively). The second cue was associated with weak evidence, with distributions of dot numbers that were closer together, centered around $M_C = 212$, $M_I = 188$. The dot number distributions were identical for the pre- and post-decision stages. Under these conditions, we found that participants showed a consistent preference towards the easier, strong-evidence task than the more difficult weak-evidence task, with a large effect size (mean proportion of easy task selections = 0.63, $t(19) = 4.17$, Cohen's $d_z = 0.93$). Though we did not expect our post-decisional manipulation to produce an effect size as large as this, with a sample size of 20 we would be able to detect an effect that is 2/3 this size ($d_z = 0.62$) with a power of .85. Power calculations were performed using the GPower3 program (Faul et al., 2007).

Stimuli and Procedure

The experiment followed the general procedure described above, with participants asked to choose on each trial between cues that were associated with subtly different perceptual decision tasks. Thus, regardless of participants' choice of cue, the difficulty of the perceptual task was the same on all trials, with fixed evidence strength in the pre-decision stage $M_C = 212, M_I = 188$. However, after participants selected the box that they thought contained more dots, post-decisional evidence differed according to the cue chosen at the start of the trial. In the boost selected box condition, 10 dots were added to the mean number of dots in the chosen box, and 10 dots removed from the unchosen box, with the stimulus continuing to update every 50 ms, as in the pre-decision stage. Thus, for this cue, strong evidence was presented following correct responses ($M_C = 222, M_I = 178$) and weak evidence was presented following incorrect responses ($M_C = 202, M_I = 198$). We expected this manipulation to increase participants' confidence (i.e., higher subjective probability that the response was correct) regardless of the accuracy of the initial decision — with strong evidence reinforcing correct choices, and weak evidence making errors less likely to be detected. In the boost alternative box condition, on the other hand, 10 dots were removed from the mean of the chosen box and 10 dots added to the unchosen box after the decision. Thus, weak evidence was presented following correct responses ($M_C = 202, M_I = 198$) and strong evidence was presented following incorrect responses ($M_C = 222, M_I = 178$), which should reduce confidence by reducing the strength of evidence in favour of initially-correct responses and increasing the likelihood that errors would be detected. Note that the underlying mean continued to favour the correct box in the post-decisional period for both conditions.

Results

Manipulation effects on performance and confidence

Since the two tasks were identical, by design, in terms of average strength of evidence presented prior to participants' first-order judgement, performance in the tasks was

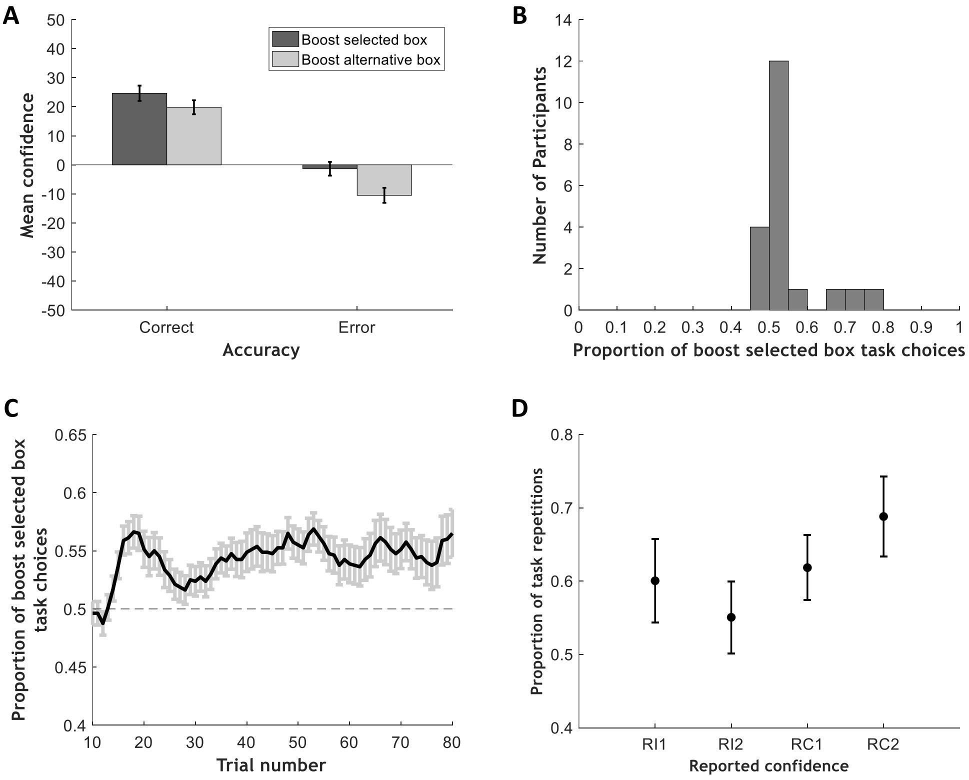

Figure 2.2: Results of Experiment 1. (A) Effects of post-decisional evidence manipulation on confidence. Mean confidence ratings for the boost selected box and boost alternative box tasks, divided according to trial accuracy. (B) Distribution of individual participants' proportion of boost selected box task choices (n=20). (C) Ten-trial running average showing the proportion of boost selected box task choices across blocks. Dashed line represents chance level. (D) Mean proportion of task repetitions according to reported confidence. Trials were first divided into groups according to reported accuracy (RI = reported incorrect, RC = reported correct), followed by a division according to median splits within the accuracy group (1 = below median confidence, 2 = above median confidence). Note: Error bas represent standard errors of the mean (SEMs).

not expected to differ. Indeed, performance levels were similar, with participants responding to a similar proportion of trials within time limits ($M = .95$, $SD = .04$ vs, $M = .95$, $SD = .03$ for boost selected and boost alternative tasks; $t < 1$). On trials with on-time responses, similar accuracy levels were observed for the boost selected ($M = 0.74$, $SD = 0.09$) compared to boost alternative task ($M = 0.73$, $SD = 0.08$), with no consistent differences across participants ($t(19) = 1.08, p = .30, d_z = 0.24$). Similarly, mean reaction times were well matched across tasks ($M = 598$ ms, $SD = 42$ vs $M = 597$ ms, $SD = 46$, for the boost selected and boost alternative

	Task1								Task2							
	Correct				Error				Correct				Error			
Exp.	*M*	Min	Max	*SD*	*M*	Min	Max	*SD*	*M*	Min	Max	*SD*	*M*	Min	Max	*SD*
1	122	87	187	25	43	17	71	14	102	55	130	20	39	11	63	14
2	120	78	177	24	46	22	84	14	103	41	125	22	38	18	60	12
3	143	80	249	45	47	19	72	14	85	8	158	38	32	2	57	15
4	129	79	196	26	37	21	56	9	108	59	141	21	33	11	58	12
5	101	60	200	30	53	33	99	14	97	11	149	25	52	4	84	17

Table 2.1: Number of correct and error trials (mean, minimum, maximum and SD) for each task in Experiments 1-5. Task 1 is the task associated with higher confidence (boost selected / correct box for Experiments 1-4 and boost incorrect box for Experiment 5). Task 2 is the task associated with lower confidence (boost alternative / incorrect box for Experiments 1-4 and boost correct for Experiment 5).

	Reported Error								Reported Correct							
	RI1				RI2				RC1				RC2			
Exp.	*M*	Min	Max	*SD*	*M*	Min	Max	*SD*	*M*	Min	Max	*SD*	*M*	Min	Max	*SD*
1	26	9	76	16	31	9	87	18	107	52	134	16	129	60	180	25
2	29	15	55	10	33	19	56	10	109	84	129	12	124	90	144	16
3	27	9	57	13	30	10	64	14	112	75	139	16	131	110	150	12
4	34	20	54	9	38	20	54	9	111	96	132	10	124	104	144	10
5	33	5	80	18	37	10	84	18	99	55	124	18	122	62	175	26

Table 2.2: Number of trials according to reported accuracy (RI = reported incorrect, RC = reported correct) and median splits (1 = below median confidence, 2 = above median confidence) within each accuracy group for Experiments 1-5. Note that for Experiment 4 trial numbers are based on objective, rather than reported accuracy. Data from participants that were excluded from analyses using this trial grouping are similarly excluded from the table.

tasks, respectively; $t < 1$). Nevertheless, manipulation of post-decisional evidence had the hypothesized effect on confidence ratings, as revealed in a 2 x 2 repeated measures ANOVA with factors of task (boost selected box, boost alternative box) and decision accuracy (correct, incorrect; see Figure 2.2A for the results and Table 2.1 for data on trial numbers). A significant main effect of accuracy indicated that, as expected, participants were more confident in their correct responses than their errors, $F(1, 19) = 90.60$, $p = .001$, $\eta^2 = 83$. Crucially, a significant main effect was also found for task, $F(1, 19) = 45.77$, $p = .001$, $\eta^2 = .71$, with higher confidence ratings in the boost selected task than in the boost alternative task. The interaction between confidence and accuracy was also significant, $F(1, 19) = 5.41$, $p < .05$, $\eta^2 = .22$, reflecting a greater confidence difference between tasks on error trials.

Task selection

Given that we successfully manipulated confidence across tasks, of critical interest was whether this confidence difference would be reflected in participants' task preferences. Consistent with this prediction, and as shown in Figure 2.2B, participants exhibited consistent preference for the boost selected task over the boost alternative one (mean proportion of boost selected choices = 0.54, range = 0.47 - 0.76, SD = 0.08). 16 of 20 participants (80%) chose the boost selected task more often than the boost alternative task, with this choice rate significantly different from chance level ($t(19) = 2.37, p = .028, d_z = .53$). The timecourse of the development of this preference (Figure 2.2C) suggested that it was established early in the block.

Error monitoring and confidence impact on task selection

To explore further this preference towards tasks associated with higher confidence, we examined how preferences emerge on a trial-by-trial basis as a function of participants' error monitoring and experienced confidence levels. Thus, we evaluated whether people tend to repeat a task more often if they believe they were correct (rather than incorrect), and separately whether they tend to repeat the same task if they experience higher confidence (vs. lower confidence) in correct decisions.

To test the effects of trial-by-trial error monitoring and confidence on task choice, trials were first divided according to reported accuracy, and then median confidence ratings were calculated separately for reported correct (RC) and reported incorrect (RI) trials, with trials in which participants reported guessing ($M = 0.04, SD = 0.03$ of trials) excluded from this analysis. Median values were used to divide RC and RI trials into subsets with higher vs. lower than median confidence ratings (RI1, RI2, RC1, RC2; trial counts are listed in Table 2). For each subset, we calculated the likelihood that participants would repeat the same task (vs. switch) on the next trial — i.e., as a function of reported accuracy, and as a function of confidence level (see Figure 2.2D). Because we define confidence as subjective p(correct), we define a trial with a 'sure incorrect' confidence rating as having the lowest possible decision confidence (even though the participant is expressing certainty in this evaluation).

A 2 x 2 repeated measures ANOVA was performed on task repetition probability, with reported accuracy (correct, incorrect) and confidence rating (below median, above median) as factors. Significant effects were found for reported accuracy ($F(1, 19) = 7.30$, $p = .014$, $\eta_p^2 = .28$), with a higher proportion of task repetitions following trials that were reported as correct than incorrect, and for the interaction between reported accuracy and confidence ($F(1, 19) = 5.62$, $p = .029$, $\eta_p^2 = .23$). There was no significant main effect of confidence ($F < 1$). Follow-up analysis revealed that for trials reported as correct, participants were significantly more likely to repeat the same task after experiencing higher confidence in a decision ($t(19) = 3.57$, $p = .002$, $d_z = 0.80$). There was no significant effect of confidence when trials were reported as incorrect ($t(19) = -1.14$, $p = .27$, $d_z = -0.26$), but interestingly the numerical trend was towards participants being more likely to repeat tasks, not less, when they experienced lower confidence (i.e. were more certain they had made an error).

Evidence strength and confidence impact on task selection

Our final analysis considered an alternative explanation of the above findings, which is that participants are showing a preference for situations associated with stronger evidence, rather than higher confidence. Our results show that confidence is increased when people receive post-decisional evidence that provides stronger support for (or less contradiction of) their decision, and that people prefer tasks with this evidence boost. With this manipulation, when choosing the boost selected box task, participants were presented with strong evidence when they responded correctly and weak evidence when they responded incorrectly. The opposite was true for the boost alternative box task. Because participants responded correctly more often than incorrectly (mean accuracy = 0.73, SD = 0.08), this means that the boost selected box task was also associated with stronger evidence overall than the boost alternative box task (mean dot difference of 28.50 vs. 20.03; SD = 1.72 vs. 1.65, respectively, $t(19) = 12.07$, $p < .001$).

To provide evidence that the emerging preference was not only the result of a bias towards stronger evidence, we fitted logistic regression models predicting task repetition on trial N+1 as a function of evidence strength and confidence on trial N, for reported correct and error trials separately, separately for each participant. Trials were divided by reported accuracy to allow the detection of differential effects of confidence, expected based on the interaction effect reported above. Evidence strength was calculated as the mean difference in dot numbers between the two boxes across frames in each trial, taking into account both pre- and post-decisional evidence. Note that although the mean dot difference between boxes was fixed across conditions, evidence strength varied across trials because of random variation in the number of dots presented in each stimulus frame around the overall mean values. Standard scores for evidence strength and confidence were computed for correct and incorrect trials separately before running the regression.

For correct trials, we observed positive slopes for confidence in 17 of 20 participants, indicating that participants were more likely to repeat task choice following decisions made with higher confidence, with the mean beta being significantly different from zero across participants (M = 0.40, SD = 0.46, $t(19) = 3.94$, $p < .001$, $d_z = 0.88$). For incorrect trials we found the reverse relationship, with participants more likely to repeat tasks when they were less confident (i.e. more sure they had made a mistake). Slopes were negative for 16 participants and the mean beta significantly different from zero ($M = -0.21$, $SD = 0.41$, $t(19) = -2.34$, $p = 0.03$, $d_z = -0.52$). Evidence strength did not significantly predict repetitions in task choice: slopes were positive for 10 of 20 participants for both reported correct and error trials, and betas were not significantly different from zero across participants ($M = 0.02$, $SD = 0.20$; $M = -0.14$, $SD = 0.64$, for correct and error trials respectively; $ts < 1$). These results support our interpretation that participants' bias for the boost selected task is a result of confidence preferences, above and beyond strong evidence. The results are also consistent with the error monitoring analysis presented above, demonstrating that confidence reliably predicts task repetitions for correct trials (or those perceived as such). For perceived incorrect trials, the effects of confidence on task choice are weaker, but the results are consistent with the numerical trend reported above.

Discussion

The results of Experiment 1 support our main hypothesis, with people choosing situations in which their confidence is higher more often. The effect overall was small (M = .54 vs. .46 for the alternative task that was associated with lower confidence) but was reliable, apparent in a large majority of participants, and was apparent even though the impact of the evidence manipulation on confidence itself was very subtle. Furthermore, at the trial level, task selection on trial N+1 was reliably predicted by confidence on trial N, with participants more likely to repeat their task choice after a perceived correct decision made with high vs. low confidence. Interestingly, for trials perceived as incorrect, participants were more likely to repeat their choice when they were more certain they had made a mistake (i.e. lower in confidence). However, these results should be interpreted with care as they were less robust, were not significant in the ANOVA, and are based on the relatively small number of trials that were perceived as errors. It is important to note that, consistent with previous studies showing that people find errors aversive (Hajcak & Foti, 2008), participants were also overall less likely to repeat their task choice following trials they judged as incorrect. Thus, if participants reported more errors in one task compared to another, this would lead to a lower proportion of trials spent on this task. We suspected that the demonstrated preference might be driven in part by these differences in error detection. Indeed, participants judged a significantly larger proportion of trials as being errors in the boost alternative compared to the boost selected box task (M = 0.23 vs. 0.15, SDs = 0.11 for the boost alternative and boost selected tasks respectively, $t(19) = 5.44$, $p < .001$, $d_z = 1.22$). Experiments 2 and 3 therefore assessed whether the reported effect would persist in situations in which error monitoring differences were less pronounced.

2.4 Experiments 2 and 3

Experiments 2 and 3 were designed to minimize differences in error monitoring across the tasks being selected, to allow us to investigate specifically the influence of confidence on task preferences. Thus, in Experiment 2, post-decisional evidence was manipulated only following correct responses, boosting evidence in favour of the correct response (i.e. in favour of participants' decisions) for one cue and boosting evidence in favour of the incorrect response (i.e. against participants' decisions) for the other. In Experiment 3, post-decisional evidence was manipulated regardless of participants' accuracy on the trial, so that post-decisional evidence was boosted in favour of the objectively correct response for one cue, and in favour of the incorrect response for the other. Since the majority of trials were correct, we expected that boosting the evidence in favour of the correct responses would still result in higher confidence overall, but also that detecting errors would be easier in this task, allowing us to test if the preference persists even under these conditions.

Methods

Participants

Twenty participants took part in Experiment 2 (12 females, ages 18-28 years, mean age = 22.8, SD = 2.77). 21 participants took part in Experiment 3, with one participant removed from the final sample due to non-compliance with instructions (always selecting the upper cue), resulting in a final sample of 20 participants (12 females, ages 18-33 years, mean age = 23.6, SD = 4.07). All participants had normal or corrected-to-normal vision. They provided written informed consent and were compensated with course credit or payment for their participation.

Stimuli and Procedure

The overall design of Experiments 2 and 3 was identical to Experiment 1, only differing in the nature of the post-decisional evidence manipulation. Thus, on all trials in both experiments, the mean strength of pre-decisional evidence was fixed at the same level as in Experiment 1 ($M_C = 212$, $M_I = 188$). In Experiment 2, evidence was manipulated only after correct responses. One cue was associated with a high-confidence task, in which post-decisional evidence was boosted after correct responses so that the difference between dot numbers grew ($M_C = 222$, $M_I = 178$)—the boost correct box task. The second cue was associated with a low-confidence task, in which post-decisional evidence was reduced following correct responses, so that the difference between the boxes decreased ($M_C = 202$, $M_I = 198$)—the boost incorrect box task. Following errors, the evidence remained the same as in the pre-decision stage in both tasks ($M_C = 212$, $M_I = 188$), so as to avoid making error detection easier in one task compared to the other. In Experiment 3, post-decisional evidence was manipulated similarly, but regardless of participants' accuracy on the trial. Specifically, one cue was associated with a task in which post-decisional evidence was always boosted in favour of the objectively correct response (the boost correct box task, $M_C = 222$, $M_I = 178$), and the second cue was associated with a task in which evidence strength was always reduced (boost incorrect box task, $M_C = 202$, $M_I = 198$), following both correct and incorrect trials (note that following correct trials, evidence manipulation in Experiments 2 and 3 was identical). With this manipulation, we expected that participants would experience higher confidence in their correct responses but lower confidence in their errors (i.e., stronger error detection) in the boost correct compared to the boost incorrect task. As in Experiment 1, for all conditions the correct square had on average more dots than the incorrect square in the post-decisional (as well as pre-decisional) period.

Results

Manipulation effects on error monitoring

Our first analyses assessed whether the manipulations in Experiments 2 and 3 were effective in minimizing differences in error monitoring between the tasks. To this end, the difference in the proportion of reported errors between the tasks was calculated for each participant. In Experiment 2, a numerically small but significant difference still remained (mean proportion of reported errors in boost correct task = .19, in boost incorrect task = .22, $t(19) = 2.30$, $p = .03$, $d_z = .52$).

A between participant t-test was performed to compare the size of these differences between Experiment 1 and 2 ($M = 0.073$, $SD = 0.06$, $M = 0.025$, $SD = 0.05$ for Experiments 1 and 2 respectively). As expected, the difference was smaller in Experiment 2 ($t(38) = 2.75$, $p < .01$, $d = 0.86$). This indicates that, though a small difference remained, the manipulation was successful in reducing the differences in error monitoring between the tasks. In Experiment 3, no significant difference was found between the proportion of errors reported in the high- and low-confidence tasks (M = .19 for both tasks, SD = .08 and .10 in the boost correct and boost incorrect tasks respectively; $t < 1$).

Manipulation effects on performance and confidence

As expected, no significant differences were found in performance between the tasks. Participants responded on time to a similar proportion of trials (Experiment 2: $M = 0.95$ vs 0.96, $SD = 0.05$ and 0.04, $t(19) = -1.68, p = 0.11, d_z = -0.38$; Experiment 3: $M = 0.96$ vs $0.97, SDs = 0.02, t(19) = -1.11, p = 0.28, d_z = -0.25$, for the boost correct and boost incorrect tasks respectively). For trials with on-time responses, no reliable differences were found in accuracy levels between the tasks (Experiment 2: M = 0.72 vs 0.73, SD = 0.07 and 0.08, $t < 1$; Experiment 3: M = 0.75 vs 0.72, SD = 0.08 and 0.07, $t(19) = 1.68, p = 0.11, d_z = 0.38$, for the boost correct and boost incorrect tasks respectively). Similarly, no reliable differences were found in mean reaction times (Experiment 2: M = 586 vs 585 ms, SD = 56 and 61 ms; Experiment 3: M = 600 vs. 604 ms, SD = 32 and 24 ms; $ts < 1$).

Figure 2.3A shows mean confidence ratings in Experiment 2, which varied across tasks as predicted. A 2 x 2 repeated measure ANOVA on these ratings revealed main effects for task ($F(1, 19) = 9.45, p = .006, \eta_p^2 = .33$) and for accuracy ($F(1, 19) = 154.34, p < .001, \eta_p^2 = .89$), and a reliable interaction between these factors ($F(1, 19) = 30.83, p < .001, \eta_p^2 = .62$). Pairwise comparisons revealed a significant difference in confidence between tasks for correct responses (M = 24.17 vs. 19.23, SD = 8.12 and 7.30, $t(19) = 5.63, p < .001, d_z = 1.26$ for the boost correct and boost incorrect tasks, respectively), but not for incorrect responses (M

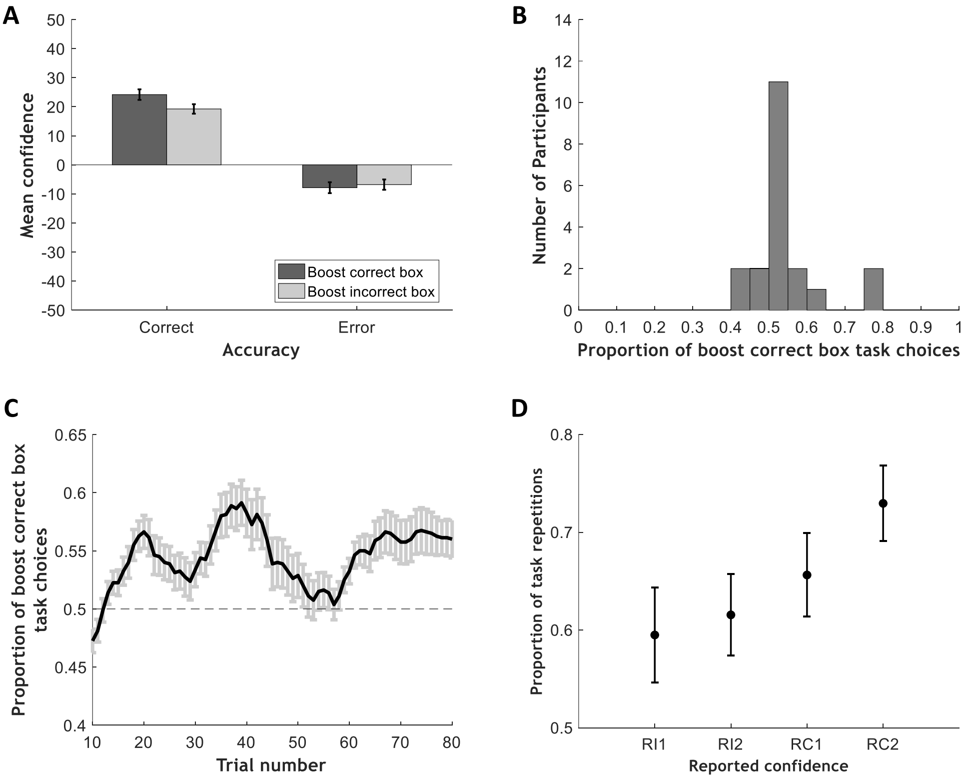

Figure 2.3: Results of Experiment 2. (A) Mean confidence ratings for the boost correct box and boost incorrect box tasks, divided according to trial accuracy. Evidence was manipulated only following correct responses. (B) Distribution of individual participants' proportion of boost correct box task choices (n=20). (C) Ten-trial running average showing the proportion of boost correct box task choices across blocks. Dashed line represents chance level. (D) Mean proportion of task repetitions according to reported confidence. Trials were first divided into groups according to reported accuracy (RI = reported incorrect, RC = reported correct), followed by a division according to median splits within the accuracy group (1 = below median confidence, 2 = above median confidence). Error bars represent SEMs.

= -7.87 vs. -6.82, SD = 8.41 and 7.94, $t(19) = -1.35, p = .19, d_z = -0.30$), as expected given that we manipulated evidence only following correct trials in this experiment. A corresponding analysis of confidence ratings in Experiment 3 (Figure 2.4A) revealed no significant main effect of task ($F(1, 19) = 1.13, p = .30, \eta_p^2 = .06$), a significant main effect of accuracy ($F(1, 19) = 148.70, p < .001, \eta_p^2 = .89$), and a reliable interaction between the two factors ($F(1, 19) = 93.92, p < .001, \eta_p^2 = .83$). Follow-up analyses revealed that, as expected, confidence was significantly higher in the boost correct task than the boost incorrect task when participants responded

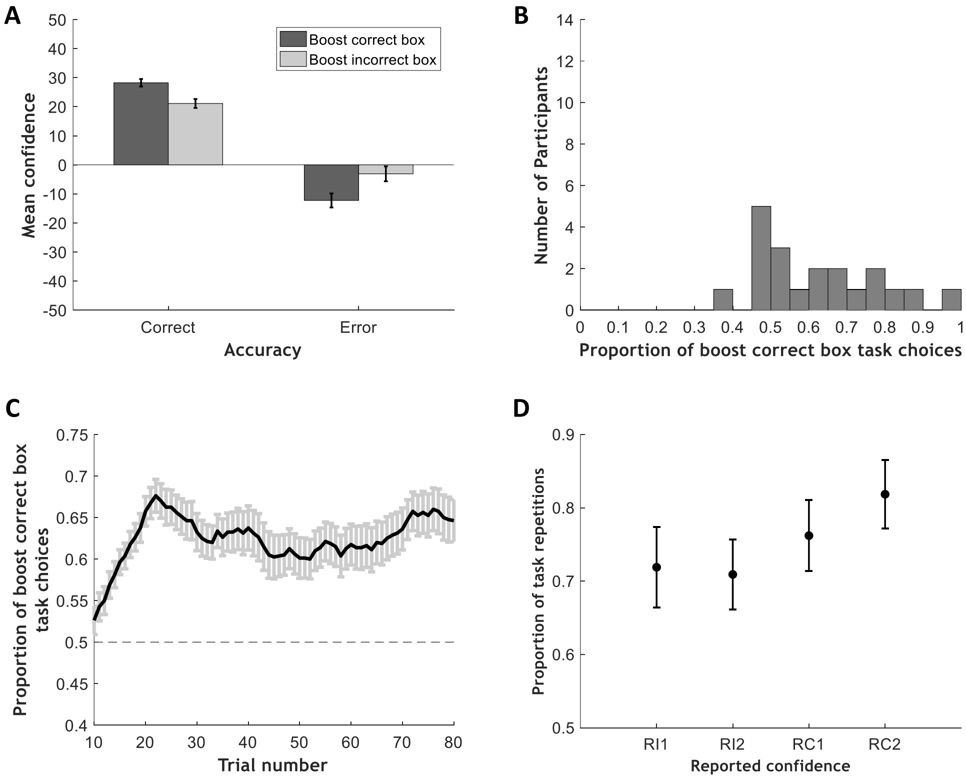

Figure 2.4: Results of Experiment 3. (A) Mean confidence ratings for the boost correct box and boost incorrect box tasks, divided according to trial accuracy. (B) Distribution of individual participants' proportion of boost correct box task selections (n=20). (C) Ten-trial running average showing the proportion of boost correct box task selections across blocks. Dashed line represents chance level. (D) Mean proportion of task repetitions according to reported confidence. Trials were first divided into groups according to reported accuracy (RI = reported incorrect, RC = reported correct), followed by a division according to median splits within the accuracy group (1 = below median confidence, 2 = above median confidence). Note: Error bars represent SEMs.

correctly (M = 28.24 vs 21.12, SD = 5.80 and 6.86, $t(19) = 6.15, p < .001, d_z = 1.38$), and this pattern reversed on incorrect trials (M = -12.26 vs -3.10, SD = 10.79 and 11.48, $t(19) = -6.63, p < .001, d_z = -1.48$). Because participants were correct on most trials (mean accuracy = 0.74, SD = 0.07), mean confidence across all trials was significantly higher in the boost correct task (M = 18.4 vs. 14.6 in the boost incorrect task, SD = 5.45 and 6.92, respectively; $t(19) = 2.85, p = .01, d_z = 0.64$). Note that one participant who showed a strong preference towards the boost correct box task only had a small number of trials to include in these analyses (8 corrects and 2 errors, as listed in Table 2.1). However, there are no meaningful changes to the results when excluding this participant.

Task selection

Results of the critical analysis of task selections are shown in Figures 2.3 and 2.4 (panels B and C), which exhibit a preference for the task associated with higher average confidence in both experiments. In Experiment 2, 16 of 20 participants (80%) chose the boost correct box task more often than the boost incorrect task (see Figure 2.2B, 2.2C). The mean proportion of boost correct choices was 0.54 (range = 0.41 to 0.78, SD = 0.09), a preference that was significantly different from chance level ($t(19) = 2.11, p < .05, d_z = 0.47$). In Experiment 3, 13 of 20 participants (65%) selected the boost correct task more than the boost incorrect one. The mean proportion of boost correct selections was 0.62 across participants (range = 0.35 to 0.97, SD = 0.16), again significantly different from chance level ($t(19) = 3.30, p < .01, dz = 0.74$; see Figures 2.4B, 2.4C). The preference for the task associated with higher average confidence was less consistent across participants in Experiment 3 than Experiment 2, but of larger numerical magnitude (and more consistent across trials of the block) in Experiment 3 (compare Figures 2.3C and 2.4C), with some participants showing an overwhelming preference for the boost correct task over the boost incorrect task.

Error monitoring and confidence impact on task selection

We performed an identical analysis to that described above for Experiment 1 to assess the trial-by-trial influence of subjective confidence on task choice: Trials were divided according to reported accuracy and then relative to each participant's median confidence rating within each of these trial subsets (RI1, RI2, RC1, RC2; guess trials were excluded; see Table 2.2 for trial counts), and then we calculated for each trial subset the likelihood that participants would repeat task choice (vs. switch) on the subsequent trial. We subjected these data to a 2 x 2 repeated measures ANOVA on repetition probability with reported accuracy (correct, incorrect) and confidence rating (below median, above median) as factors, separately for each experiment. In Experiment 2, significant effects were found for reported accuracy ($F(1, 19) =$

$21.27, p < .001, \eta_p^2 = .53$) and for confidence ($F(1,19) = 10.31, p = .005, \eta_p^2 = .35$), with participants repeating their task choice more often following correct trials, and when their confidence was high. The interaction between reported accuracy and confidence was not significant ($F(1,19) = 2.96, p = .10, , \eta_p^2 = .13$), but the numerical trend was for a larger impact of correct-trial confidence on task repetition. Because it was of interest whether effects of confidence are observed for correct and incorrect trials separately, we ran pairwise comparisons that revealed a significant difference in repetition likelihood following trials with high- vs. low-confidence on reported-correct trials ($t(19) = 5.38, p < .001, d_z = 1.20$), with participants repeating tasks more often following decisions made with high confidence. As in Experiment 1, no significant effect of confidence was found on reported incorrect trials (see Figure 2.3D, $t < 1$), and the small numerical trend was in the opposite direction to that observed in the previous experiment, so that participants repeated more after higher confidence decisions.

In Experiment 3, one participant whose data could not be binned according to median confidence was excluded from the analysis (for this participant median confidence rating for incorrect trials = -50, leading to no trials in the incorrect below-median bin). Task repetition likelihood varied reliably as a function of reported accuracy ($F(1,18) = 9.98, p = .005, \eta_p^2 = .36$), but not confidence ($F(1,18) = 2.09, p = .17, \eta_p^2 = .10$), or the interaction between accuracy and confidence ($F(1,18) = 3.75, p = .07, \eta_p^2 = .17$). Because we expected similar effects to those found in previous experiments, analyses were run to test for separate effects of confidence on correct and error trials. Indeed, here too we found a significant difference in repetition proportion following high- vs. low-confidence decisions on reported correct trials ($t(18) = 3.83, p = .001, d_z = .88$), with participants repeating their task choice more often after high confidence judgements. No such effect was found for reported incorrect trials ($t < 1$, Figure 2.4D) and, as in Experiment 1, the numerical trend was in the opposite direction.

Evidence strength and confidence impact on task selection

Similar to Experiment 1, our manipulation in the current experiments resulted in stronger evidence (greater average difference in number of dots between the two boxes) in tasks associated with higher average confidence (Experiment 2: M = 30.80 vs. 17.26; SD = 0.82 and 0.90, for the boost correct vs. boost incorrect tasks; $t(19) = 39.13, p < .001$; Experiment 3: M = 33.15 vs. 14.60; SD = 0.52 and 0.57, for the boost correct vs. boost incorrect tasks, $t(19) = 96.73, p < .001$). To test for the unique contribution of confidence while controlling for these differences in evidence strength, we ran logistic regression models using evidence strength and confidence on trial N as predictors of participants' task selection repetitions on trial N+1, separately for each participant and for reported correct and error trials, for the data from Experiments 2 and 3. In Experiment 2, for perceived correct trials, 19 of 20 participants had positive slopes for confidence, indicating they were more likely to repeat task choice following decisions made with higher confidence, with the mean beta being significantly different from zero across participants (M = .30, SD = 0.17, $t(19) = 7.66, p < .001, d_z = 1.71$). For evidence, slopes were positive in 13 of 20 participants, and numerically participants tended to be more likely to repeat their task preference following correct-response trials with stronger evidence, but the mean slope was not significantly different from zero across participants (M = .12, SD = .28, $t(19) = 1.92, p = .07, dz = 0.43$). For error trials, neither confidence (13 positive slopes, M = .07, SD = .34) nor evidence strength (14 positive slopes, M = .05, SD = .54) reliably predicted repeating task choice on the subsequent trials ($ts < 1$).

In Experiment 3, we once again found that higher correct-trial confidence was associated with increased likelihood of repeating task choice on the subsequent trial, with 18 of 20 participants having a positive beta value in the logistic regression, the mean value of which was reliably larger than zero ($M = 0.49, SD = 0.85, t(19) = 2.54, p = .02, d_z = 0.57$). As in Experiment 2, participants tended to be more likely numerically to repeat their task preference following correct-response trials with stronger evidence, but this effect was not reliable (positive betas in 14 of 20 participants, M = .19, SD = .49, $t(19) = 1.71, p = .10, d_z = 0.38$). For error trials, two participants whose data was not suitable for a logistic regression were removed from the analysis. Neither confidence (positive betas for 10 participants, M = 0.01, SD = 0.52, $t < 1$) nor evidence strength (positive betas for 10 participants, M = 0.23, SD = 0.59, $t(17) = 1.70, p = .11, d_z = 0.40$) reliably predicted repeating task choice on the subsequent trials. Thus, in line with the results from Experiment 1, Experiments 2 and 3 confirm that confidence influenced participants' task choices in a trial-to-trial manner, above and beyond the effects of trial-varying evidence strength.

Discussion

Our aim in Experiments 2 and 3 was to replicate the key findings from Experiment 1, but do so using conditions in which the effects of error monitoring should be minimized. We varied the post-decisional evidence manipulation so that errors would not be easier to detect in the task associated with lower compared to higher confidence: In Experiment 2, boosting evidence in favour of the correct box only after correct responses reduced the difference in error detection rates across conditions. In Experiment 3, boosting evidence in favour of the objectively-correct option regardless of response accuracy led to comparable rates of error detection across conditions. Even under these conditions, participants still exhibited a preference towards selecting the cue associated with higher average decision confidence, lending further support to our main hypothesis. Furthermore, our regression analyses replicated the finding from Experiment 1 that participants' trial-by-trial confidence ratings predict their subsequent task choices even when controlling for the strength of evidence presented.

An interesting question regarding the demonstrated preference is the degree to which it depends on the presence (or lack) of external feedback in Experiments 1-3. If people use confidence solely as an internal monitoring system, it might be predicted that the effects of confidence judgements would be redundant when external feedback is present. In this case, the effects previously demonstrated here might be expected to disappear. If, however, confidence acts as a cost or benefit factor intrinsic to a decision, we might still expect to see its effects even in the presence of feedback. We addressed this question in Experiment 4.

2.5 Experiment 4

To evaluate the influence of feedback on the preference for tasks associated with higher average confidence, visual feedback of participants' first-order performance ('correct', 'incorrect', 'too slow') was provided at the end of each trial in this experiment. Evidence in this experiment was manipulated in the same way as described above for Experiment 3, with a post-decision evidence boost in favour of either the objectively correct or incorrect response regardless of the accuracy of the participant's decision on the trial. We expected that participants would develop a preference for the high-confidence task (boost correct condition) even when feedback is present. This prediction was based on the results described above, indicating that people are not only biased against tasks in which they make errors, but also in favour of tasks in which they experience higher confidence in correct responses, an effect that might persist even if participants are given objective information about whether their response was correct or not.

Methods

Participants

Twenty participants took part in Experiment 4 (14 females, ages 18-26 years, mean age = 19.9, SD = 2.28). All participants had normal or corrected to normal vision. They provided written informed consent, and were compensated with course credit or payment for their participation.

Stimuli and Procedure

The design was similar to that of Experiment 3, except that feedback was given at the end of each trial. Thus, pre-decisional evidence was once again fixed across all trials ($M_C = 212, M_I = 188$), while post-decisional evidence was boosted in favour of the objectively correct response for one cue (boost correct box task,

$M_C = 222, M_I = 178$) and for the objectively incorrect response for the other (boost incorrect box task, $M_C = 202, M_I = 198$). Immediately after participants registered their confidence rating on each trial, or after the 800 ms time limit if no response was made, on-screen feedback indicated whether the participant's response was 'correct', 'incorrect', or 'too slow' for 400 ms, before the onset of the next trial.

Results

Manipulation effects on performance and confidence

Proportions of on-time responses for the two tasks were again well-matched (M = .96 SD = .02 vs. M = .96, SD = .03 for boost correct and boost incorrect tasks respectively, t < 1), as were accuracy levels and response times for trials with on-time responses (accuracy levels: M = 0.77, SD = 0.06, for both tasks, $t < 1$; response times: M = 593 vs. 588 ms, SD = 39 and 41, $t(19) = 1.53, p = 0.14, d_z = 0.34$). A 2 x 2 repeated measures ANOVA on confidence ratings (Figure 2.5A and Table 2.1) revealed main effects for task ($F(1, 19) = 4.67, p = .04, \eta_p^2 = .20$) and accuracy ($F(1, 19) = 179.65, p < .001, \eta_p^2 = .90$), and a reliable interaction between these factors ($F(1, 19) = 38.46, p < .001, \eta_p^2 = .67$). Pairwise comparisons revealed that, as expected, confidence was significantly higher in the boost correct compared to boost incorrect task for correct trials ($t(19) = 7.71, p < .001, d_z = 1.72$), but following incorrect responses this pattern was reversed ($t(19) = -4.75, p < .001, d_z = -1.06$). Participants were correct on most trials (mean accuracy = 0.77, SD = 0.06), leading to significantly higher confidence overall in the boost correct compared to boost incorrect task (M = 20.01 vs. 17.92, SD = 6.01 vs. 6.12; $t(19) = 3.12, p < .01, d_z = 0.67$).

Task selection

As shown in Figure 2.5, participants' preference for tasks in which they experience higher confidence persisted even when they received objective feedback on their performance on each trial. The preference was perhaps somewhat weaker than in the previous experiments, being evident in 12 out of the 20 participants (60%,

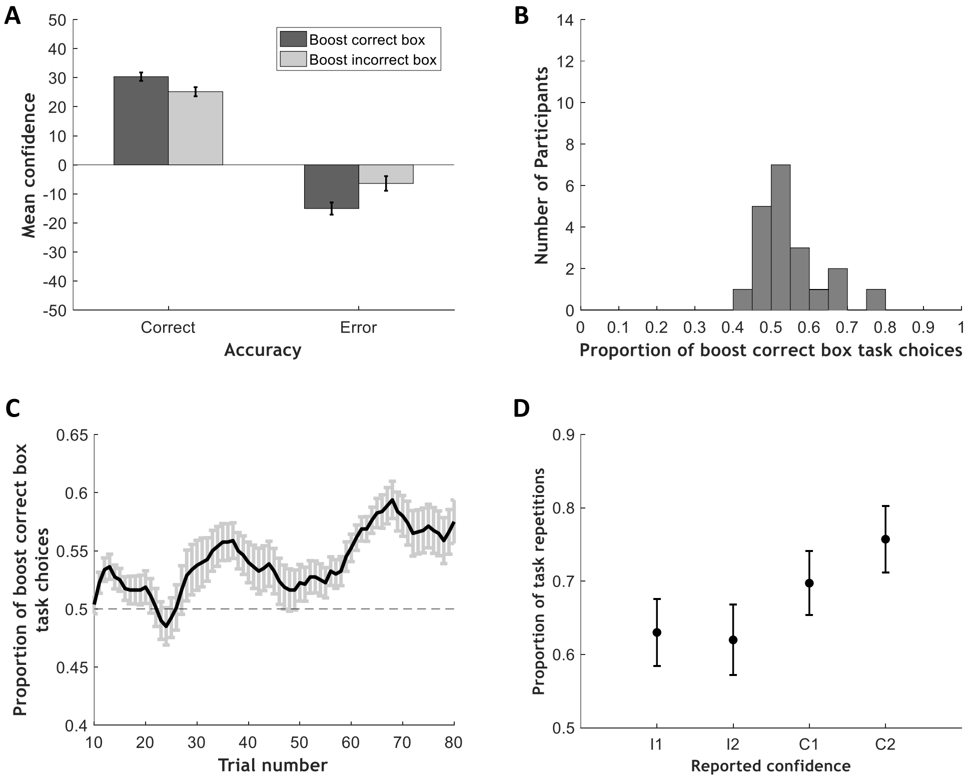

Figure 2.5: Results of Experiment 4. (A) Mean confidence ratings for the boost correct box and boost incorrect box tasks, divided according to trial accuracy. (B) Distribution of individual participants' proportion of boost correct box task selections (n=20). (C) Ten-trial running average showing the proportion of boost correct box task selections across blocks. Dashed line represents chance level. (D) Mean proportion of task repetitions according to reported confidence. Trials were first divided into groups according to objective accuracy (I = incorrect, C = correct), followed by a division according to median splits within the accuracy group (1 = below median confidence, 2 = above median confidence). Error bars represent SEMs.

Figure 2.5B) and being evident consistently only towards the end of each block (Figure 2.5C). On average, participants chose the boost correct task more often than boost incorrect task (mean proportion = 0.54, range = 0.41 – 0.77, SD = 0.08) at a rate that was significantly above chance level across participants ($t(19) = 2.18, p = .04, d_z = 0.49$).

Error monitoring and confidence impact on task selection

We performed a similar analysis to that described above for Experiments 1-3. However, since participants received feedback on their accuracy, trials were divided according to objective rather than reported accuracy. Importantly, this allowed us to assess the degree to which subjective confidence within a known accuracy level predicted task choice on the subsequent trial. The median confidence rating for correct and incorrect trials was computed, and trials divided relative to this rating (I1, I2, C1, C2, see Table 2.2 for trial counts). Two participants whose data could not be organized in this way were excluded from the analysis (median confidence rating for incorrect trials = -50). A 2 x 2 repeated measures ANOVA was then performed on repetition probability, with objective accuracy (correct, incorrect) and confidence rating (below median, above median) as factors (see Figure 2.5D). Significant effects were found for accuracy ($F(1,17) = 16.73, p < .001, \eta_p^2 = .50$), and for the interaction between accuracy and confidence ($F(1,17) = 6.24, p = .023, \eta_p^2 = .27$), but not for confidence ($F(1,17) = 4.34, p = .053, \eta_p^2 = .20$). Follow-up pairwise comparisons revealed a significant difference in repetition probability following high- vs. low-confidence decisions on correct trials ($t(17) = 4.56, p < .001, d_z = 1.11$), with participants repeating tasks more often after high confidence judgements. No significant effect was found for incorrect trials ($t < 1$), with repetition probability very similar regardless of participants' rated confidence on the previous trial.

Evidence strength and confidence impact on task selection

The evidence manipulation in Experiment 4 was identical to that used in Experiment 3, and resulted in similar evidence strengths, with stronger evidence in the task associated with higher confidence (M = 33.34 vs. 14.79; SD = 0.50 and 0.70 for the boost correct and boost incorrect tasks, respectively, $t(19) = 79.03, p < .001$). Logistic regression models predicting repetitions in task choice on trial N+1 based on the evidence strength and reported confidence on trial N were run for each participant, for objectively correct and error trials separately. 14 of 20 participants had positive slopes for confidence on correct trials, and the beta values were significantly different from zero, indicating that higher confidence was predictive of an increased likelihood of repeating task choice on the subsequent trial ($M = 0.24, SD = 0.30, t(19) = 3.58, p = .002, d_z = 0.80$), even when objective feedback was interposed. Only 10 participants had positive slopes for confidence following error trials, and this did not predict switching ($M = -0.01, SD = 0.37, t < 1$). No significant effects were found for evidence strength predicting subsequent task choice (correct trials: 14 participants had positive slopes, $t(19) = 1.55, p = .14, d_z = 0.35$; incorrect trials: 14 participants had positive slopes, $t(19) = 1.16, p = .26, d_z = 0.26$).

Discussion

This experiment evaluated whether a bias towards tasks associated with higher confidence is restricted to situations in which objective feedback is absent (and, hence, in which participants are inherently reliant on their own subjective evaluations), or whether this bias might persist even when participants are told on each trial whether their response is correct or incorrect. The results demonstrate that, even when feedback is present and accuracy levels are very closely matched, participants were biased toward selecting the task associated with higher average confidence more often than a task associated with lower confidence levels. Our results further show that even when participants know they are correct, they are more likely to choose the same task on the next trial following decisions in which they felt more confident. This result provides strong support for the role of confidence as an internal cost or benefit factor in decisions, acting as a fine-grained cue that guides task choices beyond binary error monitoring.

Experiments 1-4 all provide evidence for a consistent effect of confidence on task repetitions, above and beyond evidence. However, in all four experiments the cue associated with higher confidence was also associated with stronger evidence. In the final experiment described below, we pitted these factors against one another to examine the influences of high-confidence and strong-evidence when associated with different tasks.

2.6 Experiment 5

To study the independent effects of high-confidence and strong-evidence on task choices, in Experiment 5 we manipulated post-decisional evidence only following incorrect responses (a complementary paradigm to the one used in Experiment 2, in which evidence was manipulated only following correct responses). Here, following incorrect responses, one cue was associated with a boost in evidence strength (i.e. larger dot difference in favour of the correct answer), which should make it easier to detect errors and thus reduce overall confidence. The other cue was associated with a decrease in evidence strength, which should make it harder to detect errors and thus increase overall confidence. Therefore, evidence strength and confidence work in opposite directions in this experiment, with confidence manipulated only for incorrect responses, thus exploring the extremes of possible influences of confidence on task choice. The experiment followed the design of Experiments 1-3, in that participants did not receive feedback on each trial.

Methods

Participants

Thirty-eight participants took part in Experiment 5 (25 females, ages 18-35 years, mean age = 24.29, SD = 4.51). A larger sample was used because we expected the effect size of the task preference to be smaller in this study, given that we manipulated evidence strength on only a small subset of trials (those with incorrect responses) and pitted evidence strength and confidence effects against one another. Specifically, the chosen sample size provides power of 0.8 to detect an effect size of $d_z = 0.47$, which is the effect size observed in Experiment 2—the experiment with the weakest effect out of the four reported above, and the one with a complementary manipulation of evidence strength varying only following a subset of trials. All participants had normal or corrected-to-normal vision, provided written informed consent, and were compensated with course credit or payment for their participation.

Stimuli and procedure

The experimental design was similar to that used in Experiments 1-3, differing in only two respects. First, post-decisional evidence was only manipulated following incorrect responses in the initial dot-judgement task. For one cue, evidence strength was reduced after incorrect answers, (boost incorrect box task, $M_C = 202, M_I = 198$), which should make it harder to detect errors and thus increase overall confidence. For the other cue, evidence was increased after incorrect trials, which should lead to lower confidence (boost correct box task, $M_C = 222, M_I = 178$). Following correct responses, evidence strength remained as in the pre-decision stage. The second procedural difference in this experiment was that we used a reduced evidence strength in the pre-decision stage compared to the previous experiments, with dot number distributions centered around $M_C = 208, M_I = 192$ (which was again fixed across all trials regardless of the cue selected). We made this change because we were particularly interested in error trials, and therefore used weaker evidence to decrease initial decision accuracy.

Results

Manipulation effects on performance and confidence

As expected, accuracy was somewhat lower than in previous experiments but was closely matched between the two tasks (Ms = .65 for both tasks, SD = .08 and .07 for the boost incorrect vs. boost correct tasks, $t < 1$). Participants made on-time responses in a similar proportion of trials (M = .95 vs .94, SD = .05 and .07, $t(37) = 1.54$, $p = 0.13$, $d_z = 0.25$), with response times on these trials similar for both tasks (M = 603 vs. 604 ms, SD = 56 and 61 ms, $t < 1$). A 2 x 2 repeated measures ANOVA on confidence ratings (Figure 2.6A, Table 2.1) revealed reliable main effects of task ($F(1, 37) = 35.61, p < .001, \eta_p^2 = .49$) and accuracy ($F(1, 37) = 152.65, p < .001, \eta_p^2 = .81$), and a significant interaction between these factors ($F(1, 37) = 50.23, p < .001, \eta_p^2 = .58$). Follow-up pairwise contrasts revealed that confidence was significantly higher in the boost incorrect than in the boost correct box task for incorrect trials ($t(37) = 6.89, p < .001, d_z = 1.12$). As expected, no reliable confidence difference was found between tasks on correct trials, for which post-decisional evidence was not manipulated ($t(37) = 1.55, p = 0.13, d_z = 0.25$). When all trials were pooled together, participants expressed significantly higher confidence overall in the boost incorrect task (weak evidence) compared to boost

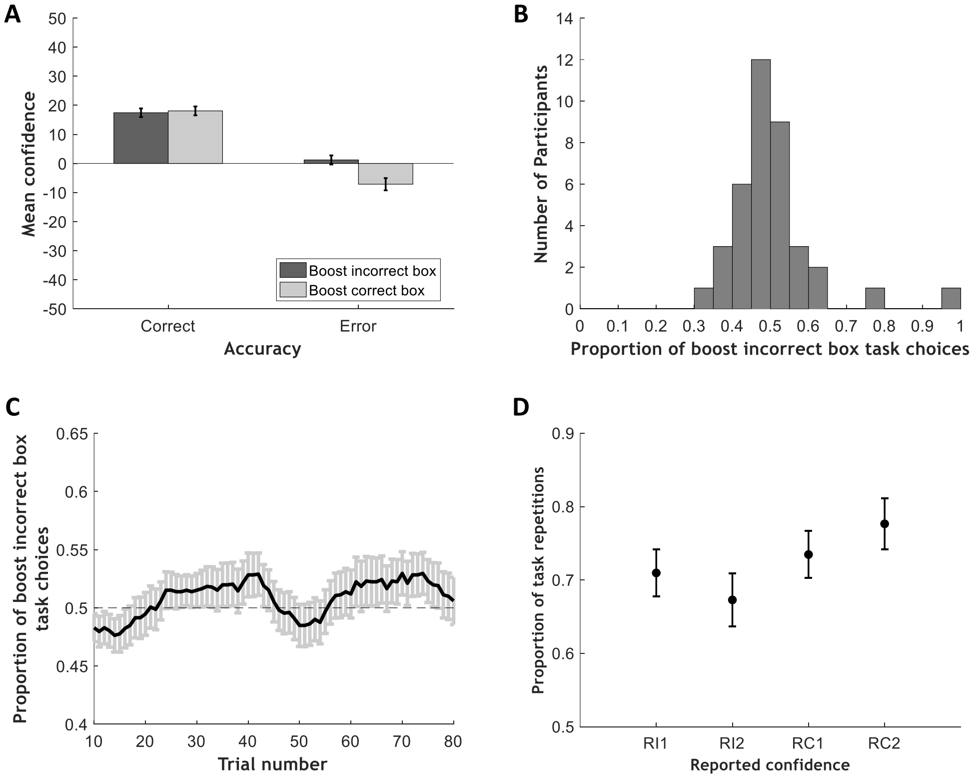

Figure 2.6: Results of Experiment 5. (A) Mean confidence ratings for the boost incorrect box and boost correct box tasks, divided according to trial accuracy. (B) Distribution of individual participants' proportion of boost incorrect box task selections (n=38). (C) Ten-trial running average showing the proportion of boost incorrect box task selections across blocks. Dashed line represents chance level. (D) Mean proportion of task repetitions according to reported confidence. Trials were first divided into groups according to reported accuracy (RI = reported incorrect, RC = reported correct), followed by a division according to median splits within the accuracy group (1 = below median confidence, 2=above median confidence). Error bars represent SEMs.

correct task (strong evidence; M = 11.76 vs. 9.41, SD = 8.14 vs. 8.95 respectively; $t(37) = 3.64, p < .001, d_z = 0.59$).

Task selection

With evidence strength and error-trial confidence acting in opposition across tasks, in this experiment we observed no consistent task preferences: 16 out of 38 participants selected the boost incorrect task, associated with higher confidence, more often (42%, 2.6B), with the overall average selection of this task not significantly different from chance level (M = 0.51, SD = 0.11, $t < 1$, 2.6C). Thus, there was no evidence for the development of a consistent bias across participants in the current experiment.

Error monitoring and confidence impact on task selection

Although participants exhibited no consistent task preference, it remains possible that trial-varying confidence nevertheless guided participants' task choices. To assess this possibility, we again divided each participant's trials according to reported accuracy and relative to the median confidence ratings within each accuracy subset (RI1, RI2, RC1, RC2, guess trials excluded; see Table 2.2 for trial counts), then calculated the proportion of trials that participants repeated task choice after each trial type. A 2 x 2 repeated measures ANOVA on repetition probability with reported accuracy (correct, incorrect) and confidence ratings (below median, above median) as factors revealed no reliable effect of confidence ($F < 1$), but a significant main effect of reported accuracy ($F(1, 37) = 16.38, p < .001, \eta_p^2 = .31$) and a reliable interaction the two factors ($F(1, 37) = 9.71, p = .004, \eta_p^2 = .21$, Figure 2.6D). Follow-up pairwise comparisons revealed a significant difference in repetition probability following high- vs. low-confidence decisions on reported correct trials ($t(37) = 3.78, p < .001, d_z = 0.84$), with participants more likely to repeat after judging a response as correct with high confidence. For reported incorrect trials, no significant effect was found ($t(37) = -1.64, p = 0.11, d_z = -0.37$) but the numerical trend pointed in the opposite direction, with participants more likely to repeat their task choice following trials in which confidence was lower.

Evidence strength and confidence impact on task selection

The average evidence strength for the boost incorrect box task was $M = 14.04$ ($SD = 0.7$), and for the boost correct task $M = 20.51$ ($SD = 1.05$), with a reliable difference found between tasks ($t(37) = 25.72$, $p < .001$). Logistic regression models predicting repetitions in task choice on trial N+1 based on the evidence strength and reported confidence on trial N were run for each participant, for reported correct and error trials separately. On reported correct trials, both confidence and evidence strength predicted choices. 32 of 38 participants had positive slopes for confidence on correct trials, and the beta values were significantly different from zero ($M = 0.31$, $SD = 0.36$, $t(37) = 5.39, p < .001, d_z = 0.87$), indicating once again that higher confidence predicted an increased likelihood of repeating task choice on the subsequent trial. 24 participants had positive slopes for evidence on correct trials, and these values were significantly different from zero ($M = 0.08$, $SD = 0.21$, $t(37) = 2.23, p = .03, d_z = 0.36$), such that participants tended to repeat their task choice following trials in which they were presented with stronger perceptual evidence (i.e., easier decisions). No significant effects were found for the regression analysis of error trials, although paralleling the numerical trend apparent in the ANOVA reported above, 23 participants had negative slopes for confidence on error trials, repeating choices more for lower confidence (M = -0.15, SD = 0.53, $t(37) = -1.72$, $p = .09$, $d_z = -0.28$; $t < 1$ for evidence strength betas).

Discussion

Experiment 5 investigated whether participants' preferences would be dominated by the strength of perceptual evidence or their degree of confidence in a task when these factors were set in opposition. If the task biases demonstrated in the previous experiments were driven mainly by a preference for strong evidence rather than high confidence, we would expect to see a bias towards the strong evidence (boost correct box) task emerge under the current conditions as well. As we did not find any such preference consistently across participants, this does not seem to be the case. Though the manipulation we used in the current experiment did not lead to the development of a consistent preference for the high confidence (boost incorrect box) task either, once again we found support for the role of confidence in task selection: As in our previous experiments, participants were more likely to repeat a task choice on the next trial if they made a decision on the current trial with higher confidence, an effect that was once again restricted to trials with correct responses. For incorrect trials we did not find that confidence or evidence strength predicted repeating task choice. These results are in agreement with participants' overall lack of a preference for either task, since in the current experiment we manipulated evidence (and thus confidence) only following incorrect trials.

2.7 Questionnaire measures in Experiments 1-5

Need for Cognition

An interesting question regarding the demonstrated biases is whether differences in their strength between individuals can be explained by trait differences in personality or cognitive style. To evaluate if preferences in the tasks were related to individuals' tendency to engage and enjoy cognitively effortful tasks, participants completed the Need for Cognition scale (Cacioppo et al., 1984). We hypothesized that participants who chose the high-confidence task more often might score lower on the need for cognition scale. To guarantee a sufficient sample size, we combined the data from Experiments 1-4 (Experiment 5 was excluded due to the different nature of the manipulation), and performed a regression analysis with NFC scores and Experiment as predictors of high-confidence task choices. However, the results did not provide support for this suggestion ($F < 1$).

Behavioral Strategies and Metacognitive Awareness

To gain further insights into participants' subjective experience of the task and explicit use of choice strategies, we administered a debriefing questionnaire adapted from Kool et al., 2010 at the end of each experiment (see Table 2.3). We did not have specific predictions about participants' responses or how these responses would relate to behaviour on the task, but report some exploratory analyses and examples we found to be of interest. Response samples are taken from participants in Experiment 1, which are representative of the responses we received across all experiments.

Questions 1 and 3 were intended to probe conscious behavioural strategies. Although most participants reported selecting randomly, or based on cues' physical attributes such as color or location at first, many participants developed a conscious choice strategy or preference that did not rely only on cues' attributes (60%, 90%, 75%, 75%, 58% of participants based on responses to questions 1 and 3 for Experiments 1-5 respectively). Reported strategies included error monitoring, perceived differences in difficulty, differences in confidence or a combination of these

1	How did you choose between the circles?
2	Did you develop a preference for specific circles? Y/N
3	If you replied yes, why? What was the difference between the circles?
4	Did you find the task associated with some circles to be easier? Y/N
5	What was the difference in the dot stimuli between tasks associated to the circles?
6	For some participants the amounts of dots changed after replying, so that one of every pair of circles was associated with a task in which dots were added to the correct square and removed from the other square (making the difference more noticeable). The other circle was associated with an opposite task in which dots were removed from the correct square and added to the other. Did it seem like this was the case for you? Y/N
7	If you answered yes to the previous question, was this something you became EXPLICITLY aware of DURING THE EXPERIMENT, or something that you realized only in retrospect?

Table 2.3: Debriefing questionnaire. This questionnaire, adapted from Kool et al. (2010) was administered at the end of each experiment to assess participants' awareness of the experimental manipulation. Note that question 6 varied slightly between experiments to represent the specific manipulation used.

(e.g. 'I started with a random colour and kept using that colour until I started making mistakes, became less sure about responses or missed a trial'). When asked explicitly if they developed a task preference, most participants responded positively (65%, 80%, 75%, 80%, 68% of participants replied yes to question 2 for Experiments 1-5 respectively). Similarly, many participants reported finding one task easier (question 4: 60%, 75%, 85%, 55%, 68% of participants respectively). To test if participants who reported experiencing differences in task difficulty were more likely to develop a preference for the high-confidence tasks, participants were grouped according to responses to question 4 (yes vs. no) and their task selections (high-confidence task selections > 50% vs high-confidence task selections < 50%). Combining the data from Experiments 1-4, 75% of participants who reported finding one task easier also demonstrated a preference for high-confidence tasks, vs. 64% of participants who did not report finding one task easier demonstrating a high-confidence preference. However, the relationship between perceived differences in difficulty and task choice was not significant ($\chi^2 < 1$), suggesting that the demonstrated preference for high-confidence associated tasks was not driven by perceived differences in difficulty (though these results might also reflect a lack of sensitivity in our questionnaire and analysis).

When asked to explain differences between the tasks in terms of differences in the dot stimuli (question 5) responses varied, with participants referring to differences in dot numbers, stimulus refresh rate, stimulus display time, size and layout of dots, or simply stating they did not know (‘I can’t really tell which is why I think the difference was mainly in my head’). Only one participant in all the experiments combined explicitly mentioned post-decisional evidence manipulation, and even when the manipulation itself was explained (question 6), most participants did not identify the manipulation as being used in their case (5%, 45%, 35%, 15%, and 26% of participants answered ‘yes’ to question 6 in Experiments 1-5, respectively). To ensure the preference effects found were not driven by participants who were explicitly aware of the evidence manipulation during the experiment, participants who responded yes to question 7 were removed from the analysis (1, 5, 4, 2, 3 participants - 5%, 25%, 20%, 10%, 8% - for Experiments 1-5 respectively) and remaining participants’ high-confidence tasks choice rates were compared to chance level. Results in Experiments 1, 3 and 4 remained significant ($t > 2, ps < .05$). In Experiment 2 results were not significant, most likely due to the large number of participants removed, but the trend remained ($t(14) = 2.02, p = .06$). In Experiment 5 the results remained non-significant ($t < 1$).

In summary, the questionnaire results suggest that most participants used selection strategies that were based on metacognitive cues such as error monitoring, difficulty or confidence. Furthermore, they provide evidence that the preference we found does not depend on awareness to the evidence manipulation. However, it is important to note that the questionnaire lacks the sensitivity to provide insight into the relationship between general metacognitive awareness to differences between the tasks and the development of a preference, due to several limitations. First, to avoid biasing the way participants perform the task, the questionnaire was administered at the end of the experiment, rather than after every block. Thus, we could not assess the relationship between developing a preference and accurately identifying tasks differences (e.g. selecting the high-confidence cue more often on a specific block and reporting that it was associated with a preferable task because it was easier, led to higher confidence or induced less errors). Second, the wording of some questions might have led participants to give specific responses. For example, asking if one task was easier might have encouraged participants to report experiencing differences in task difficulties regardless of their experience during the experiment. They then might try and think what could lead to such differences in difficulty in response to question 5, rather than reporting differences they noticed during the experiment. Thus, although the questionnaire responses provide useful insights regarding participants' strategies and metacognitive processes during the task, we are cautious in drawing any strong conclusions from these observations.

2.8 General Discussion

The present chapter examined the role of decision confidence in task selection. A series of five experiments tested the hypothesis that confidence acts as an intrinsic cost-benefit factor, biasing people towards selection of tasks in which they experience higher confidence. To this end, we employed a task-selection paradigm in which the two competing tasks were identical except for the strength of post-decisional evidence, enabling us to study the effects of subjective confidence independent of objective performance. We found that participants consistently exhibit a bias towards selecting tasks in which they experience higher confidence. Moreover, we found that participants do not only utilize confidence as an internal error-detection system, avoiding tasks in which they believe or know they have made errors, but rather use it in a graded fashion so that they tend to select tasks in which they experience higher confidence in their correct responses. This preference was apparent when controlling for objective accuracy and the strength of evidence presented, and even when external feedback was provided to render internal error-monitoring of decisions unnecessary, indicating the consistent impact of confidence on task choice.

The notion that people tend to avoid effort has been prevalent in behavioral research for many years. Historically, the 'law of least effort' (Hull, 1943) has been studied mainly in relation to physical demand, but recent advances have also provided empirical evidence for a bias against cognitively demanding tasks (Kool et al., 2010; Westbrook et al., 2013). Importantly, what exactly constitutes cognitive effort and what factors lead to its avoidance (for example, the degree to which behaviour is guided by the subjective experience of effort, as opposed to objective performance), remain unclear (Shenhav et al., 2017). Our results contribute to a growing body of evidence providing support for the involvement of metacognitive awareness in task preferences. Moreover, since there has been substantial progress in specifying confidence in computational terms (Pleskac & Busemeyer, 2010; Pouget et al., 2016; Sanders et al., 2016), linking confidence to choice behaviour is particularly meaningful in formalizing a computational framework of task preferences.

Although several other studies have examined the role of metacognitive evaluations in task preferences, none have specifically addressed this question with relation to confidence. Previous investigations have studied the influence of perceived task difficulty (Desender et al., 2017) and perceived effort (Dunn & Risko, 2016). Our findings are in line with evidence from these studies, indicating that metacognitive cues play an important role in task selection. This congruency of effects raises the question of whether the task preference displayed by participants in our study was a result of their confidence or the result of other metacognitive cues, such as perceived task difficulty or effort.

Although it is likely that participants in general base their task selections on a broad metacognitive evaluation that incorporates multiple cues (Dunn & Risko, 2016), our manipulations and data favour an interpretation of our specific results as being driven primarily by decision confidence. First, except for post-decisional evidence, the tasks were similar in all aspects including participants' accuracy and response times. It is not clear why it should be more difficult or effortful to process post-decisional evidence across tasks, given that this evidence was not required for task performance. In contrast, the normative and computational basis for the influence of post-decisional evidence on confidence is well documented (Moran et al., 2015; Pleskac & Busemeyer, 2010). Second, even allowing for the possibility that post-decisional evidence strength could somehow influence perceived effort or difficulty of an earlier perceptual decision, our regression analyses demonstrate that participants' confidence ratings account for unique variance above and beyond the evidence presented both pre- and post-decision. Indeed, in several cases, confidence was the only reliable predictor of participants' task selections from trial to trial. Thus, confidence is the most likely source of the emerging task preferences in our study. An interesting possibility to consider however, is that confidence acted on task choices by influencing perceived task difficulty. A recent study has demonstrated that response time, response repetition and response conflict all act as cues to subjective task difficulty (Desender et al., 2017). Since confidence usually correlates with response time and accuracy on tasks, it seems plausible that participants also incorporate confidence into their evaluation of task difficulty. In this case, confidence might have influenced task choices with perceived task difficulty acting as a mediator (although we did not find evidence for this suggestion in our debriefing questionnaire, but this might be due to a lack of sensitivity).

Many participants in our study reported experiencing differences in difficulty between the tasks, or developing strategies based on metacognitive evaluations such as error monitoring or decision confidence, but the majority of participants did not report awareness of the task manipulation. Only one participant identified post-decisional evidence differences between the tasks independently. Indeed, even when prompted with a description of the manipulation, most participants did not identify it as representing the experiment they had completed. Moreover, in Experiments 1, 3 and 4, we saw evidence for confidence preferences even when excluding participants who, after reading this description, reported being explicitly aware of the manipulation during the task. Thus, our results are not in line with recent studies indicating that cognitive effort affects task preferences only when people are aware of differences in cognitive demand (Desender et al., 2017; Gold et al., 2015), even if they are not aware of the source of these differences (Desender et al., 2017).

Across experiments, the preference for high confidence tasks tended to emerge relatively early in each block. However, participants did not display a persistent bias, but rather switched between the tasks throughout the blocks of the experiment, with confidence judgements predicting switches of task selections on a trial-by-trial basis. Thus, when considering the dynamics of the preference we found, it is likely that, rather than only developing a fixed and stable preference for one task driven by a general metacognitive awareness of task demands, participants also used confidence as an online monitoring system allowing them to adjust their behaviour dynamically in a strategic fashion.

In all five experiments, we found that participants were more likely to repeat their task choice after trials with decisions they judged to be correct (vs. incorrect), but that gradations in decision confidence modulated task selection even beyond this binary classification that has been well-explored in the context of reinforcement learning: Thus, our participants were more likely to repeat task choices after they made correct responses with high vs. low confidence. Notably, however, we did not find a corresponding influence of confidence on task selection following trials reported as incorrect: participants were not more likely to repeat their task choice when they were more confident. Indeed, the numerical trend was in the opposite direction in Experiments 1, 4 and 5 (albeit only reaching statistical significance in the regression performed for the first experiment), with participants more likely to repeat when they were more certain they had made a mistake. Caution is appropriate in interpreting these weak and inconsistent results, but we suspect they reflect an interaction between two competing influences on task choice following errors—a bias away from tasks associated with errors (as detected internally or signalled via feedback), and a bias away from tasks associated with uncertainty about performance (or, relatedly, weak evidence).

This reasoning is relevant to the results of Experiment 5, in which the post-decisional evidence manipulation created exactly this difference across tasks—one in which more errors were detected vs. one in which weaker evidence was presented and more uncertainty experienced—and in which no consistent task preferences were apparent. There were, nevertheless, idiosyncratic preferences apparent in individual participants in Experiment 5 (Figure 2.6B), that were stronger than can be explained by chance variation: Task selections below .45 or greater than .55 have a binomial probability $<.05$ across our 320-trial experiment, by which criterion 7 participants showed a clear preference for the task associated with higher confidence (boost incorrect task, weak evidence with fewer errors detected) and 10 preferred the task associated with stronger evidence (lower confidence, more errors detected). Thus, while it remains possible that an overall task preference was not detected due to an underpowered design (evidence was manipulated only after error trials, resulting in a smaller subset of manipulated trials compared to the previous experiments), the data suggest that this null finding is the result of individual differences in task preferences. Exploring individual differences in aversion to errors vs. uncertainty might be a useful avenue for future research, offering potential insights into differences in learning strategies, such as individual differences in information seeking and exploration vs. exploitation behaviours.

The importance of being able to monitor performance in an on-line way is intuitive, especially when external feedback is lacking. Indeed, in the influential framework proposed by Nelson and Narens (1990), the function of metacognition is viewed in terms of control of behavior, whereby we monitor in order to control. With this in mind, perhaps the most intriguing result of our study is the emergence of a preference for high-confidence tasks even when external feedback is present and objective accuracy is well-matched (Experiment 4). In such conditions, the usefulness of internal monitoring resulting in a high confidence bias seems less evident. We suggest that, given the strong association between subjective confidence and objective performance, this preference is adaptive as fine-grained confidence is sometimes likely to act as a better predictor of future performance than trial accuracy (for example, consider a two-alternative task in which you are guessing, and know that you are—in this case, knowing that your guess happened to be correct vs. incorrect may not usefully inform future behaviour). Indeed, as discussed above, we find that people do use confidence in a more sensitive way than simple error detection. This is further supported by results from Experiments 2 and 3, where a preference for a task associated with higher confidence emerged even when minimizing error monitoring differences. We conclude that confidence is used as a metacognitive cue that guides behaviour even when feedback is present, and in a more sensitive manner than binary error monitoring. This suggestion is congruent with a recent study demonstrating that confidence is automatically integrated into brain activity in ventromedial prefrontal cortex, an area associated with encoding subjective value (Lebreton et al., 2015). Taken together with recent evidence that confidence can act as a reinforcement signal for perceptual learning when external feedback is absent (Guggenmos et al., 2016), our results suggest that confidence might play an important role in learning even when external feedback is present.

Collectively, our findings add to the growing body of evidence on the role of confidence in adaptive decision making. Previous studies have demonstrated that confidence plays a role in various contexts: linking performance across sequential decisions (Van Den Berg et al., 2016), guiding adaptive information seeking (Desender, Boldt, Verguts, et al., 2019; Desender et al., 2018), preparing for upcoming choices according to anticipated difficulty (Boldt, Schiffer, et al., 2019), balancing the ecological imperatives of exploration vs. exploitation in value-based choice (Boldt, Blundell, et al., 2019) and arbitrating among competing opinions across individuals (Bahrami et al., 2010; Bang et al., 2014). Our findings are particularly congruent with recent evidence indicating that confidence is represented on a common scale across tasks of different modalities, and can be compared between tasks as efficiently and precisely as within the same task (de Gardelle et al., 2016; de Gardelle & Mamassian, 2014). De Gardelle and colleagues suggest that confidence is used as a common currency for judging our decisions, allowing different tasks to be compared and prioritized. Their results complement our findings that confidence biases decision making when all else is equal. By combining the conclusions from these studies, we can speculate that confidence is used as a general cue that guides behaviour and task preferences, even when comparing between different tasks. Future studies might extend this investigation to consider the effects of confidence preferences across different tasks and modalities. More broadly, when taken with evidence that confidence can guide behaviour flexibly across multiple contexts, the present findings support a conception whereby metacognitive processes support a variety of strategic behaviors of different kinds and spanning a range of timescales, based on a common computation (specifically here the subjective evaluation of decision accuracy).

In conclusion, the present results add to our understanding of the functional role of confidence in decision making. We find that people show a preference for situations in which they are more confident, even when controlling for other plausible influences on task selection. Since confidence and performance are strongly related, this preference can be useful for adaptive decision making in our daily life, guiding us towards tasks in which we are ultimately more likely to succeed. These results fit well with the concept of metacognition as a system in which monitoring is performed in order to enable subsequent control and adjustment of behaviour.

3

Confidence is used strategically

3.1 Introduction

The results of Experiments 1-5 in Chapter 2 demonstrated that people use their confidence to choose which tasks to engage with. Importantly, these findings, and those of previous studies examining the role of confidence in guiding behaviour, reveal a consistent mapping between confidence in a decision and subsequent behaviour when performing a specific task, so that low confidence leads to one course of action and high confidence to another. For example, when performing an information sampling task, low confidence leads to gathering additional information, whereas high confidence leads to committing to a choice (Desender et al., 2018; Desender, Murphy, et al., 2019). However, it remains unclear whether this mapping is necessarily fixed, or could exhibit some degree of flexibility and context-sensitivity. The current chapter explored the hypothesis that the relationship between confidence and subsequent behaviour is not fixed. Rather, this relationship may vary depending on the specific way in which confidence is being used, such that people use confidence in a strategic fashion according to the immediate context and goal at hand. This framing is consistent with theories that view explicit, conscious representations—such as representations of decision confidence—as serving crucially to guide behaviour in a flexible manner to support evaluation, planning, and voluntary action (Dehaene & Changeux, 2011).

To investigate this proposal, we studied the relationship between confidence and advice seeking. Previous social decision-making studies have shown that people utilise advice more when a task is difficult, and they are uncertain about their ability to make an accurate decision on their own (Gino & Moore, 2007). In these situations, confidence is used as a self-monitoring tool, signalling to a decision maker that they are unlikely to succeed and should seek the help of others. We propose that this seemingly natural relationship between low confidence and advice requests is not a fixed but rather context-dependent. Specifically, we predicted that when people want to learn about advice quality, they will request more advice when their confidence is high, rather than low.

The importance of taking advisor quality into account when deciding whether to request or follow advice is intuitive. Accordingly, studies have shown that people are sensitive to the quality of advice based on the reputation of the advisor, relative to their own level of knowledge (Yaniv, 2004; Yaniv & Kleinberger, 2000). However, we often do not have any previous knowledge about the reliability of information provided by other people. In such situations, when feedback is available, people can integrate the reliability of advisors' responses over time, updating their trust in advisors through associative learning processes (Behrens et al., 2008). Importantly, it has recently been demonstrated that people can learn about advisor reliability even when feedback is not available, by comparing advice to their own beliefs when they are certain they know the correct answer (Pescetelli & Yeung, 2020): If we are certain that a given decision is correct, we can equally be certain that anyone who disagrees with us is wrong, and accordingly reduce our trust and reliance on their advice now and in the future. Conversely, if we reach a decision but with low confidence, we should still reduce our trust in advisors who disagree, but to a lesser extent. According to this agreement-in-confidence heuristic, if objective feedback is unavailable, our best opportunity to learn about the reliability of an advisor comes when we are ourselves sure what we think about a particular decision. Based on this reasoning, we predicted that when people want to learn about advisor reliability in the absence of external feedback, they will strategically sample advice when confidence is high, breaking away from the typical relationship between low confidence and advice requests.

To illustrate this prediction, consider the situation of being unsure which of two movies to watch, and therefore seeking advice from online reviews—a straightforward example of subjective confidence guiding information seeking. Suppose you find reviews for both movies written by the same film critic, who had a strong preference for one of the two movies. However, you have never heard of this critic before, and cannot be sure you should trust their opinion. Suppose, though, that on the website you notice that they have also reviewed a movie you saw last week and loved. Now you might be motivated to read the review, even though you are confident in your opinion about this movie, so as to learn about the reviewer: If you find that the critic loved the movie too, you might infer that the critic can be trusted, and therefore follow their recommendation for tonight's film choice.

The example above demonstrates the use of confidence both as a self-monitoring tool (indicating when you need advice), and as a proxy for feedback when objective feedback is lacking (as a yardstick against which to evaluate the reliability of advice). In the present study, we tested the prediction that people can utilize confidence in such a flexible way. We expected the relationship between confidence and advice requests to vary depending on the way in which confidence is currently being utilized. Specifically, we hypothesized that frequent advice requests when confidence is low reflect the use of confidence as a self-monitoring tool, signalling that help should be requested. In contrast, advice requests when confidence is high reflect an attempt to learn about advisor quality, by using confidence as an internal feedback proxy.

We tested our predictions in three experiments using a Judge-Advisor System (JAS) paradigm. In its most basic form, the JAS consists of a decision maker (the judge), and one or more advisors who communicate their advice to the judge (J. Sniezek and Buckley, 1989; J. Sniezek and Buckley, 1995, for a review see Bonaccio and Dalal, 2006). In the present study, participants (i.e., acting as 'judge') performed a perceptual judgement task in which they reported which of two boxes contained more dots, and how confident they are in their response, similar to the task used in Chapter 2. Following each initial judgement, they chose between receiving advice from a virtual advisor or seeing the stimulus again. After receipt of the chosen information, they reported their final decision and confidence on which box contained more dots. In Experiment 6, we tested the novel prediction that, when advisor reliability is initially unknown and no objective feedback is provided, participants will ask for advice when they themselves are confident in a decision (and can therefore learn if advisors are providing accurate advice). Experiments 7 and 8 manipulated the availability of explicit information about advisor reliability (rendering learning about advisors unnecessary) and of external feedback on task performance (so that the use of confidence as an internal feedback proxy is redundant). Our data reveal that the relationship between confidence and advice requests is not fixed, but varies depending on the availability of such information. We conclude that confidence is used strategically, both as a self-monitoring tool, and as a proxy for feedback, depending on the task context and current goal.

3.2 Experiment 6

Experiment 6 aimed to test our main hypothesis, that people use confidence flexibly depending on their immediate goal. To test this hypothesis, we employed a perceptual decision-making task, in which participants had to report which of two boxes contained more dots and how confident they are in their decision, similar to the task used in Chapter 2. In the current experiment however, the stimulus did not evolve during presentation time, and no postdecisional evidence was presented. After reporting their initial response, participants chose between seeing the same stimulus again, for a longer period, or receiving advice from a virtual advisor. We predicted a U-shaped relationship between confidence and advice requests, reflecting the use of confidence for two different purposes. Specifically, we predicted that people would seek advice more frequently when confidence was very low, to improve their accuracy by using the help of others, but critically also when confidence was very high, enabling them to learn about advisor quality. Key predictions and analyses were preregistered at https://aspredicted.org/blind.php?x=n2pj9e.

Methods

Participants

26 participants took part in the experiment (17 females, ages 18-35 years, $M = 23.5$, $SD = 5.5$). All participants had normal or corrected-to-normal vision. They provided written informed consent, and were compensated with course credit or payment for their participation. One participant whose data could not be divided into quantiles (as defined below) was removed from the data analyses, as described in the results section. All procedures were approved by the University of Oxford's Medical Sciences Interdivisional Research Ethics Committee.

Task and Procedure

Participants performed a series of trials in which they first reported which of two boxes contained more dots and how confident they were in their response, and then chose to sample one of two additional sources of information, before giving a final response and confidence judgement. Figure 3.1 depicts an example trial in the main experimental blocks. Trials began with a fixation cross displayed at the center of the screen for 800 ms in total. 400 ms after the fixation cross appeared, it flashed briefly, indicating the perceptual stimulus will be presented shortly after. The stimulus, two boxes containing different numbers of dots, appeared to the left and right of the fixation cross for 150 ms. Participants reported which of the two boxes contained more dots, and how confident they are in their response on a scale ranging from 50%(guess), to 100%(sure). After the initial response was registered, images representing the advisor and the dot stimulus appeared above and below the fixation cross. Participants chose between viewing the stimulus again for a longer period (making it easier to judge which box contained more dots) versus receiving advice from a virtual advisor, by selecting the appropriate image with the mouse cursor. If participants chose to view the stimulus again, they were presented

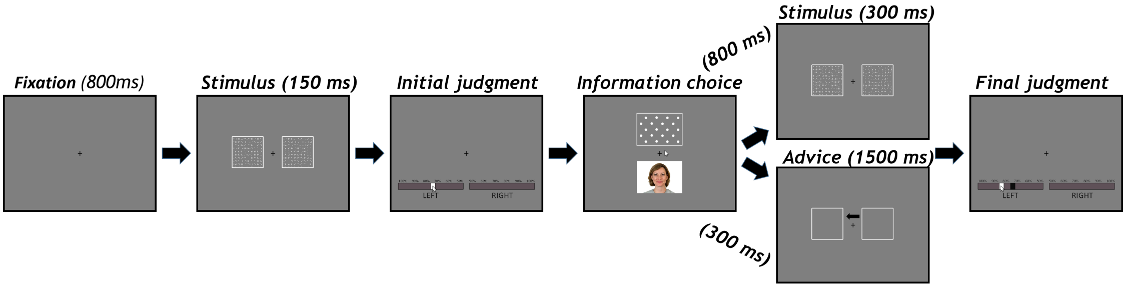

Figure 3.1: Experimental design. Participants made a speeded judgement about which of two squares contained more dots, and reported their confidence in this decision. Following the initial judgement, participants chose between viewing the stimulus again for an extended period, or receiving advice from a virtual advisor by selecting the corresponding image. After receiving the additional information they selected, participants reported their final decision and confidence on which box contained more dots.

with a fixation cross for 800 ms, followed by the same dot stimulus for an extended period of 300 ms. Following advice selections, a fixation cross appeared for 300 ms, after which the two boxes were presented to the right and left, but without the dot stimulus. If they chose advice, they saw an arrow pointing at one of the empty boxes (i.e. the advisor's recommended response), displayed for 1500 ms. Advice trials were longer than review stimulus trials to counterbalance a bias displayed towards advice selections in a pilot study. Following the additional information (stimulus re-viewing or advice), participants reported their final response and confidence on which box contained more dots, with their initial response marked on the confidence scale for reference. The next trial then began (i.e., with no feedback provided to indicate the box that objectively contained more dots).

The experiment consisted of six blocks of 60 trials each. Participants encountered a new advisor during each block, represented by a picture of a neutral Caucasian female selected from the NimStim database (Tottenham et al., 2009). Advisor accuracy varied across the 6 blocks, at 50%, 60%, 70%, 80%, 90%, and 100% correct, with the order of advisor accuracy randomised across blocks for each participant (and, across participants, randomly assigned to the 6 different advisor faces). Participants were instructed that advisors would be of different quality, but received no information about their accuracies. To encourage participants to choose according to their preference, the instructions indicated that 'It is up to you to decide on each trial if receiving advice or viewing the stimulus again for a longer duration is more useful to you'. The locations in which the advisor and view-stimulus options were presented during information choice (i.e., above vs. below the fixation) were assigned randomly for each block, with each information type appearing in each location an equal number of times throughout the experiment.

To induce a wide range of confidence judgements (that would allow us to study the relationship between different confidence levels and information choice), each block contained an equal number of trials in three difficulty levels (easy, medium, hard), with trial order determined randomly. Trial difficulty was controlled by manipulating the difference in dot numbers between the boxes, with the numbers in each box determined by the equation $nDots = 200 \pm d * l$, where $d =$ the number of dots added and subtracted to the correct and incorrect boxes correspondingly, and $l = [3, 1, 1/3]$, for the easy, medium and hard difficulties correspondingly. To achieve similar accuracy levels across participants, the medium difficulty condition was staircased to 70% accuracy using a 2-down 1-up procedure titrating the d parameter value throughout the experiment. The easy and difficult trials were defined relatively to the staircased medium difficulty by multiplying or dividing d by 3, as captured by the l parameter. The side of the box with more dots varied randomly across trials, and the locations of the dots within each box were sampled randomly on every trial. To encourage task engagement, participants received a bonus payment based on final-response accuracy during the main experimental blocks.

Before beginning the main experiment, participants viewed instructions explaining the task and performed three practice blocks. The first two blocks consisted only of the dot judgement task and contained 42 trials each. During these two training blocks, participants received feedback on their performance in the form of a tone after error trials, as well as feedback on their overall accuracy at the end of each block (based on participants' final responses). The third practice block introduced the information choice component, with participants performing 36 trials of the full task. The practice advisor had an accuracy of 80%. As in the main experiment, participants did not receive information about advisor accuracy or feedback following each trial of this block, but received feedback on their overall final response accuracy at the end of the block.

After completing the main experiment, participants completed the Concern over Mistakes, Personal Standards and Doubts about Actions subscales from the Multi-dimensional perfectionism scale (Frost et al., 1990). These surveys were included in this experiment, and in Experiment 7, to allow exploratory analysis of how perfectionism might relate to confidence and information choices. However, since no clear relationships were found, we do not report any analyses on the questionnaire data.

Stimuli and Apparatus

All stimuli were created and presented in Matlab using the Psychtoolbox3 extension (Brainard, 1997; Kleiner et al., 2007; Pelli, 1997). Stimuli were presented on a 20 inch CRT screen with grey background (127, 127, 127), a refresh rate of 60 Hz and $1600X1200$ resolution. Dot stimuli comprised two white rectangular frames in which white dots were presented in random locations in a $20X20$ grid. Advisors were represented by a picture of a Caucasian female with a neutral expression, selected from the NimStim database (Tottenham et al., 2009). The experiment took place in a dimly lit room, with participants sitting at a comfortable viewing distance from the screen.

Statistical Analyses

Confidence Quantiles. To study the relationship between subjective confidence and behaviour, trials were grouped according to confidence ratings. For each participant, N quantiles were computed on initial confidence ratings (with N depending on the specific analysis), and trials grouped into bins according to trial confidence so that $Q(n-1) < conf(t) <= Q(n)$. If certain confidence ratings were used frequently, resulting in bins with shared cut off values or containing fewer than 20 trials, new quantiles were computed above and below these values. For example, if $Q(n) = Q(n-1)$, $n-1$ quantiles were calculated on $conf <= Q(n)$, and the remaining quantiles were calculated on $conf > Q(n)$. If the last quantiles had shared cut off values, or the cutoff value was 100 (the highest confidence rating), $N - 1$ quantiles were calculated on $conf < Q(n)$, so that the highest confidence rating was binned in the last quantile.

Results

Accuracy and Confidence Judgements

To confirm that participants were complying with task instructions, our first analysis studied the differences in decision accuracy and confidence ratings between the initial and final responses. We expected that if, as instructed, participants used the additional information provided between their responses (i.e., receive advice or view stimulus again) to improve their decision accuracy, this would lead to an increase in accuracy in the final response. Consistent with this prediction, a paired t-test revealed that mean accuracy was higher on the final response, compared to the initial response ($M = 77\%$ vs. 74%, $SD = 3\%$ vs. 2%; $t(24) = 6.14, p < .001$). This increase in accuracy was mirrored in participants' confidence ratings, with participants reporting higher confidence in their final responses, compared to their initial confidence ($M = 73$ vs. 70, $SD = 8.6$ vs. 8.7; $t(24) = 7.06$, $p < .001$).

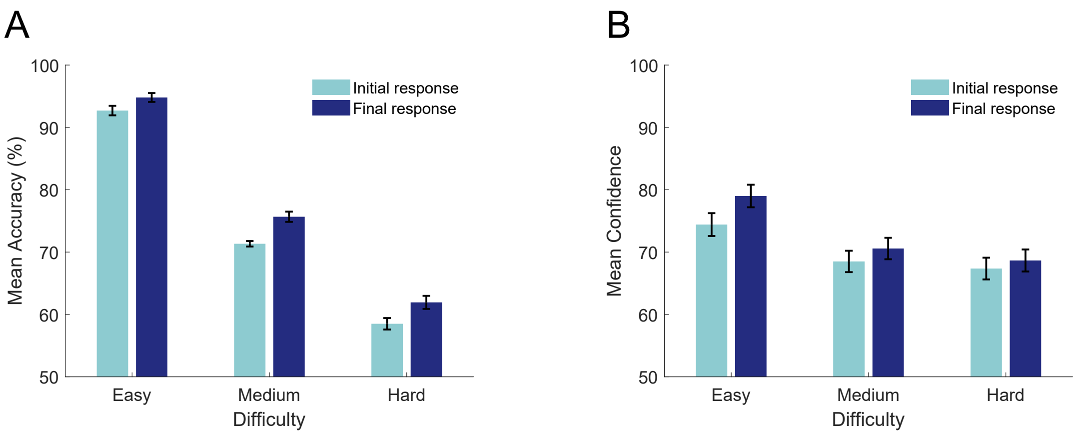

Figure 3.2: Results of Experiment 6: effects of task difficulty on accuracy and confidence. (A) Mean accuracy on the initial and final responses, divided according to trial difficulty. (B) Mean confidence ratings on the initial and final responses, divided according to trial difficulty. Error bars indicate standard errors of the mean (SEM).

Next, we wanted to confirm that participants were sensitive to the difficulty manipulation, i.e., that the varying difficulty levels induced a range of objective accuracy and subjective confidence scores. Given that our main goal in this experiment was to study the relationship between confidence and subsequent information choices, initial accuracy and confidence were of particular interest. However, for completeness, we report the effects of trial difficulty on both the initial and final responses. Mean accuracy and confidence ratings were computed as a function of trial difficulty, for the initial and final responses separately (see Figure 3.2), and subjected to separate repeated-measures ANOVAs. Results showed that, as expected, accuracy varied according to trial difficulty for both the initial and final responses ($F(2, 48) = 625.25, p < .001, \eta^2 = .96$; $F(2, 48) = 443.73, p < .001, \eta^2 = .95$, for the initial and final responses, correspondingly). Within-subject linear contrasts were significant ($F(1, 24) = 1129.18, p < .001, \eta^2 = .98$; $F(1, 24) = 805.85, p < .001, \eta^2 = .97$), with higher accuracy rates on easier trials, for the initial ($M = 93\%, 71\%, 59\%$, $SD = 4\%, 2\%, 5\%$) and final responses ($M = 95\%, 76\%, 62\%$, $SD = 4\%, 4\%, 5\%$ for easy, medium and hard trials correspondingly). Most importantly, a similar pattern emerged for confidence judgements. Confidence differed monotonically according to task difficulty (initial response: $F(2, 48) = 133.77, p < .001, \eta^2 = .85$; final response: $F(2, 48) = 152.02, p < .001, \eta^2 = .86$), with significant within-subject linear contrasts ($F(1, 24) = 154.84, p < .001, \eta^2 = .87$; $F(1, 24) = 193.46, p < .001, \eta^2 = .89$), and higher confidence ratings for easier trials on the initial ($M = 74, 68, 67$, $SDs = 9$) and final responses ($M = 79, 71, 69$, $SDs = 9$). Thus, the difficulty manipulation had the expected effects, leading to varying accuracy levels and, importantly, a range of confidence scores, with participants performing better, and reporting higher confidence, on easier trials.

Confidence-Accuracy calibration

To further study the relationship between subjective confidence and objective accuracy, we grouped trials according to confidence ratings, pooling across all trial

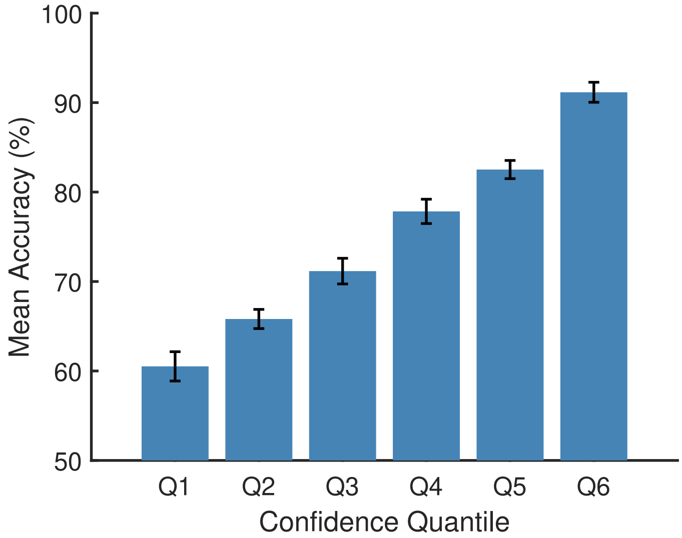

Figure 3.3: Results of Experiment 6: Mean accuracy according to reported confidence. For each participant, trials were divided into six bins according confidence ratings. Error bars indicate standard errors of the mean (SEM).

difficulties. For each participant, trials were grouped into six bins according to initial-confidence sextiles, and mean accuracy rates computed for each bin (see Figure 3.3). A repeated-measures ANOVA showed that mean accuracy varied reliably as a function of confidence ($F(5, 120) = 76.02, p < .001, \eta^2 = .76$), with accuracy increasing monotonically with the level of confidence, as captured by a significant linear trend ($F(1, 24) = 286.99, p < .001, \eta^2 = .92$; $M = 61\%, 66\%, 71\%, 78\%, 83\%, 91\%$, and $SD = 8\%, 5\%, 7\%, 7\%, 5\%, 6\%$). Thus, participants' subjective reports of confidence were well-calibrated to objective accuracy. This pattern was seen across experiments but is reported here only, for the sake of brevity.

Learning about advisor quality

Across blocks, participants encountered advisors whose accuracy varied from 50% to 100%, but no explicit trial-by-trial feedback was provided against which to evaluate the quality of this advice. Participants nevertheless learned about advisor quality when deciding whether to ask for advice: As demonstrated in Figure 3.4A , the proportion of advice requested throughout the experiment depended on advisor accuracy. A repeated-measures ANOVA on the proportion of advice requests as a function of advisor accuracy showed a significant effect ($F(5, 120) = 2.80, p = .02, \eta^2 = .818$), with a significant within-subject linear contrast ($F(1, 24) = 6.87, p = .015, \eta^2 = .71$), indicating that advice requests increased monotonically as a function of advisor quality. These results provide further evidence that people can learn about the quality of advice even when objective feedback is lacking (Pescetelli & Yeung, 2020). As described above, our hypothesis was that this learning would by driven by participants seeking advice when they were already sure about the correct answer, so that they could evaluate advisors against these firmly held beliefs, rather than only seeking advice when low in confidence as seen in previous research.

Confidence effects on information choices

Overall, participants chose to receive advice on .39 of the trials on average, with large individual differences in preferences for advice or viewing the stimulus again (proportion advice choice: $range = 0 - .91$, $SD = .26$). Both advice and viewing the stimulus again increased participants' accuracy on their final response compared to their initial response (from 74% to 77% and 72% to 74%, respectively), indicating that both types of information were of value ($t(23) = 3.97, p < .001$ and $t(24) = 4.23, p < .001$, for advice and view-stimulus trials, respectively; note that one participant never requested advice).

To answer our main question—whether people use confidence strategically depending on their immediate goal—we studied the relationship between confidence and information choices. For each participant, we divided trials into six bins according to initial-confidence sextiles, and then computed the proportion of trials in which advice was selected for each bin. We predicted a U-shaped relationship between confidence and advice selections, whereby people would choose to receive advice, rather than view the stimulus again, more often both when they were initially low in confidence (reflecting an attempt to use advice in order to improve accuracy), and when they were initially high in confidence (reflecting an attempt to learn about advisor quality). As expected, a repeated-measures ANOVA revealed a significant effect of confidence on information choices ($F(5, 120) = 5.34, p = .012, \eta^2 = .18$),

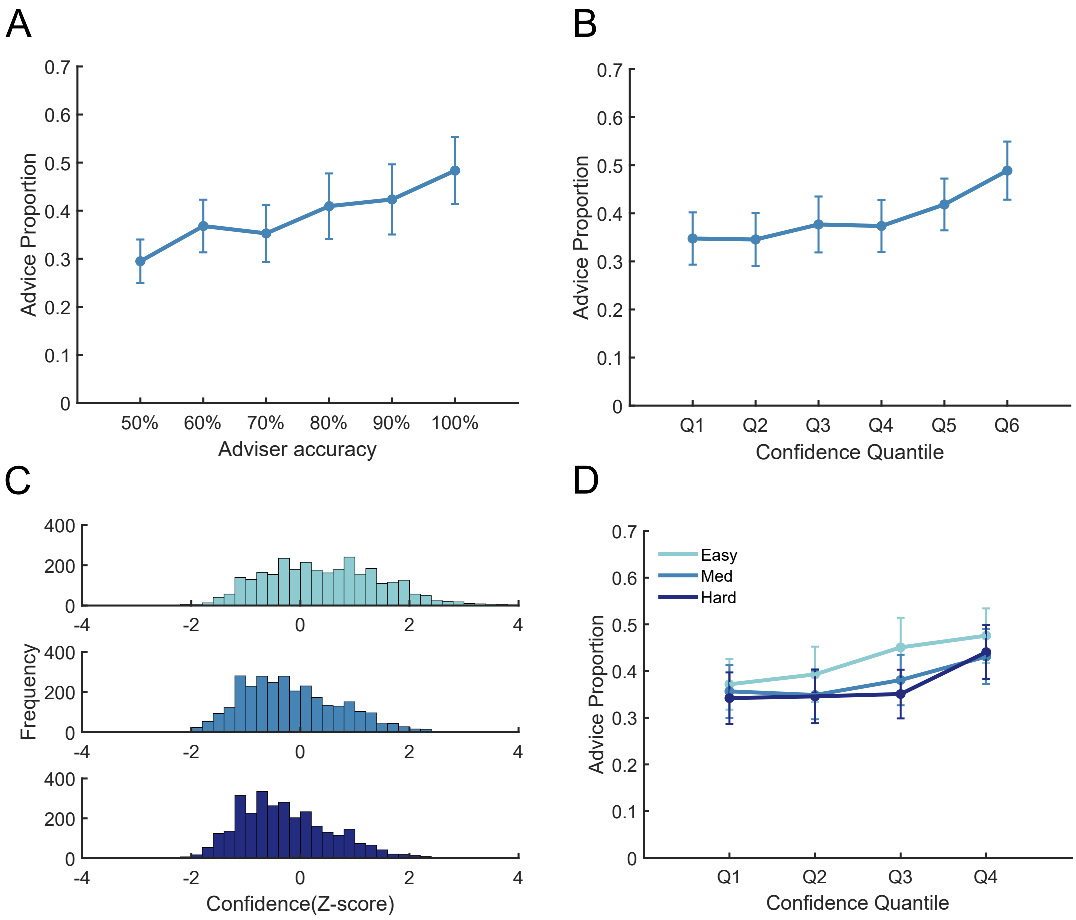

Figure 3.4: Results of Experiment 6. (A) Mean proportion of advice selections according to advisor accuracy. (B) Mean proportion of advice selections according to confidence bins across the entire experiment. (C) Normalised initial confidence ratings, divided according to trial difficulty. Confidence ratings were first normalized within each participant, across all trials regardless of difficulty. Then, normalised confidence ratings were aggregated across all participants and grouped according to trial difficulty. (D) Mean proportion of advice selections according to confidence bins and trial difficulty. For each participant, trials within each difficulty were divided into four bins according to confidence ratings. Error bars represent SEMs.

and a significant quadratic trend ($F(1, 24) = 5.82, p = .024, \eta^2 = .195$). Surprisingly, however, we found that people requested advice more often when confidence was high, but not low—the significant quadratic trend reflected a levelling off, rather than increase, in advice-seeking choices at the lowest levels of confidence. Pairwise comparisons of confidence sextile bins with Bonferroni correction revealed significant differences in the frequency of advice selections between trials with the highest confidence ratings (Q6) and medium confidence trials (Q5 and Q4, $ps < .05$), but no significant differences for the lowest confidence trials ($Q1$). Thus, while these results support the hypothesis that people use confidence as a feedback proxy to learn about advisor reliability, they do not support our initial prediction of a U-shaped relationship between confidence and advice requests (see Figure 3.4B).

Since confidence varied reliably as a function of trial difficulty, an alternative interpretation of the above results is that information choices were driven by trial difficulty, rather than confidence. Indeed, because the difficulty manipulation was introduced precisely with the intention of inducing variability in confidence ratings, trials of different difficulty levels were not equally represented in confidence bins, making it difficult to separate these effects. For example, the highest confidence bin comprised on average 36, 14 and 9 trials from the easy, medium and hard conditions, respectively. Importantly, despite these differences in confidence distributions, there was also a considerable overlap in the confidence ratings between the different difficulties. Figure 3.4C shows the distributions of initial confidence ratings for each difficulty level plotted across all participants, after normalizing confidence within each participant (across all trials regardless of difficulty). The visible overlap in confidence ratings between the different difficulty levels, renders an alternative explanation, whereby behaviour is driven by objective difficulty rather than confidence, unlikely.

To test directly whether confidence affects information seeking above and beyond trial difficulty, we performed a two-way repeated-measures ANOVA on the proportion of advice requests, with factors of trial difficulty and confidence. For each participant, trials were divided first according to task difficulty, and then based on confidence ratings. Confidence quartiles were computed for each participant within each difficulty, and trials divided into bins according to confidence (here we used four bins, rather than six, due to the smaller number of trials available within each difficulty level). We found that both trial difficulty ($F(2, 48) = 5.96, p = .019, \eta^2 = .20$) and confidence ($F(3, 72) = 4.11, p = .041, \eta^2 = .15$) reliably predicted information choices, and the interaction between them was not significant ($F(6, 144) = 1.25, p = 0.28, \eta^2 = .05$). As shown in Figure 3.4D, participants requested advice more frequently on easy trials, compared to medium and hard trials ($ps = .02$). In addition, participants asked for advice more often when they were highly confident ($p = .049, .044, .010$ for Q4 compared to Q1, Q2 and Q3 correspondingly). Notably, an increase in advice requests for the highest confidence bin is visible within all three difficulty conditions. These results indicate that confidence influences information choices above and beyond trial difficulty.

Evidence for use of confidence for learning

We predicted this high rate of advice requests at the highest levels of confidence based on the hypothesis that participants were using their own judgements as a yardstick against which to judge (and learn about) advisor quality. As a more direct test of this idea, we examined how the relationship between confidence and information choices varied during blocks, as participants presumably learned about the quality of advice offered by the current advisor. Based on our learning hypothesis, we expected more advice selections on high-confidence trials at the beginning of each block, when participants did not have any information about the current advisor's accuracy. As the block progressed, we expected less advice choices when confidence was high, since participants should have gained knowledge about advisor quality and would no longer needed to use high confidence trials to learn about advice.

Consistent with these predictions, a two-way repeated-measures ANOVA on the proportion of advice choices, with time in block (first half, second half) and confidence (Q1-Q6) as factors, revealed significant main effects for time ($F(1, 24) = 19.95, p < .001, \eta^2 = .45$) and confidence ($F(5, 120) = 5.02, p = .016, \eta^2 = .17$), and, importantly, a reliable interaction effect between these factors ($F(5, 120) = 3.95, p = .01, \eta^2 = .14$): As shown in Figure 3.5, advice choices varied more strongly as a function of confidence in the first half of blocks than the second. Follow-up one-way ANOVAs on advice choices for each block half separately revealed that confidence reliably predicted advice requests during the first time period ($F(5, 120) = 7.04, p = .003, \eta^2 = .23$) but not during the second time period ($F(5, 24) = 1.93, p = .15$).

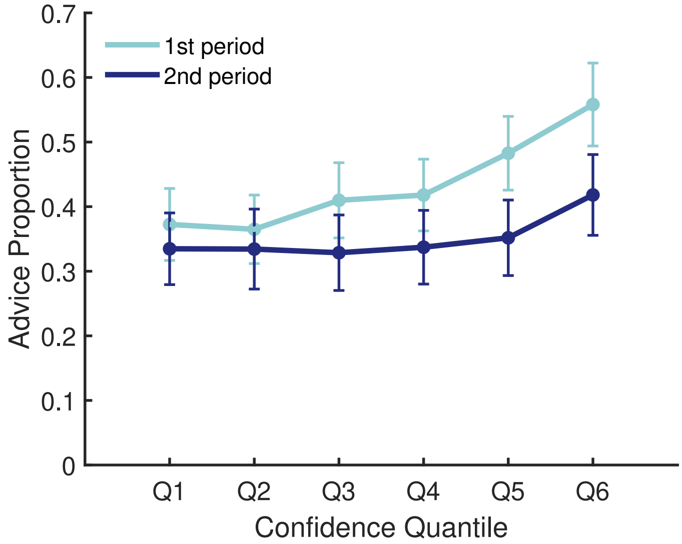

Figure 3.5: Results of Experiment 6. Mean proportion of advice selections according to confidence and time in block. For each participant, trials were divided into two time periods (first half vs. second half of block), and then into six bins according to confidence ratings, within each time period. Error bars represent SEMs.

Discussion

The results of Experiment 6 provided some support of our main hypothesis, but also produced some surprising results. As predicted, we found that participants were able to learn about the accuracy of their advisors even without objective feedback, and that they sought advice frequently when they were most confident in their own initial decisions, particularly early in each block as they were learning about the quality of advice on offer. These findings are consistent with our hypothesis that confidence is used strategically depending on peoples' immediate goal: Advice-seeking is not solely associated with situations of high uncertainty, as explored in previous research, but can also be seen in situations of high confidence, when an important goal is to learn about the current task context. Surprisingly, however, we did not find that participants were more likely to request advice when they were unconfident in their initial response, such that we did not observe the U-shaped relationship we had originally predicted between confidence and advice seeking: Instead, participants' preferences for receiving advice vs. re-viewing the stimulus were relatively stable from medium to low levels of confidence. The explanation for this pattern is unclear, and as we will see below, the effect varied across experiments, with some evidence that participants are more likely to request advice when confidence is low in Experiments 7 and 8.

Experiment 7 built on these findings to compare conditions in which participants received prior information about advice quality, rendering learning unnecessary, to conditions in which they did not receive prior information, and thus had to learn about the quality of advice. We predicted that participants would show evidence of using confidence in a strategic fashion, such that the relationship between confidence and advice-seeking would vary across these conditions.

3.3 Experiment 7

Experiment 7 tested whether confidence is used flexibly depending on the availability of external information, which we varied in a between-participants design. Using a similar task to Experiment 6, we manipulated the availability of explicit prior information about advice quality: In the 'Explicit information available' condition, participants performed the same task as in Experiment 6 but, importantly, were notified of the current advisor's average accuracy at the start of each block. The 'Explicit information absent' condition was a replication of Experiment 6, so that explicit information about advisor accuracy was not provided (and instead advisor accuracy had to be learnt). We expected the relationship between confidence and information choices to change depending on information availability. Specifically, we predicted that people will use trials on which they are highly confident to learn about advisor quality only when explicit information about advice was absent. In contrast, when explicit information about advisor quality was available, and learning about advisors unnecessary, we did not expect to see a higher frequency of advice selections when confidence was high. We did not have a specific prediction regarding the relationship between confidence and advice seeking in this case: It could be that people might request more advice when confidence is low in order to improve their immediate accuracy, but it seemed equally plausible that they might not use confidence to guide advice requests in this condition, and that advice requests would depend only on advisor quality, which was known. Key predictions and analyses were preregistered at http://aspredicted.org/blind.php?x=ay2ng3.

Methods

Participants

51 participants took part in Experiment 7. One participant whose accuracy on the initial response was more than two standard deviations below the mean accuracy was removed, resulting in a final sample of 50, with 25 participants in each condition (35 females, ages 18-33 years, $M = 21.7$, $SD = 4.2$). All participants had normal or corrected-to-normal vision. They provided written informed consent, and were compensated with course credit or payment for their participation. All procedures were approved by the University of Oxford's Medical Sciences Interdivisional Research Ethics Committee.

Task and Procedure

Experiment 7 used a mixed design, with explicit information about advisor accuracy (Absent, Available) varying between participants. For each participant, advisor accuracy varied randomly between blocks across 5 levels, from 50% - 90% (i.e., there were no blocks with 100% accurate advice), and within each block participants performed the perceptual task across three levels of task difficulty. The perceptual task and information choice were identical to Experiment 6. Thus, the 25 participants assigned to the information absent condition provided a replication of Experiment 6. The remaining 25 participants were assigned to the information available condition in which, at the beginning of every block, a message on the screen informed them of the advisor's accuracy in the current block (e.g., 'Advisor accuracy on this block: 70%'). This experiment did not include a block with 100% accurate advice, to avoid participants in the information available condition always requesting advice from the 100% accurate advisor. The end of block feedback differed slightly from Experiment 6, with participants receiving feedback on their average initial-response accuracy rate, in addition to average final-response accuracy.

Stimuli and Apparatus and Procedure

The stimuli and apparatus were identical to Experiment 6, except for advisor stimuli: In Experiment 7, advisors were represented by a picture of a smiling Caucasian female selected from the Chicago Face database (Ma et al., 2015).

Results

Accuracy and confidence judgements

Our first analysis assessed whether participants in both experimental conditions were performing the first and second order tasks in compliance with instructions. We performed a two-way mixed measures ANOVA on mean accuracy, with factors for explicit information (absent, available) and response (initial, final). The analysis showed a reliable effect for response ($F(1, 48) = 108.27, p < .001, \eta^2 = 0.69$), as well as a reliable interaction between response and external information ($F(1, 48) = 6.38, p < .02, \eta^2 = 0.12$). Follow-up tests showed that, as expected, mean accuracy was higher on the final response compared to the initial response, for participants in both experimental conditions. The difference was larger for participants who received information about advisor accuracy, likely because they could better use advice to improve their accuracy (M = 78% vs. 75%, SD = 2% vs. 3%, $t(24) = 5.90, p < .001$ and M = 79% vs. 74%, SDs = 3%, $t(24) = 8.68, p < .001$, for participants without explicit information, and with information, respectively). A similar ANOVA on confidence revealed a significant effect for response ($F(1, 48) = 51.08, p > .001, \eta^2 = .52$), but not for the interaction between response and external information ($F < 1$). Participants in both experimental groups reported higher confidence in their final responses compared to the initial responses ($M = 73$ vs. 71, and $M = 74$ vs. 71), with this difference slightly larger numerically for the group with explicit information. Thus, participants used additional information to improve their accuracy, and reliably tracked this improvement in their confidence reports. Also replicating the results of Experiment 6, mean accuracy in both experimental conditions varied reliably as a function of task difficulty, for both the initial ($F(2, 48) = 424.41, p < .001, \eta^2 = .946, F(2, 48) = 542.34, p < .001, \eta^2 = .96$)

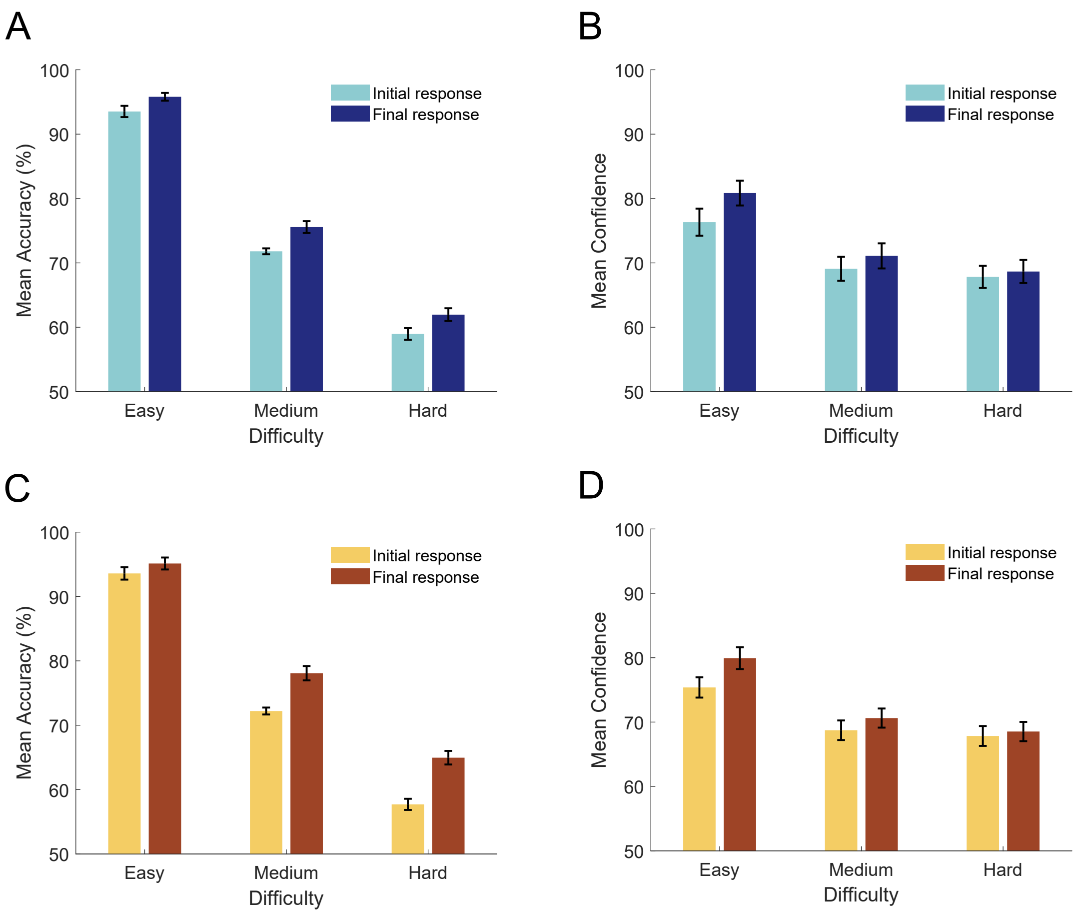

Figure 3.6: Results of Experiment 7: effects of response (initial vs. final) and task difficulty on accuracy and confidence according to the experimental condition. Participants who did not receive explicit information about advisor accuracies are depicted in blue (A-B), and participants who did receive explicit information about advisors are depicted in orange (C-D). (A,C) Mean accuracy on the initial and final responses divided according to trial difficulty. (B,D) Mean confidence ratings on the initial and final responses, divided according to trial difficulty. Error bars indicate standard errors of the mean (SEM).

and the final responses ($F(2,48) = 463.54, p < .001, \eta^2 = .95$, $F(2,48) = 240.79, p < .001, \eta^2 = .91$), for participants without external information, and with information, respectively. As shown in Figure 3.6, similar patterns were found for confidence, but are not reported for the sake of brevity.

Effects of advisor quality

A two-way mixed measures ANOVA on the proportion of advice requests with factors for advisor accuracy and external information availability showed reliable effects for advisor accuracy($F(4,192) = 42.46, p < .001, \eta^2 = .47$) and for the

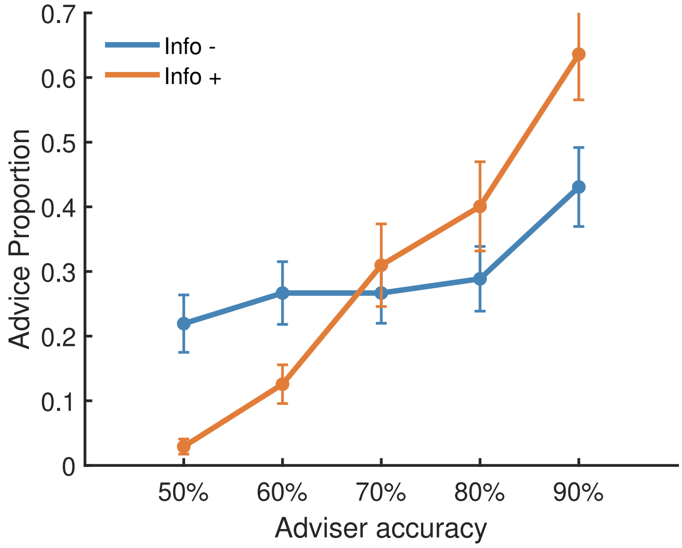

Figure 3.7: Results of Experiment 7. Mean proportion of advice selections according to advisor accuracy and explicit information availability. Participants who did not receive explicit information about advisor accuracies are depicted in blue, and participants who did receive explicit information about advisors are depicted in orange. Error bars indicate SEMs.

interaction effect between advisor accuracy and information availability ($F(4, 192) = 12.46, p < .001, \eta^2 = .21$). Critically, participants in both groups showed sensitivity to advice quality in their choices to receive advice vs. re-view the stimulus. Thus, replicating the results of Experiment 6, the proportion of advice requests was reliably predicted by advisor accuracy even when external information was not available ($F(4, 96) = 8.38, p < .001, \eta^2 = .26$; see Figure 3.7), with a reliable linear contrast ($F(1, 24) = 21.73, p < .001, \eta^2 = .475$). Not surprisingly, we found a stronger effect of advisor accuracy on the proportion of advice selections for participants who were told about advisor accuracy before each block ($F(4, 96) = 36.52, p < .001, \eta^2 = .60$), again with a significant linear trend ($F(1, 24) = 69.73, p < .001, \eta^2 = .74$).

Figures 3.8A and 3.8B, show a ten trial running average of the proportion of advice requests according to advisor accuracy and information condition, and demonstrate the differences in learning according to information availability. When explicit information was not available, participants learned about advisor accuracy throughout the block, asking for less advice from low quality advisors as the block progressed. In contrast, when explicit information about advisor accuracy

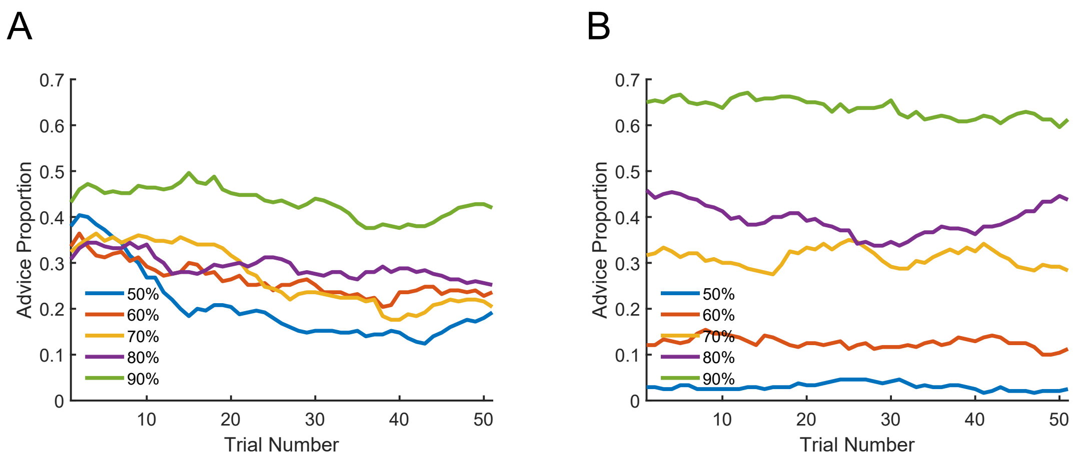

Figure 3.8: Results of Experiment 7. Ten trial running average showing the proportion of advice requests according to advisor accuracy, across blocks and participants when (A) Explicit information about advisors was not available, and (B) Explicit information about advisor accuracy was available.

was available, the proportion of advice requests varied as a function of advisor quality stably throughout the block. Thus, patterns of advisor choice reflected advisor quality for both groups of participants, but only the group without external information about advisor quality showed evidence of learning. Of interest, then, was whether confidence influenced advisor choice differently across groups, reflecting the different needs and goals of the two groups of participants.

Effects of confidence on information choices depend on explicit information availability

Our results from Experiment 6 showed that participants selected more advice when confidence was high, presumably in order to learn about advice quality. In the current experiment, we predicted that participants would use confidence differently depending on the availability of information about advisor accuracy. Specifically, we predicted that when no explicit information about advice quality was available (i.e., replicating the conditions of Experiment 6), participants would again request more advice when confidence was high, allowing them to learn about the advisors. In contrast, we did not expect to see this pattern for participants in the explicit information condition, since learning about advisor accuracy was unnecessary.

To test these predictions, each participant's trials were divided into six bins according to confidence, and the proportion of advice requests computed. We performed a two-way mixed ANOVA on the proportion of advice requests (which again varied substantially across participants: range = 0 - .88) with factors for external information (absent, available) and confidence (Q1-Q6, see Figure 3.9). As predicted, this analysis revealed a significant interaction between confidence and information availability ($F(5, 240) = 6.73, p < .001, \eta^2 = .12$), as well as a significant main effect for confidence ($F = (5, 240) = 3.54, p = .02, \eta^2 = .07$), but not for information availability ($F < 1$). Given the reliable interaction effect, indicating that confidence influences information sampling differently depending on the availability of external information about advice, we performed separate follow-up repeated measure ANOVAs, to study the relationship between confidence and information sampling within each condition.

We found that when explicit information was not available, confidence predicted information choices reliably ($F(5, 120) = 7.51$, $p < .001$, $\eta^2 = .24$), with a reliable quadratic trend found in the data ($F(1, 24) = 22.57, p < .001, \eta^2 = .49$). Importantly, consistent with the results of Experiment 6, the proportion of advice requests was highest for the highest confidence bin (Q6), with Bonferroni-corrected pairwise comparisons revealing reliable differences in the frequency of advice requests between high confidence trials (Q6) and medium confidence trials (Q2, Q3, Q4 and Q5, $ps < .05$). Interestingly, the data revealed a numerical increase in advice requests for lowest confidence (Q1) compared to Q2 ($t(24) = 2.75$), but this difference was not reliable when correcting for multiple comparisons ($p = .17$). For participants who were provided with information about advisor quality before each block, confidence reliably predicted advice ($F(5, 120) = 3.05, p = .046, \eta^2 = .11$) but, crucially, for these participants the proportion of advice requests decreased as a function of initial decision confidence, with no evidence of increased advice seeking at the highest levels of initial confidence (though note that pairwise comparisons did not reveal any reliable differences after bonferroni corrections).

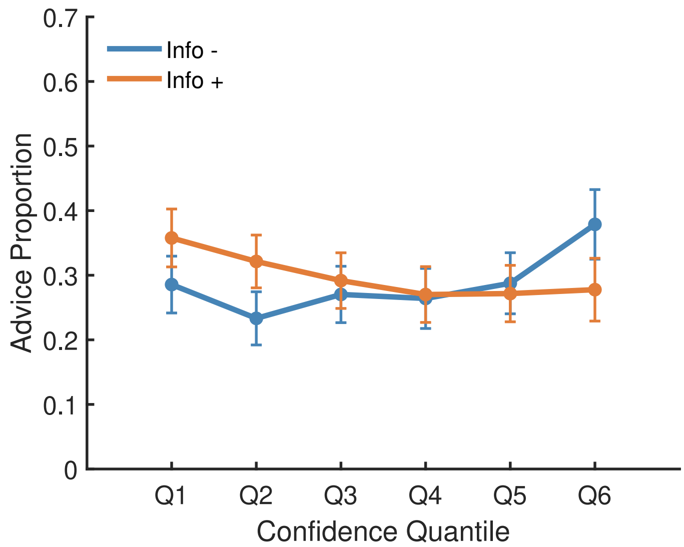

Figure 3.9: Results of Experiment 7. Mean proportion of advice choices according to explicit information availability and confidence. Participants who did not receive explicit information about advisor accuracies are depicted in blue, and participants who did receive explicit information about advisors are depicted in orange. Error bars represent SEMs.

To ensure that advice requests were driven by confidence, rather than trial difficulty, we performed two-way repeated measures ANOVAs on the proportion of advice selections as a function of confidence and trial difficulty, separately for each information condition. For each participant, trials were divided first according to trial difficulty (easy medium, hard), and within each difficulty into four bins according to confidence quartiles (Q1-Q4). When external information was not available (Figure 3.10A), results revealed a reliable interaction effect between confidence and difficulty ($F(6, 144) = 4.91$, $p < .001$, $\eta^2 = .17$), with a significant main effect of confidence on advice requests ($F(3, 72) = 6.35$, $p < .005$, $\eta^2 = 0.21$), and difficulty approaching significance ($F(2, 48) = 3.19$, $p = .053$, $\eta^2 = .12$). Since the interaction between confidence and difficulty was significant, we performed follow-up one-way ANOVAs studying the effects of confidence within each difficulty level. Results showed a significant effect of confidence only within the easiest condition ($F(3, 72) = 10.91, p < .001, \eta^2 = .31$), but not in the medium ($F(6, 72) = 2.06, p = .14, \eta^2 = .08$) or difficult trials ($F(3, 72) = 1.05, p = 0.37, \eta^2 = 0.04$). We found a reliable interaction between confidence and difficulty also when external

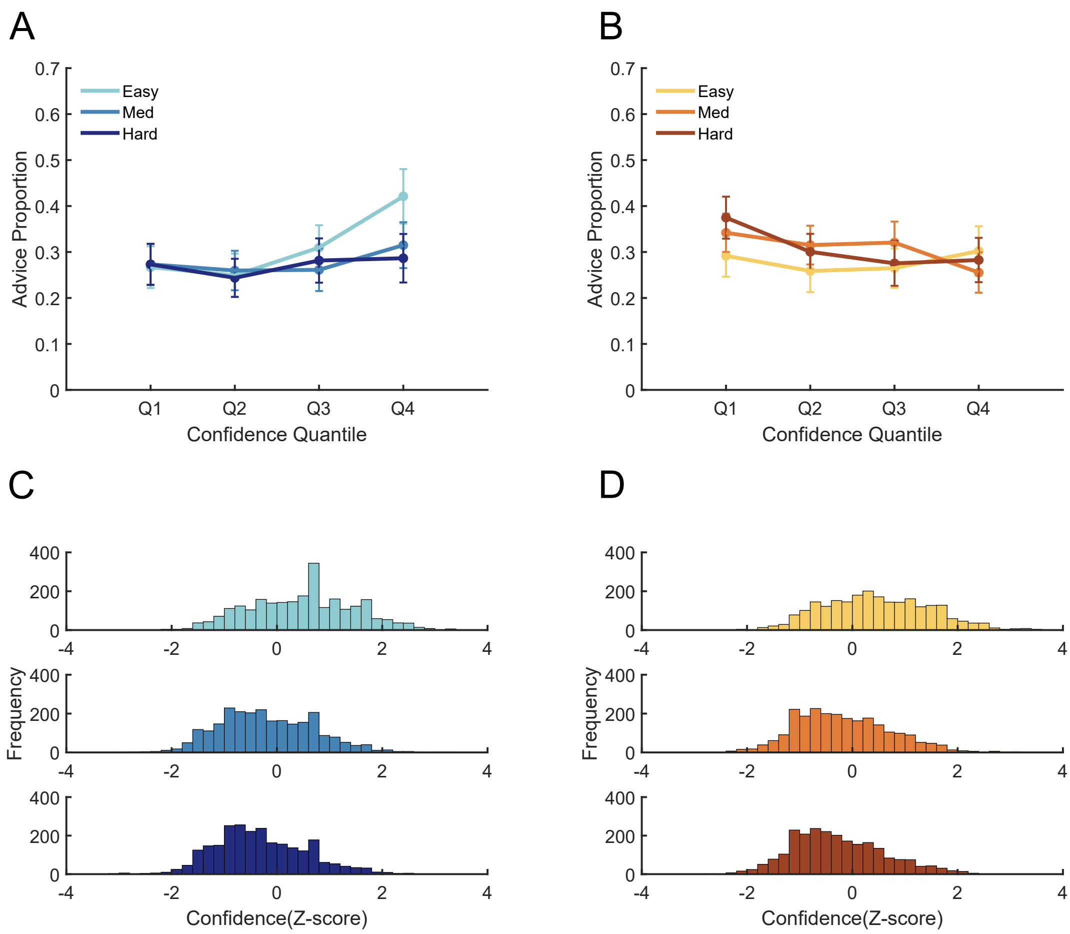

Figure 3.10: Results of Experiment 7. (A,B) Mean proportion of advice selections according to trial difficulty and confidence ratings, for each information condition. For each participant, trials were divided first according to difficulty, and then into four bins according to confidence rating within each difficulty. (C,D) Normalised initial confidence ratings, divided according to trial difficulty, for each information condition. Confidence ratings were first normalized within each participant, across all trials regardless of difficulty. Then, normalised confidence ratings were aggregated across all participants in the experimental group and grouped according to trial difficulty. Participants who did not receive explicit information about advisor accuracies are depicted in blue, and participants who did receive explicit information about advisors are depicted in orange. Error bars represent SEMs.

information about advisor accuracy was available ($F(6, 138) = 3.17$, $p = .015$, $\eta^2 = .12$), with the main effects for confidence ($F(3, 69) = 2.04$, $p = .16, \eta^2 = .08$) and difficulty ($F(2, 46) = 3.11$, $p = .06$, $\eta^2 = .12$) not significant (Figure 3.10B). Here, follow-up one-way ANOVAs examining the effects of confidence within each difficulty revealed the opposite pattern to that seen when external information was unavailable, with greatest variation in advice seeking as a function of confidence when trial difficulty was high ($F(3, 72) = 4.05, p = .03, \eta^2 = 0.14$) or medium ($F(3, 72) = 3.69, p = .03, \eta^2 = 0.13$), and less so when the task was easy ($F < 1$). These results demonstrate that information sampling decisions were guided by subjective confidence, above and beyond trial difficulty, with confidence having an effect on information choices even within a difficulty level. Since trials in the easiest difficulty led to higher confidence judgements (see Figure 3.10C), the majority of trials in which participants were certain they knew the response (i.e. the highest confidence bin) were from trials in the easiest difficulty. Thus, we see effects of confidence most clearly within the easy trials, with people requesting advice most frequently when confidence was highest. When information about advisor accuracy was available, people tended to request advice more often when their confidence was low, presumably to improve their accuracy. As shown in Figure 3.10D, low confidence ratings were more frequent in the medium and hard difficulty conditions. Thus, we see the effects of confidence only within these conditions.

Effects of time in block depend on need to learn

In Experiment 6, we demonstrated that the effects of confidence on information sampling depend on time, with people selecting advice when confidence is high more often only during the first half of each block, but not the second. We suggested that these differences are a result of learning—once participants have learned how good advisors are, they do not request advice when confidence is high. If this is the case, we should not see a change in the relationship between confidence and advice as a function of time under conditions where learning is redundant. To test this prediction, for each participant trials were divided according to time in the block and confidence ratings. We performed two-way repeated-measures ANOVAs on the proportion of advice requests, as a function of time in block (first vs. second half), and confidence bin (Q1-Q6), for each condition separately.

Replicating our results from Experiment 6, we found that when external information about advisors' accuracy was not available, advice requests were reliably predicted by confidence ($F(5, 120) = 7.49, p < .001, \eta^2 = .24$), and time

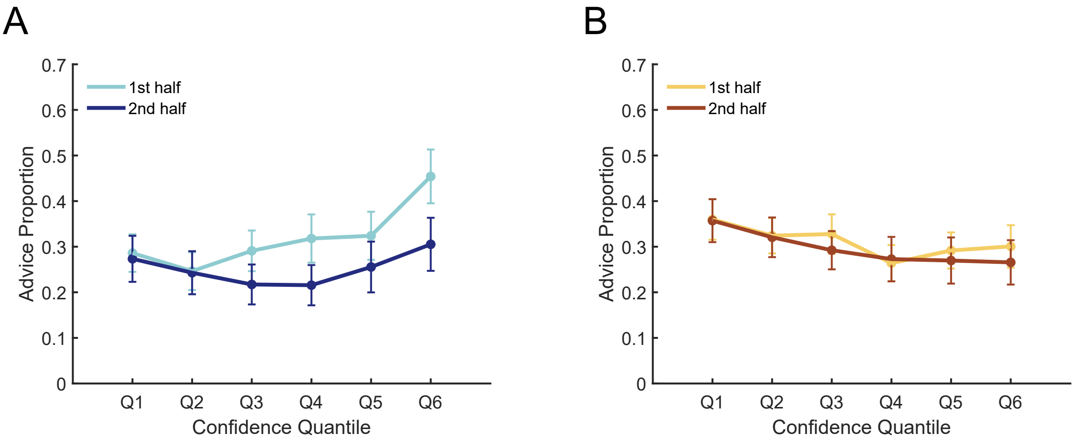

Figure 3.11: Results of Experiment 7: Mean proportion of advice selections according to information availability, confidence and time in block. For each participant, trials were divided into two time periods (first half vs. second half of block), and then into six bins according to confidence ratings, within each time period. (A) Participants who did not receive explicit information about advisor accuracy. (B) Participants who did receive explicit information about advisor accuracy. Error bars represent SEMs.

($F(1, 24) = 6.85, p = .015, \eta^2 = .22$). Importantly, the interaction term between time and confidence was also significant ($F(5, 120) = 3.72, p = .014, \eta^2 = .13$). Confidence reliably predicted advice requests during the first time period ($F(5, 120) = 8.64, p < .001, \eta^2 = .26$), but not the second ($F(5, 120) = 2.63, p = .06, \eta^2 = 0.10$). When external information was available, we found that advice requests were reliably predicted by confidence ($F(5, 120) = 3.29, p = .04, \eta^2 = .12$), but, crucially, there were no significant effects of time ($F(1, 24) = 1.95, p = .5, \eta^2 = .08$), or the interaction between time and confidence ($F < 1$, see Figure 3.11).

Discussion

Experiment 7 investigated whether, when choosing between advice and re-sampling the evidence, people use their confidence differently depending on the availability of explicit information about advisor reliability. Replicating the results from Experiment 6, we found that when explicit information about advisor quality was not available, participants requested advice more frequently when confidence was high. In contrast, when external information about advisor quality was available, participants tended to request advice more often when confidence was low. Our results further demonstrated that the relationship between confidence and information choices varied across blocks only when explicit information about advisors was not available. Importantly, participants requested advice when confidence was high only during the first half of each block. These results support our interpretation that people use their confidence as an internal feedback cue, allowing them to learn how good advice is by comparing the advice they received with their own beliefs when they are highly confident. When participants do not want to learn about advice quality, either because explicit information is available or because they have already learned earlier in the block, they do not use confidence as a feedback proxy for evaluating advisors.

Taken together, Experiments 6 and 7 provide evidence for our two key hypotheses: that confidence guides adaptive information seeking in a flexible manner, and that confidence can function as an internal feedback cue, used to evaluate the quality of advice when explicit information about advisor accuracy is not available. In Experiment 8, we aimed to provide further support for these claims by investigating how people use confidence when explicit information about advisor accuracy is not available but external feedback is provided. If, as we hypothesize, confidence was previously used as an internal feedback proxy, we would expect that its effects would change when external feedback is provided that renders this proxy redundant.

3.4 Experiment 8

Experiment 8 extended our investigation of the flexible use of confidence to conditions in which participants had to learn about advice quality, but did not need their confidence to achieve this learning. Participants performed a perceptual decision task, similar to that used in Experiments 6 and 7. To evaluate the influence of external feedback on the use of confidence, half of the participants received trial-by-trial feedback, while the other half of the participants did not. We expected that only participants who did not receive external feedback would request advice more frequently on high confidence trials, using confidence as a feedback proxy in order to learn about advisor quality. We did not expect this relationship between confidence and advice requests when participants received external feedback, since feedback would provide information about advice accuracy, regardless of subjective confidence on a given trial. Key predictions and analyses were preregistered at http://aspredicted.org/blind.php?x=8um4u6.

Methods

Participants

88 participants were tested, in order to achieve a final sample of 40 participants in each experimental group. A larger sample was recruited compared to Experiments 6 and 7, because we expected online effects might be weaker. Four participants were excluded because their data could not be divided into appropriate confidence bins, and four additional participants excluded because their accuracy was more than 2 standard deviations below the average accuracy rate. Participants were recruited via Prolific Academic (https://www.prolific.co) and were compensated for participation. They received online instructions and agreed to take part before beginning the experiment. All procedures were approved by the Oxford University Medical Sciences Interdivisional Research Ethics Committee.

Stimuli and Apparatus

Experiment 8 was an online study, run on Prolific. All stimuli were created and presented using Javascript and CSS. Stimulus sizes were defined in pixels, and so varied according to the specific screens used by each participant.

Task and Procedure

Participants performed a perceptual judgement task, similar to the task used for Experiments 6 and 7. Several changes, described below, were made to the procedure to accommodate the change to an online platform, but which did not alter the substance of the task. Each trial began with a black fixation cross at the center of a grey screen, with two white framed boxes appearing to the left and right of the fixation. After 500 ms the fixation flashed, indicating the dot stimuli will appear briefly. The dots appeared after an additional 300 ms, and remained on the screen for 150 ms. Participants had to report which box contained more dots and how confident they are in their decision, in a single response given on a scale ranging from 100% left to 100% right. If participants selected 50%, the message 'please choose one side or the other' appeared in red under the confidence scale, and they had to update their response and select one side before they could proceed. After recording their initial response, participants chose between viewing the stimulus again or receiving advice from a virtual advisor. Two boxes representing the information types appeared above and below the center of the screen, and participants selected which information to receive by pressing the corresponding box with the left mouse button. If participants chose to receive advice, the box representing the stimulus choice disappeared, and a text box emerged from the advisor box, pointing in the direction of the advisor's response, with the text 'I think it was on the RIGHT / LEFT' appearing within. After 1000 ms the confidence scale appeared below the advice, with the advice remaining on the screen until participants submitted their final response. If participants chose to view the stimulus again, they were presented once more with a fixation for 800 ms (flashing after 500 ms), followed by the same dot stimulus for an extended period of 300 ms. The confidence scale appeared immediately after the dot stimulus disappeared.

Participants were divided into the two experimental groups (feedback vs. no-feedback) randomly. For both groups, after participants submitted their final response, the box frames appeared on the screen again for 300 ms. Participants in the feedback group received feedback after every trial, in the form of the dot stimulus appearing again in the correct box for 300 ms (the incorrect box remained empty). Participants in the no-feedback group viewed the empty boxes, without the dot stimulus, for the same duration. The next trial began after a 300 ms inter-trial interval. At the end of every block, all participants received feedback on their average accuracy on the block, and could take a break. Each participant performed five blocks of 60 trials, with advisor accuracy ranged from 60% - 100% accurate, randomly ordered across blocks. As in the previous experiments, trials were in three varying levels of difficulty, with the medium difficulty staircased using a 2-down 1-up procedure.

Before beginning the main experimental blocks, participants received online instructions and training. During the training, participants performed 2 blocks of 54 trials consisting of only the dot judgement task, with trials ending after the initial response (skipping the information choice and final response sections). All participants received feedback during these blocks, so they could learn to perform the task properly. After 2 training blocks, participants viewed additional instructions explaining the entire task and introducing the information choice. Then, participants performed a 12 trial practice block of the entire task - judging which box had more dots, choosing between viewing the information again or receiving advice, and reporting a final response. Feedback in this block was contingent on the experimental condition of each participant. At the end of this block, participants were presented with a brief reminder of how to perform the task, and asked to make sure they are in a quiet and comfortable environment. To encourage task engagement, they were informed they could earn additional bonus payment depending on their accuracy levels.

Results

Accuracy and confidence judgements

As in the previous experiments, accuracy improved between the initial and final decision, for participants in both experimental conditions (M = 81% vs. 76%, SD = 4% vs. 3%, $t(39) = 11.34$, $p < .001$, for participants in the no-feedback condition, and M = 83% vs. 76%, SD = 4% vs. 3%, $t(39) = 11.50$, $p < .001$, for the feedback condition). The increase in accuracy was reflected in confidence judgements, with participants reporting higher confidence in their final judgements (M = 78 vs. 72, SD = 9 vs. 10, $t(39) = 7.33$, $p < .001$, for the no-feedback condition, and M = 81 vs. 74, SDs = 10, $t(39) = 10.14$, $p < .001$, for the feedback condition). Thus, participants used the additional information provided to improve their accuracy, and updated their confidence accordingly.

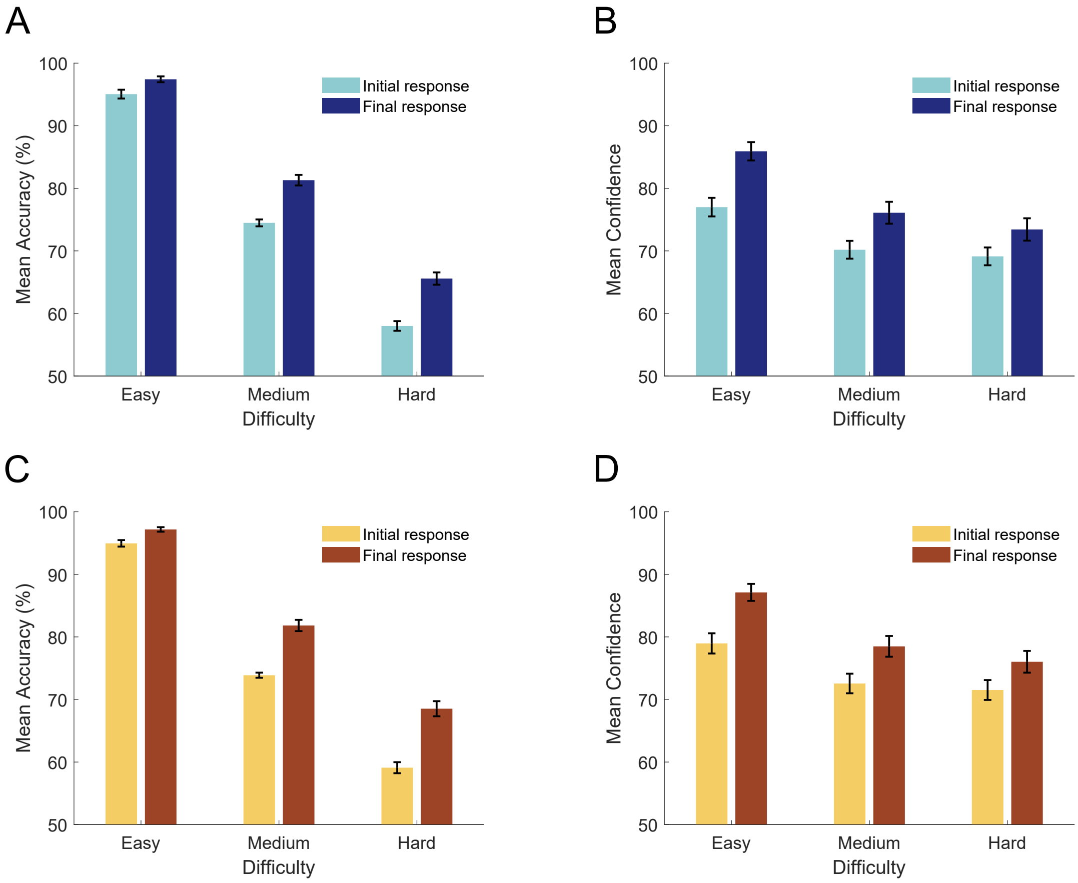

Figure 3.12: Results of Experiment 8: effects of task difficulty on accuracy and confidence according to the experimental condition. Participants who did not receive trial-by-trial feedback are depicted in blue (A-B), and participants who did receive trial-by-trial feedback are depicted in orange (C-D). (A,C) Mean accuracy on the initial and final responses divided according to trial difficulty. (B,D) Mean confidence ratings on the initial and final responses, divided according to trial difficulty. Error bars indicate standard errors of the mean (SEM).

As shown in Figure 3.12, the difficulty manipulation once again had the expected effects on accuracy and confidence, in both experimental conditions. Repeated-measures ANOVAs revealed that difficulty reliably predicted accuracy when feedback was not provided, for the initial ($F(2, 78) = 796.82$, $p < .001$, $\eta^2 = .95$) and final responses ($F(2, 78) = 570.44$, $p < .001$, $\eta^2 = .94$). Accuracy increased monotonically the easier the task, for both the initial and final responses, as demonstrated by significant linear contrasts($F(1, 39) = 1471$, $p < .001$, $\eta^2 = .97$ and $F(1, 39) = 974.99$, $p < .001$, $\eta^2 = .96$, for the initial and final responses respectively). Similar effects were found when feedback was provided, with difficulty reliably predicting accuracy, ($F(2,78) = 906.75$, $p < .001$, $\eta^2 = .96$, and $F(2,78) = 335.39$, $p < .001$, $\eta^2 = .90$), and accuracy increasing monotonically on easier trials ($F(1,39) = 1508$, $p < .001$, $\eta^2 = .97$, and $F(1,39) = 514.43$, $p < .001$, $\eta^2 = .93$), for both initial and final responses, correspondingly.

Importantly, the difficulty manipulation also led to differences in confidence judgements. In the no-feedback condition, confidence varied reliably as a function of task difficulty, for the initial ($F(2,78) = 88.52$, $p < .001$, $\eta^2 = .69$) and final responses ($F(2,78) = 99.20, p < .001, \eta^2 = 0.72$), with significant linear trends ($F(1,39) = 92.02$, $p < .001$, $\eta^2 = .70$, and $F(1,39) = 105.06$, $p < .001$, $\eta^2 = .73$, for the initial and final responses correspondingly). Similarly, when feedback was provided, difficulty predicted confidence reliably on the initial ($F(2,78) = 100.85$, $p < .001$, $\eta^2 = .72$), and final responses ($F(2,78) = 119.02$, $p < .001$, $\eta^2 = .75$), with significant linear contrasts showing that confidence increased on easier trials, for both the initial ($F(1,39) = 101.99$, $p < .001$, $\eta^2 = .72$), and final responses ($F(1,39) = 125.54$, $p < .001$, $\eta^2 = .76$). These results demonstrate that online participants performed the first and second order tasks in compliance with instructions, and that the difficulty manipulation led to a range of confidence judgements, as intended.

Learning advisor quality

Participants in both groups showed sensitivity to advisor accuracy in terms of their proportion of advice requests (Figure 3.13). A two-way mixed measures ANOVA on the proportion of advice requests with factors for advisor accuracy and feedback availability showed reliable effects for advisor ($F(4,312) = 10.08, p < .001, \eta^2 = .11$) and for the interaction effect between advisor accuracy and feedback availability ($F(4,312) = 4.93, p < .002, \eta^2 = .06$). Follow-up ANOVAs were performed to examine the influence of advisor within each group separately. For participants in the no-feedback condition, the influence of advisor accuracy on advice requests approached significance ($F(4,156) = 2.47$, $p = .058$, $\eta^2 = .06$) with a significant linear trend ($F(1,39) = 5.77$, $p = .02$, $\eta^2 = .13$), indicating that participants requested more advice from better advisors, even when feedback was not available.

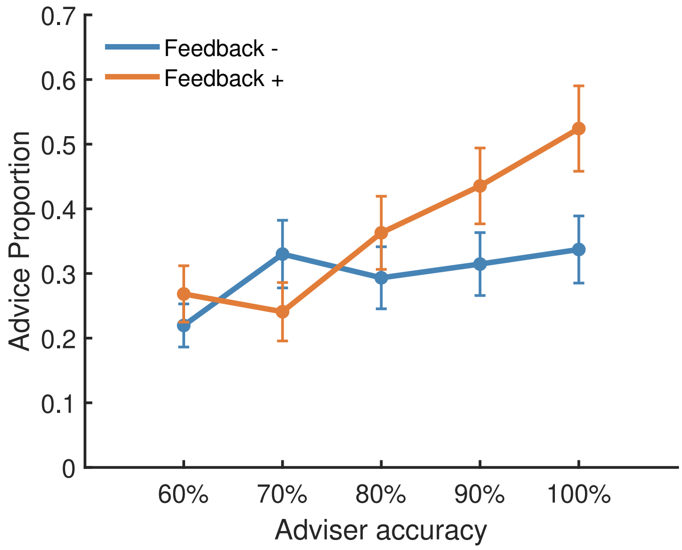

Figure 3.13: Results of Experiment 8: Average proportion of advice selections according to feedback availability and advisor accuracy. Error bars indicate standard errors of the mean (SEM).

When feedback was provided, the proportion of advice requests was reliably predicted by advisor quality ($F(4, 156) = 11.25$, $p < .001$, $\eta^2 = .22$). Participants requested more advice from better advisors, as demonstrated by a reliable linear contrast ($F(1, 39) = 31.21$, $p < .001$, $\eta^2 = .44$).

Effects of confidence on information choices depend on feedback availability

As in the previous experiments, participants displayed large individual differences in their preferences for selecting advice or viewing the stimulus again. On average, participants chose to receive advice on .30 of the trials when feedback was not available ($range = 0 - .82$, $SD = .25$), and .37 of trials when feedback was provided ($rang = 0 - .88$, $SD = .28$). Figure 3.14 shows the relationship between confidence and advice requests within each feedback condition. As predicted, a two-way mixed ANOVA on the proportion of advice selections with factors for feedback (absent, available) and confidence (Q1 - Q6) revealed a significant interaction between these two factors ($F(5, 390) = 3.36, p = .036, \eta^2 = .04$), demonstrating that confidence influenced advice-seeking differently according to the availability of feedback. No

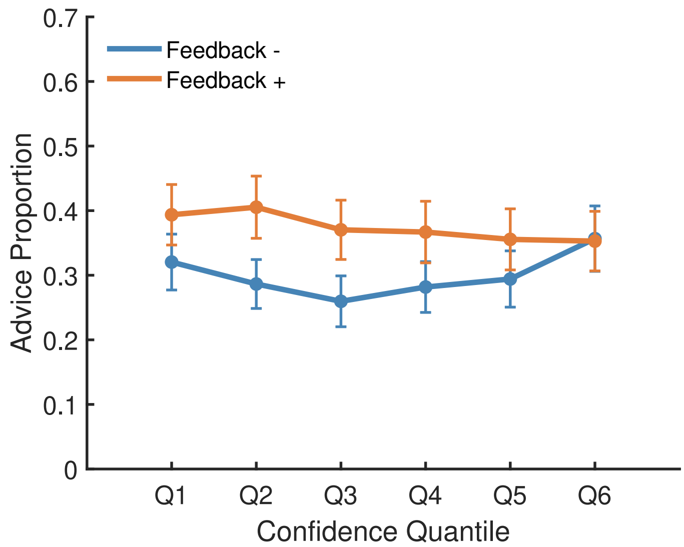

Figure 3.14: Results of Experiment 8. Mean proportion of advice choices according to external feedback availability and confidence bins.

significant main effects were found for feedback ($F(1, 78) = 1.56, p = .22, \eta^2 = 0$), or confidence ($F(5, 390) = 2.20, p = .11, \eta^2 = .03$).

To study the relationship between confidence and advice requests within each feedback condition, follow-up one-way ANOVAs were performed for each group separately. For participants who did not receive trial-by-trial feedback, we found that confidence reliably predicted advice requests ($F(5, 195) = 3.48, p = .035, \eta^2 = .08$). As shown in Figure 3.14, participants requested more advice when confidence was lowest (Q1), and highest (Q6), compared to medium confidence trials. This was captured by a significant quadratic trend ($F(1, 39) = 19.03$, $p < .001$, $\eta^2 = .33$), but pairwise comparisons did not show significant results after bonferroni corrections. Importantly, for participants receiving explicit trial-by-trial feedback, confidence did not reliably predict advice requests ($F(5, 195) = 1.82, p = .5, \eta^2 = .04$).

Next, to ensure that confidence influenced behaviour above and beyond difficulty, we performed a two-way repeated measures ANOVA studying the proportion of advice selections as a function of confidence and trial difficulty, for the feedback absent condition. For each participant, trials were divided first according to trial difficulty (easy medium, hard), and within each difficulty into four bins according to confidence quartiles (Q1-Q4, see Figure 3.15A). Results revealed a reliable interaction

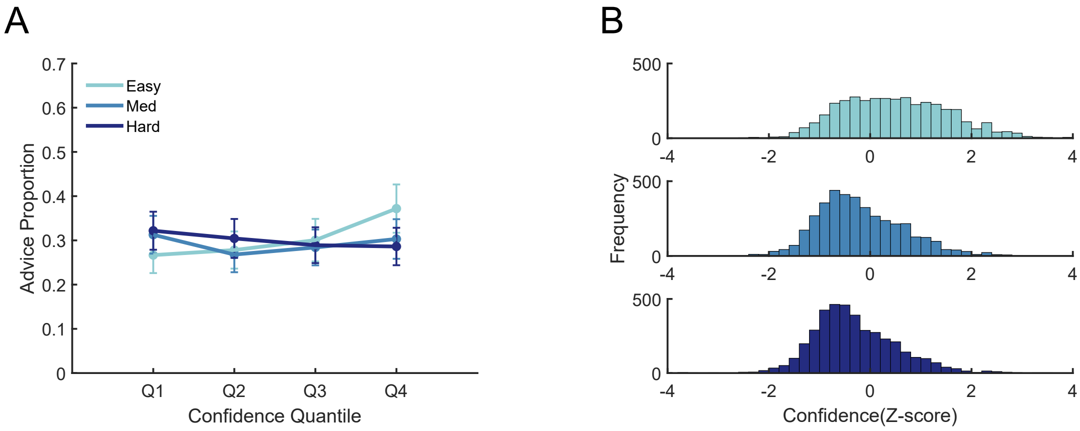

Figure 3.15: Results of Experiment 8, for the no-feedback group.(A) Average proportion of advice selections according to confidence bins and trial difficulty. For each participant, trials within each difficulty were divided into four bins according to confidence ratings. (B) Normalised initial confidence ratings, divided according to trial difficulty. Confidence ratings were first normalized within each participant, across all trials regardless of difficulty. Then, normalised confidence ratings were aggregated across all participants and grouped according to trial difficulty. Error bars represent SEMs.

effect between confidence and difficulty ($F(6, 234) = 4.56$, $p < .001$, $\eta^2 = .10$), with no significant main effects of confidence or difficulty. Follow-up one-way ANOVAs studying the effects of confidence within each difficulty level, showed a significant effect of confidence only within the easiest condition ($F(3, 117) = 4.57, p < .02, \eta^2 = .10$), but not in the medium ($F(3, 117) = 1.24, p = .30, \eta^2 = .03$) or difficult trials ($F < 1$). These results demonstrate that confidence can influence behaviour even when controlling for the effects of difficulty. Importantly, as demonstrated in Figure 3.15B, most trials in which participants reported extremely high confidence values were found in the easy difficulty. Thus, the presence of confidence effects only within this difficulty level is consistent with our hypothesis that confidence is used as an internal feedback signal to assess advisors' quality on trials where participants are very confident or even certain of the correct response.

Confidence and advice-seeking over time

Next, we examined whether confidence was used differently, depending on how much experience participants had with an advisor. Though we did not find a significant effect of confidence on information choices when feedback was available, we report the

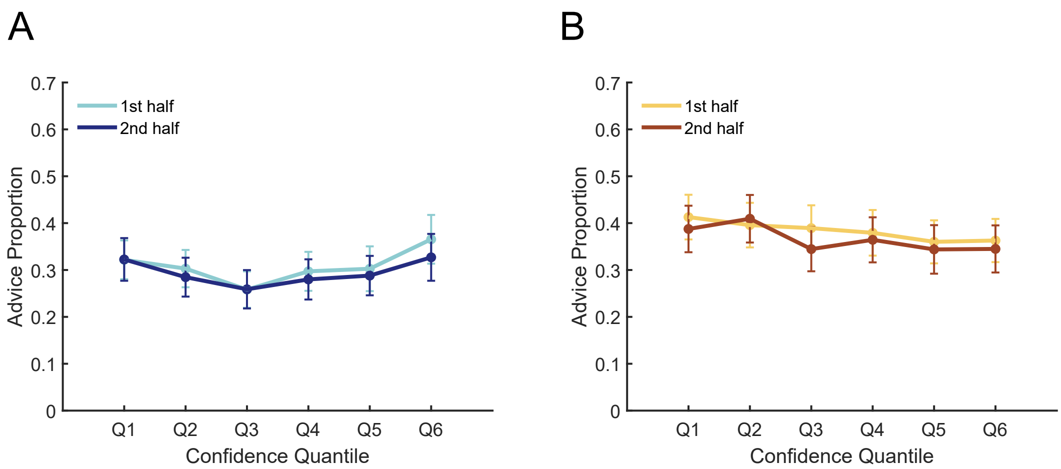

Figure 3.16: Results of Experiment 8. Average proportion of advice selections according to feedback availability, confidence and time in block. For each participant, trials were divided into two time periods (first half vs. second half of block), and then into six bins according to confidence ratings, within each time period. (A) Participants who did not receive trial-by-trial feedback, and (B) Participants who did receive trial-by-trial feedback . Error bars represent SEMs.

analysis for both feedback conditions, for completeness. Within each experimental condition, trials were divided according to time in the block (first half vs. second half), and then according to confidence bins (Q1-Q6), as shown in Figure 3.16. Two-way repeated measures ANOVAs on the proportion of advice requests, as a function of time-in-block and confidence bin were performed for each condition separately. When feedback was available, we did not find any reliable effects for time-in-block ($F(1, 39) = 1.46$, $p = .23$, $\eta^2 = .04$), confidence ($F(5, 195) = 1.91$, $p = .09$, $\eta^2 = .05$), or their interaction ($F < 1$). More surprisingly, results reveal that the interaction between time and confidence was not significant when feedback was absent ($F < 1$). The effects of time ($F(1, 39) = 1.26$, $p = .27$, $\eta^2 = .03$) and confidence were also not reliable, but confidence approached significance ($F(5, 195) = 2.85$, $p = .06$, $\eta^2 = .07$). As can be seen in Figure 3.16A, there is a trend towards participants requesting more advice on high-confidence trials, regardless of the time spent in the block. These results are consistent with the weak learning effect found in this experiment, suggesting that participants continued to use confidence to try and learn about advice quality throughout the entire block.

Discussion

Experiment 8 evaluated whether the availability of feedback modulates how people use confidence when deciding whether to request advice or re-sample stimulus information. Our aim was to extend our findings from Experiment 7 to conditions in which participants had to learn about advisor accuracy (i.e., explicit information about advice quality was not available), but did not have to rely on their confidence for learning. The results demonstrate that participants used confidence to guide information choices only when external feedback was unavailable. These results lend further support to our suggestion that, when external information is not available, confidence is used as a feedback proxy, allowing assessment of advisor accuracy.

The present results revealed weaker learning effects when external feedback was unavailable, compared with Experiments 6 and 7. Perhaps correspondingly, we also found that in this experiment participants' use of confidence was not reliably modulated by time. Instead, they continued to request advice more often when confidence was high throughout the entire block. This discrepancy might be attributed to Experiment 8 being performed on-line, such that it drew on a different (perhaps less engaged) participant sample than the in-lab studies in Experiments 6 and 7.

3.5 General Discussion

Confidence reflects a subjective estimate of the likelihood that a decision reached was a good one, and as such can play an important role in guiding adaptive behaviour. Recent studies have demonstrated the involvement of confidence in a variety of behaviours, such as learning (Guggenmos et al., 2016; Hainguerlot et al., 2018), and information seeking (Desender et al., 2018; Desender, Murphy, et al., 2019). However, these previous studies, as well as the experiments described in Chapter 2, have typically revealed consistent mappings between confidence and behaviour, whereby low confidence leads to one course of action, and high confidence to another. In the present chapter, we tested the hypothesis that the relationship between confidence and behaviour will vary strategically according to the specific goal and context, based on the notion that explicit representations of confidence and uncertainty serve crucially to guide behaviour in a flexible, goal-directed manner (Dehaene & Changeux, 2011; Nelson & Narens, 1990).

This hypothesis was tested in three experiments that examined the relationship between confidence and advice seeking. In our Judge Advisor System paradigm, participants had to choose between receiving advice and re-sampling the evidence after each initial decision. We found that, in most cases, participants' information choices were predicted by their previous confidence judgements, and, critically, that this relationship varied depending on context. Specifically, the relationship between advice seeking and confidence depended on the availability of information about advisor reliability. When participants did not have any external information about advisor accuracy, they selected advice more often when confidence was high. In contrast, when information about advisor quality was available, participants tended to request advice most often when confidence was low. These patterns of behaviour appeared in our data consistently, regardless of the type of information available about advisor quality: whether it was provided explicitly, learnable via external feedback, or learnt through exposure earlier in the block. Thus, whereas low-confidence advice requests reflect the use of confidence as a self-monitoring system—with low confidence acting as a cue that participants are unlikely to succeed in the task—high-confidence advice requests reflect the use of confidence as a feedback proxy, allowing participants to learn about advisor reliability by comparing advice to the expected answer when they are certain they know the correct answer.

Supporting the suggestion that confidence is used as a feedback proxy, we found evidence in all three experiments demonstrating that participants were able to learn about the quality of advice, and requested more advice from better advisors, even in the absence of external feedback. These results are in line with a recent study, which showed that confidence can be used to learn about the reliability of advisors when feedback is not available (Pescetelli & Yeung, 2020). Together, these results extend findings from previous studies, which have demonstrated that confidence can be used as a learning signal during perceptual tasks (Guggenmos et al., 2016), to show that confidence can also be valuable as a feedback signal when learning about social information.

The present findings complement previous research into advice taking using the Judge Advisor System. Similar to many other actions, the decision whether to seek advice can be viewed in terms of the costs and benefits associated with each line of action (Van Swol et al., 2018). Often, soliciting the help of external advisors requires the investment of time and money, particularly when advice is given by professionals or experts. Typically, the main benefit of advice is to improve performance and gains (for example when requesting advice from a banker about investments). Interestingly, there have been suggestions that advice might provide benefits in addition to improved performance, such as joint responsibility (El Zein et al., 2019; Harvey & Fischer, 1997). Here we demonstrate that advice seeking can also provide knowledge about advisor reliability. By requesting advice when cost is low (e.g. when we already know the answer), people can learn about the reliability and accuracy of others, and build up their trust in an advisor. This information can be important when deciding whether to invest in advice in the future.

Our finding that confidence is used to learn about the reliability of others resonates with the conclusions of recent studies investigating the formation of global estimates of self performance (Boldt, Schiffer, et al., 2019; Rouault et al., 2019). In these studies, the researchers demonstrated that local confidence (confidence in a single decision) is aggregated and used for building a global self-performance estimate over time. These estimates affect neural preparation for stimulus processing (Boldt, Schiffer, et al., 2019), and, importantly, allow people to decide which tasks to pursue next, based on their beliefs of how likely they are to succeed (Rouault et al., 2019). Here, we demonstrate that local confidence is used not only to form a global estimate of self-performance, but also global estimates about the performance of others. This ability to monitor and evaluate one's own performance, and compare it to the performance of others, is particularly important in social contexts, when decisions have to be made as a group, or when deciding whether to seek or use other peoples' advice. Thus, our results suggest that confidence acts as a signal that allows not only the comparison of different tasks (de Gardelle et al., 2016; de Gardelle & Mamassian, 2014) and their prioritisation, as demonstrated in Chapter 2, but also the comparison of different sources of information. Future studies might extend these findings to consider different types and sources of information beyond advisors.

Previous studies exploring advice taking have demonstrated that people use advice more often when a task is difficult (Gino & Moore, 2007), and they are unconfident (Gino et al., 2012). Based on these results, we initially predicted a U-shaped relationship between confidence and advice requests when external information about advisor accuracy was not available, a condition which we tested in all three experiments. Notably, participants consistently requested advice more often when confidence was high, but did not show a consistent effect for low confidence across the three experiments. While the data from Experiments 7 and 8 reveal a trend towards frequent advice requests when confidence was low, this pattern was not present in Experiment 6. Several methodological differences between the experiments might have caused these inconsistencies in the results, such as the testing methodology (on-line vs. in-lab testing), different advisor accuracies, and differences in the end-of-block feedback. Importantly, however, though we did not find strong evidence for this pattern of behaviour when people did not have information about advisor accuracy, our results did show that, when participants believed advice was accurate, they requested advice more often when confidence was low. These results were most evident in Experiment 7, when participants had explicit information about advisor accuracies. By demonstrating that people do not always request advice more often when their confidence is low, as we initially expected, but rather adjust their behaviour based on the specific context and goal, these results provide a degree of support to our hypothesis that confidence is used flexibly.

Seeking advice when confidence is low can be viewed as a form of cognitive offloading—a use on external resources in order to reduce cognitive demand (Risko & Gilbert, 2016). Typically, studies investigating cognitive offloading have studied this in the context of memory, looking into the use of external devices to set reminders

(Gilbert, 2015), or save information that might otherwise be forgotten (Hu et al., 2019). These previous studies have demonstrated that offloading strategies can improve memory performance and, most relevant to the present study, that people use metacognitive evaluation cues when determining whether they should rely on such external devices (Hu et al., 2019). Our study extends these findings to demonstrate that people rely on metacognitive evaluations of their performance to decide when to offload to external sources also in tasks which are not memory related. In other words, much like people would decide whether to write a shopping list based on how likely they believe they are to remember all the items otherwise, they also decide whether they should ask for advice in perceptual tasks based on how likely they think they are to succeed on their own. When people think a task is hard, they reduce the demand on themselves by using the help of others.

Critically, though the present study investigated the use of confidence for two specific purposes—as a cue that a task is hard and external assistance should be used, and as a feedback signal for learning—our main interest was not in these specific uses. Rather, the key idea which we address is peoples' ability to use confidence strategically, as demonstrated by the flexibility in the relationship between confidence and behaviour. The flexible use of confidence in our study is perhaps most evident when considering the results of all three experiments together. Results from Experiments 6 and 7 showed that, when no external information about advice was available, the relationship between confidence and advice requests changed as time progressed, with people requesting less advice when confidence was high once they had learned about the quality of advice. In Experiment 8, where participants did not learn about advice quality as effectively, participants continued requesting advice when confidence was high throughout the entire experiment. Taken together, these results show that participants changed their use of confidence strategically, in a way that advanced their current goals. These results are somewhat surprising given previous evidence that the confidence signal is integrated in the brain automatically when making decisions (Lebreton et al., 2015). Thus, it might be expected that consistent associations are learned between confidence and behaviour. For example, a learned association might be that when confidence is low external help is sought to improve performance, while when confidence is high advice is discarded. Our results demonstrate that the relationship between confidence and behaviour is more flexible than would be expected based on such associative mechanisms. An interesting question that remains open is how people are able to learn different uses of confidence and apply them to situations appropriately, particularly when situations differ only in peoples' internal state or level of knowledge, as in the experiments we describe here.

In conclusion, the results of the current chapter show that people use their confidence both as a self-monitoring tool, and as a feedback proxy allowing them to learn about the reliability of others. Together with the results of Chapter 2, these findings fit in with the concept of subjective confidence specifically, and metacognitive cues in general, acting as a tool that can be used to monitor behaviour and adjust performance accordingly (Nelson & Narens, 1990). As such, our subjective sense of confidence is a powerful tool that we carry with us constantly and use flexibly according to our current goals and needs.

4
Confidence guides attention

4.1 Introduction

The experiments reported in Chapter 3 demonstrated that people use their confidence to guide behaviour strategically, with the relationship between confidence and behaviour varying according to the current goal. While the main aim of the experiments was not to study the use of confidence for information sampling specifically, the hypothesis was tested using an information sampling task. The current chapter presents a more purposeful investigation of the link between confidence and information sampling, and aims to extend the view that representing confidence explicitly can be useful in guiding goal-directed behaviour. Specifically, my final experiment was an EEG study examining whether, in situations where different stimuli compete for processing, attention is guided by our confidence about which stimulus is task-relevant.

The amount of sensory information surrounding us constantly is immense, yet we can only process a limited amount of information at any given moment. Thus, the ability to direct our attention towards the task-relevant aspects of our environment is a ubiquitous part of adaptive behaviour. A rich body of work has investigated the mechanisms that support selective attention, particularly within the context of the visual system. Studies have shown that attention can be oriented towards various elements such as an object, feature, spatial location, or a point in time (Coull & Nobre, 1998; Desimone, 1995; Scolari et al., 2014). Additionally, it is well established that the probability that a stimulus will be selected is influenced by its physical salience, for example its intensity, and by cognitive factors, such as the current goals of the individual (Corbetta & Shulman, 2002).

To facilitate the processing of information in a goal-directed way, there is need for an existing representation of the relevant information. In other words, we must have knowledge of which information is useful for achieving our goals. When studying attention in the lab, participants are typically provided with explicit instructions regarding the task-relevant information. For example, participants might be instructed to locate a circle among diamonds. Under such conditions, it is clear that to identify the target, attention should be focussed on the shape of stimuli, and not other properties such as their colour. In natural settings, however, there is often much more uncertainty regarding which information is useful. Furthermore, some stimuli might be relevant in one context, but not in another, making it even harder to determine what to attend.

Several lines of evidence have demonstrated that people are able to learn what information is relevant through experience. For example, studies into choice behaviour have shown that, in complex environments in which options are defined by several dimensions (Gershman et al., 2010) or cues (Akaishi et al., 2016), people learn what information is relevant for obtaining rewards, and preferentially direct their attention towards this information. Of particular relevance to the present investigation, research from the tradition of associative learning has shown that attention can be conditioned in a similar way to actions, so that learning about the relationships between cues and outcomes leads to changes in attention. For example, there is substantial evidence that previously predictive stimuli are learned about faster, and fixated on for longer, than previously non-predictive stimuli, even when they currently provide equally reliable information about outcomes (Le Pelley et al., 2011; Le Pelley and McLaren, 2003; for a review on attention and associative learning see Le Pelley et al., 2016).

These studies provide evidence that previous experience leads to attentional biases towards predictive stimuli. One might ask then, whether these biases are solely driven by automatic and involuntary capture of attention, or might also be influenced by goal-directed voluntary control. Recent studies provide evidence that suggests an influence that is both top-down, and bottom-up in nature (Theeuwes, 2019). For example, stimuli and locations that have previously been associated with reward have been shown to capture attention in an automatic way, so that they are prioritised when they are no longer relevant (Anderson, 2013), and even when their selection leads to a penalty (Camara et al., 2013). On the other hand, there is evidence that attention deployment can be controlled in a more flexible way, with the effects of associative learning modulated by context and current beliefs. For example, it has been shown that learned predictiveness can be context-dependant, with previously predictive stimuli capturing attention only when encountered in the same specific context (Uengoer et al., 2013; Uengoer et al., 2018). Similarly, explicit instructions about the current reliability of a previously predictive cue can modulate the degree of attention allocated to it (Mitchell et al., 2012).

The present study aimed to extend previous findings showing that learned attention can be flexibly controlled based on our beliefs, by demonstrating that our subjective confidence in the relevance of information modulates the degree of attention allocated to it. The logic behind this suggestion is simple — the more confident we are that information is relevant, the more attention we should allocate to it (and, similarly, the more certain we are that information is not relevant, the less attention we should direct towards it). Thus, we suggest that confidence guides not only what information we attend, but also the strength we attend at.

To illustrate, imagine you are learning how to play a new instrument, such as the flute. Early on, you are struck by how unstable your tone is, but do not know the cause. You adjust various aspects of your playing, resulting in a nicer tone at times, but making it even worse at others. As time goes by, you begin to notice that certain qualities of your tone are associated with specific aspects of your playing. For example, an airy tone can usually be improved by adjusting the shape of your mouth, while a flat tone seems more related to the strength of your blowing. For a while you are still unsure of these relationships, and when hearing an airy tone you might focus slightly more on the shape of your mouth, but also consider the strength of your blowing. But, with experience, your confidence in the associations between tone and the relevant aspect of your playing grows, and by increasingly focusing your attention on the relevant information you are able to produce the rich and dynamic tone you are aiming for.

The example above demonstrates how confidence in our beliefs can be used to control the allocation of attention to competing stimuli in a graded way. Importantly, although prioritising task-relevant information has an obvious advantage over trying to process all stimuli equally, focussing on information that turns out to be irrelevant can be misleading and damage performance (e.g., if you keep trying to adjust the shape of your mouth, while the actual problem is your airflow). Thus, a system that continuously monitors our uncertainty regarding the relevance of information, and adjusts attention in a graded fashion, can be crucial for reaching the right balance between too much and too little attention selectivity (as opposed to a system that selects the information which is believed to be most relevant, and focuses on it exclusively). Following this rationale, we hypothesized that selective attention, i.e. the difference in attention resources allocated towards relevant vs. irrelevant stimuli, will be modulated by confidence in the task-relevance of a stimulus.

We tested this prediction using a perceptual decision task in which participants encountered both relevant and irrelevant stimuli, and could learn cue-stimuli associations to guide attention towards the task-relevant information. The perceptual task was a dot task in which participants had to report which of two boxes contained more dots overall, similar to the tasks used in Chapters 2 and 3. However, in the present experiment each box contained dots of two different colours. Importantly, only dots of one colour were informative in predicting the correct response on each trial. At the start of every trial, before the dot stimulus appeared, participants viewed one of two cues, indicating which colour would be relevant on this trial. They had to report which colour they believed the cue represented, and rate their confidence in this belief. By learning the associations between cues and colours, and then focusing attention on the relevant colour, participants could improve their performance in the dot task. Feedback was provided on performance in the dot task, but not on the cue-colour association judgement. This enabled us to create a situation in which, similar to learning how to play a new instrument, the relevance of information in a given context can be inferred from success in the task, and confidence in this association evolves with experience. We predicted that confidence in the task-relevance of information would modulate the degree of attention paid to it, over-and-above the basic progression of associative learning. Specifically, we expected that higher confidence in the relevance of the informative stimulus would be associated with an increase in attention to it, and a decrease in attention to the irrelevant stimulus.

To provide a neural index of attention allocation, in addition to behavioural measures of response times and accuracy, the relevant and irrelevant dots flickered at different rates. This flickering yielded corresponding steady state visually evoked potentials (SSVEP) in the scalp-recorded electroencephalogram (EEG) — brain responses at the visual cortical areas that have the same fundamental frequency as the driving stimulus (Norcia et al., 2015). Crucially, studies have shown that SSVEP amplitudes are modulated by selective attention (Andersen & Müller, 2010; Morgan et al., 1996). Thus, by studying the relationship between confidence ratings on cue-colour associations and SSVEP responses to the relevant and irrelevant dot stimuli, we could track how attention was modulated by confidence in the relevance of each colour. To anticipate, we found that attention to the relevant, but not the irrelevant stimulus, varied according to confidence.

4.2 Experiment 9

Methods

Participants

Twenty-eight participants took part in the experiment. Two participants who did not complete the full experiment due to time restraints, and one participant for whom data was missing due to technical problems, were excluded prior to analysis. The final sample consisted of 25 participants (15 females, age range = 18-35 years, M = 23, SD = 5.2). All participants had normal or corrected-to-normal vision. They provided written informed consent, and were compensated with course credit or payment for their participation. All procedures were approved by the Oxford University Medical Sciences Interdivisional Research Ethics Committee.

Task and Procedure

Our experiment combined an associative learning task and a perceptual judgement task (see Figure 4.1). The perceptual task was based on the dot task described in Chapters 2 and 3, with several modifications made to enable tracking the allocation of attention to different stimuli. As in the previous experiments, participants were instructed that their main task was to judge which of two boxes presented briefly contained more dots overall. However, in the current experiment, the boxes contained dots in two colours (blue and orange). Critically, on each trial, one of the colours was more informative of the correct box (the relevant colour), with a larger difference in dot numbers for dots of this colour. The difference in the number of dots of the other colour (the irrelevant colour) was smaller, and on average non-informative of the correct box, with the correct and incorrect boxes containing more dots of this colour with equal probability across trials. For example, on a trial in which the blue dots were relevant, the dot distributions could be 112 blue + 98 orange dots in the correct-box (220 total), and 88 blue + 102 orange in the incorrect box (190 total). Thus, although the task could be performed correctly without selective attention to one or other colour (because the correct box always had more dots overall than the incorrect box), focusing attention on dots of the informative colour (in this case blue) made the task easier. At the start of each trial, participants were presented with one of two cues, indicating which colour would be relevant on this trial, but they were not told the cue-colour mapping. However, by learning the mapping between cues and colours across the 16 trials

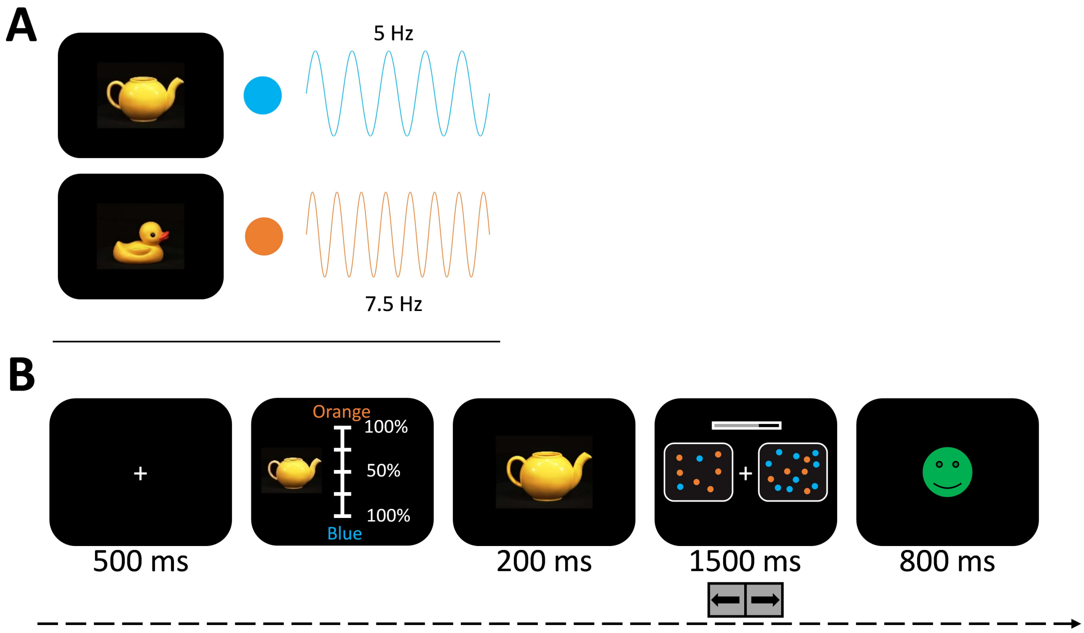

Figure 4.1: Experimental design. (A) Blue and orange dots were associated with different cues on each block. To track attention, dots of each colour flickered at 5Hz and 7.5Hz, eliciting SSVEP responses in the corresponding frequencies. (B) Trials began with a cue associated with the task-relevant colour for that trial. After rating their confidence on which colour this cue represented, the cue flashed on the screen again briefly, before the dot stimuli appeared. Participants then had to judge which of two squares containing blue and orange dots contained more dots overall. On each trial, one colour was informative of the correct response, with the correct box necessarily containing more dots of this colour.

of each block, participants could guide their attention to the relevant colour, and improve their performance in the dot task.

Each trial began with the cue (an image of an object), and a vertical confidence scale appearing at the centre of the screen, with the question 'How confident are you about which colour this cue represents?' written above. The confidence scale ranged from 50% (guess) at the centre to 100% blue and 100% orange at the opposite ends. Participants responded by adjusting the location of a slider on the confidence scale using the mouse, and pressing the left button to submit their response. The slider's starting position was randomised on every trial, so as to prevent a bias toward selecting a specific confidence rating. After a response was made (no time limit was enforced), the cue and confidence scale disappeared and a fixation cross appeared at the centre of the screen for 500 ms, followed by the cue displayed again briefly at the centre of the screen for 200 ms. After another 500 ms fixation period, the perceptual task was presented. Two boxes containing orange and blue dots appeared on the screen for 1500 ms, during which participants had to report which box contained more dots overall by pressing the corresponding mouse button (e.g. the right button if the right box contained more dots). A bar at the top of the screen indicated time progression, with progression stopping once a response was made. Regardless of response time, the dot stimulus was presented for 1500 ms, after which the boxes disappeared and only the fixation cross remained on the screen for 1000 ms. Following this, feedback on performance in the dot task was presented for 800 ms in the form of a happy, sad or neutral cartoon face, indicating if the participant was correct, incorrect, or did not respond in time. The next trial began after a 500 ms intertrial interval (see Figure 4.1). Participants could use the feedback on the dot task to learn the associations between cues and colours. For example, if a participant was sure that the box containing more dots overall contained more blue dots compared to the other box, but the reverse relationship existed for orange dots (as in the sample trial in Figure 4.1), they could learn that the cue presented at the start of the trial represented blue.

The experiment consisted of 30 blocks of 16 trials each. For each block, two new images were sampled randomly from a pool of 60 images selected from the Hemera Photo-Objects Collections (Hemera Photo Objects, Hull, Quebec, Canada). One image was associated with the colour orange, and the other image with the colour blue. The image associated with the trial-relevant colour was presented at the start of each trial, cueing the participant which colour of dots they should be paying attention to. To track attention deployment, relevant and irrelevant dots were flickered at one of two unique frequencies (5 Hz and 7.5 Hz). This flicker elicited distinguishable SSVEP responses in the corresponding frequencies, allowing us to study the degree of attention allocated to each stimulus (Norcia et al., 2015). The flicker rate for dots of each colour was consistent within a block, but varied between blocks, with each colour being tagged in each frequency on half the blocks. The relevant colour and the correct box on each trial were randomised independently, with each colour being relevant, and each box correct, on half of the trials in each block.

The number of dots in each box was determined by the equation $nDots = 200 \pm d_{tot}$, where d_{tot} is the total number of dots added and subtracted to the correct and incorrect boxes correspondingly across both colours, and the default number of dots of each colour is 100, for a total of 200 dots per box. To achieve similar performance levels of approximately 70% accuracy across participants, the d parameter was adjusted for each participant using a 2-down 1-up staircasing procedure during the training period. The total difference in dot numbers was stable throughout the experiment, but the division of this difference between the colours was determined on a trial by trial basis. The difference in dots for the relevant color d_{rel} was sampled on each trial from a normal distribution, with a mean of d_{tot} and a variance of $d_{tot}/5$, The difference in dot numbers for the irrelevant colour was then set to $d_{irrel} = d_{tot} - d_{rel}$. Thus, on average the difference in dot numbers for the irrelevant colour was centred around zero, and could be congruent with the total difference (i.e., more dots of this colour in the box with more dots overall), incongruent (i.e., more dots of this colour in the incorrect box), or neutral (i.e., with equal numbers of dots of this colour in each box).

To ensure there was an actual performance benefit to learning cue-colour associations, a pilot experiment compared dot-task performance under conditions in which participants were explicitly instructed which colour was relevant and asked to attend to it, to conditions in which they had no information about the relevant colour. The results showed that participants performed better when attending the relevant colour, confirming the suitability of our paradigm and the chosen parameters.

Since there is evidence that differences in physical salience can lead to differences in associative strength (Denton & Kruschke, 2006), we wanted to control for any stimuli-driven learning effects that could cause differences in attention allocation. To this end, pairs of blocks were matched in terms of cue sequence, the correct box order and the dot stimuli (i.e., the exact number and locations of dots presented on each trial). Comparisons between these matched blocks therefore provided a method to study effects of confidence on attention, over and above any effects that were driven more directly by stimulus properties. Block pairs were predetermined by a pseudo-randomisation procedure that ensured the first two blocks were not paired. Note that different cues were used, and the order of the cued colour reversed, between matched blocks (e.g. the sequence blue-blue-orange would be matched with orange-blue-blue), to lessen the likelihood that participants might somehow notice the matching procedure.

Before beginning the main experiment, participants performed three practice blocks that gradually introduced task elements. The first practice block consisted of 42 trials of the dot judgement task only. As in the main experiment, the boxes contained dots in two colours, but participants were only instructed to judge which box had more dots overall, with no further explanation about the different colours. Before beginning the second practice block, participants were presented with instructions explaining that on each trial one colour of dots was more informative than the other, so that tracking this colour specifically can make the task easier. They then performed another 42 trials of the dot task, during which they were cued to which colour would be relevant, much like in the main Experiment. However, the cue was a circle in the colour of the relevant dots, rather than an image, so that participants did not have to learn the associations between cues and colours. The instructions prior to the third practice block introduced the cues in image-form, and the confidence scale. Participants then performed a 16-trial practice block of the full task. To encourage task engagement, participants received a bonus payment based on dot-task accuracy during the main experimental blocks, and were informed of their dot-task accuracy at the end of each block.

Apparatus and stimuli

The experiment took place in a dimly lit, electrically-shielded booth. Participants sat at a distance of approximately 60 cm from a 22 inch screen with a 1680 x 1050 resolution and 60 Hz refresh rate. All stimuli were created and presented in MATLAB using Psychtoolbox3 extension (Brainard, 1997; Kleiner et al., 2007;

Pelli, 1997). The cues were presented at the centre of the screen, subtending a visual angle of 5°. Dot stimuli were comprised of two white frames, 6° x 6° large, containing $200 \pm d_{tot}$ orange and blue dots, each with a diameter of 0.5°, flickering at rates of 5 and 7.5 Hz.

EEG acquisition and pre-processing

EEG data were recorded using a BIOSEMI Active-Two system, with Ag-AgCl electrodes embedded in a fabric cap. The EEG was recorded from 64 channels: Fp1, AF7, AF3, F1, F3, F5, F7, FT7, FC5, FC3, FC1, C1, C3, C5, TP7, CP5, CP3, CP1, P1, P3, P5, P7, P9, PO7, PO3, O1, Iz, I1, I2, Oz, POz, Pz, CPz, Fpz, Fp2, AF8, AF4, AFz, Fz, F2, F4, F6, F8, FT8, FC6, FC4, FC2, FCz, Cz, C2, C4, C6, TP8, CP6, CP4, CP2, P2, P4, P6, P8, P10, PO8, PO4, O2, as well as the left and right mastoid. Horizontal and vertical electro-oculogram (EOG) were recorded using electrodes positioned above and below the left eye, and on the outer canthi of both eyes. Reference electrodes were positioned on the mastoid bone behind the right and left ears. The Common Mode Sense (CMS) and Driven Right Leg (DRL) electrodes were used as reference and ground electrodes. All electrodes were re-referenced to the mastoids offline. Data were continuously recorded at a sampling rate of 2048 Hz, and were re-sampled offline to 256 Hz.

Preprocessing of EEG data was carried out using custom Matlab scripts, and EEGLab (Delorme & Makeig, 2004) and Chronux toolboxes (Bokil et al., 2010). EEG signals were subjected to a Hamming-windowed finite impulse response filter (0.1 - 20 Hz). Stimulus-locked epochs were extracted beginning 1000 ms prior to stimulus onset, until 2500 ms post stimulus onset (i.e., 1000 ms after the dot stimuli disappeared). Epochs were baseline-corrected by subtracting the average activity of each channel during a period of -100 to 0 ms pre-stimulus. Trials with large artefacts were rejected by visual inspection, with an average of 16.7 trials rejected per participant. Eyeblinks and lateral eye movement artefacts were corrected by conducting an independent components analyses (ICA), and removing components based on visual inspection using SASICA (Chaumon et al., 2015). Remaining noisy channels were rejected based on the kurtosis criterion (threshold 3) in EEGLAB's channel rejection routine, and interpolated using spherical interpolation. On average, 5.2 channels were interpolated per participant.

EEG data analysis

To estimate the topography of SSVEPs and determine the appropriate electrode cluster for analysis, we first computed the power spectrum via fast Fourier transform (FFT). We used the entire stimulus presentation time, defined as a period of 0 to 1500 relative to stimulus onset. We then computed the signal-to-noise ratio (SNR) on the log of f, with noise defined as the average log amplitude in frequencies adjacent to the tagged frequencies (Norcia et al., 2015). The neighbouring frequencies were defined as $f \pm 2.125$ Hz. Figure 4.2A shows the iso-contour voltage maps of the log(SNR) amplitude for 5 Hz and 7.5 Hz. Based on these topographies, we selected a cluster consisting of channels POz, Oz, and Iz, and conducted all further analyses using the averaged data from these channels (see Figure 4.2B for the average ERP for the selected channels). Figure 4.2D depicts the average log(SNR) power spectrum, for data averaged from these channels, with clear peaks for 5 Hz and 7.5 Hz. To further inspect the time course of the amplitudes for different frequencies, we also computed the time-frequency spectrogram over our whole epochs (-1000 to 2000 relative to stimulus onset), with a moving window length of 1000 ms, and step size of 250 ms (half-bandwidth = 0.5 Hz). Responses at each frequency were baseline corrected by subtracting the average power of that frequency during a period of 1000 ms before the stimulus appeared. As expected, responses at 5 Hz and 7.5 Hz show an increase in power during stimulus presentation time, peaking at around 250-500 ms after stimulus onset (see Figure 4.2C).

Because we were interested in examining time-varying changes to SSVEP amplitudes, we also applied a narrow pass-band Gaussian filter to single-trial data. This method performs an FFT-based convolution, and can provide higher temporal resolution than other methods. Each Gaussian filter was centred at one of our tagging frequencies, with a frequency resolution of ± 1.2 Hz FWHM, resulting in

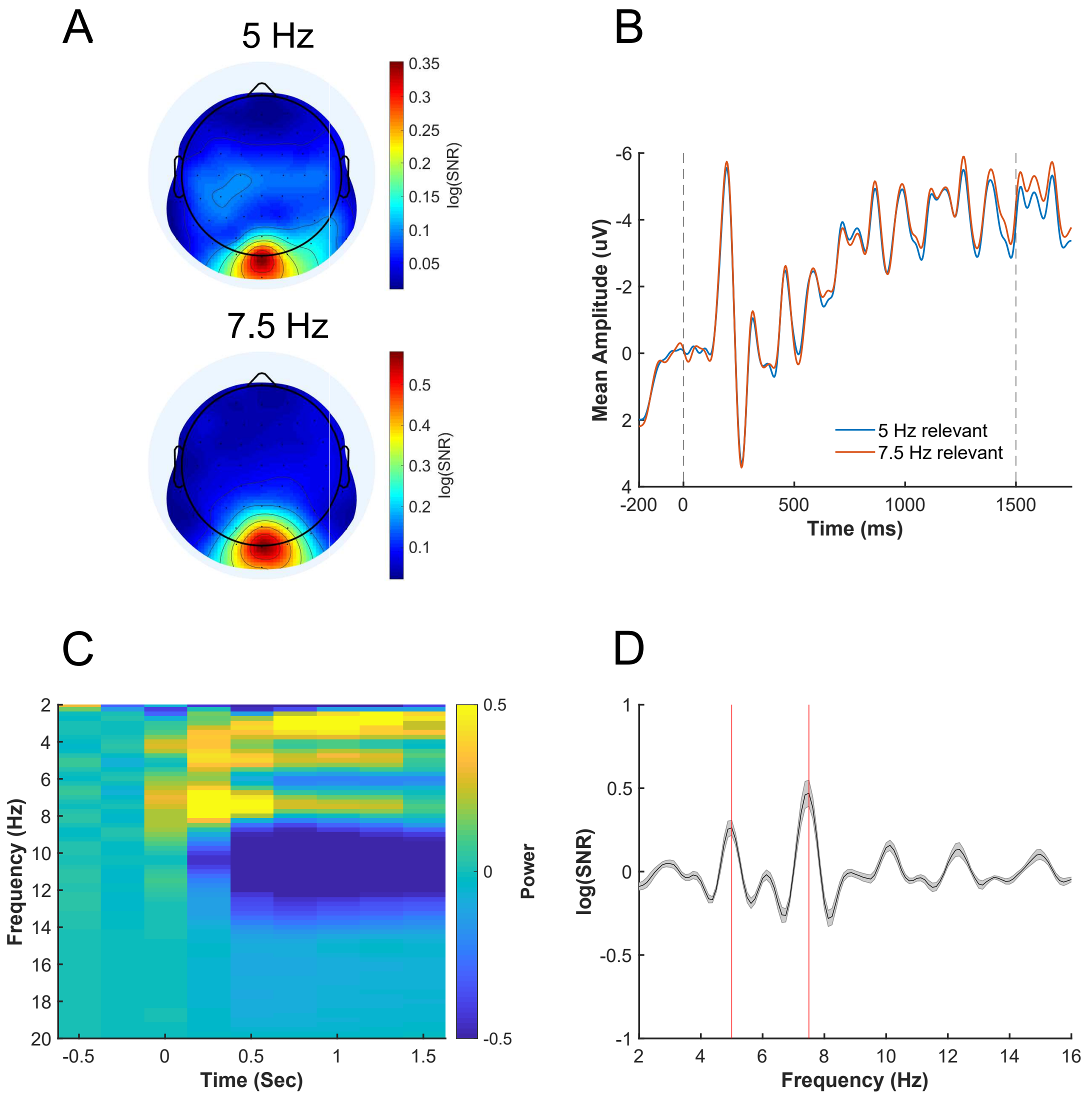

Figure 4.2: EEG data: A: Spatial distributions of log(SNR) across participants for the tagged frequencies (5 Hz, 7.5 Hz). Responses revealed the typical occipital distribution of the SSVEP signal. B: Average ERP responses across all participants, here plotted separately according to the flicker frequency of the relevant dot colour on the trial (data averaged from channels POz, Oz and Iz). Dashed lines represent the stimulus onset and offset. C: Time course of amplitudes for frequencies between 2-20 Hz. Responses at the tagged frequencies (5 Hz and 7.5 Hz) peak during stimulus presentation time. D: log(SNR) over all participants with target frequencies marked in red, averaged over Oz, POz and Iz. Shaded areas represent 1 SEM.

a time resolution of ±183.9 ms. To isolate the stimulus-locked activity from other endogenous brain rhythms, we then projected the obtained complex amplitudes from each trial onto the mean phase across all trials, for each frequency separately. This procedure allowed us to quantify single-trial activity that contributed to the mean phase-locked activity across all trials, thus reducing the contribution of single-trial non-phase-locked, and spontaneous activity at our stimulation frequencies (Andersen et al., 2008; Steinhauser & Andersen, 2019). To normalise the changes in amplitude, and allow collapsing across the two tagged frequencies, we divided the amplitude for each sample relative to a baseline amplitude computed for each frequency. As baseline amplitudes we used the mean amplitude for each frequency over a period of -700 to -200 ms pre-stimulus onset.

To quantify the effects of attention, mean SSVEP amplitudes, averaged over the entire 1500 ms of stimulus presentation time, were computed for each trial, for each frequency separately. This allowed separate measurement of the 5 Hz and 7.5 Hz SSVEPs according to whether they corresponded to the flicker frequency of the relevant or irrelevant dot colour on the trial. For analyses in which we collapsed data across the two target frequencies, we used a two-stage averaging procedure. First, we averaged data for each condition, for each frequency separately. Then, we averaged data across frequencies, separately according to whether they were relevant vs. irrelevant. This procedure ensured that data were not biased due to differences in the SSVEP responses between the frequencies. In addition to the main, stimulus-locked analyses, we also analysed the SSVEP time-course data when time-locked to the response, and found similar results to those reported below.

Results

Behavioural data

Behavioural analysis focused on two main goals. First, we wanted to assess the effectiveness of our manipulation, in terms of ensuring that our participants learned associations between cues and colours, and verifying that learning these associations had an impact on performance in the dot task. Second, we wanted to establish whether attention to the relevant information increased gradually as a function of confidence in its relevance, beyond the binary classification of a colour as relevant vs. irrelevant. Our working assumption was that enhanced attention to dots of the relevant colour would lead to improved performance in the dot task.

Thus, if attention to a stimulus is modulated by confidence in its relevance, as we hypothesized, we expected to find a positive relationship between confidence in the correct cue-colour association and accuracy on the dot task.

Learning cue-colour associations and task performance. To examine learning, we calculated the proportion of correct responses and average confidence in the cue-colour associations across the 16 trials of each block, averaging across all 30 blocks. Correct cue responses were defined as trials on which participants reported confidence of above 50% in the correct colour associated with the cue. Figure 4.3 shows the progression of accuracy and confidence in the cue association and dot-task accuracy by trial position, averaged across participants. As expected, participants typically reported guessing on the first trial ($M_{conf} = 51.7$), resulting in % correct cue responses close to zero ($M_{acc} = 13\%$). As trials progressed, there is a clear increase in cue accuracy and confidence, with accuracy reaching an average of 85% and confidence an average rating of 91 on the last trial. Thus, participants successfully learned the cue-colour associations, and reported confidence ratings that reflected this learning. Importantly, in addition to an increase in cue accuracy and confidence, performance in the dot task also improved as trials progressed, with average dot-task accuracy increasing from 76% on the first trial, to 82% on the last.

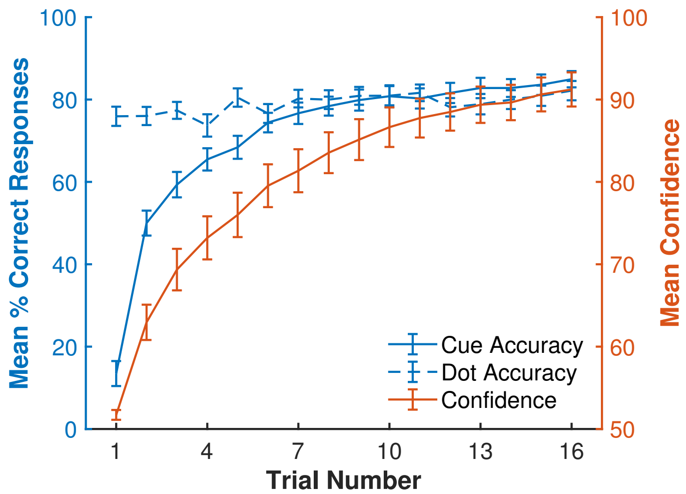

Figure 4.3: Mean cue accuracy and dot-task accuracy (blue), and confidence (orange) across trials. Error bars indicate standard errors of the mean (SEM).

To further study the relationship between learning the correct cue-colour association and attention to the dot stimuli, we examined accuracy on the dot task as a function of cue accuracy (i.e., whether participants correctly identified the colour-association for the presented cue) and differences in dot numbers for the relevant vs. irrelevant colour. Due to the method by which we sampled dot differences, the difference in dot numbers for the irrelevant colour was centred around zero and anticorrelated with the difference in the relevant colour (with the total dot difference between the boxes stable across trials). Thus, we divided trials via median split according to whether dots of the irrelevant colour favoured the incorrect response (i.e., where there was a larger than median difference for the relevant colour) or favoured the correct response (i.e., where there a smaller than median difference for the relevant colour). Of interest was whether dot-task performance depended on this division between trials, as an index of the selectivity of participants' attention to the two colours.

We subjected the mean % accuracy in the dot task to a two-way repeated-measures ANOVA, with factors for cue accuracy (correct, error, guess), and relevant dot-difference (larger vs. smaller than median). Results showed that accuracy in the dot-task varied reliably as a function of cue accuracy ($F(1, 23) = 45.55, p < .001, \eta^2 = .66$), with the interaction between the relevant dot-difference and cue accuracy also significant ($F(2, 46) = 4.84, p = .02, \eta^2 = .17$). As shown in Figure 4.4, performance in the dot task was best on trials in which participants correctly identified the relevant colour, and lowest when participants incorrectly identified the cue as representing the irrelevant colour. Furthermore, pairwise comparisons with Bonferroni corrections revealed that a large difference in dot numbers of the relevant colour was associated with reliably better performance on correct cue trials ($t(24) = 5.42, p < .001$). In contrast, the reverse relationship was found for cue error trials, with accuracy decreasing on trials with a large difference in the relevant colour ($t(24) = 2.39, p = .035$). As expected, we did not find a reliable relationship between dot difference for the relevant colour and performance on guess trials ($t < 1$), where participants are likely to be dividing attention between dots of both colours equally.

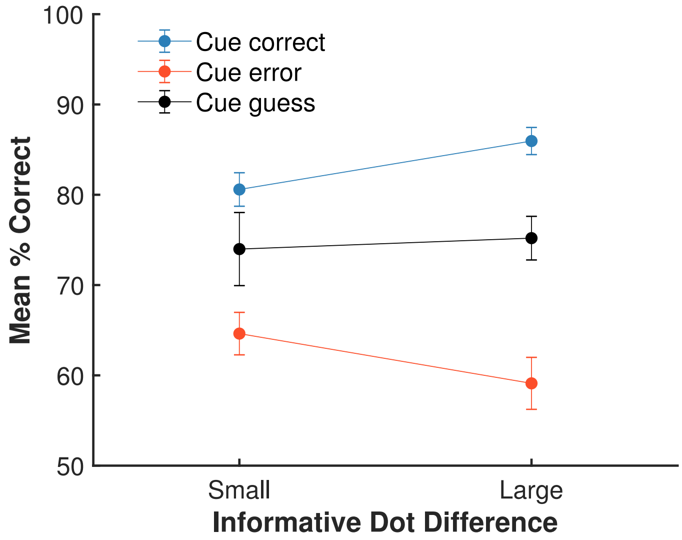

Figure 4.4: Mean accuracy in the dot-task according to cue accuracy and relevant dot differences (smaller and larger than median). Note that a larger than median difference in dot numbers for the relevant colour is also characterised by more dots of the irrelevant colour in the incorrect box. Error bars represent SEMs.

These results demonstrate that participants were performing the task as expected, and directing attention to the dots of the colour they believed was relevant. Thus, learning the correct cue-colour associations led to improved performance in the dot task, particularly in trials where focusing on the relevant colour was more helpful (i.e., when the relevant difference was large, and the irrelevant colour misleading).

Confidence effects on behavioural performance. Having established that learning the correct cue-colour associations had an influence on task performance, of critical interest was whether attention to the relevant information increased gradually as a function of confidence in its relevance (i.e. beyond its binary classification as relevant vs. irrelevant). To test this, we computed a relative-confidence measure, for which confidence ratings were converted based on cue-accuracy, so that $Conf_{rel} = CueAcc * (Conf_{orig} - 50)$, where $CueAcc = \pm 1$ for correct and error responses respectively. Thus, relative confidence ranged from -50 (highly confident that the cue represents the colour that is objectively irrelevant) to +50 (highly confident the cue represents the colour that is objectively relevant).

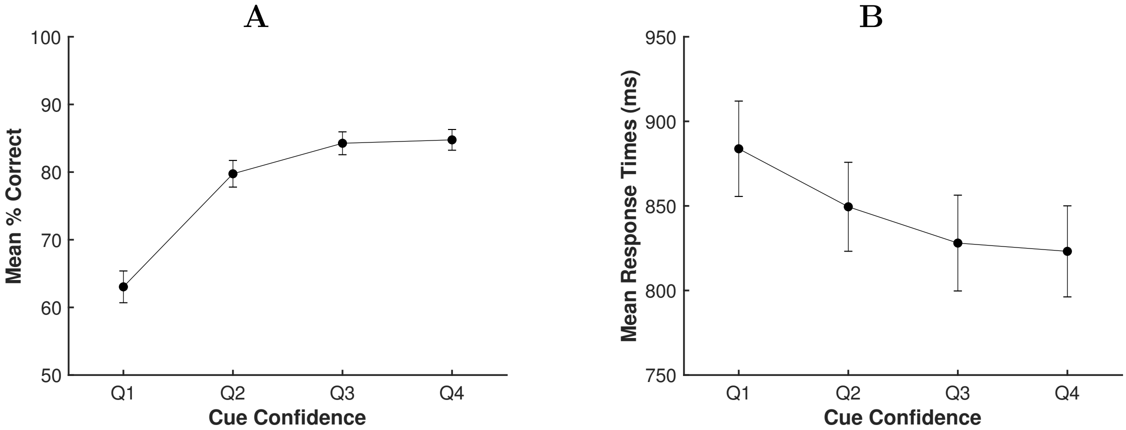

Figure 4.5: Effects of confidence on accuracy and response times. (A) Mean accuracy on the dot task, divided according to relative cue confidence. (B) Mean dot-task response times, divided according to relative cue confidence. Error bars represent standard errors of the mean (SEM).

Trials were divided into four bins according to relative confidence ratings, and the average accuracy and response time on the dot task were computed for each bin. A repeated-measures ANOVA on dot-task accuracy revealed that confidence reliably predicted accuracy ($F(3, 72) = 74.84, p < .001, \eta^2 = .76$), with accuracy increasing monotonically with confidence (see Figure 4.5a). Similarly, we also found a significant relationship between relative confidence and response times on the dot-task ($F(3, 72) = 15.58, p = .001, \eta^2 = .39$), with faster responses on high-confidence trials.

These results show that people respond to the dot task more accurately, and faster, the more confident they are in the correct cue-colour association. However, several additional factors might have contributed to this relationship. Firstly, it remains possible that the effect we found is driven simply by whether participants correctly vs. incorrectly interpret the cue-colour association. Thus, further analysis is needed to link dot-task performance to graded changes in confidence, over and above binary accuracy. Secondly, it could be that factors related directly to associative strength between cues and colours, such as the overall progression of learning and the experienced cue sequence, might have also contributed to the relationship. For example, as can be seen in Figure 4.3, confidence increases over time, as participants have been exposed more to the colour-cue associations, making it important to show the unique contribution of subjective confidence, over and above other (correlated) changes as participants gain more experience with the cue-colour associations in the current block. Thus, we performed several additional analyses aimed at establishing the influence of confidence on dot-task performance, while controlling for the effects of these alternative factors.

Logistic regression models predicting dot-accuracy as a function of cue confidence, trial position and relevant dot-difference, were fitted for each participant separately. Importantly, cue-error trials were excluded from this analysis, allowing us to assess the influence of confidence above and beyond accuracy. Meanwhile, inclusion of a regressor for trial position allows us to assess variance specifically related to confidence, while controlling for the gradual progression of learning throughout the block. This analysis revealed a positive value for the confidence regressor in 19 participants, with the mean beta value significantly different to zero, indicating that participants were more likely to respond to the dot task correctly on trials in which they were more confident in the task-relevant colour ($M = .009, SD = .02, t(24) = 2.22, p = .03$), even while adjusting for the effects of trial position (10 positive betas, ($M = -.02, SD = 06, t(24) = -1.49, p = .15$) and relevant dot difference (19 positive betas, $M = .10, SD = .14, t(24) = 3.83, p < .001$). Similarly, we fitted linear regression models predicting response times as a function of cue confidence, trial position and dot difference, for each participant. Here too, we found that confidence influenced performance, with negative beta coefficients observed in 19 participants ($M = -0.92, SD = 1.67, t(24) = -2.74, p = .01$), even after adjusting for the effects of trial position (11 negative betas, $M = 1.01, SD = 4.41, t(24) = 1.15, p = .26$) and relevant dot difference (22 negative betas, $M = -7.59, SD = 7.67, t(24) = -4.95, p < .001$). Thus, response times were shorter on trials where participants were more confident about the informativeness of the colour.

Our final behavioural analysis aimed to control further for the effects of stimulus-driven associative strength, making use of the inclusion in our design of matched blocks, which had matched sequences of cues, correct responses, and stimulus dot

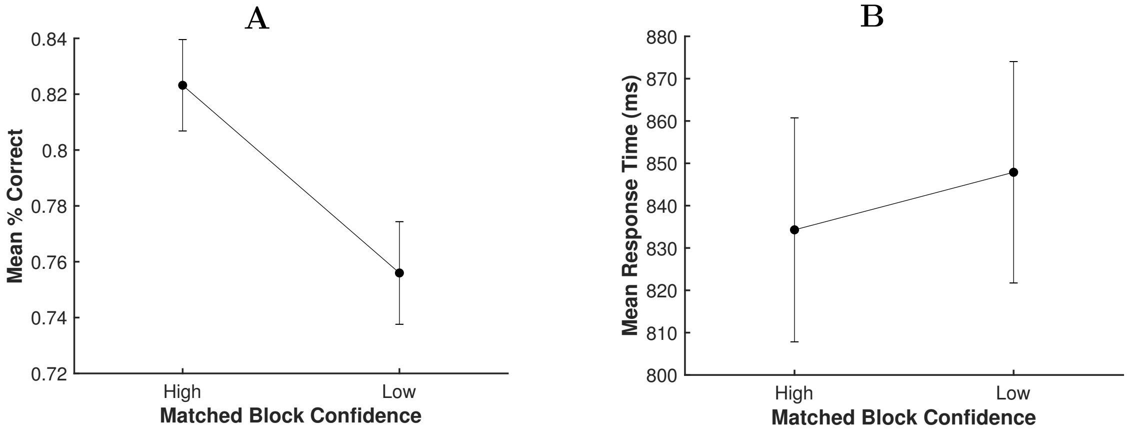

Figure 4.6: Effects of confidence on accuracy and response times. (A) Mean accuracy on the dot task, divided according to relative cue confidence. (B) Mean dot-task response times, divided according to relative cue confidence . Error bars indicate standard errors of the mean (SEM).

differences. If attention is guided mainly by stimulus-driven associative strength, performance should be similar when controlling for differences in stimuli, regardless of differing confidence levels. If, however, the effects are driven by confidence, as we hypothesized, they should persist even when associative strength is accounted for. We therefore defined each block in a matched pair as high-confidence or low-confidence blocks, based on each participant's average rated cue confidence score relative to their matched block. Pairs of matched blocks with equal confidence scores were excluded from the analysis, with ten blocks excluded overall (a pair of blocks each was excluded for five participants). Preliminary analysis revealed no significant effects of block order on matched-block classification. A paired t-test on average dot-task accuracy as a function of matched-block confidence revealed that accuracy was significantly higher in high-confidence compared to low-confidence matched blocks ($t(24) = 7.48, p < .001$). Similarly, we found that response times were reliably shorter for high-confidence blocks ($t(24) = 2.51, p < .02$).

These results support our prediction that confidence in knowing which information is task-relevant predicts performance, above and beyond the effects of alternative factors including cue accuracy, overall progression of learning through a block, and stimulus-driven associative strength. It is interesting to note however, that while the mean beta value for confidence in our regression models was significantly different from zero across participants, the beta coefficients were small, indicating that changes in confidence only led to a small change in dot-task performance when controlling for these additional factors.

SSVEP data

Our first analysis, prior to investigating the effects of confidence, aimed to establish whether we can detect more general attentional effects from the SSVEP responses at the selected electrode cluster. To evaluate attentional effects, we compared SSVEP amplitudes according to objective relevance, expecting larger SSVEP responses for relevant compared to irrelevant stimuli (Andersen & Müller, 2010; Morgan et al., 1996). Since attentional effects might be modulated by the flickering frequency used for tagging stimuli, we initially studied the relevant and irrelevant conditions separately for each frequency. For each participant, four time courses were computed by averaging the Gabor filtered response across trials, grouping by cued relevance (relevant vs. irrelevant) and frequency (5Hz, 7.5 Hz). Figure 4.7 shows the time courses, averaged across all participants. As demonstrated in these figures, SSVEP amplitudes were larger for stimuli cued as task-relevant (vs. task-irrelevant), for both frequencies, though the specific time course of this effect varied by frequency.

To assess the differences in SSVEP amplitudes, we computed the mean amplitude for each condition during stimulus presentation time (from time 0 to 1500 ms). We subjected the mean amplitude value to a two-way repeated-measures ANOVA, with factors for the condition (relevant, irrelevant) and frequency (5Hz, 7.5Hz). We found that, as expected, there was a significant difference in the strength of the SSVEP response based on the relevance of the stimuli ($F(1, 24) = 7.09, p = .014$), with larger responses to relevant stimuli. The effect of frequency on the average amplitude was not significant ($F(1, 24) = 3.67, p = 0.068$). Importantly, there was no significant interaction between frequency and relevance ($F < 1$), indicating that overall effects of stimulus relevance were similar across frequencies. Thus, in our following analyses we collapsed the data across both frequencies, computing an

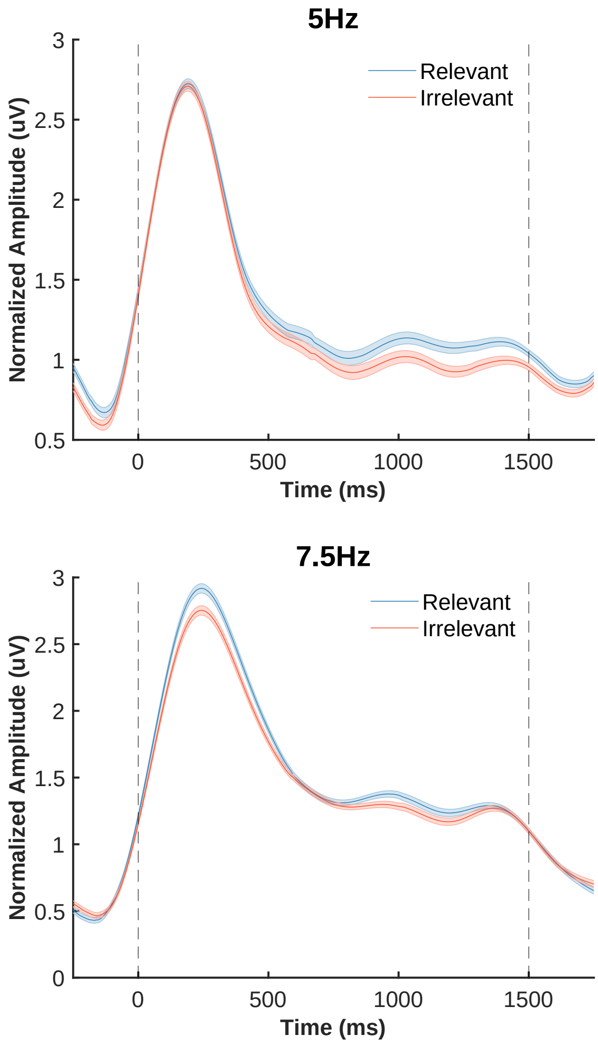

Figure 4.7: Stimulus-locked time course of SSVEP amplitudes for each frequency separately, according to task-relevance. Dashed lines represent the stimulus onset and offset. Shaded areas represent 1 SEM corrected for within-participant comparisons at the respective time point.

average SSVEP response. To ensure that both frequencies were taken into account equally across conditions, we first averaged the data for each frequency, and then averaged across frequencies (See Figure 4.8).

Confidence effects on SSVEP amplitudes. Having demonstrated that we can detect attentional effects based on stimulus relevance in our data, we were now interested in testing our main hypothesis: specifically, whether attention to a stimulus was further modulated by confidence in its relevance. Importantly, the differences in attention allocation that we observed in our data could arise from

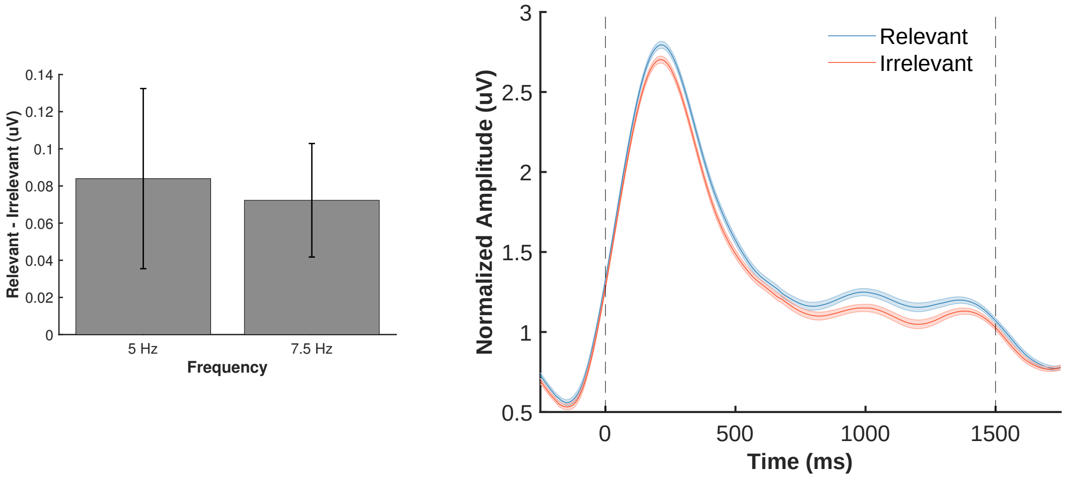

Figure 4.8: Attentional effects of relevance on SSVEPs. A: Differences in SSVEP amplitudes as a function of task relevance, for each frequency. Error bars represent SEM. B: Average time courses for the relevant and irrelevant stimuli, collapsed across both frequencies. Dashed lines represent the stimulus onset and offset. Shaded areas represent 1 SEM corrected for within-participant comparisons at the respective time point.

enhancement of the relevant stimulus, suppression of the irrelevant stimulus, or a combination of both. Our initial prediction was that we would find evidence for both processes, with attentional selectivity (i.e., the difference between activity for the relevant and irrelevant stimuli), increasing as a function of confidence. However, since there is evidence that enhancement and suppression of stimuli rely on separate attentional mechanisms, and are subject to different time-courses (Andersen & Müller, 2010), we examined the influence of confidence on attention separately for relevant and irrelevant stimuli.

To study how confidence influenced attention, as captured by SSVEP amplitudes, we followed the same approach that we applied when examining the effects of confidence on accuracy and response times in the dot-task. For each participant, trials were divided into four bins according to relative-confidence (i.e., ranging from a rating of -50, indicating high confidence in a cue-colour association that is objectively incorrect, to +50, indicating high confidence in a cue-colour association that is objectively correct). Average SSVEP amplitudes during the entire stimulus presentation time were then computed for each bin, for the relevant and irrelevant conditions separately (collapsing across the different frequencies).

A two-way repeated-measures ANOVA on the mean SSVEP amplitude as a function of stimulus relevance (relevant, irrelevant), and confidence (Q1-Q4) revealed reliable effects of relevance ($F(1, 24) = 7.78, p = .01, \eta^2 = .25$), and confidence ($F(3, 72) = 2.86, p = .049.\eta^2 = .106$), with higher SSVEP amplitude elicited by the relevant stimulus, and when confidence was high. Surprisingly, the interaction between confidence and stimulus relevance was not significant ($F(3, 72) = 2.52, p = .09, \eta^2 = .095$), although the data suggested that confidence had a different effect on the SSVEP amplitudes elicited by the relevant vs. irrelevant stimulus (see Figure 4.9). Thus, to explore the mechanism by which confidence affects attention further, we examined the relationship between confidence and the SSVEP amplitudes separately for the relevant and irrelevant stimulus. A repeated-measures ANOVA on the mean SSVEP amplitude elicited by the relevant stimulus, as a function of confidence (Q1-Q4), revealed that confidence reliably predicted SSVEP amplitude ($F(3, 72) = 4.52, p = .015, \eta^2 = .16$). As shown in Figure 4.9A, average SSVEP amplitude increased overall with confidence (though there is a numerical decrease in amplitude for the highest confidence bin, Q4, compared to Q3), as captured by a significant linear trend ($F(1, 24) = 5.91, p = .02, \eta^2 = .20$). An examination of the time courses of the SSVEP amplitudes suggests that differences in SSVEP amplitudes emerge most clearly from a period of around 500 ms after stimulus onset (Figure 4.9B. Note that, due to the Gabor filter time resolution of ± 183.9 ms, responses to the stimulus are visible before time zero). For irrelevant stimuli, the results did not show a significant effect of confidence on SSVEP amplitudes ($F < 1$).

Overall, these results support our hypothesis that attention is modulated by confidence in stimulus relevance. More specifically, the results suggest that confidence modulates attention to the task-relevant information, with people paying more attention to the relevant stimulus when they were more confident in its relevance. In contrast, we did not find a reliable effect of confidence on the task-irrelevant stimulus, contrary to our initially expectations.

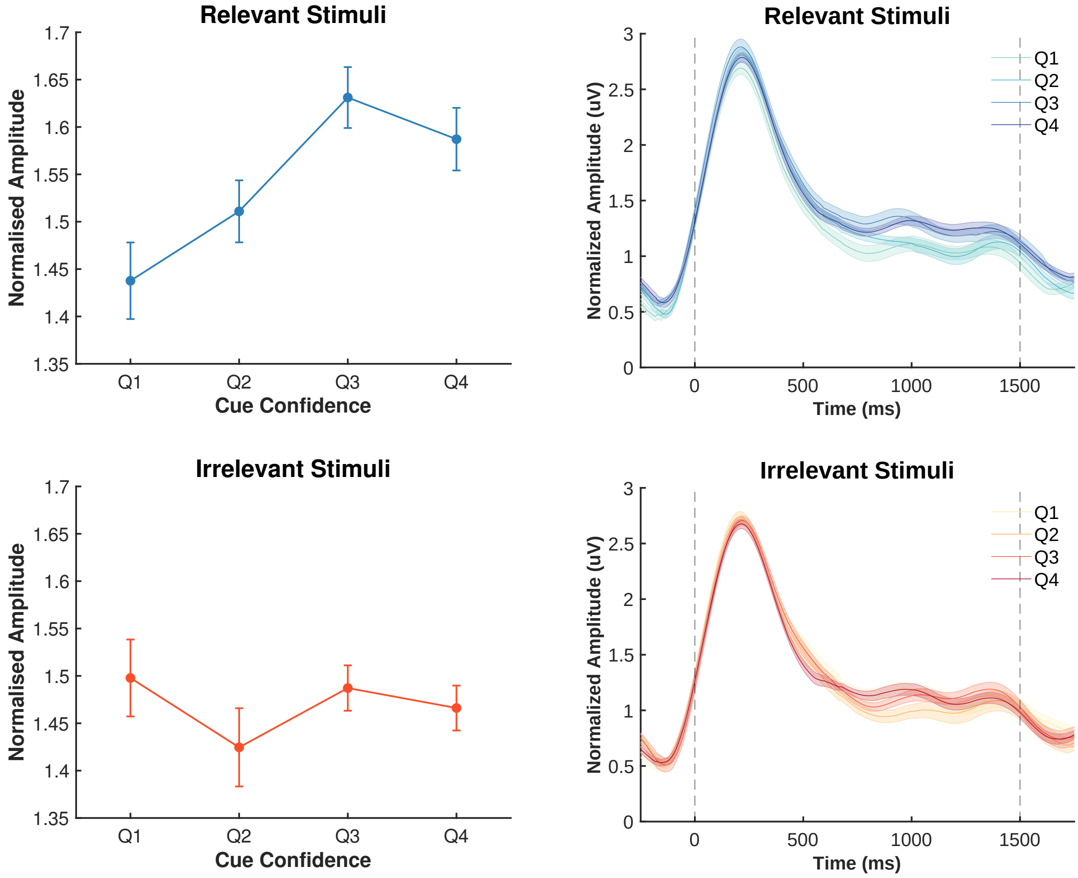

Figure 4.9: Effects of confidence on SSVEP amplitudes for relevant and irrelevant stimuli.(A) Mean SSVEP amplitude elicited by relevant stimuli, divided according to relative cue confidence bins (Q1-Q4). (B) Stimulus-locked time course of SSVEP amplitudes for the relevant frequency, according to confidence. Dashed lines represent the stimulus onset and offset. Note that, due to the Gabor filter time resolution of ±183.9 ms, responses to the stimulus are visible before time zero. (C) Mean SSVEP amplitude elicited by irrelevant stimuli, divided according to relative cue confidence bins (Q1-Q4). (D) Stimulus-locked time course of SSVEP amplitudes for the irrelevant frequency, according to confidence. Error bars indicate standard errors of the mean (SEM) corrected for within-participant comparisons. Shaded areas represent 1 SEM corrected for within-participant comparisons at the respective time point.

Factors impacting relevant-stimulus SSVEP amplitudes. To ensure that the effects we found on attention were driven by confidence, rather than alternative factors, we performed similar analyses to those described above for the behavioural results. Since reliable effects of confidence were found for the task-relevant, but not the task-irrelevant stimulus, all following analyses considered only the task-relevant stimulus.

predictor	$\beta > 0$	mean β	SD	tStat	DF	p-value
confidence	15	0.0029	0.0079	1.82	24	.081
frequency	8	-0.26	0.73	-1.79	24	.086
trial position	11	-0.0056	0.021	-1.37	24	.182
dot difference	12	0.004	0.056	0.34	24	.73
block	7	-0.004	0.014	-1.44	24	.164

Table 4.1: Results from linear regression models predicting SSVEP amplitudes for the relevant frequency. Regression models were fit for each participant separately, and a t-test comparing the average β across participants to zero performed. The number of participants for which the beta coefficient was positive is represented by $\beta > 0$.

To assess the trial-by-trial contribution of confidence to attention to the task-relevant stimulus, we fitted linear regression models predicting the SSVEP amplitude, for each participant separately. We included relative confidence, frequency, trial position, block and the relevant dot-difference as predictors. To control for the effects of binary identification of the task-relevant information, trials in which participants judged the irrelevant colour as being relevant were excluded from this analysis.

We observed positive slopes for confidence in 15 participants. However, while the mean beta value was positive, showing a numerical trend towards larger SSVEP amplitudes for trials with higher confidence, this was not significantly different to zero across participants ($M = .0029, t(24) = 1.82, p = 0.08$). Similarly, none of the other variables included in our model reliably predicted SSVEP amplitudes (see Table 4.1).

Our final analysis aimed to control for the effects of stimulus-driven associative-strength, while still using a group analysis approach (rather than trial-by-trial regression), and employing our matched-block design. For each participant, average SSVEP amplitude elicited by the relevant stimulus was computed, with trials grouped according to average block-confidence relative to the matched-block. A paired t-test on the average SSVEP amplitude as a function of matched-block confidence did not reveal a significant effect of matched block ($t(24) = 1.69, p = .10$). However, as shown in Figure 4.10, the numerical trend was in the expected direction, with higher amplitudes for high-confidence compared to low-confidence matched blocks . Figure 4.10B shows the time-courses of the SSVEP amplitudes, according

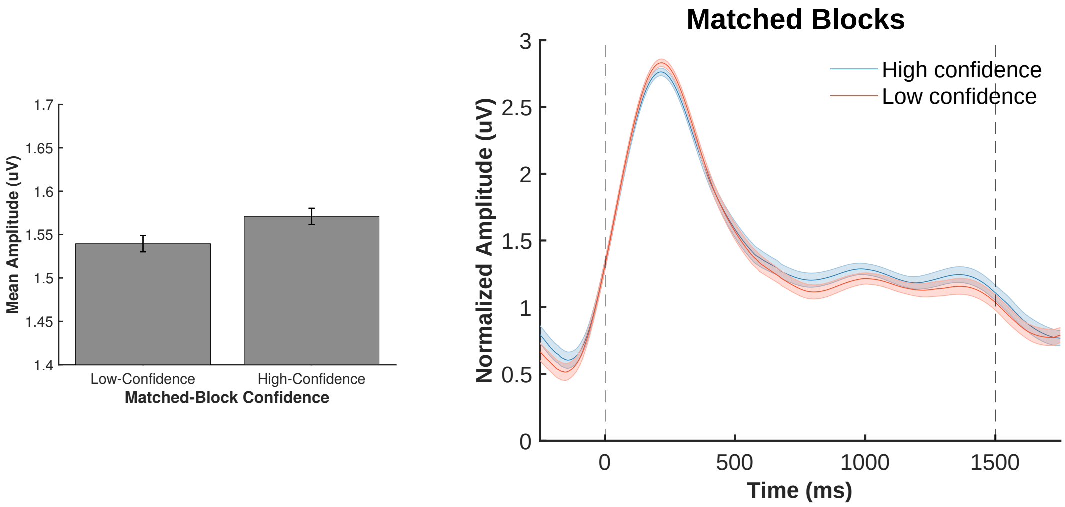

Figure 4.10: Effects of matched-block confidence on SSVEPs. A: SSVEP amplitudes according to matched-block confidence. For each participant, the average confidence was computed for each block, and blocks were then classified according to their average confidence relative to their matched block. Error bars represent SEM corrected for within-participant comparisons. B: Average time courses for the relevant stimulus, according to matched-block confidence, collapsed across both frequencies. Dashed lines represent the stimulus onset and offset. Shaded areas represent 1 SEM corrected for within-participant comparisons at the respective time point.

to matched-block confidence. Similar to the time-courses of the SSVEP amplitudes grouped by confidence quartiles (Figure 4.9B), here too the data suggest that the effects of confidence emerge at around 500 ms after stimulus onset time.

Thus, the numerical trends emerging are consistent with the ANOVA results described above, and with our prediction that attention to the task-relevant stimulus will increase with confidence in its relevance. However, the effects of confidence are not statistically significant when analysis is performed while adjusting for the influence of additional factors.

4.3 Discussion

The ability to learn which information in our environment is important, and to adjust how we allocate attention accordingly, is critical for adaptive behaviour. Previous research has demonstrated that attention to stimuli is shaped by experience, so that attention to cues that reliably predict meaningful events increases with experience, whereas attention to nonpredictive cues decreases (for a review, see Le Pelley et al., 2016). The experiment reported in this chapter examined the role of confidence in this process. Specifically, we hypothesized that confidence in the relevance of information modulates the allocation of attention to it. This hypothesis was tested using a novel experimental design that combined elements from associative learning and perceptual decision-making tasks. Mimicking real life situations, participants learned from experience which information is useful in a given context. On each trial, participants reported their confidence about which of two stimuli is task-relevant, before performing a perceptual judgement task with frequency-tagged relevant and irrelevant stimuli. Scalp EEG was recorded to track allocation of attention to each stimulus separately. We found that participants responded more accurately, and faster, when they identified the relevant information with high confidence. Crucially, these behavioural effects were accompanied by corresponding patterns in the electrophysiological data: Increased confidence in identifying the relevant information was accompanied by larger SSVEP amplitudes at the corresponding frequency. These results are consistent overall with the hypothesis, indicating increases in attention to task-relevant information as confidence in its relevance increases.

In the current experiment, both stimuli (i.e., the different dot colours) were equally predictive of the correct response overall. Thus, our finding that more attention was paid to the cued (and relevant) stimulus on a specific trial is in keeping with previous studies showing that learned changes in attention can come under contextual control (Uengoer et al., 2013; Uengoer et al., 2018). Moreover, our data demonstrate that attention to a stimulus is adjusted in a graded fashion, beyond its binary identification as informative: Higher confidence in the relevance of a stimulus led to increased attention to it, with our data suggesting that attentional adjustments occur on a trial-by-trial level. These findings resonate with growing evidence that attention is regulated flexibly to enhance performance, by increasing processing of task-relevant information in response to high conflict situations (Botvinick et al., 2001; Egner & Hirsch, 2005; Yeung, 2014), or following errors

(Steinhauser & Andersen, 2019). We propose that confidence can similarly drive adjustments in attention through cognitive control processes, based on expectancies regarding the upcoming situations. Such constant monitoring and preparatory adjustments of attention according to confidence might be important for reaching an optimal strength of attention when there is uncertainty regarding the relevance of information.

Evidence for adjustments of attention on a single-trial level emerged most clearly from our behavioural data, with higher confidence judgements predicting better accuracy and faster RT on a trial-by-trial basis. While these results are congruent with corresponding variations in attention to the relevant stimulus, it is worth noting that a direct examination of the trial-by-trial influence of confidence judgements on SSVEP amplitudes failed to reach significance. There was, nevertheless, a trend in the predicted direction. The time-courses of the SSVEP amplitudes suggest that averaging amplitudes over the entire epoch might have led to decreased sensitivity in our analysis. Consistent with this proposal, previous studies examining the trajectory of attention have demonstrated that attention shifts dynamically during task performance (Manohar & Husain, 2013). It seems plausible that shifts in attention occurred during our task, for instance, depending on whether the participant was trying to respond accurately to the current trial, or to collect information about the environment (i.e., learn the cue-colour associations), at a given time point. Thus, to strengthen the present results, further analyses might consider temporal cluster-based permutations tests to focus on time-periods of interest.

Although it is possible that the demonstrated attention effects were at least partly driven by automatic processes (such as those demonstrated by Luque et al., 2020), our data and manipulation are more in line with an interpretation of our results arising from a controlled choice, based on an explicit representation of confidence. First, attentional effects were present despite the different stimuli being equally predictive across contexts, and even when controlling for stimulus-driven associative strength through our matched block design and through regression analyses that included predictors for these effects. Second, previous studies have demonstrated that automatic attentional capture is transient and typically observed only briefly after the stimulus appears (Le Pelley et al., 2013; Luque et al., 2020). Since our stimuli appeared for an extended period (1500 ms), and we averaged SSVEP amplitudes over the entire trial, our results are unlikely to be driven by such transient effects, even if they were present in our data. Thus, our results are consistent with previous studies demonstrating a controlled aspect of predictiveness-driven biases in attention (Mitchell et al., 2012).

Our initial prediction was that, in line with the view of visual processing as a limited resource, we would observe both enhancement of attention towards task-relevant stimuli and suppression of task-irrelevent stimuli (Andersen & Müller, 2010). Notably, we did not find reliable differences in SSVEP responses to the irrelevant frequency, though the data revealed a weak numerical trend in the expected direction. It is unclear whether these results reflect a true lack of effect, or if they are the product of limitations in our experimental design and analyses. On the one hand, our results are consistent with previous evidence suggesting that cognitive control mechanisms increase selective attention by enhancing the task-relevant information, rather than suppressing the irrelevant information (Egner & Hirsch, 2005; Steinhauser & Andersen, 2019). On the other hand, as noted already, it is plausible that participants adjusted attention during stimulus presentation, for instance, if they believe they had initially selected the incorrect stimulus. Indeed, an examination of the time-course of the irrelevant SSVEP amplitudes reveals larger responses briefly after stimulus presentation on trials in which people were more confident in the irrelevant information, as expected based on our hypothesis. However, the relationship between confidence and attention reversed during stimulus presentation time. As discussed above, any rapid or dynamic shifts of attention would not be captured by our analysis, and could lead to null results.

Our study builds on previous theoretical and empirical work investigating how attention is shaped by experience. In his seminal paper, Mackintosh (1975) proposes that attention is modulated through processes of associative learning, with attention to information increasing when it predicts important outcomes, and decreasing when it does not. Numerous experimental studies have provided evidence supporting this proposal (e.g., see Le Pelley & McLaren, 2003; Le Pelley et al., 2010). Recent advances suggest that attention is influenced not only by the learned predictiveness of information, but also by its learned value (Anderson, 2013; Camara et al., 2013). To account for both phenomena, a model in which attention is modulated by the absolute associative value of stimuli has been proposed (Le Pelley et al., 2016). Consistent with this suggestion, a recent study investigating the interaction between attention and reinforcement learning showed that people can learn which dimension is relevant in complex environments, and bias attention towards it, based on the expected value associated with it(Leong et al., 2017). Our results are in line with these studies, and extend them by suggesting that modulations in attention to stimuli are driven by a subjective estimate of their relevance, rather than their objective predictability.

Several recent studies support the view that subjective estimates play a role in attention allocation (Akaishi et al., 2016; Le Pelley et al., 2016; Vossel et al., 2015). Importantly, however, these studies do not directly measure beliefs. Rather, they infer them based on behavioural observations or modelling approaches. Thus, our results are valuable in demonstrating the relationship between subjective evaluations and attention. Particularly relevant to the present investigation, a recent study found that when two stimuli are paired with the same expected reward value, but one always predicts a medium value, while the other is paired with high and low values in equal probability, attention is captured more by the latter (Le Pelley et al., 2019). The authors suggest that participants overweighted the largest outcome, leading to a higher subjective value of the second stimulus, even though the objective expected value associated with both stimuli was equal. Taken together with our results, these data suggest that attention prioritisation (or the absolute associative value in the model suggested by Le Pelley et al., 2016) is determined by subjective representations of both the reward and the predictability associated with a stimulus.

One limitation of the present study, is that it does not allow for full dissociation of the effects of confidence on attention from low-level effects of associative strength.

While our matched block design aimed to control for stimulus-driven differences in associative strength, it is still possible that associative differences existed, due to factors such as fluctuations in task engagement. To fully control for differences in associative strength, future experiments could use methods that have been shown to dissociate accuracy and confidence (following a similar rationale to that applied to the experiments in Chapter 2). One such method is a colour judgement task, in which participants report the average colour of several shapes in an array, and confidence can be dissociated from accuracy by manipulating the mean and the variability of the colours (Boldt et al., 2017). A future experiment in which a colour judgement task is used as a cue to the relevant stimulus (with the average colour indicating which colour is relevant on a trial), could extend the present findings while controlling for associative strength.

To conclude, the current study presents a first attempt at investigating how confidence can adaptively guide attention. Overall, the results were consistent with the hypothesis, and reflect an intuitive relationship, whereby the amount of attention paid to a stimulus is weighted according to confidence in its relevance. These results extended the findings of Chapters 2 and 3, demonstrating that confidence is used to monitor and adjust behaviour adaptively. However, while we initially predicted that confidence will modulate attention to both the relevant and the irrelevant stimuli, we found that confidence influenced attention only to the relevant stimuli. Future directions might consider further experimental approaches to disentangle the effects of associative strength from strategic confidence modulations, and isolate the use of confidence for guiding attention depending on the current goal.

'You can't connect the dots looking forward; you can only connect them looking backwards.'

— Steve Jobs

5 General Discussion

5.1 Overview

The research presented in this thesis examined the role of confidence in guiding adaptive behaviour, highlighting the view that explicit representations of confidence serve to guide behaviour in a flexible and goal-directed way. Each of the experimental chapters focussed on a different question related to this overarching theme: Chapter 2 asked if confidence acts as an intrinsic cost-benefit factor, guiding people when choosing between tasks; Chapter 3 examined whether people can use confidence strategically, according to the current context; and Chapter 4 tested whether confidence modulates attention to information. The three lines of research draw on, and subsequently inform, separate strands of literature. In the present chapter, I briefly review each line of research within its immediate context. Then I discuss how the separate findings come together and contribute to a wider understanding of the functional role of confidence. Finally, I discuss limitations of the present work, and directions for future research.

5.2 Summary of research

First line: confidence guides task choices

The first line of research, presented in Chapter 2, investigated whether people use their confidence as a cue to guide their choices of tasks to engage with. A wealth of past research has examined how people, and animals, choose between different tasks and lines of action. Within this literature, it is well established that actions are prioritised according to their learnt costs and benefits. These include action outcomes (e.g., Kahnemann, 1984), as well as costs and benefits intrinsic to the action itself, such as the amount of physical work (Hull, 1943; Solomon, 1948; Walton et al., 2006), and mental effort (Apps et al., 2015; Kool et al., 2010; Westbrook et al., 2019) that a task requires. Recent advances have shown that metacognitive evaluations of task difficulty (Desender et al., 2017) and effort (Dunn et al., 2016) influence task choices, raising questions regarding the extent to which learned value is determined by subjective experience, rather than objective performance (Shenhav et al., 2017).

The five experiments presented in Chapter 2 (Experiments 1-5) tested the hypothesis that decision confidence influences task choices. To this end, participants performed a task-selection paradigm in which they chose between two competing tasks, that were identical except for the strength of postdecisional evidence. Since postdecisional evidence informs decision confidence (Moran et al., 2015; Pleskac & Busemeyer, 2010), but cannot impact the decision process itself, this manipulation allowed for the dissociation of effects driven by confidence from those related to first-order task performance.

The results showed that people were biased towards selecting tasks in which they experienced higher confidence, even though their performance on the tasks was similar. Furthermore, confidence guided task selection in a more nuanced way than binary error detection; in addition to avoiding tasks in which they believed they had made an error, participants also tended to repeat a task more after they experienced higher confidence in their correct responses (compared to when they perceived performance as correct, but with lower confidence). Importantly, people showed a preference even when external feedback was provided, and when controlling for evidence strength. Thus, I concluded that subjective confidence acts as an intrinsic cost-benefit factor, guiding people's choices of which tasks to pursue.

Second line: confidence is used strategically

The findings from Experiments 1-5 contribute to a growing body of evidence showing that confidence plays an important role in guiding behaviour. However, the results of these experiments, and many of the others discussed throughout this thesis, consider a fixed relationship between confidence and behaviour: low confidence leads to one course of action, such as avoiding a task, and high confidence to another, such as choosing to engage with a task. In our daily lives however, situations are typically more complex, and usually there is no 'one size fits all' solution. Cognitive flexibility is a hallmark of human cognition, allowing us to adapt, and find new solutions when demands change (Ionescu, 2012). Furthermore, conscious representations such as confidence are thought to play a key role in this flexibility (Dehaene & Changeux, 2011). The second line of research, presented in Chapter 3, tested whether people display flexibility in their use of confidence, so that the mapping between confidence and behaviour is not necessarily fixed, but can instead vary depending on the specific context.

This proposal was studied in Experiments 6-8, with the relationship between confidence and advice seeking examined as a test case for flexible use of confidence. In a variation on the classic JAS paradigm (J. Sniezek & Buckley, 1989; J. Sniezek & Buckley, 1995), participants made an initial perceptual judgement, and then had to choose between re-sampling evidence or receiving advice from a virtual advisor, before making a final decision. The results from these experiments showed that, when no objective information about advisor reliability was available, people selected advice more often when their confidence was high, consistent with the idea that they used their confidence as a feedback proxy to learn about advisor quality. Importantly, even though feedback was absent, participants were able to learn about the value of advice and subsequently selected more advice from better advisors, indicating the value of using confidence in this way. A contrasting pattern of confidence and advice-seeking behaviour was observed when participants had prior knowledge about the reliability of advisors: Here, they requested advice more often when confidence was low, consistent with the idea that they were using confidence as a self-monitoring tool (i.e., as an internal signal that help should be solicited). These findings are consistent with confidence functioning both as a self-monitoring tool (when deciding) and as a substitute for external feedback (when learning). More broadly, the findings indicate that people use confidence in a way that is context-dependent and directed towards achieving their current goals.

Third line: confidence guides attention

While Chapters 2 and 3 focussed on ways in which confidence guides high-level cognition, Chapter 4 asked whether an explicit representation of confidence can also modulate attention and, through this, influence lower-level processing of information. Previous studies have demonstrated that attention to stimuli changes as a result of learning about them (for a comprehensive review, see Le Pelley et al., 2016), with attention to cues that reliably predict important outcomes increasing with experience, whereas attention to nonpredictive cues decreases (Le Pelley et al., 2010; Le Pelley et al., 2013). Building on these previous findings, Experiment 9, the final experiment included in this thesis, examined the role of confidence in this process. Specifically, I hypothesized that confidence in the task-relevance of information modulates the degree of attention that is allocated to it, above and beyond low-level effects of associative strength.

This prediction was tested using a perceptual-judgement task in which participants were presented with both relevant and irrelevant stimuli, and scalp EEG was recorded to provide a neural index of attention effects. The results of Experiment 9 were consistent with the hypothesis overall: The behavioural data suggested an effect of confidence on selective attention, with people responding to the perceptual task more accurately, and faster, the more confident they were that the informative stimulus was relevant. The EEG results suggested that this effect is driven by modulation of attention to the relevant, but not the irrelevant stimulus. Specifically, high confidence in the relevance of the informative stimulus was associated with greater attention towards it, and vice versa. However, the data did not show a reliable effect of confidence on attention to the irrelevant information. These results provide initial support to the suggestion that confidence can modulate attention, indicating that confidence might influence not only high-level strategic behaviour, but also what information is selected for processing at early stages. Thus, confidence might affect not only how we respond to the world around us, but also how we perceive it.

5.3 The role of confidence: past and present findings

In all 9 experiments comprising this thesis, confidence was found to influence behaviour. This was demonstrated across different contexts, with confidence reliably predicting task choices (Experiments 1-5), advice seeking (Experiment 6-8), and allocation of attention (Experiment 9). Furthermore, confidence was used in different ways—as a self-monitoring tool that guides decisions, a feedback proxy that allows people to learn about the reliability of others, and as a signal of information relevance, regulating attention weights. Taken together, these results illustrate the wide scope of behaviours on which confidence has an impact.

The view that confidence specifically, and metacognition in general, play an important role in guiding behaviour has been driving research in the domains of memory and learning (Bjork et al., 2013; Koriat, 2006; Nelson & Narens, 1990), and organisational psychology (Penrod & Cutler, 1995; Price & Stone, 2004; J. A. Sniezek & Van Swol, 2001; Zarnoth & Sniezek, 1997), for several decades. However, studies examining the functional role of confidence within the field of cognitive science have only recently begun to emerge, with interest in this area growing rapidly. Current theoretical and empirical accounts suggest that confidence acts as an important cue, guiding several aspects of behaviour (For reviews, see Meyniel et al., 2015; Shea et al., 2014; Shea & Frith, 2019; Yeung & Summerfield, 2012). For example, it has been shown that confidence influences information seeking (Desender et al., 2018; Desender, Murphy, et al., 2019), learning (Ben Yehuda, 2019; Guggenmos et al., 2016; Hainguerlot et al., 2018), and task preparation (Boldt, Schiffer, et al., 2019). Thus, the present findings contribute to a growing body of research, showing that confidence plays an important role in guiding behaviour.

Past research has demonstrated that people are able to provide confidence judgements that, although not perfect, reliably predict objective accuracy (Baranski & Petrusic, 1994). The functional value of confidence stems from its correlation with objective accuracy. However, this feature also makes it difficult to disentangle the effects of confidence from those of objective performance. Critically, in the present work, confidence was found to impact behaviour even when controlling for objective accuracy and task difficulty. The independence of effects is most evident in Experiments 1-5, where tasks differed only in the strength of postdecisional evidence, leading to a dissociation between subjective confidence and objective accuracy and response times. Therefore, the demonstrated task preferences can clearly be attributed to variations in confidence. In Experiments 6-9, confidence was not manipulated independently from accuracy, but statistical measures were used to control for the effects of task difficulty. Critically, it was found that confidence predicted information choices (Experiments 6-8), and allocation of attention (Experiment 9), even when controlling for task difficulty. Together, these data provide strong support to the view that subjective confidence impacts behaviour.

Overall, the results across experiments indicate that confidence is used to guide behaviour in a fine-grained fashion, beyond a binary classification of a response as being correct or incorrect. These findings are consistent with the influential framework proposed by Nelson and Narens (1990), whereby metacognitive monitoring exists in order to allow the online control of our cognitive processes. In this regard, the results of Experiment 4 stand out in particular, since people used their confidence to guide task choices even though external feedback was available. This finding underlines an important aspect of confidence—it provides useful internal information, that is not available from external sources. For example, consider a student sitting an exam, who did not have much time to prepare, and is aware of this. Despite not knowing a significant amount of the course material, they might receive a good mark if the questions happened to be on material that they revised. However, it would probably be wiser to rely on their low confidence and study more for the next exam, rather than study less based on their good mark. Similarly, when selecting between tasks, if a decision was made with low confidence but happened to be correct, confidence might be a better indicator of future performance than objective accuracy. Thus, it seems adaptive to self-monitor and control behaviour accordingly, even in the presence of external feedback. Importantly, previous studies have mainly focussed on how confidence influences behaviour when feedback is lacking, for example, by regulating how much information is sampled before committing to a choice (Desender, Boldt, Verguts, et al., 2019; van den Berg et al., 2016). Our results add to these past findings and suggest that confidence can further influence decision policy, even when feedback is provided.

Many of the studies investigating how confidence modulates decision-making processes have focussed on its impact on the amount of evidence that is accumulated before reaching a decision. Several aspects of this question have been examined, but the underlying logic and findings of most investigations has been similar: Experiencing low confidence in a decision leads to the accumulation of more evidence before committing to a choice, whereas higher confidence results in the accumulation of less evidence (Desender, Boldt, Verguts, et al., 2019; Desender et al., 2018; Desender, Murphy, et al., 2019; Meyniel et al., 2015). Critically, studies have typically used paradigms in which only one type of information was available (cf. Boldt, Blundell, et al., 2019). The present results extend these findings by showing that confidence can guide not only how much evidence is sampled, but also *what* evidence is sampled. The results of Experiments 6-8 show that when people are confronted with different types of information to choose from, confidence guides their selections. Importantly, it was found that participants changed the way in which they used confidence according to their current goals: When they wanted to learn about advisor quality, participants requested advice when their confidence was high. In contrast, when information about advisor quality was available, participants tended to request advice more often when their confidence was low. Thus, confidence seems to guide strategic information choices. In addition, Experiment 9 demonstrated that confidence modulates attention to information that is presented simultaneously. More specifically, it was shown that attention to the task-relevant information was weighed according to confidence in its relevance. Taken together, Chapters 3 and 4 demonstrate that confidence can influence what information is accumulated by guiding high-level strategic choices, as well as influencing low-level visual processing through modulation of attention.

The present findings are consistent with frameworks that highlight the importance of explicit confidence specifically, and conscious representations in general, in guiding behaviour in a flexible and goal-directed way (Dehaene & Changeux, 2011; Nelson & Narens, 1990; Shea & Frith, 2019). Critically, representing confidence in an explicit high-level way, that is available to multiple cognitive systems, might have several important benefits. First, it allows for planning and reasoning, while weighing information across systems within an individual or between different people, according to its reliability (Shea et al., 2014; Shea & Frith, 2019). In addition, it allows for generalising from one context to another, so that many different circumstances that resulted in the same explicit representation can lead to a single adaptive response, without having to learn the association in each situation separately. For instance, consider the situation described in Chapter 3 of being unsure which of two movies to watch, and therefore seeking advice from online reviews. Many different reasons could have led to your uncertainty—perhaps you have heard many good things about both movies and are finding it hard to choose; perhaps you heard conflicting opinions from different friends; or, alternatively, you might have no information about either movie, but these are the only two options available at the cinema. Regardless of the underlying cause, the same action, such as seeking advice, is equally useful. Therefore, it is adaptive to associate confidence levels (e.g. low confidence) with a response (e.g. seeking advice), rather than having to learn the association for each situation separately.

Taken together with previous studies, confidence emerges as a salient cue, that guides multiple aspects of our behaviour: When choosing what tasks to engage in, we might be able to use confidence as a common currency to compare tasks from different domains, so that we ultimately select those tasks that we are more likely to succeed at (de Gardelle et al., 2016; de Gardelle & Mamassian, 2014). Once we are performing a task, confidence allows us to monitor and adjust our own behaviour in an online fashion. For example, confidence guides what information we pay attention to, how much information we sample (Desender et al., 2018; Desender, Murphy, et al., 2019), whether we ask for advice, and whether we use external devices to reduce cognitive demand (Risko & Dunn, 2015; Risko & Gilbert, 2016). Furthermore, confidence can also be used to learn about the reliability of others, thus acting as an important guide for navigating the social world (Pescetelli & Yeung, 2018). Finally, it seems that people can adjust the way in which they use confidence in a goal-directed fashion, making it a particularly effective and powerful tool.

5.4 Future directions

Going beyond the dots

The experiments described in this thesis all used variations on the same perceptual task, in which participants were asked to judge which of two presented boxes contained more dots. Based on how people performed this task, I have drawn conclusions about the ways in which confidence is used to guide flexible behaviour. However, the degree to which these findings generalise remains unclear. Further exploration that goes beyond the dot task, and considers domains other than perceptual decisions, could be useful in determining the extent to which the present findings reflect the use of confidence in different settings, as discussed below.

The use of a single paradigm limits the degree to which the present findings can be generalised, since there is always a possibility that the results were driven by features specific to the task, rather than a general behavioural effect that would be found in other settings. For instance, one might argue that the task preferences we found in Chapter 2 are specific to the postdecisional evidence manipulation we used:

If people are biased away from weak evidence, then we might not find the same behavioural effects in situations where low confidence emerges due to other features, such as lack of attention, or a noisy environment. Thus, one straightforward way to establish the pervasiveness of the current findings is to replicate them using different paradigms. A significant difficulty in extending the findings of Chapter 2, specifically, is that objective accuracy and subjective confidence must be dissociated in order to detect confidence preferences. Two methods, other than the postdecisional evidence manipulation already used, might be useful for achieving such a dissociation in future studies. The first method uses a random dot motion task, in which confidence is dissociated from accuracy by manipulating the ratio of positive to negative evidence (i.e., the proportion of dots moving towards the correct vs. the incorrect direction ;Odegaard et al. (2018). The second method is a colour judgement task that was mentioned briefly in Chapter 4. In this task, participants report the average colour of several shapes in an array, and confidence can be dissociated from accuracy by manipulating the mean and the variability of the colours (Boldt et al., 2017). Demonstrating that similar preferences emerge when using a task-selection paradigm based on these tasks would show that the findings described here are not limited to the specific task that we used. The same logic can be applied to extending the findings of Chapters 3 and 4, by showing that the same information-seeking and attention effects emerge when using other perceptual tasks.

To extend our understanding of how confidence is used to guide flexible behaviour even further, future research could examine and compare the effects of confidence on related behaviours as they are observed across a range of domains and situations. For example, we might study whether people actively use confidence to learn about the reliability of others in different settings, such as general knowledge tasks, or situations in which they are interacting with another person, rather than a virtual advisor. Limiting the present investigation to perceptual tasks was a deliberate choice: As discussed in Chapter 1, perceptual decision-making tasks allow for careful control of the information on which an agent bases their decision. If we assume, as I have done, that all decisions share the same basic principles, then we can learn something about how people make complex, real-world decisions by studying simple perceptual choices under controlled conditions. Still, many differences exist between perceptual tasks and other domains, that suggest that confidence might be used to guide behaviour differently according to the setting (as demonstrated in Chapter 3, the use of confidence is flexible). For example, when performing a perceptual task, low confidence can only be used as a signal that further evidence should be accumulated before committing to a choice, because the source of the uncertainty is necessarily perceptual. In contrast, in reasoning tasks, such as a Sudoku problem (Qiu et al., 2018), uncertainty can arise from the reasoning process itself, with a likely source of uncertainty being the strategy that should be used to solve the problem. Therefore, low confidence in this context is likely to lead to exploration of different strategies, rather than a repetition of the previous strategy (such as re-viweing a stimulus). Thus, expanding the scope of the investigation to include different domains, and comparing how confidence is used to guide similar behaviours across these domains, could lead to a more comprehensive view of how people use subjective confidence. Furthermore, this exploration might shed light on the specific behaviours studied. For example, it is known that cognitive effort is costly, but can also sometimes add value to the outcome of a task, as well as to the task itself. However, when and why effort would be valuable remains unclear (for a review, see Inzlicht et al., 2018). Related to this, we might speculate that the inherent cost and effort associated with confidence that we found in Chapter 2 could vary depending on the task: It seems plausible that when performing reasoning tasks, people might prefer those tasks that lead to lower confidence, since they might find these situations challenging and rewarding. Thus, when using reasoning tasks, such as puzzles, we might expect to find the complementary effect to that demonstrated for perceptual tasks in Chapter 2. Such findings could further our understanding of the relationship between subjective confidence and the value of cognitive effort.

We have seen then, that there is much to gain by expanding the present investigation to consider and compare the use of confidence across different tasks and domains. As a final note, it is interesting to observe that despite using a single perceptual task throughout this thesis, we were still able to learn about several different uses of confidence. We saw that confidence can guide task choices, learning, advice seeking and attention. The diversity of these findings, arising from the use of a single, simple perceptual paradigm, are perhaps an indication of the wide scope and influence that confidence has on our behaviour.

Exploring individual differences

The current investigation aimed to further our understanding of shared principles in the use of confidence to guide behaviour, across individuals. Another interesting avenue for future research is the degree to which the demonstrated effects vary between individuals. Although several questionnaires were used in the present research, (Need for cognition, Cacioppo et al., 1984, and the Concern over Mistakes, Personal Standards and Doubts about Actions subscales from the Multi-dimensional perfectionism scale Frost et al., 1990), no reliable relationships between questionnaires and behaviour were found. However, several of the results presented suggested some individual differences in the ways tasks were performed. For example, in Experiment 5 people had to choose between a task that is associated with stronger evidence, but in which it is easier to notice mistakes, to a task associated with weaker evidence but higher confidence, and the results hinted at individual differences in aversion to errors versus uncertainty. Exploring such individual differences might offer insights into differences in learning strategies, such as individual differences in information seeking and exploration versus exploitation behaviours. We might speculate for example that differences in perfectionism, self-esteem, and the propensity for risk taking can all influence the degree to which people are willing to engage with low-confidence situations. Additional individual differences were seen in all the experiments presented in Chapter 3, with large variations between individuals in their overall tendency to request advice.

Studying the relationship between metacognition and advice seeking is of particular interest in terms of individual differences, and perhaps clinical implications (van der Plas et al., 2019). In Experiments 6-8, it was demonstrated that, by comparing opinions to their own beliefs when they were certain they know the answer, people are able to learn about the reliability of others. An interesting question is what happens when people are extremely overconfident. In this case, and particularly if an overconfident person also performs poorly, they might erroneously learn that other people are unreliable, and therefore rarely take advice. Since using expert advice is an important aspect of adaptive behaviour (e.g., when going to a doctor), such a bias might have significant implications on daily lives. In particular, although speculative, it seems plausible that clinical populations that show lack of insight, such as schizophrenia patients (Lysaker et al., 2014) might then also be less likely to follow medical advice. Conducting further research into the relationships between metacognition and advice seeking, and developing interventions that target metacognition (Cella et al., 2019), might therefore be important in providing better patient care.

5.5 Conclusions

Having a conscious sense of certainty in our beliefs and choices is an integral part of human experience. The work conducted for this thesis aimed to understand how an explicit representation of confidence serves to guide adaptive behaviour. The reported findings suggest that confidence plays an important role across a range of behaviours: guiding our task choices, enabling us to learn about the people surrounding us, modulating attention, and signalling when advice should be requested. Furthermore, it seems that people can change the way in which they use confidence according to their immediate aims. Together, these results support the view that an explicit representation of confidence serves to guide behaviour in a flexible, goal-directed way. Still, many open questions remain as to the ways in which confidence can help guide our way through a world with inherent uncertainty.

References

Akaishi, R., Kolling, N., Brown, J. W., & Rushworth, M. F. S. (2016). Neural Mechanisms of Credit Assignment in a Multicue Environment. *Journal of Neuroscience*, *36*(4), 1096–1112. https://doi.org/10.1523/JNEUROSCI.3159-15.2016

Andersen, S. K., & Müller, M. M. (2010). Behavioral performance follows the time course of neural facilitation and suppression during cued shifts of feature-selective attention. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(31), 13878–13882. https://doi.org/10.1073/pnas.1002436107

Andersen, S. K., Hillyard, S. A., & Müller, M. M. (2008). Attention Facilitates Multiple Stimulus Features in Parallel in Human Visual Cortex. *Current Biology*, *18*(13), 1006–1009. https://doi.org/10.1016/j.cub.2008.06.030

Anderson, B. A. (2013). A value-driven mechanism of attentional selection. *Journal of Vision*, *13*(3), 1–16. https://doi.org/10.1167/13.3.7

Apps, M. A. J., Grima, L. L., Manohar, S., & Husain, M. (2015). The role of cognitive effort in subjective reward devaluation and risky decision-making. *Scientific Reports*, *5*, 16880. https://doi.org/10.1038/srep16880

Arbuckle, T. Y., & Cuddy, L. L. (1969). Discrimination of item strength at time of presentation. *Journal of Experimental Psychology*, *81*(1), 126–131. https://doi.org/10.1037/h0027455

Asthana, H. S. (2015). Wilhelm wundt. *Psychological Studies*, *60*(2), 244–248. https://doi.org/10.1007/s12646-014-0295-1

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

Bang, D., Fusaroli, R., Tylén, K., Olsen, K., Latham, P. E., Lau, J. Y., Roepstorff, A., Rees, G., Frith, C. D., & Bahrami, B. (2014). Does interaction matter? Testing whether a confidence heuristic can replace interaction in collective decision-making. *Consciousness and Cognition*, *26*(1), 13–23. https://doi.org/10.1016/j.concog.2014.02.002

Baranski, J. V., & Petrusic, W. M. (1994). The calibration and resolution of confidence in perceptual judgments. *Perception & Psychophysics*, *55*(4), 412–428. https://doi.org/10.3758/BF03205299

Begg, I., Duft, S., Lalonde, P., Melnick, R., & Sanvito, J. (1989). Memory predictions are based on ease of processing. *Journal of Memory and Language*, *28*(5), 610–632. https://doi.org/10.1016/0749-596X(89)90016-8

Behrens, T. E., Hunt, L. T., Woolrich, M. W., & Rushworth, M. F. (2008). Associative learning of social value. *Nature*, *456*(7219), 245–249. https://doi.org/10.1038/nature07538

Ben Yehuda, M. (2019). *Agency and confidence: On the function of metacognition in action* (Doctoral dissertation). University of Oxford.

Bjork, R. A., Dunlosky, J., & Kornell, N. (2013). Self-Regulated Learning: Beliefs, Techniques, and Illusions. *Annual Review of Psychology*, *64*(1), 417–444. https://doi.org/10.1146/annurev-psych-113011-143823

Bokil, H., Andrews, P., Kulkarni, J. E., Mehta, S., & Mitra, P. P. (2010). Chronux: A platform for analyzing neural signals. *Journal of neuroscience methods*, *192*(1), 146–151.

Boldt, A., Blundell, C., & De Martino, B. (2019). Confidence modulates exploration and exploitation in value-based learning. *Neuroscience of Consciousness*, *2019*(1), 1–12. https://doi.org/10.1093/nc/niz004

Boldt, A., Gardelle, V. D., & Yeung, N. (2017). Supplemental Material for 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.supp

Boldt, A., & Gilbert, S. J. (2019). Confidence guides spontaneous cognitive offloading. *Cognitive Research: Principles and Implications*, *4*(1). https://doi.org/10.1186/s41235-019-0195-y

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), 3478–84. https://doi.org/10.1523/JNEUROSCI.0797-14.2015

Bonaccio, S., & Dalal, R. S. (2006). Advice taking and decision-making: An integrative literature review, and implications for the organizational sciences. *Organizational Behavior and Human Decision Processes*, *101*(2), 127–151. https://doi.org/10.1016/j.obhdp.2006.07.001

Boring, E. G. (1953). A history of introspection. *Psychological Bulletin*, *50*(3), 169–189. https://doi.org/10.1037/h0090793

Botvinick, M. M., Carter, C. S., Braver, T. S., Barch, D. M., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, *108*(3), 624–652. https://doi.org/10.1037/0033-295X.108.3.624

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

Brewer, N., & Hupfeld, R. M. (2004). Effects of testimonial inconsistencies and witness group identity on mock-juror judgments. *Journal of Applied Social Psychology*, *34*(3), 493–513. https://doi.org/10.1111/j.1559-1816.2004.tb02558.x

Cacioppo, J. T., Petty, R. E., & Feng Kao, C. (1984). The efficient assessment of need for cognition. *Journal of personality assessment*, *48*(3), 306–307.

Camara, E., Manohar, S., & Husain, M. (2013). Past rewards capture spatial attention and action choices. *Experimental Brain Research*, *230*(3), 291–300. https://doi.org/10.1007/s00221-013-3654-6

Cella, M., Edwards, C., Swan, S., Elliot, K., Reeder, C., & Wykes, T. (2019). Exploring the effects of cognitive remediation on metacognition in people with schizophrenia. *Journal of Experimental Psychopathology*, *10*(2). https://doi.org/10.1177/2043808719826846

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

Chaumon, M., Bishop, D. V., & Busch, N. A. (2015). A practical guide to the selection of independent components of the electroencephalogram for artifact correction. *Journal of Neuroscience Methods*, *250*, 47–63. https://doi.org/10.1016/j.jneumeth.2015.02.025

Clark, A., & Chalmers, D. (1998). The extended mind. *analysis*, *58*(1), 7–19.

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience*, *3*(3), 201–215. https://doi.org/10.1038/nrn755

Costermans, J., Lories, G., & Ansay, C. (1992). Confidence level and feeling of knowing in question answering: The weight of inferential processes. *Journal of Experimental Psychology : Learning, Memory, and Cognition*, *18*(1), 142. http://search.proquest.com/docview/1839879514/&pq_origsite=primo

Coull, J. T., & Nobre, A. C. (1998). Where and when to pay attention: The neural systems for directing attention to spatial locations and to time intervals as revealed by both PET and fMRI. *Journal of Neuroscience*, *18*(18), 7426–7435. https://doi.org/10.1523/jneurosci.18-18-07426.1998

Daniel, R., & Pollmann, S. (2012). Striatal activations signal prediction errors on confidence in the absence of external feedback. *NeuroImage*, *59*(4), 3457–3467. https://doi.org/10.1016/j.neuroimage.2011.11.058

de Gardelle, V., Le Corre, F., & Mamassian, P. (2016). Confidence as a Common Currency between Vision and Audition (S. Ben Hamed, Ed.). *PLOS ONE*, *11*(1), e0147901. https://doi.org/10.1371/journal.pone.0147901

de Gardelle, V., & Mamassian, P. (2014). Does Confidence Use a Common Currency Across Two Visual Tasks? *Psychological Science*, *25*(6), 1286–1288. https://doi.org/10.1177/0956797614528956

Dehaene, S., & Changeux, J. P. (2011). Experimental and Theoretical Approaches to Conscious Processing. *Neuron*, *70*(2), 200–227. https://doi.org/10.1016/j.neuron.2011.03.018

Dehaene, S., & Naccache, L. (2001). Towards a cognitive neuroscience of consciousness: Basic evidence and a workspace framework. *Cognition*, *79*(1-2), 1–37. https://doi.org/10.1016/S0010-0277(00)00123-2

Delorme, A., & Makeig, S. (2004). Eeglab: An open source toolbox for analysis of single-trial eeg dynamics including independent component analysis. *Journal of neuroscience methods*, *134*(1), 9–21.

Denton, S. E., & Kruschke, J. K. (2006). Attention and salience in associative blocking. *Learning and Behavior*, *34*(3), 285–304. https://doi.org/10.3758/BF03192884

Desender, K., Boldt, A., Verguts, T., & Donner, T. H. (2019). Confidence predicts speed-accuracy tradeoff for subsequent decisions. *eLife*, *8*, 1–25. https://doi.org/10.7554/eLife.43499

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

Desender, K., Calderon, C. B., Van Opstal, F., & Van den Bussche, E. (2017). Avoiding the Confict: Metacognitive Awareness Drives the Selection of Low-Demand Context. *Journal of Experimental Psychology: Human Perception and Performance*, *43*(7), 1397–1410. https://doi.org/https://doi.org/10.1037/xhp0000391

Desender, K., Murphy, P., Boldt, A., Verguts, T., & Yeung, N. (2019). A postdecisional neural marker of confidence predicts information-seeking in decision-making. *Journal of Neuroscience*, *39*(17), 3309–3319. https://doi.org/10.1523/JNEUROSCI.2620-18.2019

Desimone, R. (1995). Neural Mechanisms of Selective Visual Attention. *Annual Review of Neuroscience*, *18*(1), 193–222. https://doi.org/10.1146/annurev.neuro.18.1.193

Dunlosky, J., & Metcalfe, J. (2008). *Metacognition*. Sage Publications.

Dunn, T. L., Lutes, D. J. C., & Risko, E. F. (2016). Metacognitive evaluation in the avoidance of demand. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(9), 1372–1387. https://doi.org/http://dx.doi.org/10.1037/xhp0000236

Dunn, T. L., & Risko, E. F. (2016). Toward a Metacognitive Account of Cognitive Offloading. *Cognitive Science*, *40*(5), 1080–1127. https://doi.org/10.1111/cogs.12273

Egner, T., & Hirsch, J. (2005). Cognitive control mechanisms resolve conflict through cortical amplification of task-relevant information. *Nature Neuroscience*, *8*(12), 1784–1790. https://doi.org/10.1038/nn1594

El Zein, M., Bahrami, B., & Hertwig, R. (2019). Shared responsibility in collective decisions. *Nature Human Behaviour*, *3*(6), 554–559. https://doi.org/10.1038/s41562-019-0596-4

Faul, F., Erdfelder, E., Lang, A., & Buchner, A. (2007). G*power: A flexible statistical power analysis program for the social, behavioral and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191.

Fernandez-Duque, D., Baird, J. A., & Posner, M. I. (2000). Executive Attention and Metacognitive Regulation. *Consciousness and Cognition*, *9*(2), 288–307. https://doi.org/10.1006/ccog.2000.0447

Flavell, J. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist*, *34*(10), 906–911. https://doi.org/10.1002/bit.23191

Fleming, S. M., & Dolan, R. J. (2010). Effects of loss aversion on post-decision wagering: Implications for measures of awareness. *Consciousness and Cognition*, *19*(1), 352–363. https://doi.org/10.1016/j.concog.2009.11.002

Fleming, S. M., & Dolan, R. J. (2012). The neural basis of metacognitive ability. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1594), 1338–1349. https://doi.org/10.1098/rstb.2011.0417

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

Fleming, S. M., Van Der Putten, E. J., & Daw, N. D. (2018). Neural mediators of changes of mind about perceptual decisions. *Nature Neuroscience*, *21*(4), 617–624. https://doi.org/10.1038/s41593-018-0104-6

Folke, T., Jacobsen, C., Fleming, S. M., & De Martino, B. (2017). Explicit representation of confidence informs future value-based decisions. *Nature Human Behaviour*, *1*(1), 17–19. https://doi.org/10.1038/s41562-016-0002

Frost, R. O., Marten, P., Lahart, C., & Rosenblate, R. (1990). The dimensions of perfectionism. *Cognitive Therapy and Research*, *14*(5), 449–468. https://doi.org/10.1007/BF01172967

Fusaroli, R., Bahrami, B., Olsen, K., Roepstorff, A., Rees, G., Frith, C., & Tylén, K. (2012). Coming to Terms: Quantifying the Benefits of Linguistic Coordination. *Psychological Science*, *23*(8), 931–939. https://doi.org/10.1177/0956797612436816

Gershman, S., Cohen, J., & Niv, Y. (2010). Learning to selectively attend. https://escholarship.org/uc/item/7bn072ds

Gilbert, S. J. (2015). Strategic use of reminders: Influence of both domain-general and task-specific metacognitive confidence, independent of objective memory ability. *Consciousness and Cognition*, *33*, 245–260. https://doi.org/10.1016/j.concog.2015.01.006

Gino, F., Brooks, A. W., & Schweitzer, M. E. (2012). Anxiety, advice, and the ability to discern: Feeling anxious motivates individuals to seek and use advice. *Journal of Personality and Social Psychology*, *102*(3), 497–512. https://doi.org/10.1037/a0026413

Gino, F., & Moore, D. A. (2007). Effects of task difficulty on use of advice. *Journal of Behavioral Decision Making*, *20*(1), 21–35. https://doi.org/10.1002/bdm.539

Gold, J., Kool, W., Botvinick, M., Hubzin, L., August, S., & Waltz, J. (2015). Cognitive effort avoidance and detection in people with schizophrenia. *Cognitive, affective & behavioral neuroscience*, *15*(1), 145–154. https://doi.org/10.3758/s13415-014-0308-5

Gold, J., & Shadlen, M. (2007). The Neural Basis of Decision Making. *Annual Review of Neuroscience*, *30*(1), 535–574. https://doi.org/10.1146/annurev.neuro.29.051605.113038

Grimaldi, P., Lau, H., & Basso, M. A. (2015). There are things that we know that we know, and there are things that we do not know we do not know: Confidence in decision-making. *Neuroscience and Biobehavioral Reviews*, *55*, 88–97. https://doi.org/10.1016/j.neubiorev.2015.04.006

Guggenmos, M., Wilbertz, G., Hebart, M. N., & Sterzer, P. (2016). Mesolimbic confidence signals guide perceptual learning in the absence of external feedback. *eLife*, *5*, 1–19. https://doi.org/10.7554/eLife.13388

Hainguerlot, M., Vergnaud, J.-C., & de Gardelle, V. (2018). Metacognitive ability predicts learning cue-stimulus associations in the absence of external feedback. *Scientific Reports*, *8*(1), 5602. https://doi.org/10.1038/s41598-018-23936-9

Hajcak, G., & Foti, D. (2008). Errors are aversive: Defensive motivation and the error-related negativity: Research report. *Psychological Science*, *19*(2), 103–108. https://doi.org/10.1111/j.1467-9280.2008.02053.x

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

Harvey, N., & Fischer, I. (1997). Taking Advice Accepting Help, Improving Judgment. *Organizational Behavior and Human Decision Processes*, *70*(2), 117–133. https://pdfs.semanticscholar.org/2987/5f69c4e39bee1c415c6e9fbae7b1f02bc19b.pdf

Hauser, T. U., Allen, M., 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

Henmon, V. A. (1911). The relation of the time of a judgement to its accuracy. *Psychological Review*, *18*(3), 186–201. https://doi.org/10.1037/h0074579

Holroyd, C. B., & Coles, M. G. (2002). The neural basis of human error processing: Reinforcement learning, dopamine, and the error-related negativity. *Psychological Review*, *109*(4), 679–709. https://doi.org/10.1037/0033-295X.109.4.679

Holroyd, C. B., Yeung, N., Coles, M. G. H., & Cohen, J. D. (2005). A Mechanism for Error Detection in Speeded Response Time Tasks. *Journal of Experimental Psychology: General*, *134*(2), 163–191. https://doi.org/10.1037/0096-3445.134.2.163

Hu, X., Luo, L., & Fleming, S. M. (2019). A role for metamemory in cognitive offloading. *Cognition*, *193*(August 2018). https://doi.org/10.1016/j.cognition.2019.104012

Hull, C. L. (1943). *Principles of behavior* (Vol. 422). Appleton-century-crofts New York.

Inzlicht, M., Shenhav, A., & Olivola, C. Y. (2018). The Effort Paradox: Effort Is Both Costly and Valued. *Trends in Cognitive Sciences*, *22*(4), 337–349. https://doi.org/10.1016/j.tics.2018.01.007

Ionescu, T. (2012). Exploring the nature of cognitive flexibility. *New Ideas in Psychology*, *30*(2), 190–200. https://doi.org/10.1016/j.newideapsych.2011.11.001

Kahnemann, D. (1984). Choices, values, and frames, american psychologist, 39.

Kepecs, A., & Mainen, Z. F. (2012). A computational framework for the study of confidence in humans and animals. *Philosophical Transactions of the Royal Society 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

Kiani, R., & Shadlen, M. N. (2009). Representation of confidence associated with a decision by neurons in the parietal cortex. *Science (New York, N.Y.)*, *324*(2009), 759–764. https://doi.org/10.1126/science.1169405

Kleiner, M., Brainard, D., & Pelli, D. (2007). What's new in psychtoolbox-3?

Kool, W., McG uire, J. T., Rosen, Z. B., & Botvinick, M. M. (2010). Decision making and the avoidance of cognitive demand. *J Exp Psychol Gen*, *139*(4), 665–682. https://doi.org/10.1037/a0020198

Koriat, A. (2006). Metacognition and consciousness. *The Cambridge Handbook of Consciousness*, *3*(2), 289–326. https://doi.org/http://dx.doi.org/10.1017/CBO9780511816789.012

Koriat, A. (1993). How do we know that we know? The accessibility model of the feeling of knowing. *Psychological Review*, *100*(4), 609–639. https://doi.org/10.1037/0033-295X.100.4.609

Koriat, A. (2007). *Metacognition and consciousness.* Cambridge University Press.

Koriat, A., & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review*, *103*(3), 490–517. https://doi.org/10.1037/0033-295X.103.3.490

Le Pelley, M. E. (2012). Metacognitive monkeys or associative animals? Simple reinforcement learning explains uncertainty in nonhuman animals. *Journal of Experimental Psychology: Learning Memory and Cognition*, *38*(3), 686–708. https://doi.org/10.1037/a0026478

Le Pelley, M. E., Beesley, T., & Griffiths, O. (2011). Overt Attention and Predictiveness in Human Contingency Learning. *Journal of Experimental Psychology: Animal Behavior Processes*, *37*(2), 220–229. https://doi.org/10.1037/a0021384

Le Pelley, M. E., & McLaren, I. P. (2003). Learned associability and associative change in human causal learning. *Quarterly Journal of Experimental Psychology Section B: Comparative and Physiological Psychology*, *56 B*(1), 68–79. https://doi.org/10.1080/02724990244000179

Le Pelley, M. E., Turnbull, M. N., Reimers, S. J., & Knipe, R. L. (2010). Learned predictiveness effects following single-cue training in humans. *Learning and Behavior*, *38*(2), 126–144. https://doi.org/10.3758/LB.38.2.126

Le Pelley, M. E., Mitchell, C. J., Beesley, T., George, D. N., & Wills, A. J. (2016). Attention and associative learning in humans: An integrative review. *Psychological Bulletin*, *142*(10), 1111–1140. https://doi.org/10.1037/bul0000064

Le Pelley, M. E., Pearson, D., Porter, A., Yee, H., & Luque, D. (2019). Oculomotor capture is influenced by expected reward value but (maybe) not predictiveness. *Quarterly Journal of Experimental Psychology*, *72*(2), 168–181. https://doi.org/10.1080/17470218.2017.1313874

Le Pelley, M. E., Vadillo, M., & Luque, D. (2013). Learned predictiveness influences rapid attentional capture: Evidence from the dot probe task. *Journal of Experimental Psychology: Learning Memory and Cognition*, *39*(6), 1888–1900. https://doi.org/10.1037/a0033700

Lebreton, M., Abitbol, R., Daunizeau, J., & Pessiglione, M. (2015). Automatic integration of confidence in the brain valuation signal. *Nature neuroscience*, *18*(8), 1159–67. https://doi.org/10.1038/nn.4064

Leonesio, R. J., & Nelson, T. O. (1990). Do Different Metamemory Judgments Tap the Same Underlying Aspects of Memory? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(3), 464–470. https://doi.org/10.1037/0278-7393.16.3.464

Leong, Y. C., Radulescu, A., Daniel, R., DeWoskin, V., & Niv, Y. (2017). Dynamic Interaction between Reinforcement Learning and Attention in Multidimensional Environments. *Neuron*, *93*(2), 451–463. https://doi.org/10.1016/j.neuron.2016.12.040

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

Logan, G. D., & Crump, M. J. C. (2010). Cognitive Illusions of Authorship Reveal Hierarchical Error Detection in Skilled Typists. *Science*, *330*(6004), 683–686. https://doi.org/10.1126/science.1190483

Luque, D., Molinero, S., Jevtović, M., & Beesley, T. (2020). Testing the automaticity of an attentional bias towards predictive cues in human associative learning. *Quarterly Journal of Experimental Psychology*, *73*(5), 762–780. https://doi.org/10.1177/1747021819897590

Lysaker, P. H., Vohs, J., Hamm, J. A., Kukla, M., Minor, K. S., de Jong, S., van Donkersgoed, R., Pijnenborg, M. H., Kent, J. S., Matthews, S. C., Ringer, J. M., Leonhardt, B. L., Francis, M. M., Buck, K. D., & Dimaggio, G. (2014). Deficits in metacognitive capacity distinguish patients with schizophrenia from those with prolonged medical adversity. *Journal of Psychiatric Research*, *55*(1), 126–132. https://doi.org/10.1016/j.jpsychires.2014.04.011

Ma, D., Correll, J., & Wittenbrink, B. (2015). The chicago face database: A free stimulus set of faces and norming data. *Behavior research methods*, *47*. https://doi.org/10.3758/s13428-014-0532-5

Mackintosh, N. J. (1975). A theory of attention: Variations in the associability of stimuli with reinforcement. *Psychological Review*, *82*(4), 276–298. https://doi.org/10.1037/h0076778

Manohar, S. G., & Husain, M. (2013). Attention as foraging for information and value. *Frontiers in Human Neuroscience*, *7*(November), 1–16. https://doi.org/10.3389/fnhum.2013.00711
Metcalfe, J., Schwartz, B. L., & Joaquim, S. G. (1993). The Cue-Familiarity Heuristic in Metacognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(4), 851–861. https://doi.org/10.1037/0278-7393.19.4.851
Meyniel, F., & Dehaene, S. (2017). Brain networks for confidence weighting and hierarchical inference during probabilistic learning. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(19), E3859–E3868. https://doi.org/10.1073/pnas.1615773114
Meyniel, F., Sigman, M., & Mainen, Z. F. (2015). Confidence as Bayesian Probability: From Neural Origins to Behavior. *Neuron*, *88*(1), 78–92. https://doi.org/10.1016/j.neuron.2015.09.039
Mitchell, C. J., Griffiths, O., Seetoo, J., & Lovibond, P. F. (2012). Attentional mechanisms in learned predictiveness. *Journal of Experimental Psychology: Animal Behavior Processes*, *38*(2), 191–202. https://doi.org/10.1037/a0027385
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
Morgan, S. T., Hansen, J. C., & Hillyard, S. A. (1996). Selective attention to stimulus location modulates the steady-state visual evoked potential. *Proceedings of the National Academy of Sciences of the United States of America*, *93*(10), 4770–4774. https://doi.org/10.1073/pnas.93.10.4770
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*, e11946.
Naccache, L., Dehaene, S., Cohen, L., Habert, M. O., Guichart-Gomez, E., Galanaud, D., & Willer, J. C. (2005). Effortless control: Executive attention and conscious feeling of mental effort are dissociable. *Neuropsychologia*, *43*(9), 1318–1328. https://doi.org/10.1016/j.neuropsychologia.2004.11.024
Nelson, T. O., & Narens, L. (1990). Metamemory: A Theoretical Framework and New Findings. *The Psychology of Learning and Motivation*, *26*, 125–173. https://doi.org/10.1016/S0079-7421(08)60053-5
Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, *84*(3), 231–259. https://doi.org/10.1037/0033-295X.84.3.231
Norcia, A. M., Appelbaum, L. G., Ales, J. M., Cottereau, B. R., & Rossion, B. (2015). The steady state VEP in research. *Journal of Vision*, *15*(6), 1–46. https://doi.org/10.1167/15.6.4.doi
Odegaard, B., Grimaldi, P., Cho, S. H., Peters, M. A., Lau, H., & Basso, M. A. (2018). Superior colliculus neuronal ensemble activity signals optimal rather than subjective confidence. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(7), E1588–E1597. https://doi.org/10.1073/pnas.1711628115
Pelli, D. G. (1997). The videotoolbox software for visual psychophysics: Transforming numbers into movies. *Spatial vision*, *10*(4), 437–442.

Penrod, S., & Cutler, B. (1995). Witness Confidence and Witness Accuracy: Assessing Their Forensic Relation. *Psychology, Public Policy, and Law*, *1*(4), 817–845. https://doi.org/10.1037/1076-8971.1.4.817

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., & Yeung, N. (2018). The role of decision confidence in advice-taking and trust formation. *arXiv preprint arXiv:1809.10453*, (October). http://arxiv.org/abs/1809.10453

Pescetelli, N., & Yeung, N. (2020). The role of decision confidence in advice-taking and trust formation. *ournal of Experimental Psychology: General.* http://arxiv.org/abs/1809.10453

Pierce, C., & Jastrow, J. (1884). On small differences of sensation. https://doi.org/52,281-302

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

Price, P. C., & Stone, E. R. (2004). Intuitive Evaluation of Likelihood Judgment Producers: Evidence for a Confidence Heuristic. *Journal of Behavioral Decision Making*, *17*(1), 39–57. https://doi.org/10.1002/bdm.460

Qiu, L., Su, J., Ni, Y., Bai, Y., Zhang, X., Li, X., & Wan, X. (2018). The neural system of metacognition accompanying decision-making in the prefrontal cortex. *PLoS Biology*, *16*(4), 1–27. https://doi.org/10.1371/journal.pbio.2004037

Rabbitt, P. M. (1966). Errors and error correction in choice-response tasks. *Experimental Psychology*, *71*(2), 264–272. https://doi.org/10.1037/h0022853

Ramsøy, T. Z., & Overgaard, M. (2004). Introspection and subliminal perception. *Phenomenology and the Cognitive Sciences*, *3*(1), 1–23. https://doi.org/10.1023/b:phen.0000041900.30172.e8

Reder, L. M., & Ritter, F. E. (1992). What determines initial feeling of knowing? Familiarity with question terms, not with the answer. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(3), 435–451. https://doi.org/10.1037/0278-7393.18.3.435

Rescorla, R. A. (1972). A theory of pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. *Current research and theory*, 64–99.

Resulaj, A., Kiani, R., Wolpert, D. M., & Shadlen, M. N. (2009). Changes of mind in decision-making. *Nature*, *461*(7261), 263–266. https://doi.org/10.1038/nature08275

Richardson, J., & Erlebacher, A. (1958). Associative connection between paired verbal items. *Journal of Experimental Psychology*, *56*(1), 62–69. https://doi.org/10.1037/h0049034

Risko, E. F., & Dunn, T. L. (2015). Storing information in-the-world: Metacognition and cognitive offloading in a short-term memory task. *Consciousness and Cognition*, *36*, 61–74. https://doi.org/10.1016/j.concog.2015.05.014

Risko, E. F., & Gilbert, S. J. (2016). Cognitive Offloading. *Trends in Cognitive Sciences*, *20*(9), 676–688. https://doi.org/10.1016/j.tics.2016.07.002

Rouault, M., Dayan, P., & Fleming, S. M. (2019). Forming global estimates of self-performance from local confidence. *Nature Communications*, *10*(1), 1–11. https://doi.org/10.1038/s41467-019-09075-3

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

Samaha, J., Switzky, M., & Postle, B. R. (2019). Confidence boosts serial dependence in orientation estimation. *Journal of Vision*, *19*(4), 1–13. https://doi.org/10.1167/19.4.25

Sandberg, K., Timmermans, B., Overgaard, M., & Cleeremans, A. (2010). Measuring consciousness: Is one measure better than the other? *Consciousness and Cognition*, *19*(4), 1069–1078. https://doi.org/10.1016/j.concog.2009.12.013

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

Scolari, M., Ester, E. F., & Serences, J. T. (2014). Feature-and object-based attentional modulation in the human visual system. *The oxford handbook of attention.*

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

Shea, N., & Frith, C. D. (2019). The Global Workspace Needs Metacognition. *Trends in Cognitive Sciences*, *23*(7), 560–571. https://doi.org/10.1016/j.tics.2019.04.007

Shenhav, A., Musslick, S., Lieder, F., Kool, W., Griffiths, T. L., Cohen, J. D., & Botvinick, M. M. (2017). Toward a Rational and Mechanistic Account of Mental Effort. *Annual Review of Neuroscience*, *40*(1), 99–124. https://doi.org/10.1146/annurev-neuro-072116-031526

Siedlecka, M., Paulewicz, B., & Wierzchoń, M. (2016). But I Was So Sure! Metacognitive Judgments Are Less Accurate Given Prospectively than Retrospectively. *Frontiers in Psychology*, *7*(February), 1–8. https://doi.org/10.3389/fpsyg.2016.00218

Smith, J. D. (2009). The study of animal metacognition. *Trends in Cognitive Sciences*, *13*(9), 389–396. https://doi.org/10.1016/j.tics.2009.06.009

Smith, J. D., Zakrzewski, A. C., & Church, B. A. (2016). Formal models in animal-metacognition research: the problem of interpreting animals' behavior. *Psychonomic Bulletin and Review*, *23*(5), 1341–1353. https://doi.org/10.3758/s13423-015-0985-2

Sniezek, J., & Buckley, T. (1989). Social influence in the advisor-judge relationship. *Annual meeting of the Judgment and Decision Making Society, Atlanta, Georgia.*

Sniezek, J., & Buckley, T. (1995). Cueing and cognitive conflict in judge-advisor decision making. https://doi.org/10.1006/obhd.1995.1040

Sniezek, J. A., & Van Swol, L. M. (2001). Trust, confidence, and expertise in a judge-advisor system. *Organizational Behavior and Human Decision Processes*, *84*(2), 288–307. https://doi.org/10.1006/obhd.2000.2926

Solomon, R. L. (1948). The influence of work on behavior. *Psychological Bulletin*, *45*(1), 1–40. https://doi.org/10.1037/h0055527

Steinhauser, M., & Andersen, S. K. (2019). Rapid adaptive adjustments of selective attention following errors revealed by the time course of steady-state visual evoked potentials. *NeuroImage*, *186*(July 2018), 83–92. https://doi.org/10.1016/j.neuroimage.2018.10.059

Sumner, P., Nachev, P., Morris, P., Peters, A. M., Jackson, S. R., Kennard, C., & Husain, M. (2007). Human Medial Frontal Cortex Mediates Unconscious Inhibition of Voluntary Action. *Neuron*, *54*(5), 697–711. https://doi.org/10.1016/j.neuron.2007.05.016

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction*. MIT press.

Tenney, E. R., MacCoun, R. J., Spellman, B. A., & Hastie, R. (2007). Calibration trumps confidence as a basis for witness credibility: Research report. *Psychological Science*, *18*(1), 46–50. https://doi.org/10.1111/j.1467-9280.2007.01847.x

Theeuwes, J. (2019). Goal-driven, stimulus-driven, and history-driven selection. *Current Opinion in Psychology*, *29*, 97–101. https://doi.org/10.1016/j.copsyc.2018.12.024

Tottenham, N., Tanaka, J. W., Leon, A. C., McCarry, T., Nurse, M., Hare, T. A., Marcus, D. J., Westerlund, A., Casey, B., & Nelson, C. (2009). The NimStim set of facial expressions: Judgments from untrained research participants. *Psychiatry Research*, *168*(3), 242–249. https://doi.org/10.1016/j.psychres.2008.05.006

Uengoer, M., Lachnit, H., Lotz, A., Koenig, S., & Pearce, J. M. (2013). Contextual control of attentional allocation in human discrimination learning. *Journal of Experimental Psychology: Animal Behavior Processes*, *39*(1), 56–66. https://doi.org/10.1037/a0030599

Uengoer, M., Pearce, J. M., Lachnit, H., & Koenig, S. (2018). Context modulation of learned attention deployment. *Learning and Behavior*, *46*(1), 23–37. https://doi.org/10.3758/s13420-017-0277-y

Underwood, B. J. (1966). Individual and group predictions of item difficulty for free learning. *Journal of Experimental Psychology*, *71*(5), 673–679. https://doi.org/10.1037/h0023107

Van Den Berg, R., Anandalingam, K., Zylberberg, A., Kiani, R., Shadlen, M. N., & Wolpert, D. M. (2016). A common mechanism underlies changes of mind about decisions and confidence. *eLife*, *5*(FEBRUARY2016), 1–21. https://doi.org/10.7554/eLife.12192

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*, *26*(23), 3157–3168. https://doi.org/10.1016/j.cub.2016.10.021

van der Plas, E., David, A. S., & Fleming, S. M. (2019). Advice-taking as a bridge between decision neuroscience and mental capacity. *International Journal of Law and Psychiatry*, *67*(September 2019), 101504. https://doi.org/10.1016/j.ijlp.2019.101504

Van Swol, L. M., Paik, J. E., & Prahl, A. (2018). Advice recipients: The psychology of advice utilization.

Vickers, D. (1979). *Decision processes in visual perception*. Academic Press.

Vossel, S., Mathys, C., Stephan, K. E., & Friston, K. J. (2015). Cortical coupling reflects Bayesian belief updating in the deployment of spatial attention. *Journal of Neuroscience*, *35*(33), 11532–11542. https://doi.org/10.1523/JNEUROSCI.1382-15.2015

Walton, M. E., Kennerley, S. W., Bannerman, D. M., Phillips, P. E. M., & Rushworth, M. F. S. (2006). Weighing up the benefits of work: Behavioral and neural analyses of effort-related decision making. *Neural Networks*, *19*(8), 1302–1314. https://doi.org/10.1016/j.neunet.2006.03.005

Westbrook, A., Kester, D., & Braver, T. S. (2013). What Is the Subjective Cost of Cognitive Effort? Load, Trait, and Aging Effects Revealed by Economic Preference. *PLoS ONE*, *8*(7), 1–8. https://doi.org/10.1371/journal.pone.0068210

Westbrook, A., Lamichhane, B., & Braver, T. (2019). The Subjective Value of Cognitive Effort is Encoded by a Domain-General Valuation Network. *The Journal of Neuroscience*, *39*(20), 3934–3947. https://doi.org/10.1523/JNEUROSCI.3071-18.2019

Yaniv, I. (2004). Receiving other people's advice: Influence and benefit. *Organizational Behavior and Human Decision Processes*, *93*(1), 1–13. https://doi.org/10.1016/j.obhdp.2003.08.002

Yaniv, I., & Kleinberger, E. (2000). Advice Taking in Decision Making: Egocentric Discounting and Reputation Formation. *Organizational Behavior and Human Decision Processes*, *83*(2), 260–281. https://doi.org/10.1006/obhd.2000.2909

Yeung, N. (2014). Conflict monitoring and cognitive control.

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), 1310–21. https://doi.org/10.1098/rstb.2011.0416

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

Zarnoth, P., & Sniezek, J. A. (1997). The Social Influence of Confidence in Group Decision Making. *Journal of Experimental Social Psychology*, *33*(4), 345–366. https://doi.org/10.1006/jesp.1997.1326