Perceptual decision making, and the construction of confidence

Joshua Calder-Travis

St John's College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity 2020

Abstract

Perceptual decisions are of fundamental interest, and it is increasingly recognised that a sense of confidence supports decision-making. The drift diffusion model (DDM), provides an excellent account of decisions and response times. It also features the optimal property of tracking the difference in evidence between two options. However, the DDM struggles to account for human confidence reports. Possibly because of this, much confidence research has used non-normative models of the decision mechanism. Motivated by the idea that perceptual decision-making will reflect optimal computation, we consider 10 variants of the DDM. Motivated by the idea that the brain will not duplicate the representation of evidence, in all model variants confidence is read out from the decision mechanism.

We compare the models to benchmark findings and make 3 qualitative predictions that we verify in a preregistered study. We develop new computationally cheap predictions for confidence in DDM accounts. Using these predictions we model confidence on a trial-by-trial basis, finding that a subset of model variants provide an excellent account of the precise quantitative effects observed. Specifically, models in which confidence reflects a miscalibrated Bayesian readout perform best. Focusing on the idea of a Bayesian readout, we explore the claim that evidence associated with a selected option is overweighted in confidence. Although this idea appears to challenge Bayesian confidence, we do not find this overweighting in three previously collected datasets, and we suggest an explanation for previous results. Finally, we examine whether the DDM remains normative in settings where rapid learning is more important than the theoretical possibility of achieving maximal reward rate.

These results support the claim that confidence is based on the decision mechanism, which is itself optimal. In particular, the DDM with a miscalibrated Bayesian readout is supported on both theoretical and empirical grounds.

Acknowledgements

I am very grateful to my supervisor, Nick Yeung, for support, kindly-delivered criticism, and advice over the last four years. Just one example is when, after the department building was permanently closed with 48 hours' notice, he gave up his office so that his students could work there. I have very much appreciated his comments on this thesis. I would like to thank my co-supervisor, Rafal Bogacz, for discussions and feedback that shaped the entire course of this thesis. I am very grateful to my unofficial mentor, Wei Ji Ma, who has given me so much of his time and taught me so much. I very much appreciate the advice, work, time and ideas of my collaborators, Emma Slade, Lucie Charles, Matt Jaquiery and Nomi Carlebach.

For helpful discussions, advice, and comments, I am grateful to the members of the ACC lab (Oxford), members of the Ma lab (NYU), Andra Mihali, Heiko Schütt, and Luigi Acerbi. For plotting and significance testing code that I used when looking at EEG data, I am grateful to Fabrice Luyckx and the Summerfield lab (Oxford). In work throughout the thesis I relied on the University of Oxford Advanced Research Computing (ARC) facility. I am very grateful to the ESRC and St John's College, Oxford, for giving me this opportunity by funding my studies.

Previous publications and collaborations

Throughout the thesis I use data from other researchers. Wherever this is done it is clearly stated with a citation to the relevant publication. The project described in Chapter 5 was conceptualised by a collaboration that I set up. I coded and ran the analysis myself. The project described in Chapter 6 was a collaboration between me and another researcher. Modelling and analysis work was shared equally.

At the time of writing, none of the work in the thesis has been published. However, throughout the thesis I use material from four preprints (Calder-Travis, Bogacz et al., 2020; Calder-Travis, Charles et al., 2020; Calder-Travis & Ma, 2020; Calder-Travis & Slade, 2020). I use material that I have authored, and obtained permission from co-authors to use.

Data from Chapter 3 has been published as part of a large database of confidence studies (Rahnev et al., 2020). I am one of more than 80 authors on this paper, due to the contribution of this data.

"No part of the aim of normal science is to call forth new sorts of phenomena; indeed those that will not fit the box are often not seen at all."

"By focusing attention upon a small range of relatively esoteric problems, the paradigm forces scientists to investigate some part of nature in a detail and depth that would otherwise be unimaginable."

Kuhn, 2012/1962, p.67,68

Contents

1 Introduction

Contents

1.1 The role and relevance of confidence

How animals and humans make perceptual decisions is a fundamental question for psychology, cognitive science, and neuroscience. By "perceptual decisions" we refer

to choices made on the basis of sensory information (Usher & McClelland, 2001). In this thesis we will be specifically concerned with categorical choices. Choosing whether to cross the road on the basis of visual information about nearby objects is an example of a categorical perceptual decision. A fish choosing whether to swim away from motion in the distance is making a categorical perceptual decision. So too is someone walking to the door because they think they heard the post come through the letter box. It is increasingly recognised that a sense of confidence can support decision-making, and is used by humans in a number of ways (Bahrami et al., 2010; Drugowitsch et al., 2019; van den Berg, Zylberberg et al., 2016). Intuitively, decision confidence conveys a sense of whether a "good" decision has been reached. The precise notion of "goodness" conveyed in human confidence reports is a matter for empirical investigation, and one which we will discuss. However, unless noted, we will use the term confidence to refer to a particular statistical notion of "goodness": The probability that a decision is correct (Aitchison et al., 2015; Kiani et al., 2014; Meyniel et al., 2015; Pew, 1969; Sanders et al., 2016; Vickers & Packer, 1982).

An estimate of the probability that a decision was correct is potentially highly beneficial in a wide range of situations. Consider an owl flying after a glimpse of a grey object in the grass. The energy it exerts in the chase would sensibly be moderated by the probability it saw a mouse. In humans, confidence signals have even greater value. For example, they permit humans to communicate more information about their decisions, than simply the decision itself. In some circumstances, communication of confidence allows greater performance in group decisions, than could be achieved if members of the group only conveyed their own decision (Bahrami et al., 2010). Confidence is also highly useful in less obvious ways. For example, confidence is a key variable in a normative mechanism for learning from feedback (Drugowitsch et al., 2019).

Consistent with the apparent value of an estimate of the probability of being correct, there is strong evidence that confidence is reliably related to objective performance (Sanders et al., 2016). Moreover, there is evidence that humans use

confidence in a variety of settings. When faced with sequences of two decisions, where both decisions must be made correctly for reward, humans flexibly adjust the time spent on the second decision (van den Berg, Zylberberg et al., 2016). The adjustment is made according to the probability they were correct in their first decision, and hence, that they have a chance of obtaining reward. In the case of sequences of decisions where reward on each trial is independent of reward on other trials, confidence is still used. Here, confidence guides the balance between speed and accuracy in subsequent decisions: Low confidence suggests that more caution should be taken in future decisions (Desender et al., 2019). Considering only a single decision, there is evidence confidence is used to determine whether to seek further information (Desender et al., 2018). Additionally, confidence may be used when a specific level of performance is desired, to guide the amount of time spent deliberating (Balsdon et al., 2020). When people have a choice over the task they perform, confidence in previous decisions is used, with people selecting tasks in which they have previously experienced greater confidence (Carlebach & Yeung, 2020). When making decisions in pairs, people appear to communicate and use their confidence to reach a joint decision when there are disagreements (Bahrami et al., 2010).

Further supporting the claim that confidence plays an important role in behaviour, changes in the relationship between confidence and objective performance have been linked to psychological disorder. The relationship between perceptual information and confidence appears to be weaker in people with high compulsivity (Hauser et al., 2017; Rouault et al., 2018). This relationship appears to be reversed for other aspects of psychological disorder: Anxiety and depression symptoms actually predict that confidence reports more closely reflect perceptual information (Rouault et al., 2018). These contrasting patterns observed for different psychiatric symptoms emphasise the value of a mechanistic understanding of confidence. A mechanistic understanding of confidence may allow us to suggest how the different patterns could arise, through the effects of different forms of psychological disorder on different parts of the mechanism.

In addition to providing a detailed account of an important process, a mechanistic understanding of confidence may illuminate how the brain represents evidence, probability, and probability distributions. As we will see, there remains much debate about the representation of evidence which supports perceptual decisions. Empirical phenomena observed in confidence data may provide key constraints for distinguishing between competing theories (e.g. Zylberberg et al., 2012). Furthermore, while there is much evidence in several fields for approximately optimal computation under uncertainty (Ernst & Banks, 2002; Norris, 2006; Tassinari et al., 2006), it remains unclear when, how, and to what extent, probability distributions are represented in the brain (Ma & Jazayeri, 2014; Sahani & Dayan, 2003). Investigating confidence, which may reflect a readout of the probability of being correct (Meyniel et al., 2015; Sanders et al., 2016), appears a very natural way to ask questions about the representation of probability and uncertainty in the brain. Given this importance, there is growing interest in understanding the computations responsible for confidence.

1.2 Accounts of perceptual decision-making

We begin by considering prominent accounts of how evidence is represented in the brain, for use in perceptual decisions. Many of these models aim to account for perceptual decisions regardless of task, modality, or number of options (e.g. Green & Swets, 1966; Ratcliff et al., 2016; Tajima et al., 2019; Usher & McClelland, 2001). Later we will turn to the question of whether the same representation is used to support confidence judgements, and in what manner.

A wide range of models assume that sensory input is reduced to a simple representation of the evidence for the different options in a perceptual decision. The fact that human behaviour is variable given identical perceptual stimuli suggests that representations of perceptual evidence are corrupted by noise (Drugowitsch et al., 2016). Typically, either a single noisy variable represents the balance of evidence between the two options, or one noisy variable for each option conveys, in

some sense, the evidence for that option (Bogacz et al., 2006; Vickers & Packer, 1982). Accounts of the representation of evidence can be divided into two groups depending on whether they assume that a single static representation of the evidence is constructed, or whether they assume that the representation of evidence changes over time (Pike, 1973).

1.2.1 Static representations of evidence

Signal detection theory (SDT) is the most prominent account that assumes a static representation of evidence (Green & Swets, 1966; Palmer et al., 1993). On this view, evidence for a choice between two alternatives can be represented by a single noisy variable, regardless of the nature of the choice. For example, if the choice involves determining the presence or absence of an item in a display of many items, the theory assumes that the observer combines sensory information about all items into a single variable, which might represent the balance of evidence between the two choices, or some other closely related quantity (Palmer et al., 2000). The observer then applies a threshold to this variable, reporting option A when the balance of evidence exceeds the threshold, and reporting option B otherwise.

SDT encompasses a range of approaches, including optimal observer SDT models (Green & Swets, 1966; Palmer et al., 2000; Rosenholtz, 2001). Bayes rule provides the mathematical formalism that can be used to optimally combine noisy signals and make inferences about the environment (Ma & Jazayeri, 2014; Ma et al., 2011). The Bayesian SDT observer computes a very specific single variable from the noisy stimulus representation, namely the posterior ratio. The posterior ratio is the ratio of the probability that option A is correct and the probability that option B is correct, given the observer's measurements (Ma et al., 2011; Palmer et al., 2000; Peterson et al., 1954). Consider a case where the set of items in a display, denoted $\boldsymbol{s}$, lead to noisy measurements of those stimuli, $\boldsymbol{x}$, according to some probability distribution $p(\boldsymbol{x}|\boldsymbol{s})$. The observer's task is to determine which of two sets of stimuli, $\boldsymbol{s}_A$ or $\boldsymbol{s}_B$, are present. We often assume that the observer

has learned the statistical structure of the task, including the relationship between stimuli and measurements, and the prior probability of the two sets of stimuli, $p(\boldsymbol{s}_A)$ and $p(\boldsymbol{s}_B)$ (Ma & Jazayeri, 2014; Ma et al., 2011). The observer can then compute the posterior ratio, Θ, using Bayes rule:

$$\Theta = \frac{p(\boldsymbol{s}_B|\boldsymbol{x})}{p(\boldsymbol{s}_A|\boldsymbol{x})} = \frac{p(\boldsymbol{x}|\boldsymbol{s}_B)}{p(\boldsymbol{x}|\boldsymbol{s}_A)} \frac{p(\boldsymbol{s}_B)}{p(\boldsymbol{s}_A)} \tag{1.1}$$

The posterior ratio, Θ, then becomes the noisy variable to which a threshold is applied to determine the response (Ma et al., 2011). The optimal observer uses a specific threshold, and selects the option that maximises expected reward (Ma & Jazayeri, 2014). When the two options are equally rewarded, this will simply be the option with the highest probability of being the correct option.

SDT models have proved successful models of choices in a wide range of domains, ranging from visual search to memory (Palmer et al., 2000; Wixted, 2007). The idea that we use Bayesian inference to make decisions has received support in domains as diverse as multi-sensory decisions (Ernst & Banks, 2002), and word recognition (Norris, 2006). Another major advantage of SDT models is their *relative* simplicity. As they posit only a single static representation of the stimulus and associated evidence variable, it is often computationally feasible to make trial-by-trial predictions for responses, given the specific stimuli presented (see Appendix G for a study using trial-by-trial SDT modelling that I have done during my DPhil, but which falls outside the scope of this thesis). Because these models predict behaviour on a trial-by-trial basis using the full stimulus, rather than a summary of the stimulus, we can generate predictions for the relationship between any stimulus statistic and any behavioural statistic. In contrast, as we will see, when modelling data with more complex models, computational cost often prohibits trial-by-trial modelling, and necessitates condition-by-condition modelling. That is, for the purpose of modelling, all trials in a condition are treated as the same. Such an approach discards the rich properties of stimuli that are commonly used.

The main disadvantage of SDT models as an account of perceptual decisions is that they appear to neglect a whole set of phenomena. In particular, as these

models are based on a static representation of evidence they make no predictions, and do not account for, patterns in response time data (Pleskac & Busemeyer, 2010). Of course we could supplement SDT models with the claim that response times are determined by an additional random variable unrelated to the decision process, such as random fluctuations in the duration of motor preparation (Luce, 1986; Pike, 1973). However, response times do not appear to vary completely randomly (Pike, 1973). Instead, response times are systematically related to accuracy and confidence (Pleskac & Busemeyer, 2010). This suggests some (possibly indirect) causal relationship between the representation of evidence, which informs choices and confidence, and response time. Moreover, humans can trade speed for accuracy (Garrett, 1922; Heitz, 2014; Wickelgren, 1977). One sensible explanation for this phenomenon is that a stream of evidence measurements are received over time, and are averaged to reduce noise and increase performance (Bogacz et al., 2006; Ratcliff & McKoon, 2008; Ratcliff & Rouder, 1998). Appropriately and flexibly trading speed for accuracy is arguably of fundamental importance to survival: The longer a seal waits, the more likely it is to correctly identify the fish in the distance. However, waiting is itself costly. In the case of the seal, the longer it waits, the more likely it is that the fish (whether predator or prey) will spot them. Hence, "stopping rules" (Ferguson, 2011), not just decision rules, are important to survival. An adequate account of perceptual decisions will need to be able to describe and explain the stopping rules that are actually used.

1.2.2 Dynamic representations of evidence

The DDM

The idea that, instead of a static representation of evidence, evidence samples are continuously received over time, is central to many models of perceptual decision making (Bogacz et al., 2006). It is often assumed that, given two options, two streams of evidence measurements are generated. For example, in a binary choice regarding which of two arrays contains most dots, the observer might receive two

sets of noisy evidence measurements, one corresponding to the number of dots in the left array, and one corresponding to the number of dots in the right array.

The drift diffusion model (DDM) is one of the most prominent models of two-alternative decisions, from a family of models in which observers receive noisy measurements of evidence for the two options (Fig. 1.1; Bogacz et al., 2006; Ratcliff and McKoon, 2008). In the DDM, observers track the difference in evidence measurements between the two options. That is, for each sample, observers subtract the measurement for option B from the measurement for option A, and add this to a running total (Ratcliff & McKoon, 2008). The average rate of accumulation is called the drift-rate. The drift-rate is determined by the difference between the evidence signals for the two alternatives (not the noisy measurements of those signals). Whilst the drift-rate determines the average path of the accumulator, noise in evidence measurements causes the accumulator to fluctuate randomly around this path. The accumulator begins near zero, and a decision is triggered when the accumulator reaches either a positive or negative threshold. The threshold reached determines the choice, and the time taken to reach it the response time.

While there are important exceptions (Rafiei & Rahnev, 2020), the DDM enjoys 50 years of empirical support (Ratcliff et al., 2016). The model, and extensions, accounts well for the shape of response time distributions, and how accuracy and response time change with task difficulty (Ratcliff, 1978; Ratcliff & McKoon, 2008). Assuming trial-to-trial variability in drift-rate, and in the starting point of the accumulator, the model can also explain why in some cases error responses are faster than correct responses, but in other cases correct responses are faster than errors (Ratcliff & McKoon, 2008; Ratcliff et al., 1999). It has proved a unifying framework, explaining data in perceptual decisions (Ratcliff & McKoon, 2008), value-based decisions (Milosavljevic et al., 2010), and memory tasks (Ratcliff, 1978). Crucially, in relation to our discussion of SDT, the DDM can naturally account for how humans trade speed for accuracy: In a context where accuracy is more important than speed, observers can set decision thresholds that are far from the

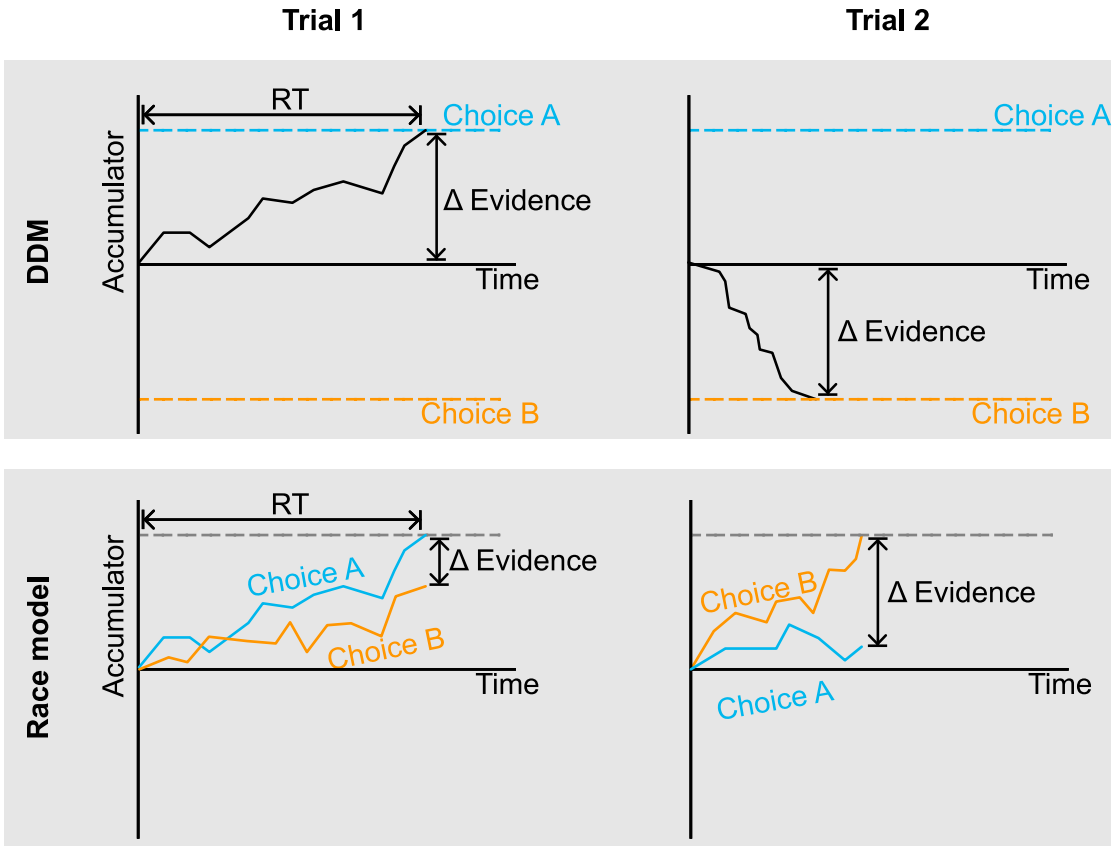

Figure 1.1: In the drift diffusion model (DDM; Ratcliff and McKoon, 2008), the observer accumulates the difference in evidence measurements for two options. When the difference in evidence reaches a threshold value, a response is triggered. In the race model, the observer separately accumulates evidence measurements for choice A and for choice B (Bogacz et al., 2006; Vickers, 1970). When either of the two accumulators reach a threshold value, a response is triggered. Due to the criterion used for triggering a response, in the DDM, observers end every trial with the same difference in evidence between the chosen and unchosen alternative (Yeung & Summerfield, 2014). In the race model, the final difference in evidence varies, because the unchosen accumulator can be anywhere below the decision threshold at the time of decision (Kiani et al., 2014).

starting point of the accumulator (Ratcliff & Rouder, 1998). In this case, it will take longer to reach a threshold, but when reached, a great deal of evidence will have been accumulated in favour of the decision made. The reverse holds when speed is more important than accuracy.

In addition to accounting for key patterns in data, the DDM is also a normative model of decision making. Throughout the thesis, by "normative model" we mean a model in which the observer makes optimal decisions on the basis of the information that is assumed to be available to them (Rahnev & Denison, 2018). This contrasts with models, that we will encounter below, which assume the observer has access to a certain representation of evidence but they use this evidence sub-optimally to determine their decision. We use this definition of "normative" because claims about the optimality of an observer are meaningless unless we make assumptions about what information is available to that observer (Rahnev & Denison, 2018). We assume this definition of "normative" throughout the thesis, and return to further discuss this point in Section 7.3.3.

To see why the DDM is a normative model in the sense described, we must first note that if evidence measurements for the two options are equally reliable and signal strength is known then, under standard noise assumptions, the optimal algorithm involves comparing evidence measurements for the two alternatives, and accumulating this difference over time, until a specific accumulated difference in evidence is reached (Bogacz et al., 2006; Gold & Shadlen, 2007; Moran, 2015; Tajima et al., 2019; Wald & Wolfowitz, 1948). This is exactly the policy which the DDM represents. This policy is optimal in the sense that, on average, no other policy can achieve the same level of performance with fewer evidence measurements (Wald & Wolfowitz, 1948). It makes intuitive sense that the optimal policy involves subtracting evidence measurements for one option from evidence measurements for the alternative: If there are only two alternatives, information which makes one option more probable makes the other less probable. Indeed, in the context discussed, the DDM is equivalent to tracking the posterior probability of each

option until a fixed threshold on these probabilities is reached (Bitzer et al., 2014; Gold & Shadlen, 2007; Moran, 2015).

When signal strength is unknown, a time-dependent threshold on the accumulated difference in evidence is required for optimality, in the sense of maximising reward rate (Drugowitsch et al., 2012; Tajima et al., 2019). Nevertheless, tracking the difference between the two accumulators remains part of the optimal algorithm. For an intuition of why a decreasing threshold can be optimal when signal strength is unknown, consider a case in which the observer hasn't made a decision after lengthy deliberation. The observer must be accumulating evidence very slowly, suggesting to them that signal strength is very low. If the observer thinks signal strength is very low, there is almost nothing to gain from collecting more evidence measurements, and doing so would cost the observer time, so they should lower their decision threshold and make an immediate decision (Malhotra et al., 2017).

Alternatives to the DDM

While the DDM is arguably the most widely used model of perceptual decisions, there are a large number of alternative models featuring properties that in some way deviate from the normative DDM.

In one alternative to the DDM, which we refer to as the race model, instead of accumulating the difference between evidence measurements, evidence measurements for option A are accumulated separately from evidence measurements for option B (Fig. 1.1; Bogacz et al., 2006; Vickers, 1970). In the race model, a decision is triggered when either of these two accumulators reaches a threshold, with the choice determined by the accumulator to reach the threshold first. A number of other models have similar accumulation mechanisms in the sense that response and response time is determined by a race between two accumulators. We will refer to this family of models as those with "race-like accumulation". One example of a model in this family is the accumulator model (Smith & Vickers, 1988). In this model, the decision mechanism does not receive pairs of evidence measurements,

one for each option, but single values reflecting a noisy measure of the difference in evidence. A comparison is conducted using this noisy sample (Ratcliff & Smith, 2004). If the evidence sample is greater than some reference value, then accumulator A is incremented by the absolute difference between the sample and the reference. Accumulator B does not change. The reverse holds if the sample is smaller than the reference value. In another model featuring race-like accumulation, the Poisson counter model (Pike, 1973; Ratcliff & Smith, 2004), it is again assumed that the values of the accumulators never decrease. In this model the accumulators function as counters, with the count of an accumulator increasing by one, at random times in accordance with a Poisson distribution. If a stimulus provides more evidence for option A, then the rate of accumulator increments will be higher, making it more likely that accumulator A will reach the decision threshold first.

The range of models with race-like accumulation complicates empirical and theoretical comparison to the DDM, but we can get an intuitive understanding for why all race-like accumulation dynamics are non-normative. Consider a case in which both accumulators are close to the threshold at the time of decision. The observer has almost the same amount of evidence for both options, and is effectively responding randomly, yet the observer makes a response. Guessing can be an appropriate behaviour, for example, if one has been deliberating for a long time (Drugowitsch et al., 2012; Malhotra et al., 2017) but, in models with race-like accumulation, guessing can happen at any time point. The opposite situation can also arise. Consider a case in which accumulator A has a great deal more evidence than accumulator B. Hence, the observer can be almost certain option A is correct. In models with race-like accumulation, regardless of how confident the observer can be, if accumulator A hasn't reached its threshold the observer continues deliberating, wasting time. In a sense, models with race-like accumulation imply that observers waste half the information available to them, when setting their response time.

In terms of empirical comparison, tests of specific models with race-like accumulation have been conducted, along with tests which apply fairly generally

to the whole family of models. Ratcliff and Smith (2004) compared the DDM with the accumulator model and the Poisson counter model. The Poisson counter model performed poorly. The accumulator model performed better, but could not account for a key qualitative effect observed in response time data under certain conditions: Error response times which are on average faster than correct response times. At a broader level, Teodorescu and Usher (2013) note that for any model with race-like accumulation, and in which the input to one of the two accumulators can be manipulated without affecting the input to the other, we can make a very strong prediction. Specifically, if we increase the input to one accumulator without affecting the input to the other, models with race-like accumulation predict there will never be a slow down in response times. The reason is that if we present more evidence for option A, then accumulator A will reach the decision threshold faster. If it reaches the threshold before accumulator B, then it will speed up the response. If not, then because accumulator B is unaffected by accumulator A, accumulator B will reach the threshold at the time it was going to anyway, and there will be no change in response time (Miller & Ulrich, 2003; Teodorescu & Usher, 2013). Teodorescu and Usher (2013) found this prediction did not hold: They observed a slowing of response times when more evidence was presented for the incorrect option. We applied this test to data collected in our experiments and also found that the prediction of race-like accumulator models did not hold (Appendix A.1; note we also found some evidence that predictions of the DDM did not hold).

The race model and the DDM lie on a continuum of models (Bogacz et al., 2006). At one end is the race model, with two completely independent accumulators, at the other end is the DDM. The DDM is equivalent to a model in which there are two completely anti-correlated accumulators (and hence only one of these needs to be tracked). Between these two extremes lie intermediate models that feature two partially anti-correlated accumulators, where evidence measurements for option A have some effect reducing the accumulator for option B. However, the measurements for option A are not weighted as heavily as the measurements for option B, in the accumulator for option B (Bogacz et al., 2006; Moreno-Bote, 2010). The reverse

holds in the accumulator for option A. These intermediate models inherit the same criticisms regarding suboptimality which apply to race-like models, but they do not inherit the same empirical weaknesses. Models with partially anti-correlated accumulators have been successfully used to model behavioural data (Kiani et al., 2014; van den Berg, Anandalingam et al., 2016; Zylberberg et al., 2012), and don't make the strong prediction that increasing evidence for one option will never slow down response times. There have been few direct comparisons between the DDM and models with partially anti-correlated accumulators. One reason for this may be the high computational cost of making predictions from models which feature a dynamic representation of evidence. In many cases, the equations that describe model predictions have no simple solution, and must be solved numerically (Shinn et al., 2020). (Although, for a specific set of circumstances, Shan et al. (2019) recently found solutions which do not require use of the usual numerical methods.) The strength of anti-correlation between accumulators has sometimes been fit as a free parameter (Kiani et al., 2014), and the result of such fits provide some limited evidence in this debate. Teodorescu et al. (2016) reported that the value of the fitted anti-correlation parameter suggested perfect anti-correlation. In contrast, Zylberberg et al. (2012) found that partially anti-correlated accumulators could account for some features in confidence data that the DDM could not (Section 1.6 for further detail, and Chapter 5 for an alternative explanation).

Beyond the DDM to race-like continuum, there are two classes of perceptual decision models that deserve mention. First, there are models related to the leaky competing accumulator (LCA) model (Usher & McClelland, 2001). The LCA features the leaky accumulation of evidence, in other words evidence is lost from each accumulator over time, and features the inhibition of one accumulator by the other, meaning that when one accumulator has a high value it has a suppressive effect on the other accumulator. These two mechanisms have somewhat opposite effects with leaky accumulation meaning that evidence gathered early in a trial is largely lost and has little impact on a decision, while competition in the form of inhibition can have the reverse effect (Usher & McClelland, 2001). If one accumulator gains a

small advantage at the beginning of a trial, inhibition can lead to the exaggeration of this advantage, and hence, the relative importance of evidence gathered early in a trial. The LCA has performed well, although not necessarily better, when compared to the DDM (Ratcliff & Smith, 2004; Teodorescu et al., 2016). One early argument, which suggested that the LCA might provide a better description of choice and response time data, involved the claim that only the LCA could adequately capture the pattern observed in accuracy as stimuli are presented for increasingly long durations (Usher & McClelland, 2001). However, the DDM seems to perform just as well as the LCA, if not better, when we assume that participants do not observe a stimulus indefinitely, even if it continues to be presented (Ratcliff, 2006).

A second class of models that differs from both the DDM and the race model similarly assumes that observers receive a noisy stream of evidence, but discards the assumption that this evidence stream is accumulated. Instead, as evidence measurements are made, they are compared to some criterion and then quickly discarded (Cisek et al., 2009; Thura, 2016). Such models are related to models which feature leaky evidence accumulation: Here this leak is extremely high (Thura, 2016). It is surprisingly difficult to distinguish between behaviour resulting from the perfect integration of evidence and behaviour resulting from the use of only the most recent evidence, with many patterns in response time and choice data explainable from both frameworks (Stine et al., 2020). However, integration of evidence may best describe perceptual decision making in humans (although not in macaques; Hawkins, Wagenmakers et al., 2015). Recently Stine et al. (2020) showed that a strong test can be constructed by using two conditions, one in which participants are free to respond when they are ready, and one in which the researcher sets the time of response. (We will use exactly this combination of trials, as it also generates a very strong test for confidence models.) The results supported extended evidence integration, although there was some evidence of leak (Stine et al., 2020).

1.3 Accounts of confidence

Confidence reports necessitate an additional layer of modelling, in addition to the modelling required for perceptual decisions. Given the success of models of response times and choices in perceptual decisions, much work has explored whether these models can be extended to additionally account for confidence (e.g. Kiani et al., 2014; Vickers & Packer, 1982; Zylberberg et al., 2012). We will first discuss the possibility that confidence is constructed using the same representation of evidence that supports the decision, before considering the possibility that confidence is determined by additional variables separate to the decision mechanism.

1.3.1 Confidence as determined by the decision mechanism

Confidence as a function of evidence

An obvious candidate for a decision mechanism variable that could be used to determine confidence is the amount of evidence gathered (Vickers & Packer, 1982; Yeung & Summerfield, 2014). If the balance of evidence strongly favours the selected option, then confidence in that option may be high. A question that arises in models with a dynamic representation of evidence regards the timing of the readout of confidence from the evidence accumulation process. A key debate in the confidence literature has been between accounts of confidence with a "decisional locus", which propose confidence is determined by the evidence accumulated at the time of the decision, and accounts of confidence with a "post-decisional locus" (Yeung & Summerfield, 2014). The models in the latter category claim that evidence accumulation continues beyond a decision, and confidence reports may reflect the state of the accumulation process at a later time.

For all the success of the DDM explaining patterns in choices and response times, the model struggles to explain confidence reports, and this problem is particularly acute if we subscribe to a decisional locus view of confidence (Yeung & Summerfield, 2014). The reason for the difficulty is that in the DDM a decision is triggered

when the accumulator crosses a fixed threshold. Therefore, at the time someone commits to a decision, they will always have the same accumulated difference in evidence favouring the selected option (Fig. 1.1). This apparently leads to the conclusion that people and animals should have equal confidence in all attempts at a task. If evidence accumulation continues following a decision, and confidence is not read out at this time, then the DDM might be able to account for variability in confidence (Pleskac & Busemeyer, 2010). Consistent with this idea, recent research has supported the post-decisional locus view of confidence. These findings have come from studies which presented evidence following a decision, and found such evidence has a clear effect on confidence (Charles & Yeung, 2019; Moran et al., 2015), or from studies that have manipulated the time between decisions and confidence report (Yu et al., 2015). However, it is not clear that post-decisional evidence accumulation is enough to render the DDM an adequate account of confidence: Typically, perceptual decision-making experiments have used stimuli which clear on response. Even in this case, confidence remains related to performance (Sanders et al., 2016).

One successful approach to explaining confidence using the DDM has focused on the idea that people may continue to accumulate evidence measurements following a decision, even in the absence of continued stimulus presentation (Moran et al., 2015; Pleskac & Busemeyer, 2010; van den Berg, Anandalingam et al., 2016). Research on changes of mind strongly suggests that evidence accumulation does indeed continue following a decision, regardless of ongoing availability of stimulus information (Resulaj et al., 2009; van den Berg, Anandalingam et al., 2016). Specifically, both Resulaj et al. (2009) and van den Berg, Anandalingam et al. (2016) asked participants to make a perceptual decision, and to indicate their response by moving a handle to one of several targets. Importantly, in both cases, the stimulus cleared as soon as movement began. Nevertheless, Resulaj et al. (2009) observed changes of mind following the clearing of the stimulus, in the form of trials where participants initially began moving the handle to one target, before switching to the other. van den Berg, Anandalingam et al. (2016) collected confidence reports using this procedure, and also observed changes in confidence. These results can be explained

by taking into account the fact that sensory and motor processing takes time (Ratcliff & McKoon, 2008; Resulaj et al., 2009; van den Berg, Anandalingam et al., 2016). As a result, at the time of a response, there will be information about the stimulus in these processing "pipelines" (Resulaj et al., 2009; van den Berg, Anandalingam et al., 2016). A considerable amount of evidence appears to be in processing pipelines – results of previous studies suggest the pipeline contains approximately 400ms of information from the stimulus – and evidently enough to completely reverse the initial decision (Ratcliff & McKoon, 2008; Resulaj et al., 2009). This suggests that, even in the absence of continued stimulus presentation following a decision, the existence of pipeline evidence may allow the DDM to explain variability in confidence. Importantly, on trials in which an observer is correct, evidence measurements in the processing pipeline will tend to support the decision (Pleskac & Busemeyer, 2010). On the other hand, if the initial decision was incorrect, evidence measurements in the processing pipeline will tend to contradict it. As a result, people will have greater confidence in correct responses than in errors.

A relationship between confidence and accuracy is not the only relationship that an adequate model of confidence needs to capture. For example, confidence shows clear relationships to response time, decreasing with response time when response time is under the control of the participant, and increasing with response time when this is set by the researcher (Pleskac & Busemeyer, 2010). Pleskac and Busemeyer (2010) proposed a DDM account of confidence, which featured a post-decisional locus for the confidence readout (Fig. 1.2). By also including the idea of trial-to-trial variability in drift-rate in their model, named two-stage dynamic signal detection theory (2DSD), they were able to account qualitatively for the relationship between confidence and response time. The assumption of drift-rate variability is common, and reflects the hypothesis that an identical stimulus will not always generate identical signal strength in the associated internal evidence measurements (Ratcliff & McKoon, 2008; Ratcliff & Smith, 2004; Ratcliff et al., 1999; Voskuilen et al., 2016). On a trial in which the drift-rate is high, the threshold will be reached quickly triggering a quick response, but this high drift-rate will also

drive rapid accumulation following a decision. The accumulator will reach a large value, generating high confidence. In contrast, when drift-rate is low, it will take a long time for the accumulator to reach a decision threshold, and little evidence will be gathered following a decision, leading to reduced confidence.

Although Pleskac and Busemeyer (2010) showed that 2DSD explains an impressive number of patterns in confidence data, Pleskac and Busemeyer (2010) note that the model only provides a partial explanation for the relationship between confidence and response time. Specifically, 2DSD underestimates how strongly confidence and response time are related (Pleskac & Busemeyer, 2010, p.881). A further difficulty for DDM accounts of confidence is that confidence appears to show systematic relationships with response time, and the stimulus presented, even in experimental setups which aim to force a decisional locus for confidence reports. Kiani et al. (2014) used an elegant task design in which single ballistic eye movements permitted the collection of simultaneous decision and confidence reports. Even in this case, confidence varied with response time and the difficulty of the task. The implication of this design and result is that evidence measurements following a decision – so crucial in 2DSD, and DDM models of confidence more generally – likely had little role in confidence reports.

Perhaps as a result of these considerations, many researchers have studied models that provide clearer explanations of confidence, even if this means using non-normative decision-making mechanisms (De Martino et al., 2013; Kiani et al., 2014; Moreno-Bote, 2010; van den Berg, Anandalingam et al., 2016; Zylberberg et al., 2012). If we discard the DDM assumption that the cumulative difference in evidence for the two options is tracked, and that a criterion on this value triggers response, then the "balance of evidence" will no longer always be the same at decision time (Vickers & Packer, 1982; Yeung & Summerfield, 2014). For example, the race model and models with partially anti-correlated accumulators can explain variability in confidence reports, even in experimental setups that ensure a decisional locus for confidence. In these models the balance of evidence at decision time is

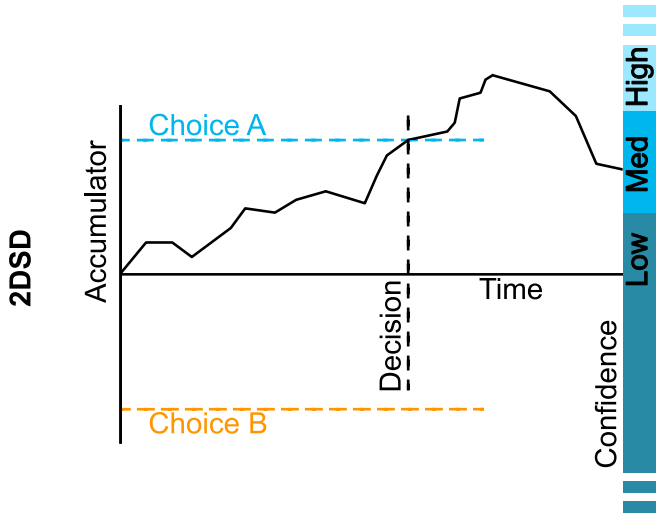

Figure 1.2: Two-stage dynamic signal detection theory (2DSD; Pleskac and Busemeyer, 2010). The model features a post-decisional locus for confidence. Following a decision, evidence continues to accumulate. It is the state of the accumulator, once additional information has been incorporated, that determines the level of confidence in the decision.

variable, being given by the difference between the two accumulators (Fig. 1.1; Vickers, 1979; Vickers and Packer, 1982). Whereas the winning accumulator will always be at the threshold at the time of the decision, the other accumulator could be anywhere below the threshold (Kiani et al., 2014). Triggering a decision on the basis of only one of two accumulators leads to variability in confidence reports but, as discussed, this property renders the decision-making mechanism non-normative, and commits us to the claim that humans effectively ignore half the information available to them when terminating decisions.

In practice, such models have proved to be successful at accounting for patterns in confidence data. Kiani et al. (2014) combined partially anti-correlated accumulators, with an optimal readout for confidence that reflected the probability of being correct. The model could predict the relationship between confidence and stimulus discriminability for both correct and error trials. van den Berg, Anandalingam et al. (2016) found that a similar model provided a good account of changes of mind and changes of confidence. Partially anti-correlated accumulators have also proved successful in other studies (Zylberberg et al., 2012). Nevertheless, the model proposed by Kiani et al. (2014) embodies a peculiar claim. As discussed, decisions

are terminated by one accumulator, irrespective of the state of the other. In this sense, the decision-making mechanism ignores information. On the other hand, confidence is based on the state of both evidence accumulators. Recall that the task used by Kiani et al. (2014) ensures decisions and confidence are computed at the same time. We are led to the highly counter-intuitive conclusion that the decision mechanism discards highly-relevant information when, at the same moment, this information is available for use by another cognitive process.

Confidence as a function of time

Another decision variable that could plausibly be used by observers to determine confidence is response time. Indeed, a sensible initial suggestion is that confidence could be a function of time alone (Petrusic & Baranski, 2009). For example, on the basis of relationships observed in data, we might hypothesise that confidence is a monotonically decreasing function of response time (Henmon, 1911; Petrusic & Baranski, 2009). Zylberberg et al. (2012) compared models in which confidence reflected a read out of the final balance of evidence between the alternatives, and models in which confidence reflected the amount of time taken to make a decision. Models in which confidence was a function of decision time provided a much better account of the effect, on confidence, of evidence fluctuations in the stimulus. However, the view that confidence is only a function of response time fails to explain the very different relationships observed between response time and confidence under different experimental conditions (Pleskac & Busemeyer, 2010). When observers are free to set the time of response, confidence decreases with response time. However, when the response time is set by the researcher, confidence increases with response time. Furthermore, confidence is known to vary as a function of the difficulty of the task, even when only trials with the same response time are considered (Kiani et al., 2014).

A solution to this tension might be that confidence is a function of both the balance of evidence and response time. This is true of the models discussed in the

previous subsection, used by Kiani et al. (2014) and van den Berg, Anandalingam et al. (2016). When a decision was triggered, the observer took the difference between the two accumulators, along with the time spent accumulating evidence, and using a Bayesian readout, mapped these values to the probability of being correct. We saw in Section 1.2.1 how an SDT observer could map stimulus measurements to the posterior ratio between option A and option B. The same principles carry over to the case where the representation of evidence is dynamic. The difference in the dynamic case is that instead of receiving a single measurement of evidence for the two options, the observer receives a sequence of measurements and, in integration models, accumulates these measurements. However, under standard assumptions, only time and the final state of the accumulators are relevant, and the accumulators' path up to that point adds no further relevant information (Moreno-Bote, 2010). In the situation where the observer only has two options, the observer can calculate the probability of either option A or option B from the posterior ratio between options A and B, because the probability of these two options must sum to 1. Hence, the observer can calculate the probability, given their measurements, that the decision they made is correct.

If the difficulty of the task is unknown to the observer, and if we consider a fixed value for the balance of evidence, then in a range of settings Bayesian confidence decreases with time (Moran, 2015; Moreno-Bote, 2010). An intuition for this effect is the following. The average rate at which you have accumulated evidence gives you information about the difficulty of a trial. If you accumulate 10 units of evidence favouring your choice in 1 second, you are accumulating faster than if you accumulate 10 units in 2 seconds. Hence, in the first case you are more likely to be on an easier trial than in the second case. As a result, you can be more confident in your response in the first case. We refer to this effect as "difficulty discounting". On the other hand, and intuitively, a greater balance of evidence in favour of the selected choice is associated with a greater probability of being correct, and hence with higher Bayesian confidence.

We can now explain the different relationships between confidence and time in conditions where observers are free to respond when they want, and conditions where the researcher sets the time of response by, for example, cuing the observer when to respond (Pleskac & Busemeyer, 2010). When the researcher sets the time of response, a longer response time means more evidence is presented to the observer, and hence they will end the trial with a greater balance of evidence favouring their choice, and higher confidence (Pleskac & Busemeyer, 2010). In contrast, when the observer sets the time of the response, we know that while no response has been made, neither accumulator has crossed a decision threshold. In effect, the thresholds limit the amount of evidence that can be gathered prior to a response, and hence the balance of evidence will grow more slowly in this case. The effect of difficulty discounting may be stronger than the slow increase in the balance of evidence, leading to decreasing confidence with response time.

One subtlety in the preceding discussion is that we have spoken as if our brains are actually performing Bayesian inference. Behaviour may be Bayes optimal even if the brain is not representing or computing with the probability distributions in 1.1, simply because the observer has learned an optimal mapping on the basis of feedback (Kiani & Shadlen, 2009; Ma & Jazayeri, 2014). Considering our specific case, where the observer receives a series of evidence measurements, makes a decision after a certain amount of time, and must report a level of confidence, Bayes rule prescribes the optimal way to map evidence measurements and response time to confidence. If provided feedback, the observer could learn the optimal mapping through the association of accumulator state and time with success or failure (Kiani & Shadlen, 2009). Even in the absence of feedback on a specific task, it is conceivable that the observer has learned from other experiences a mapping from evidence and time to confidence, that is approximately correct in a new task. The way to distinguish between an observer who has learned a Bayesian mapping via association, and a Bayesian observer who performs inference using representations of the relevant probability distributions, is to change one element of the task (Ma & Jazayeri, 2014). If we change, for example, the two sets of stimuli

which the observer is aiming to distinguish (the $\boldsymbol{s}_A$ and $\boldsymbol{s}_B$ in eq. 1.1), an observer who has slowly learnt the mapping between evidence, time, and confidence, will have to relearn this mapping. On the other hand, an observer who is performing Bayesian inference using representations of probability distributions will be able to flexibly adapt, by updating only the representation of the two sets of stimuli. Other representations, such as the representation of the likelihood function (the $p(\boldsymbol{x}|\boldsymbol{s}_A)$ in eq. 1.1) will not need changing and can immediately be applied in the new task. We will not commit to one view or the other, but note that a reader who does not believe animals and humans represent probability distributions, could nevertheless agree that confidence is Bayes optimal.

1.3.2 Confidence as informed by additional variables

While it is appealing to think that confidence is simply a readout from the same process that led to a decision, there is empirical evidence suggesting that the construction of confidence is at least somewhat more complicated than this.

Metacognitive noise

Confidence appears to be based on different information to the decision due to the mechanics of the process that reads out the state of the accumulator, in order to determine confidence (Fleming & Daw, 2017; Maniscalco & Lau, 2012). Specifically, this readout is noisy, and confidence reports are subject to additional noise that decisions are not subject to. There is a great deal of evidence for this "metacognitive noise". First, confidence reports appear to be subject to more noise than decisions (Maniscalco & Lau, 2012). Second, consideration of metacognitive noise has proved necessary when modelling choice and confidence data (De Martino et al., 2013; van den Berg et al., 2017). Third, metacognitive noise makes it possible to explain the counter-intuitive finding that as sensory noise increases, confidence judgements are corrupted by a smaller amount of noise, relative to the amount of noise corrupting decisions (Bang et al., 2019).

Note that metacognitive noise on its own can only explain variability in confidence reports, not why confidence reports are related to performance, or why confidence changes systematically with time taken to decide, and with presented evidence. Hence, the existence of metacognitive noise does not help the DDM explain systematic patterns in confidence data.

A separate accumulation

The race model, and partially anti-correlated accumulators, avoid the problems of the DDM by assuming that the confidence readout has access to more information than the decision process. Another group of evidence integration models involve the idea that the confidence readout has access to different information from the decision process because there are two separate sets of accumulators (or static representations of evidence), one set that governs the decision made and one set that determines the degree of confidence in that decision (Balsdon et al., 2020; Fleming & Daw, 2017; Jang et al., 2012). According to these models, both sets of accumulators receive the same input but are subject to at least partially independent noise. One set of accumulators behaves exactly as those described in Section 1.2, and accumulates evidence until a decision threshold is reached, triggering a response. Confidence is read out from the second set of accumulators. Because noise corrupting the two accumulators is to some extent independent, the state of the second set of accumulators is not completely determined by the first set of accumulators. This has important implications for DDM models. The DDM decision accumulator would, as discussed, always be in the same state when a decision is triggered. However, the confidence accumulator would no longer always be in the same state, generating variability in confidence at the time of a decision.

Although this idea may provide an account of confidence data, it commits us to the claim that the brain effectively has two perceptual systems, both accumulating noisy versions of the same signals, while costing twice as much in terms of energy consumption (Lennie, 2003). Moreover, if the observer did have two noisy versions

of an accumulator, the observer could improve performance by averaging the information in these two accumulators, leading to a single accumulator with reduced noise. It is unclear why the brain would not use this available information.

Beyond evidence accumulation

Confidence also appears to be informed by variables that seem quite unrelated to perceptual decision-making mechanisms, duplication of these mechanisms, or readouts from these mechanisms.

Throughout we have been using the term "confidence" to refer to a sense of the probability that a correct decision has been made. However, a complete theory of confidence may need to take into account the potentially very different purpose of confidence in social contexts (Hertz et al., 2017). For example, confidence reports may be used in a way so as to ensure influence over others (Hertz et al., 2017). In this case, the computational goal of the confidence system would not just be an estimate of the probability of being correct, but also to maintain or increase social influence. Additionally, general levels of confidence appear to be influenced by individual differences. For example, a person's average level of confidence may be a trait (Pallier et al., 2002), that is correlated with optimism (Ais et al., 2016). Psychological disorders also appear to be related to between-participant differences in overall levels of confidence (Rouault et al., 2018).

Although the role of confidence in social contexts, and individual differences in general levels of confidence, are fascinating questions, we aim to focus our research on the perceptual component of confidence. To do this, throughout, we will only study the ordering of confidence reports, and we will only look at that ordering within-participants. We will not study the mapping between observers' estimates of probability, and objective probability (Lichtenstein & Fischhoff, 1977), nor will we compare general levels of confidence across participants. A benefit of this approach is that we do not need to worry about the possibility that observers

scale, shift or stretch the readout before reporting confidence on the scale provided (Aitchison et al., 2015; Festinger, 1943).

Some variables appear to have within-participant effects on confidence, but initially seem unrelated to any evidence accumulation mechanisms. One example is the contribution of the motor system to confidence. Transcranial magnetic stimulation (TMS) of the premotor cortex has been found to affect confidence, in the absence of changes in performance (Fleming et al., 2015). Electromyographic (EMG) activity, recorded from the thumbs prior to response, is related to increased confidence but not increased performance (Gajdos et al., 2019). Moreover, when participants are cued to make a motor action that is irrelevant, but similar to the movement they will make in an upcoming choice, confidence in that choice is increased (Siedlecka et al., 2020). Although these results point to a role of the motor system in confidence, such a role may complement the idea that confidence is a product of the decision making mechanism, rather than conflicting with it (Fleming & Daw, 2017). Specifically, it may be that confidence reflects a function of evidence and time, but that this function is read out from the motor system. Consider again the example at the outset, of an owl flying after a glimpse of an object in the grass. If the speed of the owl is sensibly modulated by the confidence of the owl, then confidence can be read out from the motor system. The fact that the motor system affects confidence does not invalidate the claim that, at the algorithmic level, confidence is a function of evidence and decision time (Marr, 2010/1982). Nevertheless, it does highlight that a complete understanding of confidence will also require an account of the relevant algorithm's implementation.

Another feature which at first appears inconsistent with the idea that confidence is a function of time and evidence, is history effects that have been observed in confidence data. Confidence in a decision on one trial has been found to affect confidence in the next trial (Rahnev et al., 2015). This "confidence leak" effect was observed even when looking at confidence in two different tasks, and could not be accounted for by the idea that performance on one trial affects performance on the

next trial. One response to this finding would be to suggest that confidence is not just a function of time and evidence, but is also directly a function of confidence on previous trials. However, an alternative explanation is that confidence on a trial has an indirect effect on the next trial, by altering the function of time and evidence used to read confidence out. Rahnev et al. (2015) found that the observed effect could be well accounted for by a model in which it was the readout of confidence which was altered. Specifically, the model featured the idea that confidence is used to update the observer's belief about the difficulty of the task: Low confidence on one trial suggests the task is more difficult than thought. Taking this into account, confidence on the next trial will be reduced. Again, we see that the influence of additional variables does not invalidate the claim that within-participant confidence is simply a noisy readout of response time and evidence accumulated.

1.4 Accounting for confidence with the DDM

1.4.1 How?

In the preceding discussion we have seen that a great deal of work has gone into building computational models of decisions and confidence (e.g. Balsdon et al., 2020; Kiani et al., 2014; Pleskac & Busemeyer, 2010; van den Berg, Anandalingam et al., 2016), but that models of confidence have often incorporated assumptions inconsistent with the dominant, normative, model of decision making (De Martino et al., 2013; Kiani et al., 2014; van den Berg, Anandalingam et al., 2016; Zylberberg et al., 2012). We saw that the DDM, coupled with a confidence readout based on accumulated evidence, struggles to account for key patterns in confidence data even in the most successful version of this model, 2DSD. We now sketch the outline of a DDM account that may provide an adequate account of confidence data. We provide full details of the model in Chapter 2. The aim of this exercise is to resolve the tension between the fact that the DDM provides an excellent model of choice and response time data, and the finding that it struggles to

explain confidence, without resorting to two parallel evidence accumulators that respectively support decisions and confidence.

Building on the work of Pleskac and Busemeyer (2010) and van den Berg, Anandalingam et al. (2016) showing that post-decision evidence has an important role to play, we retain this feature. We consider a number of other possible extensions, which are partly motivated by the strong relationship between confidence and response time. First, we consider the possibility that, as in 2DSD, drift-rate variability helps account for the relationship between time and confidence (Pleskac & Busemeyer, 2010). Second, we consider the possibility that the decision threshold used to trigger a decision may decrease over time (Drugowitsch et al., 2012; Malhotra et al., 2017). Finally, we consider three different ways that confidence could be read out: Confidence could reflect the final state of the accumulator (Pleskac & Busemeyer, 2010), a calibrated Bayesian readout that will take into account the state of the accumulator but also reflect difficulty discounting (Kiani et al., 2014; Pew, 1969; Vickers and Packer, 1982), or a miscalibrated Bayesian readout (Drugowitsch et al., 2014).

A clear relationship between confidence and response time would arise if decision thresholds decrease over time, meaning a smaller value of accumulated evidence is required to trigger a decision for later vs. earlier decisions (Drugowitsch et al., 2012; Malhotra et al., 2017). As later decisions are made with a smaller balance of evidence in their favour, confidence will be lower. As discussed in Section 1.2.2, decreasing decision thresholds can be optimal when the difficulty of the task is unknown to the observer (Drugowitsch et al., 2012; Malhotra et al., 2018; Tajima et al., 2019; Tajima et al., 2016). Note that the difficulty of the task can be unknown to the observer due to external factors such as variability in stimuli across trials, or due to internal factors such as drift-rate variability (Moran, 2015).

Empirical evidence regarding the question of whether decision thresholds are flat or decreasing is mixed. There is some evidence that decision thresholds decrease in some situations (Malhotra et al., 2017; Palestro et al., 2018), although there is

also conflicting evidence (Hawkins, Forstmann et al., 2015; Pardo-Vazquez et al., 2019; Voskuilen et al., 2016), and mixed results (Evans et al., 2020). One possible explanation for these discrepancies is that people only use decreasing thresholds when this is optimal. Evans et al. (2020) conducted a particularly informative study by exploring decision thresholds in settings where it is optimal to use decreasing thresholds. Perhaps surprisingly, participants did not learn to maximise reward rate through the use of decreasing thresholds when the difficulty of the task was unknown to them and when provided with trial-by-trial feedback. In contrast, Palestro et al. (2018) observed strong evidence for decreasing thresholds in an experiment where trial-by-trial feedback was provided. Evans et al. (2020) did find evidence for decreasing thresholds when an explicit deadline was imposed on the time taken to respond, suggesting that people do use decreasing thresholds when this is clearly advantageous. Another distinction that has been drawn is between fast individual perceptual decisions, and slow "expanded judgement" tasks, where a series of stimuli are presented and a judgement must be made about the collection of stimuli (Voskuilen et al., 2016). Strong evidence has been found for decreasing thresholds in expanded judgement tasks (Malhotra et al., 2017; these authors note that increasing thresholds can also be optimal). The tasks we use below are not easily categorised in this way. In any case, it seems prudent to consider the possibility of decreasing thresholds as, if they are used, they would generate a systematic relationship between confidence and time.

While a subtlety, it is important to distinguish two effects of time on confidence: within and between conditions (Vickers & Packer, 1982). Within a condition, decreasing thresholds will cause confidence to decrease with elapsed time. However, if in a new experimental condition we ask participants to slow down, their response times, accuracy, and confidence may all increase (Pleskac & Busemeyer, 2010; Vickers & Packer, 1982). From the perspective of the model, asking participants to slow will cause them to shift their decision thresholds outwards (Ratcliff & McKoon, 2008). The thresholds may still collapse over time, but they will now have a greater absolute value at every time point, meaning more evidence is required

to trigger a decision. This example shows that, even with decreasing thresholds, we do not expect increasing response time across conditions to decrease confidence across conditions. Only within a single condition do we expect a strong negative relationship between response time and confidence.

Another extension to the DDM that can generate a clear relationship between confidence and time is an idea considered in the previous section. Specifically, the idea that confidence may reflect a noisy Bayesian readout based on the time spent accumulating evidence, and the evidence accumulated (Aitchison et al., 2015; Kiani et al., 2014; Meyniel et al., 2015; Pew, 1969; Sanders et al., 2016; Vickers & Packer, 1982). This contrasts with the confidence readout used in 2DSD which is based on the accumulated evidence alone. As discussed in Section 1.3.1, Bayesian observers often show an effect in their confidence reports which we refer to as difficulty discounting, where the longer spent deliberating, the lower confidence in a fixed amount of evidence becomes (Moran, 2015; Moreno-Bote, 2010). This effect occurs when the observer does not know the difficulty of the task, which can happen when a stimulus is variable, or in the presence of drift-rate variability (Moran, 2015). We also consider the possibility that the difficulty of the task is unknown to the observer because they hold incorrect beliefs about the task (Drugowitsch et al., 2014), and specifically, have not yet learned the variability in task difficulty (relative to other sources of variability). Variability in task difficulty might in fact be zero, but if the observer does not know this, difficulty discounting will apply. This possibility is especially likely where limited feedback is provided to participants, but may also be true in cases where extensive feedback is provided: Several studies suggest that humans are bad at estimating and compensating for noise associated with stimulus variability (de Gardelle and Mamassian, 2015; Herce Castañón et al., 2019; Zylberberg et al., 2016; Zylberberg et al., 2014).

1.4.2 Why?

A sensible question to ask is why we should even try to reconcile the normative DDM with patterns observed in confidence data, when there are alternative models – such as models with partially anti-correlated accumulators – that have provided good models of confidence data. This question is particularly important as, within this thesis, we will not attempt to provide a direct comparison between the performance of the DDM and other accounts of perceptual decision making. Instead we suggest that the DDM is superior on theoretical grounds. Hence, if it can be shown that the DDM provides an excellent account of confidence data, in addition to choice and response time data, it would become the best model of perceptual decisions and confidence.

The ultimate reason we favour the DDM is that it is a normative model of perceptual decisions (in the sense of the term "normative" described in Section 1.2.2). We acknowledge that there is overwhelming evidence that human behaviour is suboptimal in many ways (Rahnev & Denison, 2018), and that the implementation of any algorithm is also likely to be suboptimal (Section 1.3.2; consider the recurrent laryngeal nerve in the giraffe; Shorrocks, 2016, p.61). Nevertheless, we suggest that there are at least three theoretical reasons for favouring the DDM over other models, even though we know this optimal model will not perfectly describe perceptual decision making.

First, the suboptimal algorithm and implementation used, will be used in virtue of the fact that they approximate the computational goal of the system, subject to neural and cognitive constraints (Lieder and Griffiths, 2020; Marr, 2010/1982, especially p.337-338). This is especially likely to be true for perceptual decisions, which we have been making for well over 500 million years (Lee et al., 2011). Second, there are several features that make an account a good scientific theory. Accuracy is vital, but the scope of the theory and its simplicity are also important (Kuhn, 1978). An account proposing that cognitive processes generally conform to normative principles, is by its nature broad in scope. In the context of decisions and confidence, a DDM account of confidence provides the possibility of a unified

explanation of choices, response times and confidence. Whereas it is less clear that this is the case for decision models which are specifically motivated because they can explain confidence data. Third, focusing on optimal algorithms, and attempting to falsify these first, allows us to constrain the direction of our search. We have seen the huge range of mechanisms proposed to account for perceptual decisions and confidence. The number of possible accounts is even greater when we consider that most models of perceptual decisions can be combined with most models of confidence, generating a very large number of possible combinations. In this situation, even if we don't believe perceptual decision-making or confidence will be close to optimal, this seems like a sensible place to begin.

For these reasons, we attempt to build a satisfactory account of decisions and confidence by extending the DDM, retaining its optimal accumulation mechanism. The reader unconvinced by these arguments may nevertheless be able to appreciate the explanatory power of the resulting model, and will hopefully come to view the resulting model as equally good as other non-normative models which can also account for confidence data.

1.5 Objections to the claim DDM is optimal

We have argued that the DDM is a normative model of perceptual decision-making. Tracking the difference in evidence, as in the DDM, is a feature of the algorithm which maximises reward rate (Section 1.2.2). However, reasonable objections can be raised about whether the DDM is optimal in a relevant sense. For instance, it could be argued that the relevant notion of optimality is a resource rational one: The optimal algorithm is not the algorithm which maximises reward rate but the biologically plausible algorithm which balances maximising reward rate, with the cost of devoting neurons to a task (Lieder & Griffiths, 2020; Marr, 2010; van den Berg & Ma, 2018). Another objection could be that the DDM is optimal in a context which is completely unlike those faced in the real world (Rahnev & Denison, 2018). The DDM applies to choices between two options, but animals

and humans frequently have to make choices between multiple options. Animals and humans, also don't repeatedly perform the same task, but are constantly faced with new "tasks", with new statistics.

We will not address the question of biological plausibility, or whether the DDM represents an efficient use of neural resources, except for a couple of points. First, some models of perceptual decision-making, unlike the DDM, have been specifically motivated on the grounds that they are biologically plausible. The leaky competing accumulator (LCA), discussed in Section 1.2.2 is one example: Usher and McClelland (2001) linked leak and competition between accumulators to the decay of activity observed in populations of neurons, and the inhibition between such populations. While we acknowledge this is an attractive way to build a model, we would suggest that such a close link between the workings of an algorithmic-level theory and the function of neurons does not need to be shown for a model to be biologically plausible. Second, regarding the efficient use of neural resources, we have already noted that having confidence read out from the same mechanism which generates perceptual decisions, would roughly half the number of neurons needed to support these processes, and hence, would half the energy cost (Lennie, 2003).

1.5.1 Multiple alternatives

Models of choices between more than two alternatives often have a similar structure to the models discussed in Section 1.2.2. These models typically assume that each of the N options in a decision making task generate a noisy signal in the brain, representing the evidence for that option (Bogacz et al., 2006; Usher & McClelland, 2001). As in the case of two choice options, the signal that corresponds to the correct alternative has a higher mean, and these evidence signals are then accumulated in some form.

It is not immediately clear how the DDM would scale up to the situation in which there are more than two alternatives (Usher & McClelland, 2001). The key property of the DDM is looking at the difference in evidence between two options,

but when there are more than two options, it is not clear what difference we should track. Hence, at first sight it might seem like the DDM is not optimal in the sense of maximising reward rate in the context of the decisions that animals and humans regularly face. It is easier to see how other models would scale up. For example, the race model could be extended to the case of multiple alternatives simply by adding more evidence accumulators (Usher & McClelland, 2001). Similarly, from the framework of the LCA, we can add more accumulators which will all compete with each other.

The multihypothesis sequential probability ratio test (MSPRT) is a model for multi-alternative choice, and is closely related to the DDM. The MSPRT claims that observers track the posterior probabilities for each of the options, and a response is triggered when a posterior probability crosses a threshold (Baum & Veeravalli, 1994; Bogacz & Gurney, 2007). The MSPRT is in some ways similar to the race model: A value for each option is tracked, and the first value to cross a threshold triggers the corresponding response. However, the fact that posterior probabilities are tracked means that when evidence for one option is received, this is counted as evidence against the other options, reminiscent of the DDM treating evidence for one option as evidence against the other. In fact, in some situations, and as mentioned in Section 1.2.2, the DDM is equivalent to tracking the posterior probability on the two options until one crosses a threshold, just as in the MSPRT (Bitzer et al., 2014; Gold & Shadlen, 2007; Moran, 2015). While the MSPRT uses posterior probabilities for stopping, perhaps surprisingly, it turns out to only be optimal under conditions in which the observer makes very few errors, or the cost of errors are very small (Dragalin et al., 1999).

The stopping rule which maximises reward rate in more general conditions has been derived (Drugowitsch et al., 2012; Tajima et al., 2019; Tajima et al., 2016). In the N-dimensional space defined by the N evidence accumulators, the optimal stopping policy is to terminate deliberation when the point in this space which represents the state of the N accumulators, reaches a time-dependent curved bound

(Tajima et al., 2019). Crucially, however, the sum of the evidence for the N options is irrelevant to the question of when to stop, according to the optimal policy. When there are only two options, if we ignore the sum of evidence for the two options, the only variability being monitored must be the difference in evidence. Hence, while it is not clear how to scale up the DDM, it is clear that the DDM (with possibly time-dependent bounds) can be viewed as the scaled down version of the optimal stopping rule for multiple alternatives.

1.5.2 Computational cost and changing tasks

Although we have seen that we cannot criticise the optimality of the DDM on the grounds that it only applies to two-alternative choices, we might still be worried that optimality, in the sense of maximising reward rate, is irrelevant because it may never be achieved. There are at least two reasons why it would be difficult to maximise reward rate in realistic settings.

First, calculating the optimal decision bound involves first calculating the value associated with each point in the N-dimensional space defined by the N evidence accumulators. This "value function" describes the reward expected in the future, if one behaves optimally. Unfortunately, computing the value function involves considering some distant future time, and then working backwards, calculating the value function time step by time step (Drugowitsch et al., 2012). The computational intensity of computing the optimal stopping rule makes it less plausible that animals explicitly compute the exact rule. An alternative is that animals gradually learn a mapping from the value of the accumulators and elapsed time, to stopping vs. waiting. Unfortunately, this is also likely to be challenging, especially when the number of choice options is high. In this case the animal must learn to map from a high dimensional vector (with one entry for each accumulator).

Second, in real environments, the "task" is constantly changing, along with the associated statistics. When this happens, the optimal stopping rule will also change (Drugowitsch et al., 2012). Hence, in naturalistic settings, the optimal stopping rule

has to be constantly relearned. The challenge will be all the greater because of the computational cost of learning the optimal stopping rule in even a static environment.

If maximal reward rate will never be achieved, then optimality in this sense is arguably irrelevant. Instead, what is important is fast learning in new situations. In Chapter 6 we test the learning speed of an approximately optimal stopping rule, which reduces to the DDM with time-dependent thresholds in the case of two alternatives.

1.6 Empirical challenges to Bayesian confidence

A large number of results appear to contradict the idea that confidence reflects a Bayesian readout. For example, there is evidence that people take into account perceptual uncertainty in their confidence reports, but not in the precise manner that a Bayesian observer would (Adler & Ma, 2018a). Both in two-alternative and three-alternative choices there is evidence that confidence reports may reflect quantities which pertain to uncertainty, but not specifically the probability of a correct choice (Li & Ma, 2020; Navajas et al., 2017). We will return to these studies in Chapter 7. Here we focus on two other challenges to the idea of Bayesian confidence: Differences between performance and confidence (Lichtenstein & Fischhoff, 1977; Rollwage et al., 2020), and the finding that decision congruent evidence is frequently weighted more heavily than decision incongruent information in confidence reports (Peters et al., 2017; Zylberberg et al., 2012).

It is worth noting at the outset that our intuitions about the behaviour of Bayesian observers can be misleading. It is certainly intuitive to think that, for a Bayesian observer who reports the probability they are correct, their performance and average confidence will align. This turns out not to be the case in general (Drugowitsch et al., 2014). In particular, as soon as we consider performance and confidence as a function of a variable unknown to the observer, then average confidence reported by a Bayesian observer need not match their performance.

The hard-easy effect – the finding of overconfidence on difficult trials and underconfidence on easy trials – provides a good illustration of why this is the case (Drugowitsch et al., 2014; Lichtenstein & Fischhoff, 1977). Consider an experiment in which the observer makes a choice between two alternatives, and in which there are two difficulty levels. In the more difficult condition, the difficulty of the task is so high that it is impossible. Clearly, if a cue informed the observer that they were in the impossible condition, they should always report 50% confidence, which would then match their chance performance. However, if we do not provide the observer a cue they will not know which condition they are in. They can make a probabilistic inference, but they will not know with complete certainty the condition. A confidence report of 50% is not appropriate, as there is always some chance they are not in the impossible condition. The Bayesian observer will report greater than 50% confidence on trials in the impossible condition, even though their performance will be at chance. Hence, differences between performance and confidence do not necessarily invalidate the idea of a Bayesian readout for confidence.

Another phenomenon which appears to contradict the idea of Bayesian confidence has been observed in several studies. This phenomenon can be observed in settings where we can classify some evidence provided by a stimulus as "decision-congruent", and some as "decision-incongruent" (Peters et al., 2017; Zylberberg et al., 2012). Decision-congruent (decision-incongruent) evidence is evidence pertaining to the selected (unselected) option. Consider a task in which the aim is to determine if there are more dots on the left-hand or right-hand side of the display. If the observer selects left, then the dots presented on the left-hand side constitute the decision-congruent evidence, and the dots on the right the decision-incongruent evidence. While choices appear to be determined by the balance of evidence for the two options, confidence appears to be more closely, or entirely, determined by decision-congruent evidence (Peters et al., 2017; Zylberberg et al., 2012; although see Charles and Yeung, 2019; note we will analyse the data collected by Charles and Yeung, 2019). This effect has been observed in a variety of stimuli, and when using both behavioural and neural measures of evidence. Additionally, several studies have

capitalised on this result to manipulate confidence without changing performance (Koizumi et al., 2015; Rollwage et al., 2020; Samaha et al., 2016). At first sight, this result suggests that (a) human confidence is determined by a simple heuristic which neglects much of the information available, and that (b) confidence and decisions are based on different representations of evidence (Peters et al., 2017). Both these ideas clearly conflict with the ideas and model we have set out in this chapter.

In Chapter 5, we examine these challenges to Bayesian confidence in detail. As in the case of differences between confidence and performance, we will see that our intuitions regarding the link between Bayesian confidence and decision-congruent evidence are misleading.

1.7 Existing methodological approaches

In Section 1.2.1 we noted one advantage of models that feature a static representation of evidence: The relative tractability with which we can make precise trial-by-trial predictions from these models. In contrast, it is often very computationally expensive to make trial-by-trial predictions from the DDM, featuring as it does, a dynamic representation of evidence.

There are several existing techniques that can be used for calculating the probability, according to the DDM, of different responses and response times (Brown et al., 2006; Chang & Cooper, 1970; Cox & Miller, 1965; Diederich & Busemeyer, 2003; Drugowitsch, 2016; Navarro & Fuss, 2009; Smith, 2000; Tuerlinckx, 2004; Tuerlinckx et al., 2001; Voss & Voss, 2008). Importantly, approaches have been developed that can handle the dynamic stimuli used in the experiments reported in this thesis (i.e., stimuli that change over the course of a trial) and time-dependent thresholds. One approach involves using finite difference methods to approximate the evolution of the probability distribution over accumulator state (Chang and Cooper, 1970; Voss and Voss, 2008; Zylberberg et al., 2018). Time and space are discretised and, working forward from the first time step, we solve a set of simultaneous

equations at each time step, to find the evolution of the probability distribution over accumulator state. If we are only interested in the probability distribution over response times and choices, we can use expressions described by Smith (2000). To evaluate these expressions we only need to discretise time, not space. Again working forward from the first time step, we can calculate the probability of deciding at each time step. Both approaches require that we discretise the time course of a trial into small time steps. Hence, in both approaches, we must perform a large number of computations, on the order of the number of time steps considered.

This computational cost becomes important if we want to leverage trial-by-trial variability in dynamic stimuli. Typically, computation time for calculating predictions is reduced by making predictions on a condition-by-condition basis (e.g. Kiani et al., 2014; Ratcliff & McKoon, 2008; van den Berg, Anandalingam et al., 2016; Zylberberg et al., 2016). We design experiments so that there are a small number of different conditions (e.g. levels of contrast), then we treat all trials from a single condition as the same, and make predictions for behaviour in each condition, rather than for each stimulus individually. This approach does not capitalise on trial-by-trial variability of the dynamic stimuli which are often used. These random fluctuations in evidence make every trial unique. Hence, a method that reduced the computational cost of making trial-by-trial predictions for dynamic stimuli could be of great benefit. A simple mathematical expression for confidence predictions would provide additional advantages, such as insight into the mechanisms and properties of confidence in DDM models.

Moreno-Bote (2010) derived expressions for confidence that cover DDM models, take into account time-dependent thresholds, and that could be extended to account for dynamic stimuli of the kind we consider. However, these derivations use two assumptions about the computation of confidence that are not in line with findings we have discussed above. First, Moreno-Bote (2010) made the intuitive assumption that decisions and confidence are based on the same information. However, as we discussed in Section 1.3.1, sensory and motor processing takes time, and there will

be stimulus information in these processing "pipelines" that does not contribute to the initial decision, but that does inform confidence (Charles & Yeung, 2019; Moran et al., 2015; Ratcliff & McKoon, 2008; Resulaj et al., 2009; van den Berg, Anandalingam et al., 2016). Moreover, information in the processing pipeline is affected by trial-to-trial fluctuations in drift-rate, and this may have important effects on confidence (Pleskac & Busemeyer, 2010). Second, we discussed the substantial evidence that the process which "reads out" confidence into a behavioural report is corrupted by "metacognitive noise" (Section 1.3.2; Bang et al., 2019; De Martino et al., 2013; Maniscalco and Lau, 2012; van den Berg et al., 2017).

In Chapter 4 we develop a new approach, which allows us to capitalise on the within-trial and between-trial fluctuations in stimuli that are commonly used, when predicting confidence. Importantly, we take into account pipeline evidence, and metacognitive noise.

1.8 Thesis outline

Having reviewed models of perceptual decision making, and confidence, argued for a DDM model of confidence, and sketched how this could be achieved, we now turn to testing these ideas.

In Chapter 2 we take the extensions to the DDM sketched in Section 1.4.1, which were motivated theoretically, empirically, or in both ways, and combine them into a number of variants of the core DDM model. Our goal is to determine (a) which DDM model variant fits confidence data best, and (b) whether any of the models can explain the range of qualitative and quantitative effects observed in confidence data. Of particular interest will be whether any of the variants can overcome a critical limitation of the 2DSD model in accounting for the strength of the relationship between confidence and time. In the remainder of Chapter 2 we test the power of the models by considering how they fare against benchmark findings regarding confidence and its relation to other variables. Additionally, we examine

whether they can reconcile findings which would appear to conflict. We derive qualitative predictions from the model variants, which we test using a previously collected dataset (Charles & Yeung, 2019).

Chapter 3 focuses on a new study that provides a preregistered test of the qualitative patterns identified in Chapter 2. We then use computationally cheap expressions, which we developed, to fit the models using trial-by-trial predictions for confidence. This allows us to examine whether any model can account for the precise quantitative effects observed in confidence data. If one of the model variants can account for qualitative and quantitative effects in confidence data, it will establish that we can provide an empirically adequate account of confidence using the DDM. Hence, we can maintain the view that one of the most basic cognitive functions, perceptual decision making, is generated through a mechanism with optimal properties. While this is the main goal, we will end with a parsimonious account of decisions and confidence, which will provide a unified explanation for a large range of empirically observed effects.

The predictions for confidence used in Chapter 3 are derived in Chapter 4. Specifically, we derive predictions for the probability distribution over confidence reports, given the response and response time on a trial. We derive computationally cheap predictions which make the trial-by-trial modelling of dynamic stimuli used in Chapter 3 feasible. We chose to present the derivations following their application in Chapter 3 because the method, results and analysis in Chapter 3 can be understood without reference to the technical developments in Chapter 4. At the same time, the modelling in Chapter 3 provides context and motivation for the approach developed in Chapter 4. Beyond the practical motivations, in Chapter 4 we also aim to gain insight into the confidence of optimal observers.

In Chapter 5 we address a recently-reported finding that appears to invalidate the idea that confidence reflects a Bayesian readout of the probability of being correct. The finding is that decision-congruent evidence is over-weighted relative to decision-incongruent evidence in confidence reports, but not perceptual decisions

(Peters et al., 2017; Zylberberg et al., 2012). Matching the intuition that a Bayesian should use all evidence available – both decision-congruent and decision-incongruent – we present simulations confirming over-weighting of decision-congruent evidence is not predicted by the models developed in Chapter 2 - 4. However, analysing three existing datasets (Carlebach & Yeung, 2020; Charles & Yeung, 2019) ideally suited to evaluating the separate contributions of decision-congruent and decision-incongruent evidence, we do not find a strong over-weighting of decision-congruent evidence in confidence. We explore reasons for this apparent discrepancy.

In Chapter 6 we address the concern that the DDM is not optimal in a sense which is relevant in naturalistic settings. We consider a model which reduces to the DDM with time-dependent thresholds in the case of choices between two alternatives. We motivate the model as an approximation to the optimal stopping rule for terminating deliberation when faced with multiple alternatives (Tajima et al., 2019). Using neural networks as model learners we examine the speed of learning in the context of choices between two or more alternatives. Compared to other models, the model which reduces to the DDM led to faster learning, and consequently greater reward, even before the theoretical maximum for reward rate had been reached.

Chapter 7 concludes with an integrative discussion of the combined modelling and empirical investigations reported in this thesis.

1.8.1 Beyond the scope of the thesis

In Appendix E and G we report work conducted during the DPhil, but which falls outside the scope of this thesis. In one study, we collected an EEG dataset and looked at the EEG correlates of evidence accumulation, before and after perceptual decisions. In another study, we used a Bayesian SDT model to explore the effects of distractor statistics in visual search.

1.9 Open science

For each project described in this thesis, all code written for that project, and all new data collected for that project, will be made freely available online, upon journal publication.

2

Models and qualitative tests

Contents

In Chapter 1 we saw that the DDM, while a normative and highly successful model of decisions and response times, struggles to account for confidence reports (Bogacz et al., 2006; Ratcliff et al., 2016; Yeung & Summerfield, 2014). One of the most successful DDM accounts of confidence, 2DSD, can explain variability in confidence reports, and the qualitative pattern in the relationship between confidence and time, but fails to capture the strength of this effect (Pleskac & Busemeyer, 2010). Frequently, models featuring alternatives to the normative decision-making mechanism of the DDM, or models in which confidence relies on a second separate evidence accumulation have been used (Balsdon et al., 2020; De Martino et al., 2013; Jang et al., 2012; Kiani et al., 2014; van den Berg, Anandalingam et al., 2016; Zylberberg et al., 2012). We considered whether it was possible to take a different

approach, and retain the DDM but make sensible extensions to this model. The ideas of a decision threshold which decreases with time (Drugowitsch et al., 2012; Malhotra et al., 2017), and a Bayesian readout for confidence (Kiani et al., 2014; Pew, 1969; Vickers & Packer, 1982), appeared particularly promising. We start by setting out in detail the models that we will investigate. We then assess whether the models can explain benchmark findings regarding confidence and its relationship to other variables such as response time, evidence and accuracy. We will also see that the models allow us to account for patterns in confidence following errors that have previously been difficult to explain. We finish the section by describing three qualitative predictions that are largely shared by all the models considered. As a preliminary test of these models, we test one of the predictions in a previously collected dataset (Charles & Yeung, 2019) in Section 2.2.

2.1 Models

We consider 10 models which are all variants of a core model. The core model has as its central assumption the idea from the drift diffusion model (DDM; Ratcliff and McKoon, 2008), that observers track the total difference in evidence between alternatives. We apply this idea in two contexts termed the "interrogation" and "free response" conditions (Bogacz et al., 2006). In the "interrogation" condition the observer does not have control over the time of the response, or the duration of stimulus viewing. Instead the observer must monitor the stimulus until it clears, at which point the observer can respond. In this case the observer does not need to set a criterion for stopping evidence accumulation and responding, as the end of the trial is determined for them. Performance can be maximised by observing and using all information presented. Once all information from the stimulus has been processed, the observer can simply pick the option favoured by the accumulated evidence (Fig. 2.1; Bogacz et al., 2006). As observers accumulate evidence until all information from the stimulus is processed, if the stimulus is

presented for 600ms, in the model the evidence accumulation lasts 600ms (not including sensory and motor processing delays).

Under "free response", an observer is free to set the time of response. Given this flexibility, they must use some policy to determine when to stop accumulating evidence and make a decision. Following the DDM, we assume the observer sets thresholds on the accumulated evidence, one for each response option (Fig. 1.1). When the accumulator reaches either threshold a response is triggered. As discussed in Chapter 1, sensory and motor processing takes time. As a result, some information is still in these processing pipelines at the time of response, and although it will not contribute to the initial decision (Ratcliff & McKoon, 2008; Resulaj et al., 2009; van den Berg, Anandalingam et al., 2016) it will be processed immediately following the decision and can therefore inform any subsequent confidence report given (Fig. 2.1; Moran et al., 2015; van den Berg, Anandalingam et al., 2016).

The time spent accumulating evidence following a decision and prior to a confidence report has itself been investigated (Baranski & Petrusic, 1998; Moran et al., 2015; Pleskac & Busemeyer, 2010). We do not apply time pressure to confidence judgements from participants, and make the parsimonious assumption that evidence accumulation ends once all pipeline evidence has been processed (Moran et al., 2015). For example, if it takes 400ms for information to pass through sensory and motor processing pipelines, and the stimulus clears on response, information presented in the stimulus in the 400ms prior to response will not have been processed. Therefore, evidence accumulation continues for 400ms following the crossing of the decision threshold. This way of modelling confidence reports is both plausible given the experiment design, and adds no new parameters to the models (see Section 3.5 for further discussion).

In the core model, confidence is based on a readout of the final state of the accumulator, which tracks the difference in evidence between the two choices (Pleskac & Busemeyer, 2010; Vickers, 1979). In all models we allow the possibility that the readout is corrupted by metacognitive noise (Maniscalco & Lau, 2012, 2016), which

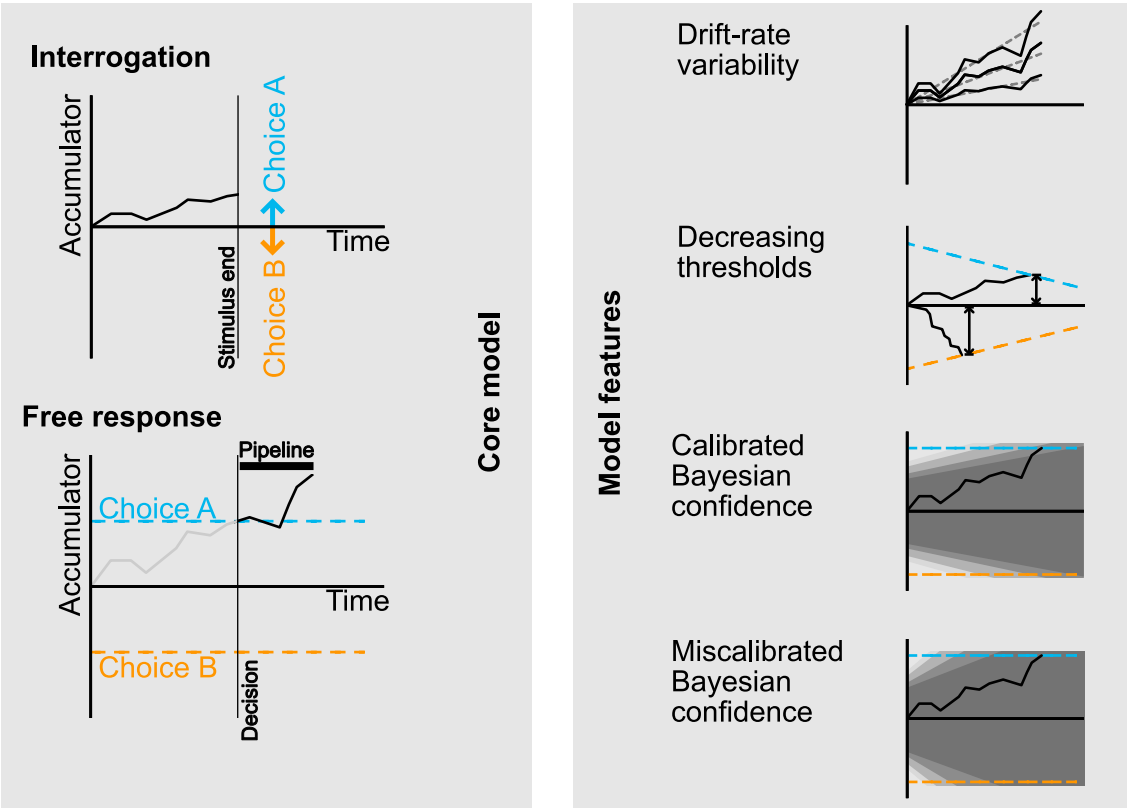

Figure 2.1: All models considered are built from a core model in which observers track the total difference in evidence between alternatives (Bogacz et al., 2006). When the researcher sets the time of the response (interrogation condition), observers accumulate evidence until all information from the stimulus is processed (Bogacz et al., 2006). When observers set the time of response (free response condition), observers accumulate evidence until the accumulator reaches one of two decision thresholds (Ratcliff & McKoon, 2008). Evidence accumulation continues for a short time after a decision, as sensory and motor processing pipelines mean there will be additional information which did not contribute to the decision (Resulaj et al., 2009). Model variants are constructed by adding combinations of four model features to the core model.

we assume to be normally distributed (see Section 7.3.2 for further discussion of this choice). As discussed in Chapter 1, in all analyses we treat confidence as an ordinal variable because we aim to focus on the perceptual component of confidence, rather than individual differences in confidence or in the use of confidence scales (Ais et al., 2016; Aitchison et al., 2015; Festinger, 1943).

This core model, based on the DDM, would struggle to account for confidence reports for the reasons described in Chapter 1. Principally, with a constant drift-rate, flat decision thresholds, and confidence readout based on the accumulated evidence, the model has no way of accounting for the relationship between confidence and response time (Pleskac & Busemeyer, 2010). We consider four features that could be added to the core model to better account for observed features of human confidence reports (Fig. 2.1).

The first feature is drift-rate variability (Chapter 1; Ratcliff et al., 1999). We operationalise drift-rate in a different way to the usual approach (Appendix A.2). Typically, evidence provided by the stimulus is constant over a trial, or approximated as such, generating a fixed drift-rate which is also constant over a trial (Bogacz et al., 2006; Ratcliff & McKoon, 2008; Ratcliff et al., 2016). Drift-rate variability can then be operationalised as trial-to-trial fluctuations in this rate. In the experiments considered in this thesis, we use dynamic stimuli in which evidence provided by the stimulus varies over time, changing every 50ms "frame". This procedure provides rich datasets with which to test the models, and presumably a time-varying drift-rate as different frames from the stimulus – providing different levels of evidence – are processed in turn (Smith, 2000). As a result, we need a new operationalisation of drift-rate in this context. We want to capture the idea that trial-to-trial quality of information processing varies, such that the effect of presented evidence on the accumulator varies. We use the term "standardised drift-rate" to refer to a coefficient which has a multiplicative effect on the strength of the relationship between evidence presented and changes in the accumulator (Appendix A.2). When standardised drift-rate is 1, the effect of evidence on the accumulator is at its

average level. When standardised drift-rate is higher (or lower), evidence presented in the stimulus has a larger (or smaller) effect on the evidence accumulation.

The second feature that we used to extend the core model is decision thresholds that are not constant but rather decrease over time (Chapter 1; Fig. 2.1; Drugowitsch et al., 2012). Decreasing decision thresholds generate a clear relationship between confidence and response time as slower decisions are made with a smaller balance of evidence in favour of the selected choice.

The third feature considered is that, instead of confidence reflecting the state of the accumulator, confidence is a Bayesian readout of the probability of being correct (Section 1.3.1; Kiani et al., 2014; Pew, 1969; Vickers and Packer, 1982). For such a readout, the observer uses the final state of the accumulator, and the time spent deliberating (Moreno-Bote, 2010). As discussed in Chapter 1, in many settings Bayesian confidence decreases with time, in the absence of changes to accumulated evidence (Moran, 2015; Moreno-Bote, 2010). This effect, which we refer to as difficulty discounting, arises because the rate at which you have accumulated evidence provides information about the difficulty of the trial.

The fourth feature considered is confidence reflecting a miscalibrated Bayesian readout (Chapter 1; Drugowitsch et al., 2014). Specifically, the observer makes incorrect assumptions about the relative magnitude of different sources of variability. This idea is consistent with the finding that human observers are poor at learning and dealing with noise associated with stimulus variability (de Gardelle and Mamassian, 2015; Herce Castañón et al., 2019; Zylberberg et al., 2016; Zylberberg et al., 2014). Incorrect estimation of the magnitude of different sources of variability will affect the confidence of a Bayesian observer, as a Bayesian observer explicitly takes into account the sources of variability that can affect the accumulator (Appendix A.2). Incorrect estimation of this kind will affect the strength of difficulty discounting, i.e. the decrease in confidence with response time for a Bayesian observer who doesn't know the difficulty of the task (Chapter 1). We refer to the two Bayesian readouts as calibrated and miscalibrated Bayesian confidence. In the case of Bayesian confidence,

Model number	1	2	3	4	5	6	7	8	9	10
Drift-rate variability	-	✓	-	✓	✓	✓	-	✓	-	✓
Decreasing thresholds	-	-	✓	✓	-	✓	-	-	✓	✓
Bayesian confidence										
Calibrated	-	-	-	-	✓	✓	-	-	-	-
Miscalibrated	-	-	-	-	-	-	✓	✓	✓	✓

Table 2.1: Ten model variants were constructed using the core model and combinations of the model features.

we assume that normally distributed metacognitive noise corrupts the observer's estimate of the log posterior ratio between the two alternatives.

We combined the core model with these features to produce 10 models (Table 2.1). We considered all possible combinations of the features, which led to 2 (drift-rate variability yes or no) $\times$ 2 (flat or time-dependent decision thresholds) $\times$ 3 (confidence reflects accumulator, or calibrated Bayesian readout, or miscalibrated Bayesian readout) = 12 models. In the absence of drift-rate variability, and when there is only a single stimulus difficulty, a calibrated observer knows the difficulty of the task so difficulty discounting does not apply. The effect of this is that the calibrated Bayesian readout becomes identical to a readout of the state of the accumulator (when confidence is treated as an ordinal variable; Bogacz et al., 2006; Moran, 2015). Consequently, we removed the two Bayesian models that exactly duplicated the predictions of the corresponding non-Bayesian models.

As discussed above, the core model (model 1) is a baseline DDM that we expect to struggle to account for confidence reports (Table 2.1). It is included for comparison. Model 2 is closely related to the 2-stage signal detection model (2DSD; interrogation version; Pleskac and Busemeyer, 2010). 2DSD is a DDM account which captures many key patterns in confidence data (Pleskac & Busemeyer, 2010). However, 2DSD does not capture the strength of the relationship between confidence and response time (Pleskac & Busemeyer, 2010). For all models, we do

not consider trial-to-trial variability in the start point of the accumulator, a source of variability considered by Pleskac and Busemeyer (2010). (Note also the difference in operationalisation of drift-rate variability above.) There is no reason why start point variability could not be included in the core model. However, the inclusion of start point variability is usually justified on the grounds that it permits models to account for certain patterns in response times (Ratcliff et al., 2016). As our main focus was understanding confidence, and to keep the models as constrained as possible, we did not consider this additional complexity. Models 3 - 10 represent different combinations of the possible extensions to the DDM identified in Chapter 1. A major motivation for these extensions was that they may allow the DDM to account for the strength of the relationship between confidence and time.

2.1.1 Explaining key findings

There are a range of phenomena that a satisfactory account of decisions and confidence would need to explain. We will explore whether these effects can be explained using the idea of difficulty discounting, which is present in models with a miscalibrated Bayesian readout (models 7 - 10, Table 2.1), and those with a Bayesian readout in combination with drift-rate variability (models 5 - 6, Table 2.1; Moran, 2015). (Difficulty discounting refers to the fact that in many settings, if accumulated evidence is held constant, Bayesian confidence decreases with time because the rate of evidence accumulation provides information about the difficulty of the trial; Chapter 1; Moran, 2015; Moreno-Bote, 2010.) In anticipation of later results, we focus on the implications of difficulty discounting, and in particular, on the core model with a miscalibrated Bayesian readout for confidence (model 7; Table 2.1). However, we will also note where decreasing decision thresholds (models 3-4, 6, 9-10; Table 2.1) provide a clear explanation for an effect.

In their evaluation of 2DSD, Pleskac and Busemeyer (2010) set out phenomena observed in decisions, response times and confidence as “empirical hurdles”. They found that their 2DSD model met the hurdles identified. Model 7 differs from

2DSD in that it doesn't feature trial-to-trial variability in the start point of the accumulator, or drift-rate variability, but it does feature a Bayesian readout of confidence not present in 2DSD. The Bayesian readout of confidence does not affect the decision stage. As noted, trial-to-trial variability in start point could easily be incorporated into the model, as could drift-rate variability. In this case the decision mechanism would be identical to 2DSD, and would inherit all decision-related properties identified by Pleskac and Busemeyer (2010). The focus of this thesis is confidence, and we search for the simplest model which can explain confidence reports. We do not doubt that variability in start point and drift-rate are required to explain decisions (Ratcliff & McKoon, 2008), but here we focus on finding the minimal model required to understand confidence. Accordingly, we focus on the hurdles identified by Pleskac and Busemeyer (2010) which relate to confidence.

The first confidence-related hurdle identified by Pleskac and Busemeyer (2010), hurdle 2, is that confidence increases as the strength of signal provided by the stimulus increases. A stronger signal leads to a higher drift-rate and hence, on average, a greater amount of evidence in the pipeline favouring the correct response, and increased confidence in the free response condition (Pleskac & Busemeyer, 2010). Difficulty discounting will exacerbate the effect because higher drift leads to faster responses, and as a result, higher confidence. Similarly, decreasing decision thresholds may exacerbate the effect because faster responses are generated when the decision threshold is reached sooner, and hence these decisions will be made with a greater balance of evidence in their favour. Increased signal strength will also increase the evidence accumulated in the interrogation condition.

We also use the explanation set out by Pleskac and Busemeyer (2010) for the fact that choice accuracy and confidence are positively correlated within individuals, even when considering trials of a fixed difficulty (hurdle 3 in Pleskac and Busemeyer, 2010). In all the models we consider (Table 2.1), in the free response condition, confidence is affected by pipeline evidence. This evidence will, in general, support the correct option, adding further evidence to the accumulator if the correct option

was chosen. The reverse holds in the case of an incorrect choice. On error trials in the interrogation condition, presented evidence favours the unchosen option, so accumulated evidence in favour of the chosen option is likely to be low, generating low confidence. In both cases the overall result is greater confidence on correct trials, and hence, a correlation between confidence and accuracy.

We can also explain relationships identified between confidence and time. It is known that in free response tasks, within a condition, confidence is higher for faster responses than slower responses (hurdle 4 in Pleskac and Busemeyer, 2010). However, confidence increases with response time when comparing different conditions in which there is different emphasis on trading speed for accuracy, and when stopping is enforced at a particular time such as in the interrogation condition (hurdle 5 in Pleskac and Busemeyer, 2010). The latter findings are easily explained within any DDM account of confidence (Ratcliff & McKoon, 2008). In the free response condition when speed is emphasised, observers use lower decision thresholds such that less evidence is required to trigger a decision. This leads to faster, less accurate choices, but also to lower confidence. In the interrogation condition, the state of the accumulator simply grows with time as more evidence measurements are received, explaining an increase in confidence (provided the effect of difficulty discounting is not too strong).

In contrast, the model variants considered here offer differing accounts of within-condition variation in confidence as a function of response time. When the decision threshold is flat, the final state of the accumulator is determined by the height of this threshold, and any pipeline evidence (Pleskac & Busemeyer, 2010). Pleskac and Busemeyer (2010) noted that 2DSD struggles to explain the strength of the negative relationship between decision time and confidence, beyond the weak dependency introduced by drift-rate variability. In contrast, models that include decreasing decision thresholds inherently create this strong dependency between confidence and decision time, because slower decisions are made with a smaller balance of evidence in their favour. Difficulty discounting may also help us explain why the

relationship between decision time and confidence is so strong: Observers who apply difficulty discounting take into account not just the balance of evidence, but also the time spent deliberating, in their confidence readout. We will see in computational modelling below, whether difficulty discounting can account for the strength of the effect of response time on confidence. Note that the explanations provided by difficulty discounting and decreasing decision thresholds still apply when considering trials of a single level of difficulty (Kiani et al., 2014).

Another benchmark finding is that confidence is a stronger predictor of accuracy following free responses which are speeded, compared to free responses where participants are asked to emphasise accuracy (hurdle 8 in Pleskac and Busemeyer, 2010). One mechanism through which 2DSD can account for this effect is if people take longer between decisions and confidence reports under speeded conditions (Pleskac & Busemeyer, 2010). The longer this interjudgement time, the greater the effect of pipeline evidence, which increases confidence on correct trials, and decreases it on incorrect trials. The effect of speed instructions can be explained through other mechanisms. For example, this effect can be accounted for using the idea of difficulty discounting, even if the time between decisions and confidence reports is constant across conditions. Simulating using model 7, we observe that the correlation between confidence and accuracy decreases within each simulated condition as a function of time (Fig. 2.2; Appendix A.3 for details of the simulations). When the decision threshold is flat, accumulated evidence at the time of decision is the same whether a correct or error response is made: It is at the level of the decision threshold. Differences in confidence between correct and error trials are driven by pipeline evidence, but will also be affected by difficulty discounting. In a range of settings, the effect of difficulty discounting is to divide the accumulated evidence by a function of time (Drugowitsch et al., 2012; Moreno-Bote, 2010; Appendix A.2), decreasing the effect of evidence on confidence. At slow response times, difficulty discounting will be stronger, decreasing the effect of pipeline evidence. As pipeline evidence is responsible for the difference in confidence between correct and error trials, this difference will decrease.

This mechanism does not account for the fact that in Fig. 2.2 the correlation between confidence and accuracy also decreases across conditions, when accuracy is emphasised relative to speed. This between-condition effect was also present when simulating data using a DDM with decreasing decision thresholds (model 3), and the core DDM (model 1). We suggest this between-conditions effect arises because of differences between the conditions in terms of the proportion of correct and error trials. Both the point-biserial correlation coefficient and Kendall's tau are known to decrease as the proportion of trials in each category of a dichotomous variable (here accuracy) moves away from a 50%-50% split (Babchishin & Helmus, 2016). In the speeded condition accuracy is lower. As a result, there is a more even split between the number of correct trials and the number of error trials, leading to a large correlation. When accuracy is high, most trials are correct leading to a lower correlation. While this effect could be viewed as an artefact, we have seen that models featuring difficulty discounting can also provide a mechanistic explanation for the stronger relationship between accuracy and confidence under speeded conditions.

We noted that 2DSD struggles to explain confidence reports when confidence is based on exactly the same evidence as the decision (Chapter 1; Yeung and Summerfield, 2014). This is because, with flat decision thresholds, the accumulated evidence at the time of the decision is always the same. Therefore, it is important to ask whether difficulty discounting can help account for confidence in this case. Kiani et al. (2014) used an experimental setup which minimises the accumulation of pipeline evidence after a response, by eliciting choices and confidence reports simultaneously (see also Adler and Ma, 2018a; Peters et al., 2017; Zylberberg et al., 2012). Response time and evidence were both related to confidence, even when the other variable (evidence or response time) was held constant. Difficulty discounting (Chapter 1), which applies whether or not there is accumulation of pipeline evidence, accounts for the observed effect of response time on confidence. Similarly, as we have seen, decreasing decision thresholds can account for decreasing confidence with time. However, none of the models considered (Table 2.1) can account for an effect of evidence which is independent of the effect of time, when there is no

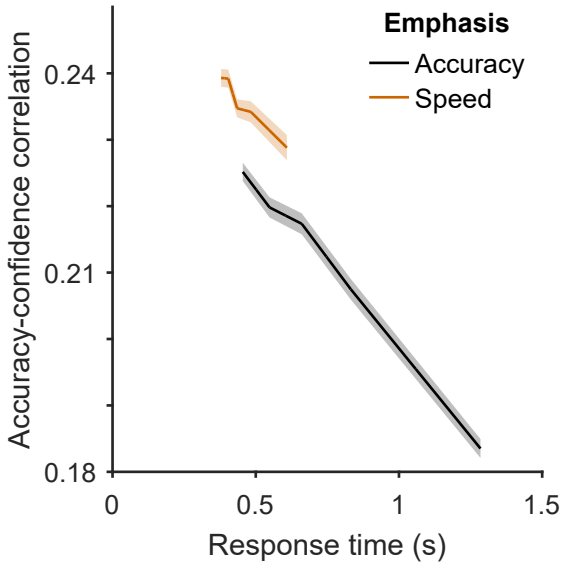

Figure 2.2: Rank correlation between accuracy and confidence in a dataset simulated using model 7 (which features a miscalibrated Bayesian readout for confidence). Two conditions were simulated, in which observers set a lower or higher decision threshold, due to the observer emphasising speed or accuracy. Consistent with previous findings, confidence was more strongly related to accuracy under speeded conditions. Kendall's Tau was computed after dividing data for each condition into 5 bins using response time. Shading represents ±1 SEM of the mean. While the results of the simulation could be made arbitrarily precise by simulating more data, SEM is shown in the figure to indicate the precision resulting from simulations of the size used. Simulation details in Appendix A.3, plotting details in Appendix A.4.

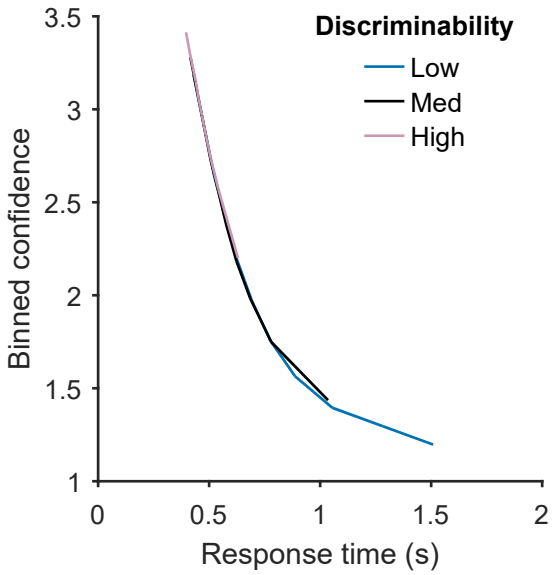

Figure 2.3: The effect of response time and stimulus discriminability on confidence, in the absence of variability in the duration of processing pipelines. Error shading (which is very narrow) represents ±1 SEM. Plotting details in Appendix A.4.

accumulation of pipeline evidence. These models attribute any effect of evidence on confidence – beyond effects mediated by decision time – to the accumulation of pipeline evidence following a decision.

We verified that model 7 as it currently stands cannot account for a unique effect of evidence. To do this we simulated trials with three different levels of stimulus discriminability. We then simulated the behaviour of observers following model 7 (Appendix A.3). In these simulations confidence was read out at the time of the decision, to reflect the simultaneous choices and confidence reports used by Kiani et al. (2014). We plotted confidence as a function of response time, separately for the 3 levels of stimulus discriminability. There was no evidence that, at fixed response time, confidence also varied with the evidence provided by the stimulus (Fig. 2.3).

However, even if there is a very small amount of variability in the duration of sensory and motor processing pipelines (Luce, 1986; Ratcliff et al., 2004; Ratcliff et al., 2016), then the model can account for an effect of evidence over and above the effect of time. In this case, the amount of time spent accumulating evidence up

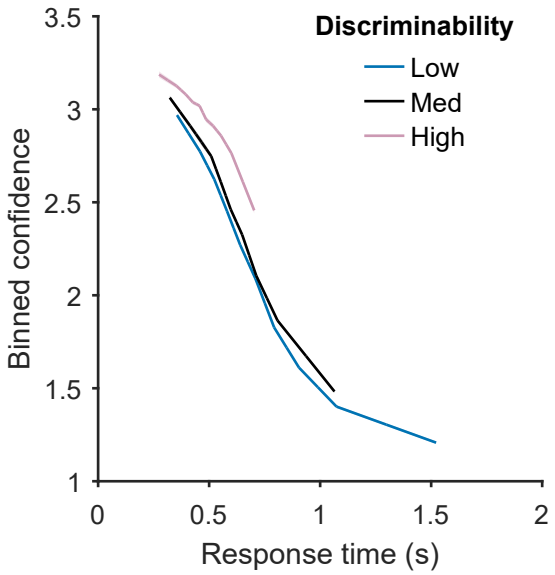

Figure 2.4: The effect of response time and stimulus discriminability on confidence, in the presence of variability in the duration of processing pipelines. Error shading (which is very narrow) represents ±1 SEM. Plotting details in Appendix A.4.

to the decision is no longer known by the researcher. As a result, the strength of difficulty discounting, which depends on the amount of time spent accumulating evidence, is unknown. Evidence becomes important (from the perspective of the researcher), and predicts confidence, because it provides information about this timing, and therefore, about the strength of difficulty discounting. We ran the same simulation as that used in Fig. 2.3, but this time included variability in the duration of the processing pipeline, with standard deviation of 0.1s (Appendix A.3). Even with this small amount of variability, the model now predicted that confidence reports formed simultaneously with decisions would vary as a function of stimulus discriminability, at a fixed response time (Fig. 2.4). We found analogous results when using decreasing decision thresholds: Such models cannot account for a unique effect of evidence in the absence of variability in the duration of processing pipelines, but they do feature this effect when variability is included.

In addition to benchmark findings such as those identified by Pleskac and Busemeyer (2010), with the difficulty discounting and flat decision threshold of model 7 we can explain some results that have previously appeared difficult to

reconcile. Sanders et al. (2016) observed that confidence in errors decreases as a task becomes easier and stimuli more discriminable. However, Kiani et al. (2014) found that, when confidence reports were collected simultaneously with decisions, confidence in errors increased with stimulus discriminability. These findings can be reconciled by considering the timing of decision and confidence reports (Desender et al., 2020; Kiani et al., 2014). Sanders et al. (2016) used confidence reports which followed decisions, while Kiani et al. (2014) used confidence reports which were made simultaneously with decisions. If the latter setup eliminates the accumulation of pipeline evidence following the crossing of the decision threshold, then evidence accumulation will always end at the decision threshold, and the final accumulated difference in evidence will always be the same. On the other hand, decision time will vary between trials. Increased signal strength leads to faster correct responses, but crucially, it also leads to faster error responses (when the decision threshold is flat, and without drift-rate variability; Shadlen et al., 2006). Faster responses are associated with reduced difficulty discounting, and greater confidence. Hence, in the absence of pipeline evidence, confidence in errors will be greater when a task is easier (Desender et al., 2020). When confidence reports follow a decision, and pipeline evidence is accumulated, pipeline evidence generally favours the correct option, reducing confidence in errors. When signal strength is greatest, this effect is strongest, leading to lower confidence in errors when a task is easier. In making these claims, we have relied on intuition. Once we have derived mathematical expressions for confidence in Chapter 4, we will verify these intuitions for the behaviour of model 7.

While it is important to account for benchmark findings, and findings that have been difficult to reconcile, stronger tests are whether a model can make novel predictions, and whether a model can account for a wide range of effects observed in real data, using a single set of parameter values. In Section 2.1.2 we make three key qualitative predictions using the core model and variants (Table 2.1). We examine one of these predictions in a previously collected dataset in Section 2.2 (Charles & Yeung, 2019), as a preliminary test of the family of models. In Chapter 3 we test

all three predictions in a new dataset, before fitting the models to data to examine whether the models can also account for the precise qualitative effects observed.

2.1.2 Qualitative predictions

We sought qualitative predictions that would provide a test of the entire set of model variants considered. All models considered share a set of strong claims, hence it is feasible to make and test distinctive predictions. Verification of such predictions in data would provide support and justification for the assumptions made when constructing this family of models. All models assume that confidence reflects a (possibly Bayesian) readout of the evidence accumulated, and that the same accumulator used for the decision is used to read out confidence. Additionally, the models assume a decision threshold is used in the free response condition, but not in the interrogation condition.

In the free response condition, the idea that decisions and confidence are the result of the same accumulation generates distinctive predictions. If we know the shape of the decision threshold, we can work out the state of the accumulator at the time of the decision, simply by looking at the position of the threshold at this time. Using this information would make evidence processed prior to a decision (i.e. all evidence apart from pipeline evidence) irrelevant to confidence. Even if we don't know the shape of the decision threshold, "pre-decision" evidence can only affect confidence by changing the time at which the decision threshold is reached. In practice it will be difficult to determine precisely which evidence is processed prior to the decision, and which evidence is in the pipeline. However, we can make the strong prediction that once the effect of time has been accounted for, pre-decision evidence will have a smaller effect on confidence than pipeline evidence. Moreover, we predict that this relationship will not hold in the interrogation condition, where we hypothesise that observers will not use decision thresholds and will instead weight all evidence equally in their confidence report.

We can also make specific predictions for the relationship between confidence and time from the idea that the same accumulator is used for decisions and confidence. In the free response condition, the existence of a decision threshold ensures that accumulated evidence at the time of the decision is either constant, in the case of a flat threshold, or falls with the time spent deliberating in the case of decreasing thresholds. This will generate confidence which is either constant or decreasing with response time. Additionally, in the presence of drift-rate variability, confidence will fall with response time, as low drift rates lead to long response times, and a decreased effect of pipeline evidence (Chapter 1; Pleskac and Busemeyer, 2010). Confidence will also fall with time if the observer uses a Bayesian readout, due to the effect of difficulty discounting. Hence all variants of the core model, but not the core model itself (i.e. model 1 in Table 2.1), predict that confidence should fall with response time.

In contrast, confidence may increase with time in the interrogation condition. This is because we hypothesise no decision thresholds are used in this case, so more time allows the accumulation of more evidence. Confidence will increase with time provided the effect of difficulty discounting is not too great. In any case, we predict the relationship between time and confidence will be more negative in the free response condition, due to the presence of decision thresholds in this case, and their absence in the interrogation condition.

These considerations led us to three qualitative predictions:

A Under free response, once the effect of response time as been accounted for, the effect of pre-decision evidence on confidence will be smaller than the effect of evidence gathered after the point of commitment to a decision.

B Once the effect of response time as been accounted for, the effect of pre-decision evidence on confidence will be stronger in the interrogation condition than in the free response condition.

C Confidence will be more negatively related to response time under free response, than in the interrogation condition.

These predictions are shared by all models.

2.2 Preliminary study

We used a pre-existing dataset for an initial test of one of the qualitative predictions made by all 9 model variants (Charles & Yeung, 2019). Using the pre-existing dataset we could test prediction A.

2.2.1 Method

Experiment Participants were asked on each trial to report whether, on average, there were more dots in an array to the left of fixation or in an array on the right (Fig. 2.5). The number of dots in each array was changed every 50ms frame, and fluctuated around either a high mean value (correct array) or low mean value (incorrect array). There were two conditions in the experiment that determined how long participants had to respond: One condition emphasised speed of responding with a short response deadline of 800ms; the other condition emphasised accurate responding and correspondingly used a liberal response deadline of 3000ms. A full description of the methods can be found in the original paper (Charles & Yeung, 2019). We only use data from the slow, 3000ms response time condition, because here participants have a large amount of freedom over when to respond, and hence this case is similar to the free response condition discussed above. Following a response, the stimulus continued to be presented for a further 1000ms.

Analysis One frame-by-frame measure of evidence for the chosen option can be found by subtracting the number of dots in the unchosen array from the number of dots in the chosen array. This measure of evidence will be highly correlated with accuracy: On trials in which participants are correct, the number of dots in the

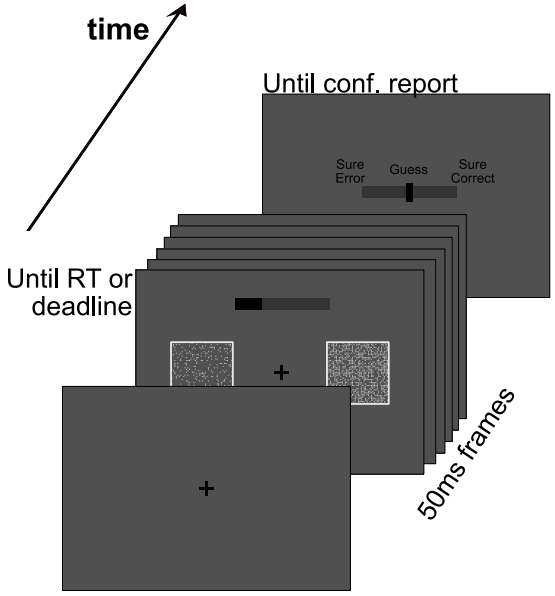

Figure 2.5: The task used by Charles and Yeung (2019). Participant's task was to determine which array contained more dots on average. The number of dots changed every 50ms. Participants could respond when they liked, but a deadline of either 800ms or 3000ms was imposed. There was no time pressure on the confidence report at the end of each trial. The figure is a schematic that is not too scale.

chosen array is drawn from a distribution with a high mean, and the number of dots in the unchosen array is drawn from a distribution with a low mean. The reverse holds on trials in which participants are incorrect. We wanted to assess the contribution to decisions and confidence of evidence presented at different time points, but if evidence at each time point is correlated with accuracy, evidence at one time point will be correlated, to some extent, with evidence presented at any other time point. In order to generate a measure of evidence which was independent across frames, we instead looked at fluctuations in the number of dots around the mean of the relevant distribution (Charles & Yeung, 2019; Resulaj et al., 2009; Zylberberg et al., 2012). That is, for each frame and array, we looked at the number of dots presented, and subtracted off the mean of the distribution from which this number was sampled. We then, for each frame, subtracted the resulting value for the unchosen array, from the resulting value for the chosen array. We refer to this measure of evidence as the "evidence fluctuations".

We combined our frame-by-frame measure of evidence fluctuations into three quantities: Pre-decision evidence, which is evidence processed before the participant made their decision (at the time of threshold crossing), pipeline evidence, evidence in sensory and motor processing pipelines at the time of the physical response (Resulaj et al., 2009), and post-response evidence, evidence presented following a response. To compute these quantities we needed to know which frames were processed prior to a decision, and which frames were presented before the physical response but were only processed after the decision. We inferred which evidence was processed prior to decisions by looking at which evidence had an effect on the choice made (Resulaj et al., 2009; van den Berg, Anandalingam et al., 2016). If evidence in a frame has an effect on the choice made, in general, evidence fluctuations in the frame will support that choice. Hence, average evidence fluctuations in the direction of the chosen option (referred to as "evidence fluctuations for choice") will be above zero. Separately for each participant, we looked at the average evidence fluctuations for the chosen option in the 12 frames preceding responses. Adapting the approach taken by van den Berg, Anandalingam et al. (2016), we fit a step function to the

average evidence residuals. We fit the initial value of the function (in the 12th frame prior to response), and the frame in which the function dropped down to zero by minimising the root-mean-square-error. The frame in which the function drops was constrained to be between 1 and 10 frames prior to the response, inclusive of 1 and 10. The frame in which the step function drops provides an estimate of the number of stimulus frames immediately prior to a response which do not inform the decision, but are in sensory and motor processing pipelines. The mean number of frames estimated to be in the processing pipeline was 6 (interquartile range, 3). Pre-decision evidence, pipeline evidence, and post-response evidence on each trial were calculated by summing the evidence residuals in the relevant frames.

We fit a probit ordinal regression model to the data (Long, 1997), with confidence reports binned into 5 categories as the outcome variable. We arbitrarily broke ties in which the same confidence level was reported, before binning confidence reports separately for each participant. Ordinal regression is very similar to logistic regression, but can handle cases in which the outcome variable has several ordered categories, rather than only two categories as in logistic regression (Long, 1997). Predictors in the regression were response time, pre-decision evidence, pipeline evidence, post-response evidence, and accuracy, individually z-scored. We applied the reverse transformation to the coefficients produced by the regression, to return them to the units they would have had if z-scoring was not applied (Appendix A.5). For all tests of predictions A - C (Section 2.1), we verified the same results were obtained without any z-scoring.

To examine whether evidence fluctuations and response time predicted confidence, we performed statistical tests on the coefficients produced by the ordinal regression onto confidence. These coefficients reflect the strength of the relationship between the predictors and the outcome. As we performed a ordinal regression for each participant, for each predictor the regression produced one coefficient per participant. We used these coefficients in t-tests to look for differences from zero, or other coefficients. As the analysis in this section of the paper is exploratory we did not

apply a correction for multiple comparisons. We measured effect size using the one-sample variant of Cohen's d (Cohen, 1988), computed by dividing the mean of the values being compared to zero, by the estimated population standard deviation.

For the regression analysis, we excluded any trial without a confidence report (because the response was too slow), and any trial in which the number of frames prior to response was less than, or equal to, the estimate of the pipeline duration for the participant. That is, the estimate for the participant of the number of frames which were in the processing pipeline at the time of responses. We excluded participants which met any of the following conditions: (a) prior to binning, any particular confidence value was reported on more than 30% of trials; (b) after predictors were z-scored and combined into a single array, the 2-norm condition number of this array was greater than 1000; (c) there were fewer than 70 included trials. These criteria led to datasets from 5/23 participants being excluded from the regression analysis.

To visualise the relationship between variables we quantile binned variables on the x-axis of plots. For details see Appendix A.4.

2.2.2 Results

We computed measures of pre-decision evidence, pipeline evidence (evidence in processing pipelines at the time of physical response), and post-response evidence. We regressed these quantities onto confidence, together with response time and accuracy. We compared the regression coefficients to each other and to zero using t-tests.

Confidence decreased with response time: Testing the regression coefficients for time, we found that time significantly negatively predicted confidence across participants ($t(17) = -4.0, p = 0.0010, d = -0.93$). This result matches previous findings for the case when participants are free to set the time of response (Pleskac & Busemeyer, 2010, empirical hurdle 4).

Looking at evidence fluctuations in the direction of the choice made, at different time lags relative to the response (Fig 2.6, A; Resulaj et al., 2009), there is an

apparent peak and then decline in the strength of the relationship between evidence fluctuations and choice. Evidence received just before the response appears to have little effect on the choice made, consistent with the idea of a processing pipeline. In contrast, evidence fluctuations just before the time of response appear strongly related to confidence (Fig 2.6, B), supporting the idea that pipeline evidence is processed following a response and influences confidence (Moran et al., 2015; van den Berg, Anandalingam et al., 2016). Moreover, 1 second prior to response, evidence residuals appear to have little effect on confidence (Fig. 2.6, B), suggesting that pre-decision evidence may not be a strong predictor of confidence. Note that our measure of pre-decision evidence is likely to show some relationship with confidence, regardless of the true effect of pre-decision evidence, because the procedure for dividing evidence into pre-decision and pipeline evidence will be subject to noise.

To determine the reliability of these patterns, we looked at the coefficients for evidence which resulted from the regression of time, evidence and accuracy onto confidence. Comparing the regression coefficients to zero, we found that pre-decision ($t(17) = 2.8, p = 0.013, d = 0.65$), pipeline ($t(17) = 6.1, p = 1.2 \times 10^{-5}, d = 1.4$), and post-response evidence ($t(17) = 3.7, p = 0.0018, d = 0.87$), all had a significant effect on confidence. Importantly, and consistent with prediction A, paired t-tests found that the effect of pre-decision evidence on confidence was significantly smaller than the effect of pipeline evidence ($t(17) = -3.1, p = 0.0067, d = -0.73$). However, the effect of pre-decision evidence was not significantly different from the effect of post-response evidence ($t(17) = -1.3, p = 0.23, d = -0.30$). We would expect post-response evidence to have a similar effect to pipeline evidence: From the perspective of the observer, it is just further evidence processed following the decision threshold crossing, which can be used to inform confidence. When stimuli continue to be presented following a response, participants may use some rule to determine how long they will continue to accumulate evidence (Moran et al., 2015; Pleskac & Busemeyer, 2010). In the experiment, evidence continued to be presented for 1000ms after a response. Participants may have stopped paying attention to the stimulus before

the end of this period, reducing the effect of post-response evidence. The declining effect of post-response evidence in Fig. 2.6, B, provides some support for this view.

2.3 Discussion

This chapter has introduced a family of models that retain the optimal property of the DDM, tracking of the difference in evidence between two alternatives (Bogacz et al., 2006; Tajima et al., 2019), while allowing for different ways that confidence in the decision reached might be derived from the difference in evidence. This family includes a core DDM, which fails in predictable ways to account for human confidence reports, and 9 variants of this model which were motivated by theory and previous empirical research. The confidence "read out" is different in the different models, but all models share the idea that the evidence accumulator which generates decisions also generates confidence. The 9 variants include some with a feature, drift-rate variability, that has previously been explored in DDM accounts of decision and confidence (Pleskac & Busemeyer, 2010), but also two features that have received less attention in terms of their effect on the DDM's ability to account for confidence: Decision thresholds which decrease over time (Drugowitsch et al., 2012; Malhotra et al., 2017), and a calibrated or miscalibrated Bayesian readout (Desender et al., 2020; Drugowitsch et al., 2014; Kiani et al., 2014).

We saw that with decreasing thresholds, or with the difficulty discounting that affects Bayesian confidence in many settings, we could account for benchmark findings regarding the relationship between confidence and trial difficulty, accuracy, speed emphasis, confidence report timing, and response time. Focusing on a model in which confidence reflects a miscalibrated Bayesian readout, we looked at how the model explains findings – regarding changes in confidence on error trials with task difficulty – that have been difficult to reconcile.

We next identified 3 qualitative predictions, made using the shared assumptions of all model variants under consideration. For one of these predictions, we reasoned

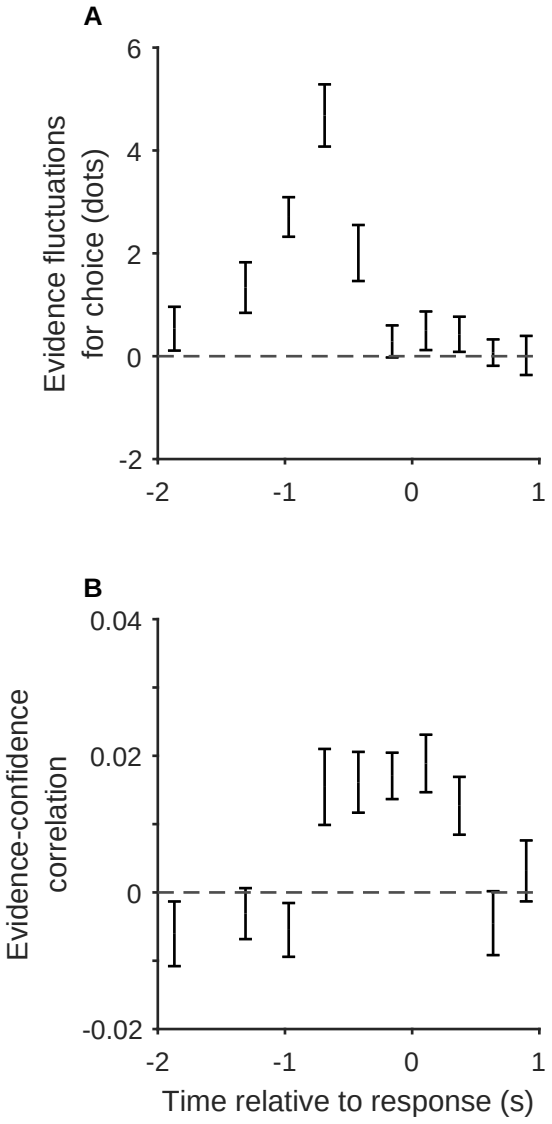

Figure 2.6: The effect of evidence fluctuations on choices and confidence in the preliminary study. For subplot B we computed the rank correlation (Kendall's Tau) between evidence fluctuations and confidence. Consistent with predictions, pre-decision evidence had a weaker effect on confidence than pipeline evidence. Pipeline evidence is evidence received immediately prior to response, and does not contribute to the initial choice due to sensory and motor processing delays (Resulaj et al., 2009). Error bars represent ±1 SEM of the mean. Plotting details in Appendix A.4.

that pre-decision evidence can only have an effect on confidence by changing the time at which the decision threshold is reached. At this point, the accumulator is at the height of the decision threshold, and the path taken up to this point is irrelevant. In an existing dataset, we verified that when response time is taken into account the effect of pre-decision evidence is smaller than the effect of pipeline evidence. The effect of pre-decision evidence was not smaller than the effect of evidence presented after a response. This may have been because evidence accumulation does not continue indefinitely after a response (Pleskac & Busemeyer, 2010).

We replicated the benchmark finding that, when response time is under the control of the participant, confidence decreases with response time (Henmon, 1911; Petrusic & Baranski, 2009; Pleskac & Busemeyer, 2010). The core DDM (model 1 in Table 2.1) has no mechanism to account for this effect (Section 2.1). While model 2, which is closely related to 2DSD, can account for this effect qualitatively, in Chapter 3 we explore whether this model can also account for the strength of this effect using trial-by-trial modelling. 2DSD has struggled to account for the strength of this effect before (Pleskac & Busemeyer, 2010). Chapter 3 also provides a preregistered test of all three qualitative predictions made here.

3

Quantitative tests

Contents

In Chapter 2 we described a core DDM, and 9 variants of this model featuring combinations of drift-rate variability, decreasing decision thresholds, and calibrated or miscalibrated Bayesian confidence. We made three qualitative predictions from these models, and tested one of these in an existing dataset. We next sought to test all three of the qualitative predictions (A-C, Section 2.1), in a preregistered, adequately powered study. The motivation for making these qualitative predictions was to provide a test for the assumptions underlying all the models. In particular, in all models confidence is some form of readout of the accumulation which is also used to make the decision. We also made the assumption that a decision threshold is used when participants are free to set the time of response (free response condition), but not when response time is cued by the experimenter (interrogation condition).

We reasoned that the effects of pre-decision evidence on confidence should be diminished if the evidence accumulator used for confidence also triggers a decision by passing through a decision threshold (Section 2.1.2). Hence, we expect a reduced effect of pre-decision evidence in the free response condition, both relative to the effect of pipeline evidence in this condition (Prediction A), and relative to the effect of pre-decision evidence in the interrogation condition (Prediction B). We also predict a more negative relationship between confidence and response time in the free response condition, relative to the interrogation condition. We assume the observer only uses a decision threshold in the free response condition. This threshold provides a bound on the amount of evidence that can be accumulated prior to a response (Prediction C).

In the later sections of this chapter, we turn to test whether the models can account for precise quantitative effects in the data. We wanted to create a particularly strong test for the models by requiring the models to fit a very wide range of effects. We have seen that the models make very different predictions for confidence in the interrogation and free response conditions. Hence, using both conditions in the same experiment, and requiring the models to simultaneously fit the very different patterns from these conditions, provides a particularly strong quantitative test of the models (Ratcliff, 2006). Our main questions will be whether any DDM variant can simultaneously account for the precise quantitative patterns observed in the two conditions, and if so, which DDM performs best.

3.1 Experimental method

Participants. 49 participants were recruited to take part in the study. Consistent with a preregistered recruitment schedule (see below), only data from the first 48 participants were analysed. Gender, age and handedness information is provided in Appendix A.7. The study was approved by the appropriate university ethical review body, and all participants gave informed consent.

Apparatus. A CRT monitor at 60Hz refresh rate, a resolution of 1600 x 1200, and physical viewable size 36.4 x 27.3 cm was used to present stimuli. A windows PC running Psychtoolbox-3 on MATLAB was used for the experiment (Brainard, 1997; Kleiner et al., 2007; Pelli, 1997).

Stimuli. Stimuli were two circular arrays of dots, one presented 2.9 degrees of visual angle (DVA) to the left and one 2.9 (DVA) to the right of the centre of the screen (assuming stimuli were viewed at approximately 60cm). The arrays had no outline, but had an imaginary circular boundary at a radius of 2.0 DVA. Within this imaginary circle, 3096 non-overlapping dots were defined. Each time new stimuli were created (every 50ms) only a subset of these dots was actually presented. Each dot was square, 0.043 DVA in height, and all dots were separated from their neighbours by 0.022 DVA . Dots and the background were grey scale with RGB colour value 200 and 80 respectively.

Procedure. Similarly to the preliminary study, on each trial participants' task was to determine which of two arrays contained the most dots on average. Participants used the left or right mouse button to indicate their response, before reporting their confidence. The study used both free response trials (participants can respond when ready), and interrogation trials (participants cued when to respond; Section 2.1; Bogacz et al., 2006), with a blocked within-participants design (Fig. 3.1).

In the free response condition participants were asked to be "as accurate and fast as possible". After a response, the stimulus cleared (unless the response was before the stimuli appeared, in which case a warning message was displayed for 6 seconds and the trial ended). In the interrogation condition, participants had to wait for the stimuli to disappear, at which time the fixation cross would turn from white to red, and they could respond. They had 1 second to respond, or they would receive a message, "Too slow" and the trial would end after a further 2 seconds. If they responded before the stimuli disappeared they would receive a message, "Too early", and the trial would end after a further 2 seconds. The

time at which the stimuli disappeared was random and drawn from a truncated normal distribution. Prior to truncation, the normal distribution had a mean and standard deviation matched to the mean and standard deviation in the most recent free response block (and mean=750ms, standard deviation=400ms on the very first block). The distribution was truncated at 200ms and 4000ms. Each block of the experiment only contained trials from one of the two conditions, and participants were informed of the condition before the start of each block.

The number of dots in the two arrays was re-sampled every 50ms from two independent truncated normal distributions, one for each array. On each trial, prior to truncation, one of the normal distributions had a mean 90 dots higher than the mean of the other distribution. Each distribution had a standard deviation of 220 dots. These distributions were then truncated to ensure that the number of dots was always positive, and always below or equal to the maximum number of dots that could be displayed in an array at once. The half-way point between the means of the two distributions was itself randomly and independently set on a trial-by-trial basis, to reduce the risk that participants would focus on a single array of dots (Charles & Yeung, 2019). The half-way point was drawn from a truncated normal. Prior to truncation it had a mean of 1000 dots, and standard deviation of 100 dots. The distribution was truncated to be between 500 and 1500 dots. In each block there were an equal number of trials in which the correct answer was left and right, and the order was randomly shuffled.

After a legitimate response (i.e., the response did not incur a too slow/fast error message), participants were asked to report their confidence. The cursor would appear at the centre of the screen, along with an arc which formed a semi-circle around the cursor. Participants reported their confidence by clicking on the arc. This setup was chosen so that participants did not have to move the mouse further to report some levels of confidence, than others. The left of the arc was marked “Definitely wrong”, the centre marked “Don’t know”, and the right marked

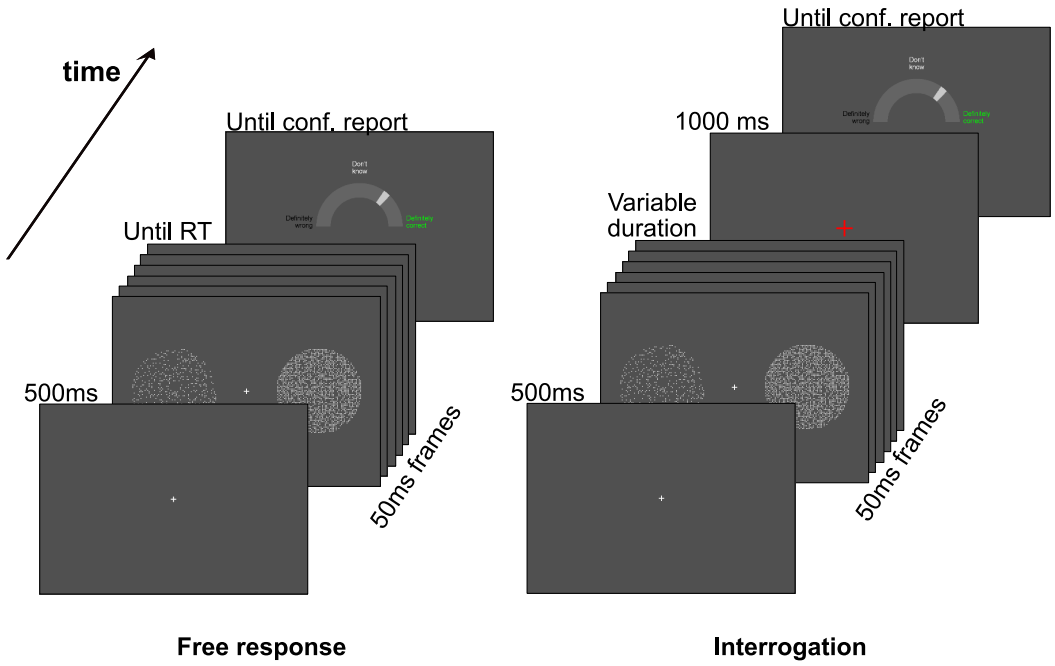

Figure 3.1: Participant's task was to determine which array contained more dots on average. The number of dots changed every 50ms. In the free response condition participants could respond when they liked. In the interrogation condition participants had to respond after a red cross appeared, and within 1 second of it appearing. There was no time pressure on the confidence report at the end of each trial. Trial-by-trial feedback was not provided.

"Definitely correct". Participants were asked to use the full range of the scale. There was no time pressure when reporting confidence.

Following a confidence report the next trial would begin after a further 400ms. The next trial would not begin until all mouse buttons were released. Each trial began with a 500ms presentation of the fixation cross only.

In the main experiment there were 16 blocks of 40 trials, and participants had 2 blocks of training prior to this. Every two blocks there was a free response, and an interrogation block, but the order within each pair was randomised. Whilst participants did not receive feedback on a trial-by-trial basis, at the end of each block participants received their score on the block, along with their high scores in each condition.

Analysis. The analysis used was similar to the analysis in the preliminary study (Section 2.2), so we only describe differences between these analyses. Unlike in the preliminary study, in this experiment we did not present evidence following responses, hence there is no post-decision evidence, and we did not include a predictor for this in the ordinal regression. Confidence binned into 5 categories was again used as the outcome variable. In addition to binning confidence separately for each participant, we binned separately for each condition. We did not arbitrarily break ties between confidence reports before binning, because we collected confidence on a continuous scale. We ran the ordinal regression analysis on the free response and interrogation conditions separately.

So we could test the qualitative predictions made in Section 2.1, we again divided the evidence presented into pre-decision evidence and pipeline evidence, by estimating the number of frames in sensory and motor processing pipelines at the time of response (Section 2.2; Resulaj et al., 2009; van den Berg, Anandalingam et al., 2016). We only used data from the free response condition to estimate the number of frames in processing pipelines. In the interrogation condition, if the manipulation is successful, decisions are made after the stimulus has cleared from the screen. Hence, no extra evidence is gathered between the time of the decision and the response. Nevertheless, we artificially divide up the evidence stream in the interrogation condition by designating frames prior to the end of the stimulus as "pipeline" frames. The number of frames artificially designated as part of the pipeline was matched, on a participant-by-participant basis, to the number of pipeline frames inferred in the free response condition. The motivation for this approach is that it allows us to check for artefacts: In the interrogation condition there should be no difference between effects of the pre-decision evidence and the "pipeline" evidence.

For the regression analysis, additional preregistered exclusion criteria were used, compared to the analysis in the preliminary study: We also excluded (a) participants with fewer than 60% correct responses, where this is calculated using included trials (Section 2.2), (b) participants where a confidence bin, after binning with MATLAB's

'quantile' function, contained $< 5\%$ of included trials. All exclusion criteria were applied separately to both conditions, and if met in either condition, data from the participant were excluded. These criteria led to the dataset from 1/48 participants being excluded from the regression analysis.

To visualise the relationship between variables we quantile binned variables on the x-axis of plots. For details see Appendix A.4.

Preregistration. All analysis details, including exclusion criteria, were decided prior to data collection and recorded in a pre-registration (Open Science Framework registration, https://osf.io/wzrhm). Preregistered sample size was justified with a power analysis (see pre-registration; Faul et al., 2007). We preregistered four statistical tests on the coefficients produced by the ordinal regression analysis (pre-registration of a test is indicated in the results section). As these tests were specified a priori we made no correction for multiple comparisons.

It should be noted that we originally made all preregistered predictions on the basis of model 3 only (Table 2.1). Later we realised a range of models are consistent with the predictions (models 2-10 in Table 2.1). We compare all models using computational modelling in later sections. There is a typographical error in the preregistration such that the first time prediction B (as lettered in that document) is stated it does not make sense. Our intent is clarified by looking at the stated statistical test for prediction B. We made one small deviation from the preregistration (Appendix A.6).

3.2 Experimental results

We first tested the 9 variants of the core model (Table 2.1), using the three qualitative predictions made in Section 2.1. We turn to computational modelling of the data in later sections.

Averaging across participants, 98% of trials ended with valid confidence reports, and accuracy on these trials was 77%. Average response time was 2.08 seconds in the free response condition, and 2.19 seconds in the interrogation condition. Distributions over response times and confidence reports are provided in Appendix A.8. Using the method described in Section 2.2 we estimated the number of 50ms frames in sensory and motor processing pipelines at the time of response. Across participants the median value was 6 frames (interquartile range, 1). We predicted accuracy in an ordinal regression using confidence (and an intercept term). Running one regression for each participant and comparing the resulting coefficients to zero, we found that accuracy and confidence were reliably correlated ($t(47) = 13, p = 5.8 \times 10^{-17}, d = 1.9$). Taken together, these results suggest participants understood the task and reported a meaningful value for confidence.

We first looked at the effect of response time on accuracy. Longer response times were associated with lower accuracy in the free response condition but greater accuracy in the interrogation condition (Fig. 3.2, A). To explore this pattern further, for each participant we looked at the mean response time in correct trials, and the mean response time in error trials. We compared the difference to zero across participants using a t-test. In the interrogation condition, errors were faster than correct choices ($t(47) = -6.0, p = 2.4 \times 10^{-7}, d = -0.87$), while in the free response condition errors were slower than correct choices ($t(47) = 3.6, p = 0.000\,78, d = 0.52$). In the interrogation condition, if the observer pays attention throughout the course of the trial, longer response times for correct responses are expected. This is because observers receive more evidence when the stimulus is presented for a long time, and will therefore, be more accurate. The expected pattern is very different in the free response condition, where the observer must set a decision threshold and determine response time themselves. Both drift-rate variability and decreasing decision thresholds can lead to errors which are slower than correct choices (Ratcliff & McKoon, 2008; Shadlen & Kiani, 2013), providing evidence that one of these processes may be present.

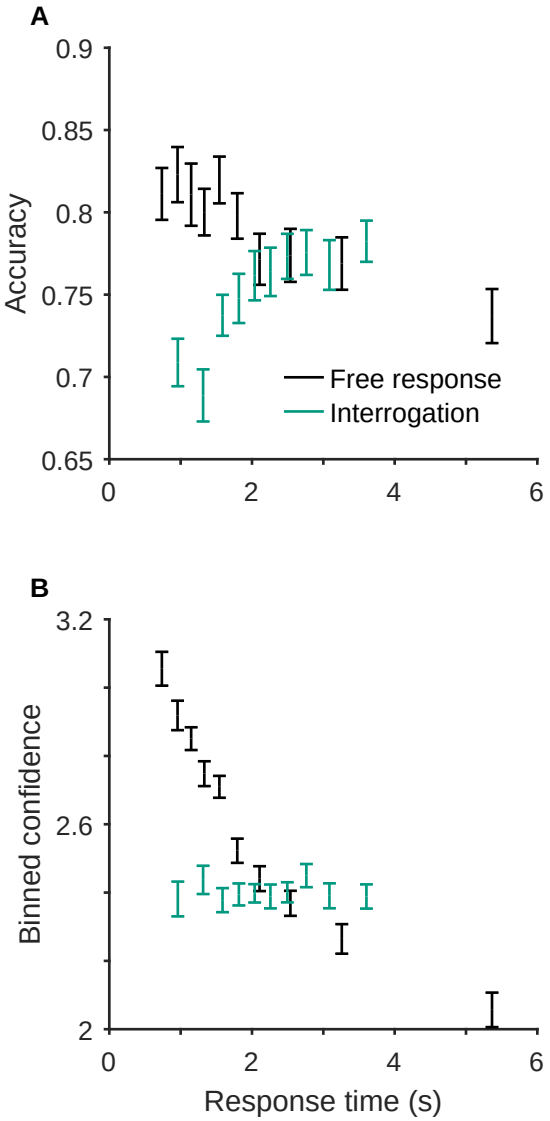

Figure 3.2: The effect of response time on accuracy and confidence in the main study. Consistent with predictions, confidence decreased with response time in the free response condition, and the relationship between time and confidence was more negative in the free response condition than in the interrogation condition. Error bars represent ±1 SEM of the mean. Plotting details in Appendix A.4.

We next explored the effect of response time on confidence. In the free response condition, confidence decreased with time (Fig. 3.2, B). For each participant, and separately for the two conditions, we ran a regression to predict confidence using pre-decision evidence, pipeline evidence, response time, and accuracy (Section 2.2 and 3.1). To test the apparent effect of response time in the free response condition, we compared the regression coefficients for time to zero using a t-test. Confidence was negatively related to response time ($t(46) = -8.8, p = 9.7 \times 10^{-12}$, one-tailed, $d = -1.3$, preregistered), consistent with the results from the preliminary study and previous findings (Pleskac & Busemeyer, 2010, empirical hurdle 4). This finding is consistent with all variants of the core model, but not the core model itself (Section 2.1).

In contrast, in the interrogation condition, response time did not have a significant effect on confidence ($t(46) = -0.90, p = 0.37, d = -0.13$; Fig. 3.2, B). The core model and all of its variants predict a more negative effect of response time in the free response than in the interrogation condition, because observers only use a decision threshold in the free response condition (Section 2.1). Indeed, a paired t-test comparing the relevant regression coefficients showed this was the case ($t(46) = -8.1, p = 1.1 \times 10^{-10}$, one-tailed, $d = -1.2$, preregistered, prediction C). Nevertheless, we were surprised that confidence did not increase with response time in the interrogation condition. This result conflicts with previous findings (Pleskac & Busemeyer, 2010, empirical hurdle 5), and the intuition that if the stimulus is presented for longer, more evidence will be gathered, and confidence will be higher.

One explanation for this result is that, in the interrogation condition, observers only monitor the stimulus for a short initial period of time. There is evidence that in interrogation tasks observers set "implicit" decision thresholds (Balsdon et al., 2020; Ratcliff, 2006). These thresholds trigger decisions which are stored by the observer until the response cue, with evidence received following the decision ignored. The finding that accuracy increases with response time in the interrogation condition provides evidence against this interpretation (Fig. 3.2, A). Additionally,

we matched response times across the two conditions with some success (Appendix A.8), which may have minimised the number of times implicit thresholds were reached even if they were present.

Analyses looking at the effect of evidence at different time lags allowed us to further investigate the possibility that observers used implicit decision thresholds in the interrogation condition. We looked at the average evidence fluctuations in the direction of the choice made, as a measure of the strength of the relationship between evidence presented at different time lags, and the response made (Fig. 3.3, A and B; Section 2.2; Resulaj et al., 2009). Looking at average evidence fluctuations at time-lags relative to the onset of the trial (Fig. 3.3, B), it appears that evidence presented near to the start of the trial has a stronger effect on the decision. A number of mechanisms have been proposed to account for such primacy effects, including implicit decision thresholds (Kiani et al., 2008; Tsetsos et al., 2012). However, an implicit threshold seems inconsistent with the pattern observed in average fluctuations at time-lags relative to response (Fig. 3.3, A). This pattern suggests all evidence in interrogation trials, whenever it is presented, is weighted equally. When looking at time lags relative to response (Fig. 3.3, A), any stimulus onset effects are presumably averaged out due to variability in trial duration. Model-free analyses of evidence weighting time courses must be interpreted cautiously because they are affected by both the weight given to sensory information received at different time points, and the mechanics of the decision-making process itself (Okazawa et al., 2018). Accordingly, we return to these effects when we consider the fit of the models to the data (Section 3.4).

An alternative explanation for the lack of a relationship between confidence and response time in the interrogation condition is that observers, if they can be construed as using Bayesian inference, have an incorrect generative model of the environment (Drugowitsch et al., 2014). A mismatch between the observer's model of the environment and the true model would also explain the difference noted above, between the way confidence and accuracy change with response time (Fig. 3.2, A

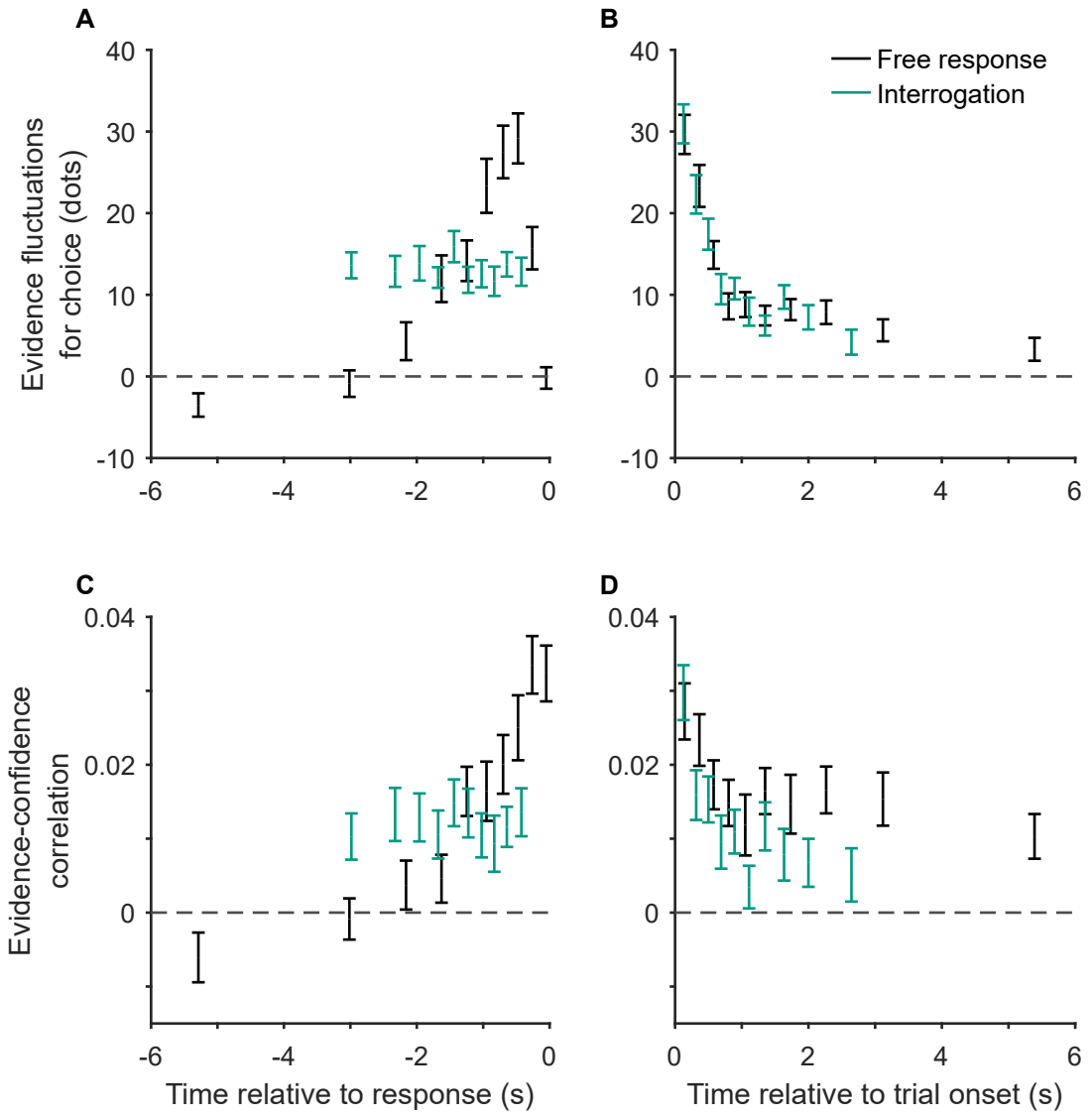

Figure 3.3: The effect of evidence fluctuations on choices and confidence in the main study. In subplots A and B we plot the average evidence fluctuations in the direction of the choice made. This serves as a measure of the effect of evidence on the choice made (Section 2.2; Resulaj et al., 2009). For subplots C and D we computed the rank correlation (Kendall's Tau) between evidence fluctuations and confidence. At time lags relative to response, all evidence appeared to be weighted equally in the interrogation condition. However, there was evidence that frames occurring at the onset of the stimulus were especially strong predictors of responses in both conditions. As predicted, pre-decision evidence had a stronger effect on confidence in the interrogation than in the free response condition. Within the free response condition, pre-decision evidence had a weaker effect on confidence than pipeline evidence. Error bars represent ±1 SEM of the mean. Plotting details in Appendix A.4.

and B), whereas the idea of an implicit decision threshold would not. Care must be taken when interpreting differences between belief and performance, because such differences can occur even when the observer has a correct generative model, if we consider belief and performance as functions of variables which are unknown to the observer (Drugowitsch et al., 2014). Here we consider accuracy and confidence as a function of response time, which are in principle accessible to the observer.

From the core model and its variants, we also made predictions for the effect of evidence on confidence. We looked at the rank correlation between evidence fluctuations and confidence at different time lags (Fig. 3.3, C and D). We see a similar pattern in the effect of evidence on confidence, as in the effect of evidence on response. Again, all evidence appears to be weighted equally in the interrogation condition, when looking at time lags relative to response (Fig. 3.3, C), but a primacy effect becomes apparent when looking at the effect of evidence relative to trial onset (Fig. 3.3, D). On the basis of the core model and variants, we reasoned that in the free response condition, evidence presented prior to the time of a decision would have a reduced effect on confidence, once the effect of time had been accounted for (Section 2.1). We can explore this possibility by returning to the coefficients produced by predicting confidence in an ordinal regression (Section 2.2 and 3.1). A t-test on the coefficients showed that pre-decision evidence had a stronger effect in the interrogation condition than in the free response condition ($t(46) = -2.4, p = 0.010, \text{one-tailed}, d = -0.35$). This effect was even clearer when we baselined the pre-decision evidence coefficients relative to the pipeline evidence coefficients (separately for participant and condition), as we had preregistered ($t(46) = -11, p = 3.9 \times 10^{-14}, \text{one-tailed}, d = -1.5$, preregistered, prediction B). Comparing coefficients within the free response condition, we found that pre-decision evidence had a smaller effect than pipeline evidence ($t(46) = -6.2, p = 7.3 \times 10^{-8}, \text{one-tailed}, d = -0.90$, preregistered, prediction A).

In summary, in Section 2.1 we made three qualitative predictions using the variants of the core model (predictions A - C also hold for the core model itself). We

confirmed two of these predictions in a pre-existing dataset, and then confirmed all three in a preregistered test. These results provide strong support for our modelling approach. One possibility which we have not been able to completely rule out is that, in the interrogation condition, people use an implicit decision threshold and stop accumulating evidence after this threshold is reached. A number of results argue against this idea, but we did observe a primacy effect such that evidence presented early in a trial is over-weighted in the decision and confidence report. We return to this question in the next sections, as we use computational modelling to test the quantitative predictions of the models, and attempt to determine which model variant fits the data best.

3.3 Modelling method

We now seek a stronger, quantitative test of the models: A key motivation of our study was to provide a DDM model of confidence that could account for all relationships observed in confidence data, and the strength of these effects. We also hope to determine which model variant provides the best account of the data.

Our focus is accounting for confidence reports. Accordingly, we do not fit the models to response times and responses themselves. Instead we fit the models to confidence, given the stimulus, response and response time. Fitting to confidence reports allows us to use simple mathematical expressions, which are computationally cheap to evaluate (Chapter 4). Response times and responses effectively become held out data, which the models are not fit too. We can ask the models to predict this data, providing a very strong test (Kiani et al., 2014).

Following the approach described in Section 2.1, we model confidence reports as ordinal data. For computational modelling we divide confidence reports into four bins on a participant-by-participant basis.

Making predictions

In Chapter 4 we derive approximate expressions for the confidence of DDM observers under both interrogation and free response conditions, allowing for drift-rate variability, time-dependent thresholds and metacognitive noise. The approximations used only introduced small discrepancies between the derived predictions, and simulated confidence reports (Chapter 4). Chapter 4 covers a set up that matches the one used in this study, with a choice between two equally probable options, where the stimulus provides two evidence signals, one for each option. The evidence signals vary over the course of a trial, but are constant within short "frames". The fluctuations in evidence in each signal are normally distributed around a mean value, and the absolute difference between the means for the two options is constant across trials. In each trial, the observer has to determine the evidence signal with the greatest mean. In the experiments used here, the number of dots in the two arrays can be viewed as the two evidence signals provided by the stimulus. Using the derivations in Chapter 4 we capitalise on the dynamic stimuli in our experiment – which vary within and between trials (Fig. 3.1) – to make trial-by-trial predictions for confidence.

We operationalise drift-rate variability in the manner described in Section 2.1. In Chapter 4 we only consider a calibrated Bayesian readout of confidence, not a miscalibrated Bayesian readout or, as in the core model, a readout of the final state of the accumulator. We provide the necessary extension in Appendix A.2. This extension makes the models no longer identifiable: Different sets of parameters can perfectly trade off with each other without affecting the model predictions. We constrain model parameters in a way that has no effect on the patterns the models can or can not explain (Appendix A.2).

We make a number of other adjustments to the derivations in Chapter 4. We model decision thresholds as symmetric and flat, or symmetric and linearly decreasing, in the case of time-dependent thresholds. We allow for the possibility of lapses in confidence reports where confidence does not follow from the usual

process. Specifically, we assume that confidence lapses occur with equal probability regardless of response, response time, and evidence presented. A lapse generates a confidence report in any of the four confidence bins with equal probability. To keep the model as simple as possible, we do not consider separate types of lapses for responses and confidence (Adler & Ma, 2018a), but we return to this point in Section 3.4. In addition to this, on free response trials where the response time is faster than the estimated duration of sensory and motor processing pipelines (a free parameter in all models), we model confidence reports as certainly the result of a lapse. For numerical reasons, we also treat trials in the interrogation condition as lapses if there is very low metacognitive noise, and the estimated probability of the response given is extremely small. Further details on lapses are provided in Appendix A.2. A number of other low-level differences from the derivations in Chapter 4 are also described.

In Chapter 4 we assume that observers ignore the fact that evidence is constant within each 50ms frame, and ignore the increased or decreased importance of stimulus variability when the stimulus is being processed better or worse (i.e., high or low standardised drift-rate). In Chapter 4 we test the effects of both assumptions and find that objective accuracy and subjective confidence remain closely related. Here we make one additional assumption. In our study we presented an equal number of trials in which the correct answer was the left array, and the correct answer was the right, in each block of 40 trials. In principle, if observers were counting, they could track an estimate of the number of trials in which left and right had been the correct answer. We ignore this possibility here, which is especially unlikely given the absence of trial-to-trial feedback.

Model parameters

The models described in Section 2.1 have between 8 and 11 free parameters (Table 3.1). All models had a parameter for the standard deviation of accumulator noise, σ_{acc}. Accumulator noise corrupts the evidence measurements which drive changes

Model		1	2	3	4	5	6	7	8	9	10	**All**
Accumulator noise	σ_{acc}											1
Metacognitive noise	σ_m											1
Threshold height	a											1
Pipeline duration	I											1
Lapse rate	λ											1
Conf. bin bounds	d_i											3
Drift-rate variability	σ_φ		1		1	1	1		1		1	
Decision thresh. slope	b			1	1		1			1	1	
Estimated variability ratio	Γ							1	1	1	1	
Total		8	9	9	10	9	10	9	10	10	11	

Table 3.1: Parameters in the models. All models shared the 8 parameters of the core model. Model variants with additional features contained additional parameters. Drift-rate variability introduced a parameter (standard deviation of distribution over standardised drift-rate), time-dependent thresholds introduced a parameter (threshold slope), and a miscalibrated Bayesian readout for confidence also introduced one additional parameter (estimated variability ratio). Models 5 and 6 have the same parameters as models 2 and 4 respectively, but are not the same model. Models 5 and 6 feature a Bayesian readout of confidence (which introduces no additional parameters; Table 2.1). Details of the role of the parameters in the computational model are provided in Appendix A.2.

in the accumulator, and it affects each evidence measurement independently. All models also had a parameter for the standard deviation of metacognitive noise, σ_m (Section 2.1). We fit the height of the decision threshold, a, (with one decision threshold at a and the other at $-a$,) and the duration of sensory and motor processing pipelines, I. Finally, all models had a lapse rate parameter, λ, and three parameters describing the bounds between the four confidence bins, d_i. In the model, confidence reports are determined by the location of a continuous variable x_c. If x_c falls between the boundaries d_i and d_{i+1} then a confidence report falling in bin i is given. Further details of the roles of the parameters in the computational model are provided in Appendix A.2.

In addition to these parameters, specific model variants had additional parameters. All models that included drift-rate variability had a parameter describing its standard deviation, σ_φ. All models with time-dependent decision thresholds included a parameter describing the slope of the decision threshold as a function of time, b. (Upper and lower thresholds were then given by $a - bt$ and $-a + bt$ where t is the time spent accumulating evidence.) Models using a miscalibrated Bayesian readout of confidence included one additional parameter, Γ, accounting for the incorrect beliefs of the observer. Specifically, Γ, is related to the observer's belief about the ratio between drift-rate variability, and total variability from drift-rate, accumulator noise, and from the stimulus. Incorrect estimation of any of these sources of variability will lead to a value for Γ that differs from the value used by a calibrated Bayesian. For details of the exact role of the parameters see Appendix A.2.

Note that we fit data from both the free response and interrogation condition simultaneously, and all parameters relevant to both conditions were shared between them. (Decision threshold height and slope, and pipeline duration are only relevant in the free response condition.) This provides a very strong test for the models because we are requiring the models to fit two very different sets of patterns using the same parameter values (Ratcliff, 2006).

Fitting

At any hypothesised value of the parameters, $\boldsymbol{\theta}$, we can compute the likelihood. Assuming the confidence report given on a trial is conditionally independent of the confidence report given on any other trial, the likelihood of the parameters is given by,

$$L(\boldsymbol{\theta}) = p(C^{(1)}, C^{(2)}, ...|\boldsymbol{\theta}, \mathbf{E}^{(1)}, R^{(1)}, t_r^{(1)}, \mathbf{E}^{(2)}, R^{(2)}, t_r^{(2)}, ...) \tag{3.1}$$

$$= \prod_i p(C^{(i)}|\boldsymbol{\theta}, \mathbf{E}^{(i)}, R^{(i)}, t_r^{(i)}) \,, \tag{3.2}$$

where $C^{(i)}$, $\mathbf{E}^{(i)}$, $R^{(i)}$, and $t_r^{(i)}$ are the confidence report, stimulus, response, and response time on trial i. We fit the models to the data by maximising the log-likelihood. All data were used in either the fitting or fit evaluation; all participants,

and every trial in which a confidence report was obtained were included. The lapse rate parameter was included to account for any random responses, so exclusions are not necessary.

A single fit began with the evaluation of the likelihood function at 200 randomly drawn sets of parameter values. The parameter set with the highest log-likelihood was used as the start point for MATLAB's fmincon optimiser ('Matlab Optimization Toolbox', 2017). For details of the limits applied to parameters during optimisation, and the way the 200 candidate sets of parameters were drawn, see Appendix A.12.

We repeated this entire process 40 times for every fit (that is, for each participant, model, and cross validation fold; see below). Rerunning fitting many times improves the chance of finding the true maximum likelihood, rather than getting stuck in local maxima, and allows one to perform a heuristic assessment of the optimiser's performance (Acerbi et al., 2018). For our assessment see Appendix A.9.

Model comparison

We compared the fit of the models using 5-fold cross validation, which automatically penalises flexibility (Lewandowsky & Farrell, 2011), and does not require explicit estimation of a penalty as in the Akaike information criterion (AIC), and the Bayesian information criterion (BIC). Every fold, 4/5 of the trials were used for training and 1/5 used as test trials. We fit the models to the training trials, and evaluated the models using the average negative log-likelihood in test trials (Honig et al., 2020). We computed this value for each fold, and averaged across folds to provide a measure of the performance of each model for each participant. We refer to this measure as the negative cross-validated log-likelihood (–LLcv). A more negative –LLcv value indicates that the model was better at predicting the test data.

To determine the overall best fitting model we computed the mean –LLcv across participants. To assess the reliability of the difference in fit between this and other models we computed, participant-by-participant, the difference between the –LLcv

of all models and that of the best fitting model. We then computed the mean –LLcv difference, along with 95% confidence intervals found by bootstrapping 10000 times.

We also fit the entire dataset without dividing into training and test data. Evaluating these fits using the AIC and BIC gave similar results to those of the cross validation (Appendix A.10). To further satisfy ourselves that the model fitting was recovering the true underlying model, we ran a model recovery analysis, simulating data from each of the 10 models, fitting each set of simulated data with all 10 models, and then evaluating the AIC and BIC to confirm the correct model was recovered. This analysis suggested that the AIC and BIC could be too conservative in the current context, but otherwise the results were as expected (Appendix A.11).

To plot the behaviour of the fitted models, we ran a simulation for each participant using the fitted parameters. We use the parameters resulting from the fits to the entire dataset (with no training-test split applied; further details in Appendix A.3).

3.4 Modelling results

Using computational modelling, we aimed to determine which of the 10 DDM models of confidence fit the data best (Table 2.1), and whether any DDM model could provide an adequate quantitative account of confidence.

Model comparison

We fit the models and evaluated their performance using cross-validation. Model 7, in which the observer uses a miscalibrated Bayesian readout of confidence, provided the best fit to the data (Table 2.1 and Fig. 3.4). However, the fit was not reliably better than for model 9, which also includes time-dependent decision thresholds. The fit for model 8, which features drift-rate variability as well as a miscalibrated Bayesian readout, was nearly as good. Models in which confidence reflects the final state of the accumulator, or in which confidence reflects a calibrated Bayesian

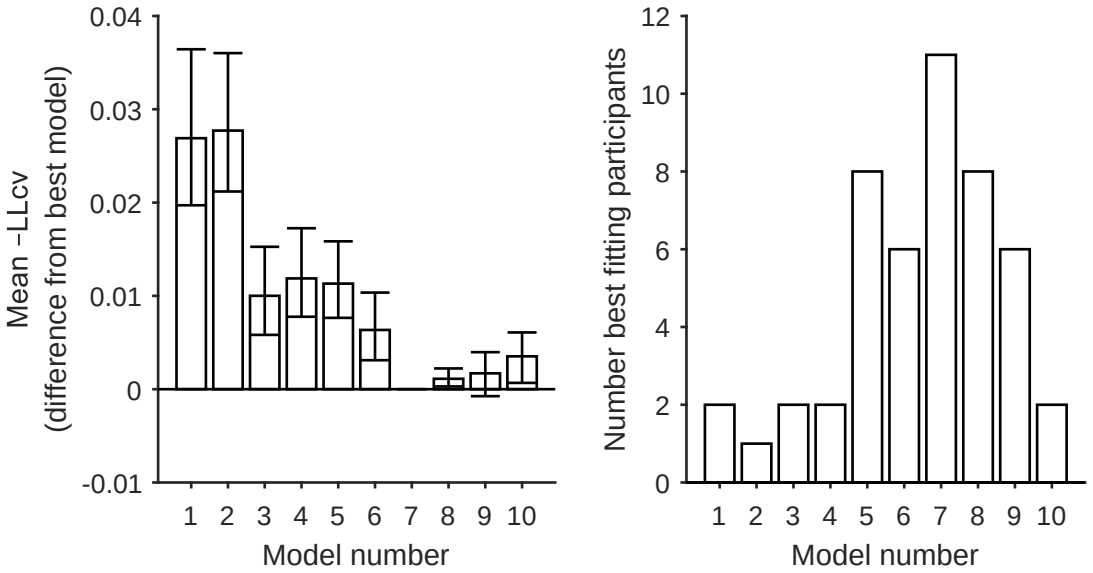

Figure 3.4: Negative cross-validated log-likelihood (–LLcv) relative to overall best fitting model, and number of participants for which each model provided the best fit. A lower value of –LLcv indicates better fit. Models in which confidence reflects a miscalibrated Bayesian readout fit best (models 7–10). Unlike in other plots, error bars represent 95% bootstrapped confidence intervals.

readout, models 1-6, performed worse. These results suggest that confidence reflects a miscalibrated Bayesian readout, but it is not clear whether observer's used time-dependent thresholds, or whether drift-rate variability is present. A similar picture emerged using AIC and BIC for model comparison (Appendix. A.10). Considering fits to individual participants, model 7 was the best fitting model for the greatest number of participants (Fig. 3.4). Two other models with a miscalibrated Bayesian readout also best fit large numbers of participants (models 8 and 9). Interestingly, the two models with a calibrated Bayesian readout (models 5 and 6) provided the best fit for many participants.

We next sought to understand why some models fit better than others. Two-stage dynamic signal detection theory (2DSD) struggles to account for the strength of the relationship between confidence and time (Section 1.3.1; Pleskac and Busemeyer, 2010). Model 2 is very similar to 2DSD. While model 2 can account for some decrease in confidence with time in the free response condition, because it contains drift-rate

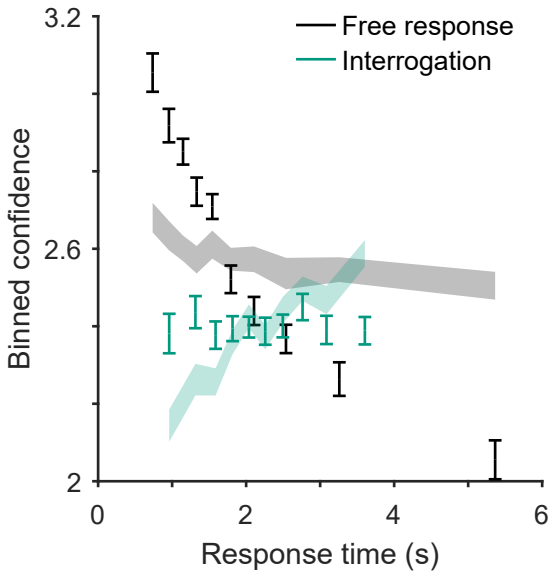

Figure 3.5: Effect of response time on confidence in the data (error bars), and in model 2 (shading). Model 2 did not capture the strength of the effect of response time on confidence in the free response condition. Error bars and shading represent ±1 SEM. Plotting details in Appendix A.4.

variability (Section 1.3.1; Pleskac and Busemeyer, 2010), it clearly struggles to capture the strength of this relationship (Fig. 3.5). This is very similar to the conclusion drawn by Pleskac and Busemeyer (2010) in relation to 2DSD. In the interrogation condition model 2 does not even match the qualitative pattern in the data: Model 2 predicts increasing confidence with response time but in the data confidence does not change with response time.

It may be more surprising that models using a calibrated Bayesian readout for confidence (models 5 and 6) also struggle to account for the relationship between confidence and time. As discussed in Chapter 2, these models may be able to capture a stronger relationship between confidence and response time due to difficulty discounting, so we might expect them to perform better. Model 5, which uses a calibrated Bayesian readout, only slightly underestimates the decrease in confidence with response time in the free response condition, but it struggles to account for the finding that confidence changes little with response time in the interrogation condition (Fig. 3.6). Accuracy does increase with response time in the interrogation

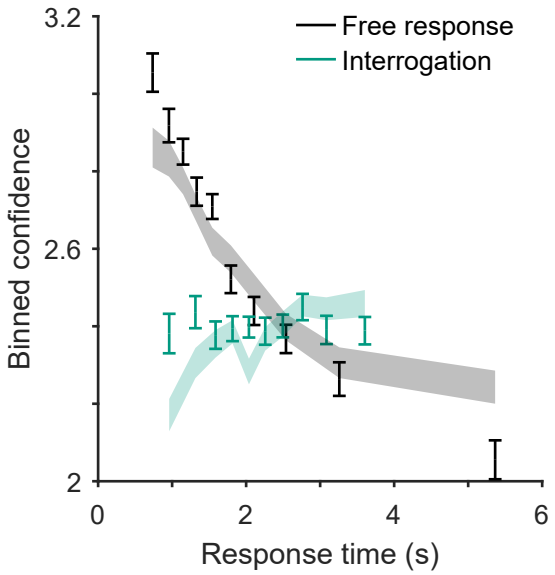

Figure 3.6: Effect of response time on confidence in the data (error bars), and in model 5 (shading). Model 5 slightly underestimated the effect of response time on confidence in the free response condition. Error bars and shading represent ±1 SEM. Plotting details in Appendix A.4.

condition (Fig. 3.2, A), suggesting a mismatch between the observers' model of the environment and the true generative model (Section 3.2; Drugowitsch et al., 2014). Specifically, this reasoning suggests that participants may be applying a level of difficulty discounting that is too strong, and reducing their confidence too rapidly with response time. The reduction may be so strong that it completely cancels the effect on confidence of the accumulation of evidence over time.

Consistent with this reasoning, the best fitting model (model 7) has a miscalibrated Bayesian readout for confidence. This model is capable of simultaneously capturing the effects of response time in the two conditions, with confidence decreasing as a function of response time in the free response condition while remaining relatively constant in the interrogation condition (model 7; Fig. 3.7, A).

Fits of the best model

Having seen that model 7 outperformed other models in the cross validation, and explored why this was the case, we next asked whether this model could account for a

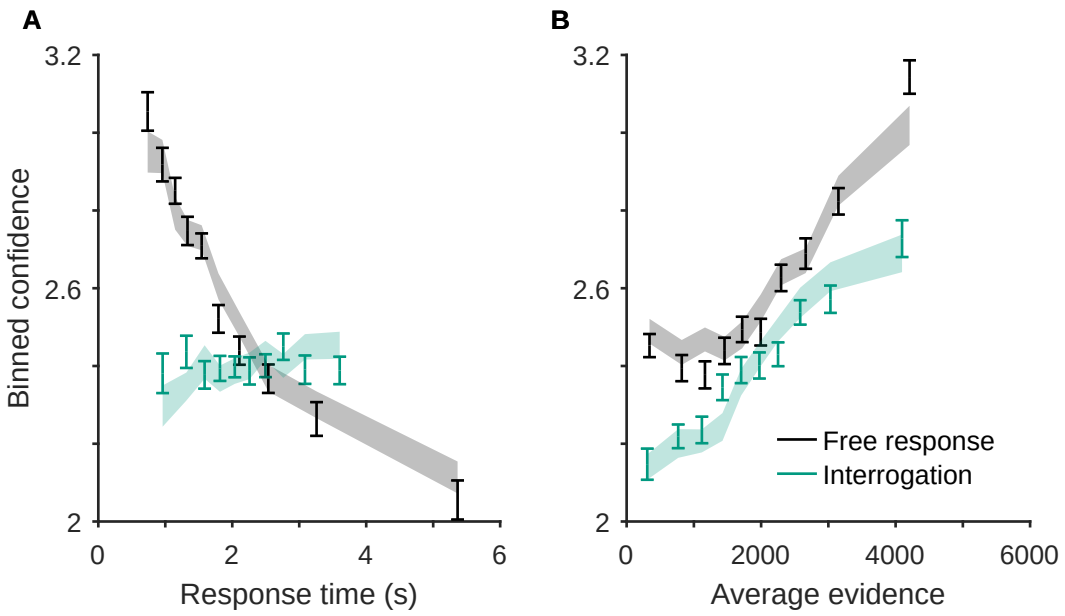

Figure 3.7: Effect of response time and average evidence on confidence in the data (error bars), and in the best fitting model, model 7 (shading). Model 7 accounted well for quantitative patterns in the effects of both response time and average evidence, and in differences between the two conditions. Average evidence is computed by summing, over all frames, the difference in dots presented in the two arrays, before taking the absolute value and dividing by the time the stimulus was presented for. Error bars and shading represent ±1 SEM. Plotting details in Appendix A.4.

wide range of patterns in confidence data. For example, in addition to looking at the effects of response time on confidence, we also wanted to see if model 7 could account for the effect of evidence. We computed the absolute value of the average evidence presented over the course of each trial, and plotted confidence as a function of this measure. The model also accounts well for the effect of average evidence, except at the very highest values (Fig. 3.7, B). Importantly, the model again captures how this relationship differs in the free response and interrogation conditions.

We have seen that model 7 can account for the effect of response time, and separately, that it can account for the effect of evidence. A potential concern is that we have not shown that the model is capturing the unique effects of these variables. For example, it could be that model 7 largely predicts a relationship between evidence and confidence because these both correlate with response time, but that in the data evidence has a strong effect even at a fixed response time.

To examine whether the model can account for the independent effect of evidence and response time, we plotted these simultaneously (similar to the plots of time and task difficulty in Kiani et al., 2014). On a participant-by-participant basis, and separately for the two conditions, we grouped trials into 3 bins based on the absolute value of the total evidence presented, then assessed the relationship between response and confidence for each bin separately.

In Fig. 3.8, A and B, we plot for the free response and interrogation conditions respectively, confidence as a function of response time, separately for low, medium and high total evidence trials. Model 7 captures the effect of response time at all levels of evidence, and also accounts for increased confidence with greater evidence at a fixed response time. The model captures differences between the two conditions, such as the stronger effect of evidence at fixed response time in the interrogation condition. There appear to be some systematic deviations for the fastest and slowest response times within an evidence bin, but in general the model provides a very good fit. Two mechanisms allow the model to account for these independent effects of response time and evidence. First, difficulty discounting leads to an effect of time independent of the effect of evidence. Second, the accumulation of pipeline evidence allows for an effect of evidence which is independent of any effect of response time.

In addition to the effects of response time and evidence, another key pattern for any model of confidence is the relationship between accuracy and confidence. The model provided an excellent fit to the accuracy at different confidence levels in both conditions, with one exception (Fig. 3.9). The model over estimated accuracy at the lowest level of confidence in the free response condition. Nevertheless, even here the model correctly fitted the qualitative pattern of greater accuracy in the free response condition, than in the interrogation condition, for this lowest level of confidence.

As our aim is to build a DDM model of confidence which unifies and explains a wide range of phenomena, we wanted to look at whether the model could also account for the precise nature of the effect of evidence on confidence. To do this we looked at whether model 7 could account for the way confidence varies with the

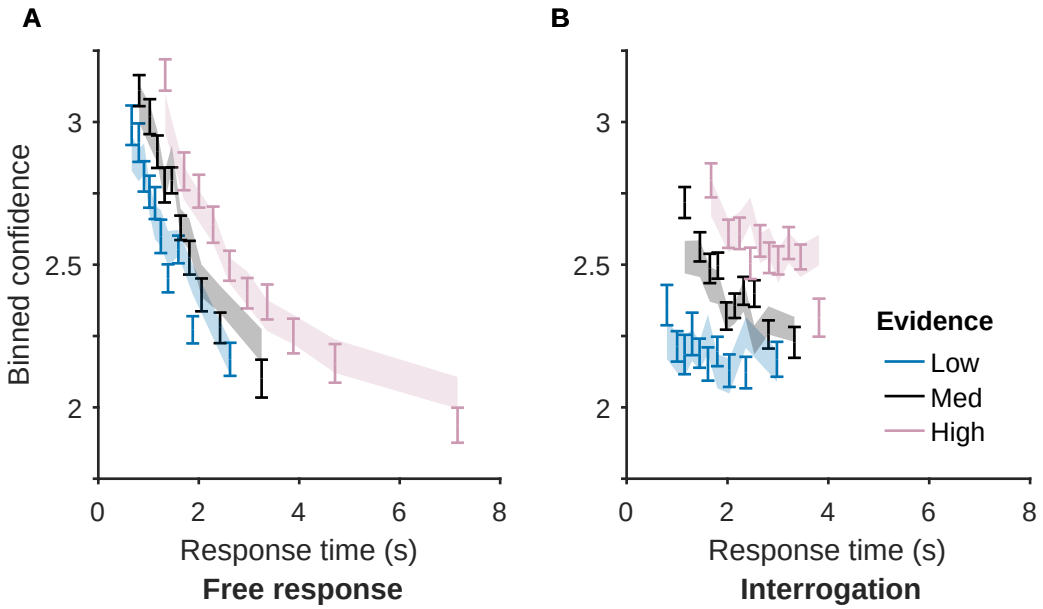

Figure 3.8: Effect of response time and evidence, considered simultaneously. Effect on confidence shown for the data (error bars), and in the best fitting model, model 7 (shading). Except at the longest and shortest response times, model 7 accounted well for the simultaneous effect of time and evidence. Evidence is computed by summing, over all frames, the difference in dots presented in the two arrays, before taking the absolute value. We binned trials according to this value, separately for the two conditions. Error bars and shading represent ±1 SEM. Plotting details in Appendix A.4.

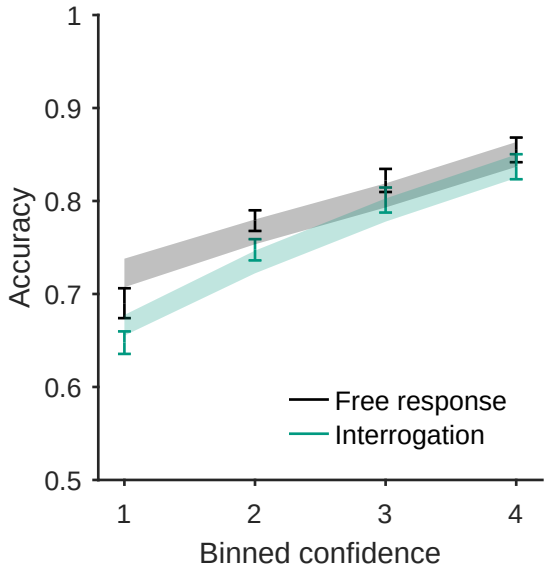

Figure 3.9: The relationship between confidence and accuracy in the data (error bars), and in the best fitting model, model 7 (shading). Generally, the model captured quantitative and qualitative patterns well. Error bars and shading represent ±1 SEM. Plotting details in Appendix A.4.

evidence presented at different time lags relative to the response and trial onset. Considering first the effect of evidence at time lags relative to the response (Fig. 3.10, A), we see that in the free response condition the model successfully accounts for the weak correlation between evidence presented well before a response, and the response made. It also accounts for the strong effect of evidence presented immediately prior to a response, which will be in sensory and motor processing pipelines at this time (Section 1.3.1; Resulaj et al., 2009). Unlike pre-decision evidence, which only affects confidence by changing the time at which the decision threshold is crossed, in the model pipeline evidence can directly affect the final balance of evidence (Chapter 2). While the pattern in the free response condition is largely captured, there are some small deviations of the model from the data: For evidence presented around 1.5 seconds prior to response, the model appears to underestimate the effect of this evidence on confidence. One possibility is that these differences arise because there is significant between trial variation in the duration of the processing pipeline, a commonly made assumption (e.g. Ratcliff & McKoon, 2008; van den Berg, Anandalingam et al., 2016). In the interrogation condition the model captures the observed effects well, with evidence at all time points contributing approximately equally to confidence (Fig. 3.10, A). The model slightly overestimates the effect of evidence received close to a response. In the model evidence from all time points is weighted equally in the observer's decision. However, on short trials, the effect of individual frames on confidence is likely to be greater because fewer frames are used to determine confidence. Note that in Fig. 3.10, A, short trials will only contribute to plotted time points close to the response. Finding that the model accounts well for the effect of evidence in the run up to a decision supports the model's assumptions, including the assumption that participants continue accumulating evidence until the end of each trial, and that participants do not use an implicit decision threshold (Section 3.2).

The pattern is different when we look at the effect of evidence at time lags relative to the onset of each trial (Fig. 3.10, B). Here, both in the free response and interrogation condition, the model underestimates the effect of evidence presented

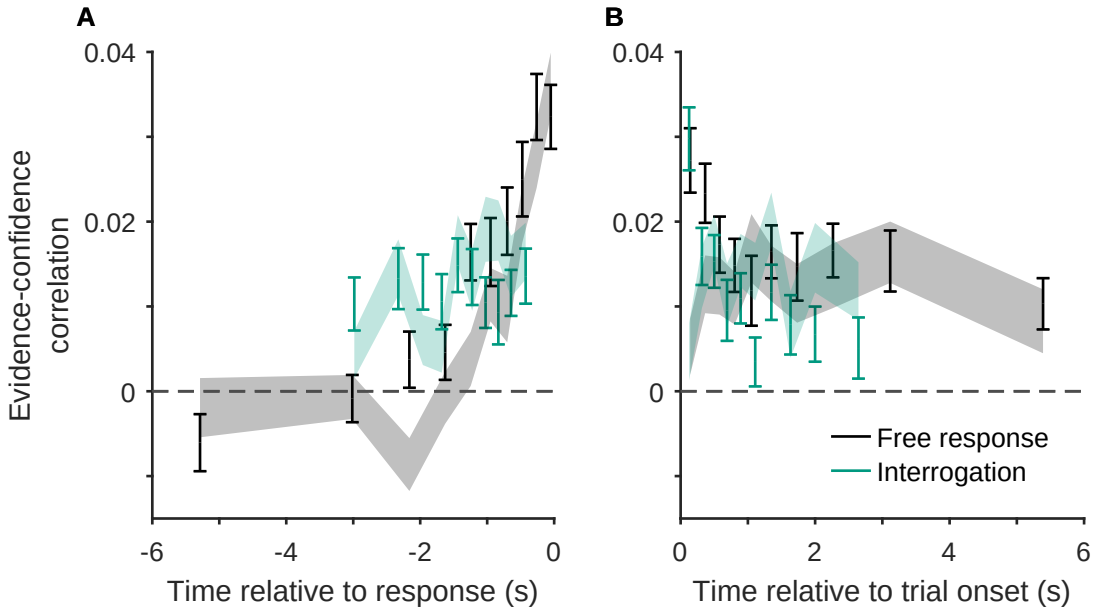

Figure 3.10: The effect of evidence fluctuations on confidence in the data (error bars), and in the best fitting model, model 7 (shading). To measure this effect we computed the rank correlation (Kendall's Tau) between evidence fluctuations and confidence. The model accounted well for the effect of evidence at time lags relative to response, in both conditions. However, the model failed to capture the strength of the effect of evidence presented at the onset of the stimulus. Error bars and shading represent ±1 SEM. Plotting details in Appendix A.4.

at the beginning of a trial. The model assumes all evidence presented over the course of a trial is given equal weight as it is accumulated. Therefore, the deviation of the model from the data provides further support for the idea that there may be an over-weighting of evidence presented early in a trial (Section 3.2).

See Appendix A.13 for the fitted parameter values of the models on which we have focused.

Simulating responses and response times

We have fit the models to confidence reports given the response, response time and evidence presented on each trial, finding that models with a miscalibrated Bayesian readout of confidence can provide an excellent fit to most aspects of the data. While we have not fit responses and response times, a core hypothesis motivating all models used is the idea that the same process that generates decisions

also generates confidence (Fig. 2.1). Given that we now have estimates for all parameters of the evidence accumulation process that leads to confidence (Table 3.1), if our hypothesis is correct, we should have estimates of all the parameters required to simulate responses and response times (see Kiani et al., 2014 for related approach). It may be that some parameters have a large effect on accuracy, but small effect on confidence, and so are relatively poorly estimated from fits to confidence data alone. A good example of such a parameter is decision threshold height. It is clearly weakly constrained by confidence data as, in the parameter fits for model 7, the 75th percentile across participants is nearly twice the median value (Appendix A.13). On the other hand, bound height likely has a very strong effect on response times and accuracy, and hence would be well constrained if we had directly fit to these quantities instead of aiming to predict them. Nevertheless, we expect simulations using parameters from confidence fits to at least approximate properties of responses and response times.

Without any additional fitting, we simulated new trials and stimuli using the parameters from the fits to confidence data. We simulated entire diffusion processes, from onset, through decision, to confidence report (Appendix A.3). The predictions for accuracy, and how accuracy changes with response time were sensible, although there were clear differences between the model and the data (Fig. 3.11). For example, accuracy was overestimated. We have not included in the models a lapse rate parameter for responses. We estimated a lapse rate for confidence reports, the proportion of trials on which a random confidence report is given, but there could be more response lapses than confidence lapses. Supporting this view, the median fitted confidence lapse rate parameter for model 7 was very low (0.17 %), suggesting that decision lapses have not been fully captured. The general overestimation of accuracy might therefore be a result of not fitting a decision lapse rate parameter. However, accuracy appeared to be specifically overestimated at late response times (Fig. 3.11), suggesting other explanations. It may be that drift-rate variability is important in accounting for decisions and response times, but has relatively little effect on confidence. In free response tasks, drift-rate variability causes slow

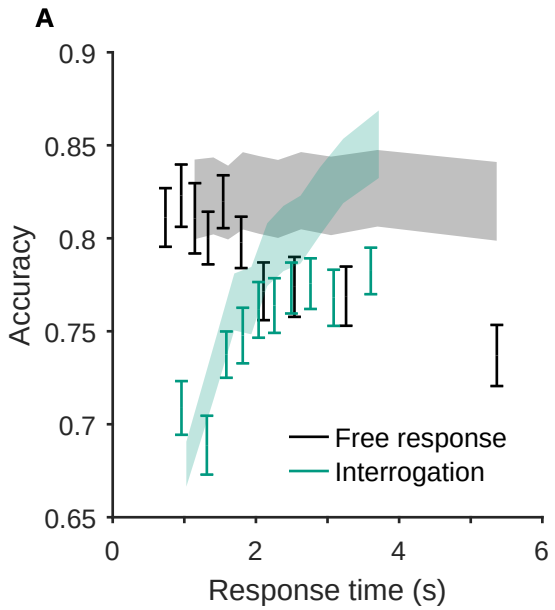

Figure 3.11: The effect of response time on accuracy in the data (error bars), and as predicted by the best fitting model, model 7 (shading). The predictions of the model were sensible, although there were clear differences to the data. The model predictions for accuracy were too high at long response times. Error bars and shading represent ±1 SEM of the mean. Plotting details in Appendix A.4.

errors, and hence would predict decreased accuracy in trials with slow responses (Ratcliff & McKoon, 2008). Drift-rate variability also limits the performance that can be achieved in interrogation tasks at large response times (Ratcliff, 1978). While drift-rate variability doesn't appear to be needed to account for confidence data, this line of reasoning suggests that if we fit to responses, response times and confidence simultaneously, we may find more evidence for models featuring drift-rate variability.

We can also look at whether the diffusion mechanism assumed by the model, and fitted using confidence data, can account for the relationship between evidence and response at different time lags. As was the case for the relationship between evidence and confidence, the model underestimated the effect of evidence presented at the beginning of a trial (Fig. 3.12, B). The fact that the model fails in this regard for both the evidence-response and evidence-confidence relationships is consistent with the idea that an over-weighting of early sensory evidence could be included in the model to improve the fit. The model accounted reasonably well for the effect of

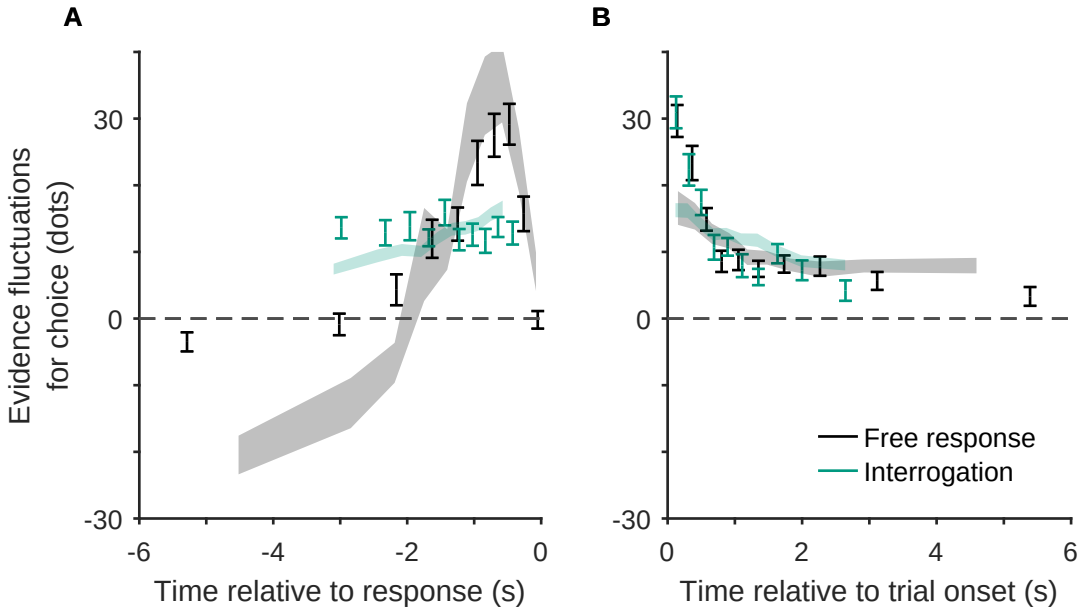

Figure 3.12: The effect of evidence fluctuations on choices, in the data (error bars), and as predicted by the best fitting model, model 7 (shading). The model predictions were generally reasonable, and captured some key qualitative effects. The model predictions did not capture the strength of the effect of evidence presented at the onset of the stimulus (subplot B). Error bars and shading represent ±1 SEM of the mean. Plotting details in Appendix A.4.

evidence at time lags relative to response including, in the free response condition, the increase and decrease in effect in the run up to a response (Fig. 3.12, A). The model predicted that, in the free response condition, evidence fluctuations at lags far before response would favour the unchosen option, presumably because initial evidence fluctuations favouring the alternative is the only situations in which very long trials occur. However, data from participants did not clearly demonstrate this effect. This may be because in reality very long trials are the result of lapses. (Note that there was a trend in the data for the very longest responses; see also Fig. 5 of Charles and Yeung, 2019)

Taken together, these model predictions do not contradict the view that the same underlying model, the model that we fit to confidence reports, also explains response times and decisions. Of course the model would align closer with the data if we had fit to decisions and response times in addition to confidence. However, using parameters estimated by fitting to confidence reports, and without fitting

any additional parameters, we can generate reasonable predictions for patterns in response and response time data.

3.5 Discussion

In this chapter we built on the work in Chapter 2 showing that DDM models of confidence can account for key qualitative patterns in the relationship between confidence and difficulty, accuracy, speed emphasis, and response time. Here we collected a new dataset and, in a preregistered test, compared patterns in the data to three of the qualitative predictions made by 10 DDM variants (A-C, Section 2.1). The predictions concerned the relationship between confidence, evidence, and response time, as a function of whether the choice over when to respond lies with the decision maker or is externally imposed. All three predictions were verified providing support for the assumptions on which they were based. The assumptions used were that there is a single evidence accumulator for both decisions and confidence, and that observers use a decision threshold in the free response condition but not in the interrogation condition.

Given this validation of the assumptions shared by all DDM variants considered, we next tested the ability of the DDM variants – which feature combinations of drift-rate variability, decreasing thresholds, and Bayesian confidence – to account for precise quantitative patterns in confidence reports. A particular strength of the modelling was that we used trial-by-trial predictions for confidence, given the response time, response, and evidence shown in each trial. As a result, the fits of the models reflect how well they account for the effect of each and every stimulus shown. In this way we capitalised on variability of the dynamic stochastic stimuli used. Furthermore, through simulations, this approach allowed us to explore the effect of specific features of stimuli, and the model fits to these features.

DDM variants that featured a miscalibrated Bayesian readout of confidence fit the data best. We found the best fitting model provided an excellent account of the

relationships between confidence and response time, and confidence and evidence. We set up a very strong test for the models by requiring them to simultaneously fit data from the free response and interrogation conditions, with all parameters relevant to both conditions shared (Ratcliff, 2006). This test is especially strong because we expected these two conditions to generate very different relationships between confidence and time (Chapter 2; Pleskac and Busemeyer, 2010). Indeed, this was the case. Nevertheless, a DDM variant with a miscalibrated Bayesian readout for confidence provided an excellent fit to the data in both conditions.

There are wide implications of the finding that the DDM – which accounts well for decisions and response times (Ratcliff & McKoon, 2008; Ratcliff et al., 2016) – can capture key patterns in confidence data. In particular, we need not abandon the claim that the brain uses normative decision mechanisms for one of the most basic cognitive functions, perceptual decision-making. We can explain confidence reports from within a framework in which the difference in evidence is accumulated, instead of needing to rely on non-normative models such as those in which evidence for the two alternatives is separately tracked (Section 1.3.1; De Martino et al., 2013; Kiani et al., 2014; van den Berg, Anandalingam et al., 2016; Zylberberg et al., 2012). The results also show that we can account for patterns in confidence data while maintaining that decisions and confidence result from the same evidence accumulation, instead of hypothesising one accumulator for decisions and a second accumulator for confidence (Section 1.3.2; Balsdon et al., 2020; Fleming and Daw, 2017; Jang et al., 2012). This is attractive from a theoretical point of view because it means we are not committed to the idea that the brain accumulates noisy versions of exactly the same information twice. As a result the brain only needs a single population of neurons to track evidence, reducing energy consumption (Lennie, 2003).

There are specific implications of the success of models with a miscalibrated Bayesian readout. The idea that confidence reflects a Bayesian readout has proved successful in the context of non-normative models of perceptual decisions (Kiani

et al., 2014; van den Berg, Anandalingam et al., 2016). We have seen that the idea of a Bayesian readout is also successful when using the normative framework of the DDM, building on work looking at the theoretical implications of such a model (Moreno-Bote, 2010), and complimenting recent empirical findings (Desender et al., 2020). That a DDM with a miscalibrated Bayesian readout can account for a wide range of patterns observed in previous research, and in this study, indirectly supports the view that confidence is Bayesian in the ongoing debate about this claim (Adler & Ma, 2018a; Li & Ma, 2020; Meyniel et al., 2015; Navajas et al., 2017; Peters et al., 2017; Sanders et al., 2016). This result also provides a resolution to the debate discussed in Chapter 1 between models in which confidence reflects the evidence accumulated (e.g. Pleskac & Busemeyer, 2010; Vickers & Packer, 1982), and those in which confidence reflects response time (e.g. Petrusic & Baranski, 2009; Zylberberg et al., 2012). Under a Bayesian readout, confidence is a function a both evidence and response time (Moran, 2015; Moreno-Bote, 2010). Under a miscalibrated Bayesian readout, confidence remains a function of time even if this is not normative given the structure of the task.

An important question to ask is about the way in which observers are miscalibrated, such that their confidence is a function of time, even when this is not normative. The observer we considered could only be miscalibrated in a specific way: By incorrectly estimating the magnitude of different sources of variability (Chapter 2; Section A.2.2). This form of miscalibration is consistent with the finding that humans are poor at dealing with stimulus variability (de Gardelle and Mamassian, 2015; Herce Castañón et al., 2019; Zylberberg et al., 2016; Zylberberg et al., 2014). More specifically, what seems to be key for the miscalibrated Bayesian observer model to explain the data, is the possibility that observers incorrectly believe there is a substantial amount of trial-to-trial variability in drift-rate. The importance of this possibility is clearest when we consider model 7. In this model there is no internal trial-to-trial variability in drift rate and in such a situation, for the stimuli we presented, a Bayesian readout does not use time, only the final state of the accumulator (Section A.2.2; Moreno-Bote, 2010). However, if the observer

incorrectly believes that there is trial-to-trial variability in drift rate, then their Bayesian readout of confidence will take into account both the final state of the accumulator and the time spent accumulating evidence.

While we have thought of drift rate variability as something that arises internally (Chapter 1; Ratcliff and Smith, 2004), trial-to-trial variability in drift-rate can equally be produced by a stimulus that varies in the average level of evidence it provides from trial-to-trial (Ratcliff & Smith, 2004). Note our stimuli provided different amounts of evidence on each trial because they contained frame-by-frame variability in evidence. This is different from trial-by-trial variability which affects all frames in a trial in the same way, for example, boosting the evidence provided by every frame in a trial. Hence, one possible explanation for the results suggesting observers incorrectly estimate drift-rate variability is that observers incorrectly believe there is trial-by-trial variability in the stimuli, perhaps confusing this with the frame-by-frame variability which is indeed present in the stimuli used.

Our results build on the two-stage dynamic signal detection theory (2DSD) of Pleskac and Busemeyer (2010) by showing that DDM variants can not only explain decreasing confidence with response time, but can successfully account for the strength of this relationship. 2DSD can predict some relationship between confidence and response time due to the presence of drift-rate variability, but struggles to account for the strength of this relationship (Section 1.3.1; Pleskac and Busemeyer, 2010). It is surprising that model 2, which is closely related to 2DSD and contains drift-rate variability, did not outperform model 1, which lacks drift-rate variability (Table 2.1). However, the models were not just fitting the decrease in confidence with response time that was observed in the free response condition. With the same parameter values for those parameters relevant to both conditions, the models needed to simultaneously account for the very different relationship between confidence and response time in the interrogation condition. In attempting to also fit data from the interrogation condition, parameters in model

2 may have been driven to values at which any advantage of the model over model 1 in the free response condition was lost.

Although model comparison results were clear regarding the type of readout used for confidence – i.e., that models including a miscalibrated Bayesian readout outperform all others – it was unclear whether decreasing decision thresholds or drift-rate variability are also important. As discussed, studies investigating whether humans use time-dependent thresholds have provided mixed results (Evans et al., 2020; Hawkins, Forstmann et al., 2015; Malhotra et al., 2017; Palestro et al., 2018; Pardo-Vazquez et al., 2019; Voskuilen et al., 2016). One explanation for why we did not find a clear result regarding decreasing decision thresholds, is that humans use a decision threshold which only slightly deviates from flat, making time-dependence difficult to detect (Voskuilen et al., 2016). Our motivation for considering decreasing decision thresholds was that such thresholds can be optimal (Drugowitsch et al., 2012; Malhotra et al., 2018). However, we did not compare flat decision thresholds to the optimal decision thresholds. Instead we compared flat thresholds to decreasing thresholds, where we fit the slope of the threshold as a free parameter. Removing this free parameter by computing the optimal decision thresholds might allow us to draw stronger conclusions.

Drift-rate variability is a central component of many DDM models (Ratcliff & McKoon, 2008; Ratcliff et al., 2016; Ratcliff et al., 1999). It may be surprising then that we did not find clear evidence in favour of models featuring drift-rate variability. The explanation for this difference may be that we fit only to confidence data. The inclusion of drift-rate variability in DDM accounts has been justified on the grounds that it explains key empirical phenomena regarding the speed of correct and error responses (Ratcliff & McKoon, 2008; Ratcliff et al., 2016; Ratcliff et al., 1999). Thus, drift-rate variability may be an important part of explanations for patterns in responses and response times, but not an important part of understanding patterns in confidence.

It is important to note that some features of the data remained unexplained by the best fitting model. The model couldn't explain why evidence presented at the onset of the stimulus had a stronger effect on responses and confidence than evidence presented later. We speculate that this effect arises because initial sensory samples are over-weighted, but acknowledge that changes to the decision mechanism itself might be able to account for this effect (Okazawa et al., 2018; Tsetsos et al., 2012). There are presumably numerous perceptual and cognitive phenomena which could be found in the data, but which we have not modelled, such as Weber's law (Cobb, 1932), adaptive gain control (Cheadle et al., 2014), and confidence leak (Rahnev et al., 2015). From the perspective of models that don't incorporate these features, the data may look more noisy, with more unexplained variance, but the presence of other phenomena does not immediately invalidate an approach.

None of the models considered include the idea of a confidence threshold, following the decision threshold, that determines the time of the confidence report (Moran et al., 2015; Pleskac & Busemeyer, 2010). Instead, we simply made the parsimonious assumption that the entire stream of evidence presented in the stimulus is processed, and once this is complete, confidence is reported (Section 2.1; Moran et al., 2015). This is a strength of the model: We can explain the rich pattern in confidence data without needing to postulate additional mechanisms. However, it does mean that the models would need to be extended to make predictions for the timing of confidence reports (Moran et al., 2015; Pleskac & Busemeyer, 2010). In the special case of simultaneous responses and confidence reports, the way to extend the models is clear. In this situation confidence presumably reflects a (possibly Bayesian) readout of evidence accumulated up to the time of the decision (Desender et al., 2020; Kiani et al., 2014). In other cases a "confidence boundary", which is similar to the decision threshold but determines the time of the confidence report, could be included in the model (Moran et al., 2015).

A broader concern is with the nature of the miscalibrated Bayesian observer models we used. Specifically, we may worry that these models, in which we do not

constrain the observer's generative model to match the true generative model, are too flexible (Bowers & Davis, 2012; Jones & Love, 2011; Rahnev & Denison, 2018). We allowed the possibility that observers incorrectly estimate the magnitude of difference sources of variability. This idea is supported by previous findings that humans deal poorly with noise introduced by stimulus variability (de Gardelle & Mamassian, 2015; Herce Castañón et al., 2019; Zylberberg et al., 2016; Zylberberg et al., 2014). Additionally, where noise is due to external sources rather than internal ones, it is necessarily the case that an observer will use an incorrect model of this noise, until they have had time to learn the relevant statistics. Finally, model comparison using cross validation showed that the miscalibrated Bayesian observer models were best at predicting data not used for training, suggesting that the extra flexibility of these models is warranted.

Notwithstanding the limitations discussed, we have seen that models in which decisions and confidence are generated by the same evidence accumulator, an accumulator which tracks the difference in evidence between two alternatives, can account for a wide range of qualitative and quantitative patterns in confidence. Hence, we do not need to abandon the idea that perceptual decision-making has a normative basis, or the idea that the brain will save on neural hardware when it can. Throughout we have seen that one idea in particular, the idea of a miscalibrated Bayesian readout, provides a powerful framework for understanding and predicting confidence.

4

A new modelling method

Contents

4.1 Introduction

In this chapter we present a new method for making trial-by-trial predictions for confidence. In Chapter 1 we noted that trial-by-trial modelling using SDT is often feasible because these models posit a single static representation of evidence that is not too computationally expensive to simulate. In contrast, in models featuring a dynamic representation of evidence, the computational cost of simulating behaviour can be extremely high because the entire evolution of the evidence

representation must be simulated. If simple mathematical expressions can be found for the predictions of a model, then these can be used instead of simulating the model's predictions. Here the aim was to derive simple mathematical expressions for confidence in DDM accounts, thus avoiding the need to simulate a dynamic representation of evidence, and making trial-by-trial modelling computationally feasible. We used the resulting derivations for the modelling work in Chapter 2. We chose this counter-intuitive order of presentation because the modelling in Chapter 2 provides both context and motivation for the technical work in this chapter.

In Appendix F we summarise an initial approach we took to this problem, in which we aimed to simplify the modelling problem by directly inferring the shape of the decision thresholds using recently developed methods (Malhotra et al., 2017). Unfortunately, using simulations to test this initial approach we found it to be unreliable in the context of our studies, but in interesting ways. In particular, in the presence of a decision threshold, there is no simple relationship between the amount of evidence presented and the internally accumulated evidence (while the method assumes there is a simple relationship). Prior to a decision, even if changes to the state of the evidence accumulator are normally distributed over small intervals of time, with a mean determined by the evidence presented, the probability distribution over the state of the accumulator will not be normal. This is because, having reached a time t without a response, we know that the accumulator is not beyond either threshold, nor has it been at any point up to t (otherwise the observer would already have made a decision; Moreno-Bote, 2010). This constraint results in non-normal probability distributions over accumulator state (Fig. 4.1; Cox and Miller, 1965; Smith, 2000).

Our contribution is not to solve this issue and provide new expressions for the probability distribution over response times and decisions in diffusion models (see Section 1.7 for a discussion of existing approaches). Instead, we note that even though there is not a simple relationship between presented evidence and changes in the internal evidence accumulation prior to a decision, there is still a simple

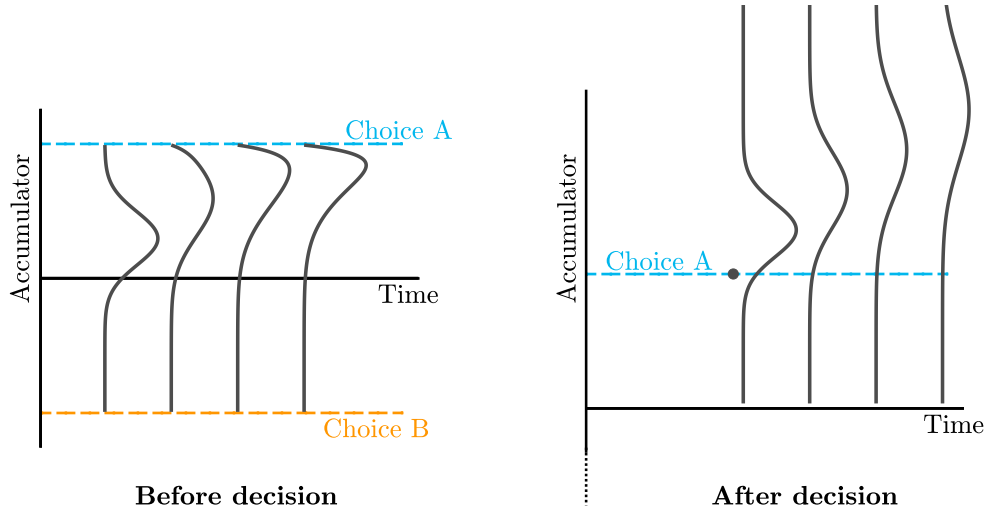

Figure 4.1: Probability distributions over accumulator state before and after a decision. Even if increments in the accumulator are normally distributed prior to a decision, the probability distribution over the state of the accumulator quickly becomes non-normal. This is because, if we reach time t without a decision, we know the accumulator has not been beyond either decision threshold prior to t. Following a decision, normally distributed increments in the accumulator lead to a normal distribution over accumulator state because there are no longer decision thresholds.

relationship following a decision. Following a decision, decision boundaries are no longer relevant, hence normally distributed changes in the state of the accumulator will lead to a normal distribution over this state (Fig. 4.1). If we can compute the mean and variance of this distribution we will have a very simple expression, that can be evaluated rapidly, for the probability distribution over accumulated evidence.

Using this insight, we derive approximate expressions for the probability, according to the DDM, of different confidence reports, given the response and response time on a trial. As discussed, the framework of the DDM with possibly time-dependent thresholds includes the optimal decision policy, whether or not signal strength is known by the observer. Confidence is allowed to be a noisy readout of the probability of being correct, and we account for the effects of pipeline evidence, and variability in signal-to-noise ratio. The derived expressions only need evaluating once per trial, instead of at each time step, and allow for dynamic stimuli of a certain form, thereby making trial-by-trial modelling of such stimuli feasible.

Symbol	Meaning
S	Stimulus (1 or 2)
R	Response (1 or 2)
C	Confidence
μ_i	Mean stimulus evidence for option i
μ	Difference between mean stimulus evidence for the two options
$\overline{\mu}$	Reference value (mid-point between mean stimulus evidence for the two options)
$\Delta\mu$	Magnitude of the difference between the two means
E_{ij}	Evidence for option i presented in stimulus frame j
δE_{ij}	Evidence presented for i in time step j. $E_{ik} = \sum_j \delta E_{ij}$ (sum over time steps in frame k)
E_i	Evidence for option 2 minus evidence for option 1 in frame i
E	Sum over all E_i
$\boldsymbol{E}$	Vector of all E_i
$\overline{E}$	Average difference in evidence between the two options, prior to decision
ΔE	Sum over $\delta E_{2i} - \delta E_{1i}$ for time steps following the decision
φ	Standardised drift rate (Quality of evidence processing)
δx_{ij}	Noisy measurement of option i in time step j
δx_i	Difference between the measurements for the two options ($\delta x_{2i} - \delta x_{1i}$)
$\delta \overline{x}_i$	Sum of the measurements for the two options ($\delta x_{2i} + \delta x_{1i}$)
$\boldsymbol{\delta x}$	Vector containing δx_i for every i
x	Accumulated difference in measurements, $\sum_i \delta x_i$
x_d	The value of x at t_d (i.e. height of relevant decision threshold at this time)
Δx	The change in x following a decision (at t_d)
x_{ll}	Log posterior ratio
x_c	Noisy measurement of x_{ll} which determines confidence
d_i	Boundaries on x_c which separate confidence reports falling into different bins

Table 4.1: Part 1 of 2. Symbols used in the derivations, along with key abbreviations.

Symbol	Meaning
σ_E	Standard deviation of evidence in a frame
σ_{acc}	Standard deviation of accumulator noise
σ_φ	Standard deviation of φ
σ_m	Standard deviation of metacognitive noise
t_f	Duration of a frame
t_e	Duration of evidence presentation
t_d	Duration from onset of accumulation to first crossing of a decision threshold
t_r	Response time
I	The interval $t_e - t_d$, the duration of the evidence pipeline
ν	$\varphi(\mu_2 - \mu_1)/t_f$
$\overline{\nu}$	$\varphi(\mu_2 + \mu_1)/t_f$
ν_0	$\Delta\mu/t_f$
σ_ν	$\nu_0\sigma_\varphi$
s^2	$\sigma_{acc}^2 + (\sigma_E^2/t_f)$
$\theta(t)$	$(s^2 + t\sigma_\nu^2)/2\nu_0$

Table 4.2: Part 2 of 2. Symbols used in the derivations, along with key abbreviations.

4.2 Model

4.2.1 Overview

Our aim is to derive expressions for the probability distribution over confidence, given the response and response time on a trial. Such expressions may provide direct insight into how confidence changes with other variables (Section 4.4), and make trial-by-trial modelling of confidence in DDM accounts computationally feasible (as we saw in Chapter 3). To derive these predictions, we must first specify a model for decisions and confidence.

We consider a situation identical to that used in previous chapters. Observers must make a choice between two alternatives when presented with a stimulus that provides two evidence signals, one for each option, with the evidence provided by the stimulus possibly varying over time (Fig. 4.2; Bogacz et al., 2006; Moreno-Bote, 2010). Considering an example from the studies in the previous chapters, the observer might be presented with two clouds of dots, with the number of dots

in each cloud constantly changing. The observer's task might be to determine which of the two clouds contains the most dots on average (Charles & Yeung, 2019). Here the dots in the two clouds would correspond to the two streams of evidence. We assume that the observer only receives noisy measurements of the presented evidence. In our model, consistent with the DDM, the observer takes the difference between the noisy measurements of evidence for the two options, and accumulates this difference (Fig. 4.2).

We again consider the "free response" and "interrogation" conditions (Section 2.1, McMillen and Holmes, 2006). As before, in the free response condition we assume the observer sets two thresholds on the accumulator state, one for each choice (Bogacz et al., 2006; Ratcliff & McKoon, 2008). When the accumulator reaches one of these thresholds, the corresponding response is triggered (Fig. 4.2). As discussed (Section 1.3.1), at the time of response, measurements of evidence that have recently been presented will still be in sensory and motor processing pipelines, and hence will not contribute to a decision (Resulaj et al., 2009). However, these measurements will be processed immediately following a response, and can be used to inform confidence (van den Berg, Anandalingam et al., 2016). In the interrogation condition, once the stimulus clears, the observer uses the final state of the accumulator to determine their response (Fig. 4.2). Unlike in other chapters, in this Chapter the observer always uses a Bayesian readout of confidence (Kiani et al., 2014; Moreno-Bote, 2010).

In the remainder of this section we set out a model for confidence. First, we describe the observer's task, and the beliefs held by the observer. Second, we find the rule a Bayesian observer would use to map evidence measurements to a decision and confidence. Third, we test the effect of the beliefs we ascribe to the observer. Fourth, we describe the "read out" process that determines confidence. In Section 4.3, we turn to the main aim of the chapter, which is to use the drift diffusion framework to derive simple expressions for the probability distribution

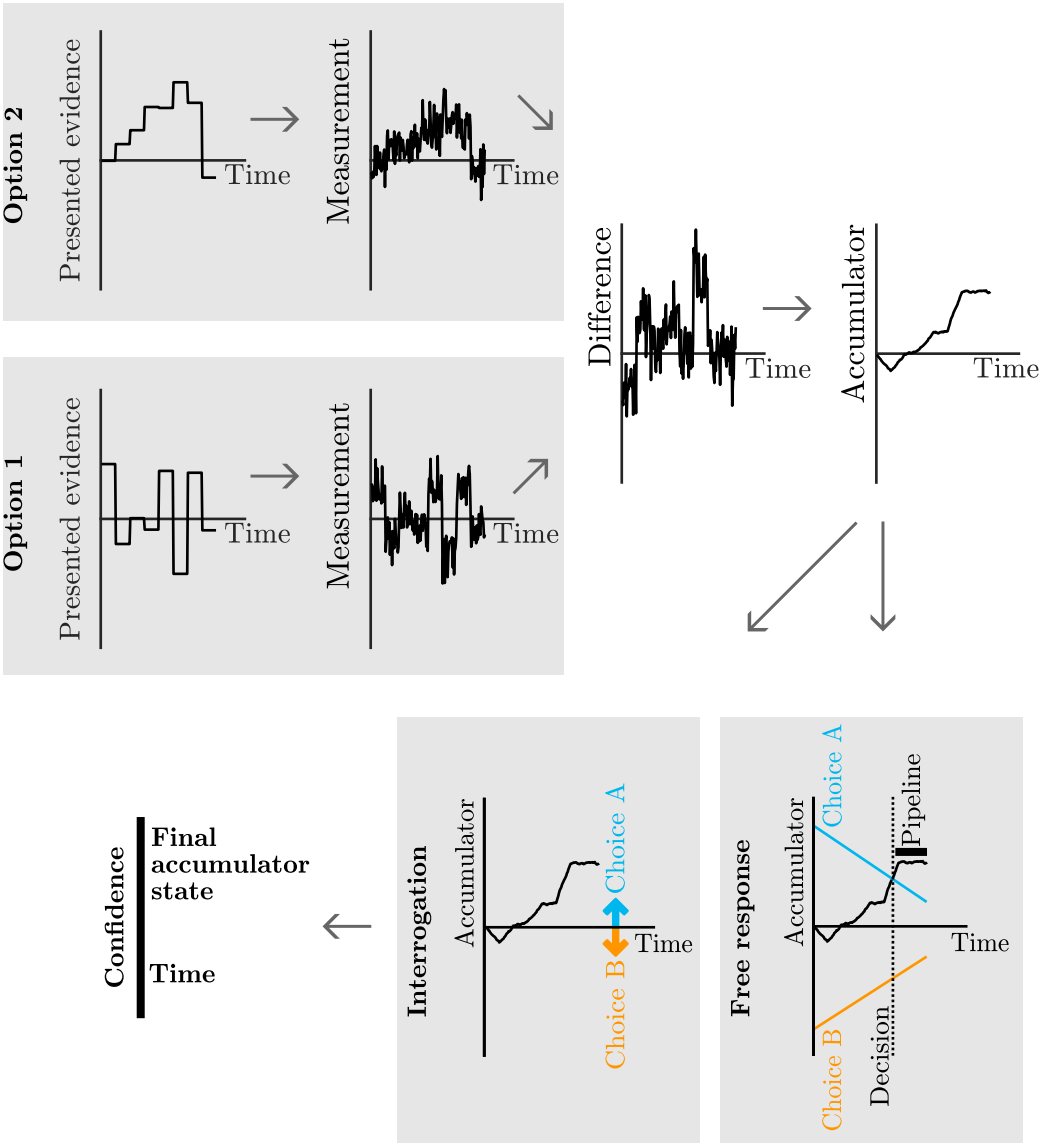

Figure 4.2: The model of confidence and decisions. Time varying evidence is presented for the two response options. The observer only receives noisy measurements of the evidence presented. The observer takes the difference between the evidence measurements and accumulates this difference. In the interrogation condition the duration of the stimulus is set by the researcher. The observer accumulates noisy measurements until the stimulus ends and all evidence is processed. Then the observer simply picks the option which is favoured by the accumulated measurements. In the free response condition, the observer uses decision thresholds, one for each option, which trigger a response. Following a decision, evidence measurements in the processing pipeline are accumulated, and confidence is informed by the accumulator state at the time of threshold crossing, plus changes to the accumulator caused by evidence measurements from the pipeline.

over confidence reports, given a response and response time. A summary of symbols used in the derivations can be found in Table 4.1 and Table 4.2.

4.2.2 Task and observer beliefs

Our first step is to describe the observer's task mathematically, and to describe beliefs about the task that we assume the observer holds. We consider a generic perceptual decision-making task and make the standard assumption that the observer receives two streams of noisy evidence measurements. The observer aims to determine the correct response by inferring which evidence stream is drawn from the distribution with the greater mean (Bogacz et al., 2006; Moreno-Bote, 2010). We consider the case where the two options are equally probable. Denote the mean evidence for the two options, μ_1 and μ_2, and the difference between these means, $\mu = \mu_2 - \mu_1$. $\Delta\mu$ is the magnitude of the difference between the two means. This value is fixed, but we incorporate variability in signal strength below (see discussion of standardised drift rate variability, φ). We also allow for the possibility that the sum of the means for the two streams, $\overline{\mu} = \mu_1 + \mu_2$, independently varies according to some distribution $p(\overline{\mu})$. These assumptions give us,

$$p(S = 1) = p(S = 2) = \frac{1}{2} \tag{4.1}$$

$$p(\mu|S = 1) = \delta(\mu + \Delta\mu) \tag{4.2}$$

$$p(\mu|S = 2) = \delta(\mu - \Delta\mu)\,. \tag{4.3}$$

S denotes the stimulus with the greater mean evidence, and hence the correct answer.

We make the standard DDM assumption that, in the free response condition, the observer sets decision thresholds to trigger a response (Ratcliff & McKoon, 2008). Denote the time spent accumulating measurements prior to the first crossing of a decision threshold, t_d, and the time of the response relative to the beginning of the trial t_r. These are not the same because there is an inevitable lag between the point of internal commitment to a decision and the point at which that decision is externally registered via an overt movement (Luce, 1986; Resulaj et al., 2009). Therefore, at

the time of response, recent evidence measurements are still in sensory and motor processing pipelines. Hence, information presented in the stimulus immediately prior to the response will not contribute to the decision, but can still be accumulated to influence any subsequent readout of the decision process, such as a confidence report. Denote the duration of the stimulus immediately prior to the response that does not contribute to the decision, because corresponding measurements are still being processed, I. In both the free response and interrogation conditions we denote the total time during which the stimulus is presented t_e. In the free response condition, $t_e = t_r$, as a response triggers the end of the stimulus. In the models, observers use all the information the received in t_e to inform their confidence.

We consider here the general case of time-varying evidence in the stimulus. Our derivations also apply to time-constant evidence as a special case of time-varying. We consider evidence which is piecewise-constant within short stimulus "frames" of duration t_f. Consider a discretisation of time into very short time steps (much shorter than the duration of a frame). We use E_{ij} to denote the evidence presented for option i during frame j. If the evidence presented in each frame is normally distributed around the underlying mean, then we have,

$$p(E_{ij}|\mu_i) = N(E_{ij}; \mu_i, \frac{\sigma_E^2}{2}) \,, \tag{4.4}$$

where $\sigma_E^2/2$ is the variance over presented evidence. Note that μ_1 and μ_2 do not correspond to anything the observer directly observes. μ_1 and μ_2 are the underlying mean evidence for the two options, and are constant throughout a trial. The actual presented evidence, E_{ij}, is what varies over the course of a trial, and is drawn from a distribution centred on the underlying means.

Although in principle observers could leverage knowledge that evidence is constant throughout each frame, we assume that observers ignore this additional structure. This will be a valid approximation when evidence frames used in an experiment are sufficiently short. In this case, the approximation will lead to little discrepancy between the observer's estimate of the probability of being correct, and the true probability. We test this claim below, once we have derived an

expression for the probability of being correct (Fig. 4.3, and equation 4.27). Instead of accounting for the constant evidence over the course of a frame, we assume that observers treat the fraction of evidence received in a small time step, j, of duration δt, as generated according to,

$$p(\delta E_{ij}|\mu_i) = N(\delta E_{ij}; \mu_i \frac{\delta t}{t_f}, \frac{\sigma_E^2}{2} \frac{\delta t}{t_f}) \,, \tag{4.5}$$

where, δE_{ij} indicates the evidence presented for option i in time step j only, not over the course of a frame ($E_{ik} = \sum_j \delta E_{ij}$ where the summation is taken over all time steps in frame k). Importantly, the observer incorrectly assumes that the δE_{ij} at different time steps, j, are independent (i.e., the observer ignores the fact that evidence is constant through all time steps in a single frame).

As in the DDM, the presented evidence drives an internal evidence accumulation which is subject to normally distributed noise (Ratcliff & McKoon, 2008). We also take into account the fact that the stimulus varies over time. Over small time step, j, the incremental change in the state of the accumulator for option i, denoted δx_{ij}, can be described by,

$$p(\delta x_{ij}|\delta E_{ij}, \varphi) \;\; = \;\; N(\delta x_{ij}; \delta E_{ij}\varphi, \frac{\sigma_{acc}^2}{2}\delta t) \,. \tag{4.6}$$

where σ_{acc} is the variance of noise in the accumulation (Drugowitsch et al., 2012). φ is a random variable which accounts for variability in drift rate. "Drift rate" refers to the rate at which evidence presented in the stimulus drives the accumulation of evidence measurements (Ratcliff & McKoon, 2008). Where a stimulus, and hence drift rate, is constant over the course of a trial, drift rate variability is trial-to-trial variability in this rate (Ratcliff & McKoon, 2008; Ratcliff & Smith, 2004; Voskuilen et al., 2016). Here, where stimulus evidence, and hence drift rate, varies over the course of each trial, we operationalise drift rate variability as a multiplicative factor that determines how well stimulus information is processed. To distinguish this operationalisation from the usual operationalisation, we refer to the multiplicative factor as standardised drift rate, because the mean value of this variable will be 1, regardless of the strength of the presented evidence. When standardised drift rate

is high, there is a strong signal from the stimulus, and evidence from the stimulus is accumulated rapidly. Noise in the accumulation is unaffected, hence, a higher standardised drift rate also leads to a higher signal-to-noise ratio.

It is usually assumed that drift rate variability follows a normal distribution (Ratcliff & McKoon, 2008). We make the same assumption here,

$$p(\varphi) = N(\varphi; 1, \sigma_\varphi^2)\,. \tag{4.7}$$

We assume that observers correctly model the effect of internal noise (accumulator noise, and standardised drift rate variability). Hence, observers infer that increments in the accumulators are related to the underlying signal as follows,

$$\begin{aligned} p(\delta x_{ij}|\mu_i, \varphi) &= \int d(\delta E_{ij})\ p(\delta x_{ij}|\delta E_{ij}, \varphi) p(\delta E_{ij}|\mu_i) && (4.8)\\ &= \int d(\delta E_{ij})\ N(\delta x_{ij}; \delta E_{ij}\varphi, \frac{\sigma_{acc}^2}{2}\delta t) N(\delta E_{ij}; \mu_i \frac{\delta t}{t_f}, \frac{\sigma_E^2}{2}\frac{\delta t}{t_f}) && (4.9)\\ &= N(\delta x_{ij}; \mu_i \varphi \frac{\delta t}{t_f}, \frac{\delta t}{2}(\sigma_{acc}^2 + \varphi^2 \frac{\sigma_E^2}{t_f}))\,, && (4.10) \end{aligned}$$

To reach the final line we rearranged, and applied the result in Appendix B.1. In words, this equation states that evidence measurements have a mean which matches the underlying stimulus signal multiplied by standardised drift rate, but are variable due to both internal accumulator noise, and the fact that the evidence presented in the stimulus is itself variable. It is interesting to note that the standardised drift rate, φ, not only affects the mean increment, but also the variability of increments, through the term $\varphi^2\sigma_E^2/t_f$. This effect arises because high standardised drift rate increases the effect of the stimulus on the accumulation, increasing the effect of both the stimulus mean, and variability in the stimulus. We assume that participants ignore this difference between stimulus variability and internal variability (which is not affected by standardised drift rate). Hence we assume that observers believe evidence measurements are corrupted by the same amount of variability regardless of the level of standardised drift rate:

$$p(\delta x_{ij}|\mu_i, \varphi) = N(\delta x_{ij}; \mu_i \varphi \frac{\delta t}{t_f}, \frac{\delta t}{2}(\sigma_{acc}^2 + \frac{\sigma_E^2}{t_f}))\,. \tag{4.11}$$

We test the effect of this assumption on the calibration of confidence below (Fig. 4.3).

It is simpler to perform calculations with transformed variables $\delta x_i = \delta x_{2i} - \delta x_{1i}$ and $\delta \overline{x}_i = \delta x_{2i} + \delta x_{1i}$, than with δx_{2i} and δx_{1i}. This is because the observer's task involves detecting the difference between the means of the stimuli, μ_1 and μ_2, so the difference between accumulator samples is likely to be the most important variable. In Section 4.2.1 we stated that we will use the standard idea from the DDM that the observer tracks the difference in evidence between alternatives (Bogacz et al., 2006; Ratcliff & McKoon, 2008). However, we continue to consider the sum of evidence measurements here, until we can show that they provide no relevant information and that the DDM observer loses nothing by discarding them. As δx_{1i} and δx_{2i} are normally distributed, we can compute their sum using standard formulas. Using (4.11), this gives,

$$p(\delta x_i|\mu_1, \mu_2, \varphi) = N(\delta x_i; \varphi(\mu_2 - \mu_1)\frac{\delta t}{t_f}, \delta t(\sigma_{acc}^2 + \frac{\sigma_E^2}{t_f})) \tag{4.12}$$

$$p(\delta \overline{x}_i|\mu_1, \mu_2, \varphi) = N(\delta \overline{x}_i; \varphi(\mu_2 + \mu_1)\frac{\delta t}{t_f}, \delta t(\sigma_{acc}^2 + \frac{\sigma_E^2}{t_f})) \tag{4.13}$$

To simplify these expressions we change variables to

$$\nu = \frac{1}{t_f}\varphi(\mu_2 - \mu_1) = \frac{1}{t_f}\varphi\mu \tag{4.14}$$

$$\overline{\nu} = \frac{1}{t_f}\varphi(\mu_2 + \mu_1) = \frac{1}{t_f}\varphi\overline{\mu} \tag{4.15}$$

$$s^2 = \sigma_{acc}^2 + \frac{\sigma_E^2}{t_f}\,. \tag{4.16}$$

This gives,

$$p(\delta x_i|\nu) = N(\delta x_i; \nu\delta t, s^2\delta t) \tag{4.17}$$

$$p(\delta \overline{x}_i|\overline{\nu}) = N(\delta \overline{x}_i; \overline{\nu}\delta t, s^2\delta t) \tag{4.18}$$

We can compute the probability distributions over ν and $\overline{\nu}$ given S using (4.2), (4.3)

and standard formulas for distributions over products of random variables, giving,

$$p(\nu|S=1) = N(\nu; -\frac{\Delta\mu}{t_f}, \frac{\Delta\mu^2}{t_f^2}\sigma_\varphi^2) = N(\nu; -\nu_0, \sigma_\nu^2) \quad (4.19)$$

$$p(\nu|S=2) = N(\nu; \frac{\Delta\mu}{t_f}, \frac{\Delta\mu^2}{t_f^2}\sigma_\varphi^2) = N(\nu; \nu_0, \sigma_\nu^2) \quad (4.20)$$

$$p(t_f\overline{\nu}|S) = \int d\overline{\mu}\; p(\overline{\mu})p(\varphi = \frac{t_f\overline{\nu}}{\overline{\mu}})\frac{1}{|\overline{\mu}|} \quad (4.21)$$

where $\nu_0 = \frac{\Delta\mu}{t_f}$, and $\sigma_\nu = \frac{\Delta\mu}{t_f}\sigma_\varphi$. Note that $\overline{\nu}$ is independent of S.

In summary, observers are aiming to infer S, which affects ν via,

$$p(\nu|S=1) = N(\nu; -\nu_0, \sigma_\nu^2) \quad (4.22)$$

$$p(\nu|S=2) = N(\nu; \nu_0, \sigma_\nu^2)\,. \quad (4.23)$$

They receive two independent streams of evidence which they assume are related to ν via,

$$p(\delta x_i|\nu) = N(\delta x_i; \nu\delta t, s^2\delta t) \quad (4.24)$$

$$p(\delta\overline{x}_i|\overline{\nu}) = N(\delta\overline{x}_i; \overline{\nu}\delta t, s^2\delta t)\,. \quad (4.25)$$

As $\delta\overline{x}$ only depends on $\overline{\nu}$, which is independent of S, $\delta\overline{x}$ provides no information regarding S, and can be ignored by the observer without any loss of relevant information. Hence, as described in the model overview, the observer only tracks the difference between evidence measurements, not their sum.

4.2.3 The Bayesian observer

Having specified the properties of the task and the beliefs of the observer, we can now infer the rule used by a Bayesian observer to translate all evidence measurements gathered into a decision and confidence. A Bayesian observer uses all measurements that have been received and processed to determine their response and confidence. (Evidence measurements in sensory processing pipelines will not contribute.) We denote the vector of all evidence measurements, $\delta x_1, \delta x_2, ..., \delta x_N$ as $\boldsymbol{\delta x}$. Inferring the posterior probability over S, given $\boldsymbol{\delta x}$, in the environment described by (4.22), (4.23),

and (4.24), is an inference problem which has been considered before. Moran (2015), using a result from Drugowitsch et al. (2012), derived the posterior probability over the two options $S = 1$ and $S = 2$. We also provide a derivation in Appendix B.2, and simply state the result here. Using Bayes rule with all evidence samples, the log posterior ratio is given by (Appendix B.2; Moran, 2015),

$$x_{ll} = \log \frac{p(S = 2|\boldsymbol{\delta x})}{p(S = 1|\boldsymbol{\delta x})} = \frac{2x\nu_0}{s^2 + t\sigma_\nu^2} = \frac{x}{\theta(t)}\,, \tag{4.26}$$

where $x = \sum_{i=1}^{N} \delta x_i$, while x_{ll} and $\theta()$ provide abbreviations for the purpose of making future equations less cluttered, and t is the time spent accumulating evidence.

A Bayesian observer would report whichever option is more likely. Hence, they will report $S = 1$ when $x_{ll} < 0$, which is the same as $x < 0$. Denote this report $R = 1$, and a report for $S = 2$ as $R = 2$. In the interrogation condition, the observer simply has to wait for the stimulus to end at t_e. Once the observer has accumulated all evidence measurements, taking t_e, the observer can respond according to the sign of the accumulator state, x (Fig. 4.2). In the free response condition, the observer uses a decision threshold for triggering a response (Bogacz et al., 2006; Ratcliff & McKoon, 2008). This threshold describes an accumulator magnitude, $|x|$, which when reached, triggers a response. We allow the threshold to vary with time (Drugowitsch et al., 2012). As discussed, at the time of response, measurements corresponding to recently-presented evidence will still be in the processing pipeline (Resulaj et al., 2009). The response will be based on x at the time of the decision, t_d, while confidence will incorporate additional pipeline evidence measurements (Fig. 4.2). Processing will continue until measurements from the full duration of stimulus presentation, t_e, have been processed.

The observer can use the log posterior ratio in the computation of confidence, as it is monotonically related to the probability they are correct through,

$$p(\text{correct}|\boldsymbol{\delta x}) = p(S = 1|\boldsymbol{\delta x}) = \frac{1}{1 + e^{x_{ll}}} = \frac{1}{1 + e^{\frac{x}{\theta(t_e)}}}\,, \tag{4.27}$$

if the observer reports $R = 1$. A similar expression holds for $R = 2$, the only difference being that x is replaced with $-x$, and x_{ll} with $-x_{ll}$.

4.2.4 Testing the observer's beliefs

Above, we assumed that observers approximate the true generative model for the evidence measurements. Specifically we assumed (a) that observers ignore the fact that evidence is constant within each frame, and (b) that observers ignore the increased effect of variability in the stimulus, when standardised drift rate is high, and the decreased effect when standardised drift rate is low. While the assumptions may be plausible, it is also plausible that, if given feedback, observers could learn to closely approximate the mapping from x and t_e to probability correct (Kiani & Shadlen, 2009). To adequately capture this situation we need to ensure that, under our assumptions about the observer, confidence remains closely related to performance.

To test the two approximations, we simulated the diffusion process using small time steps. At each time step the accumulator increment, δx_i, was determined by drawing a value consistent with (4.6), until a decision threshold was crossed (Tuerlinckx et al., 2001). Accumulation continued until all evidence measurements had been processed, at which point a simulated confidence was determined in accordance with (4.27). Full details of the simulation can be found in Appendix B.3.

Fig. 4.3 shows accuracy against confidence for two cases, one in which there is no variability in standardised drift rate, φ, and one with variability. Full details of the process used to plot the data is also provided in Appendix B.3. Only one approximation is relevant when there is no variability in standardised drift rate, approximation (a). This approximation appears to have little effect on the mapping between accuracy and confidence. An additional approximation is relevant when there is variability in standardised drift rate, assumption (b). We can see that this causes a moderate mismatch between accuracy and confidence (Fig. 4.3).

It is interesting to note that the latter approximation, that observers ignore the effect of drift rate on stimulus variability, is implicitly present in much previous work. This is because dynamic stimuli are often used, but stimulus variability is not treated

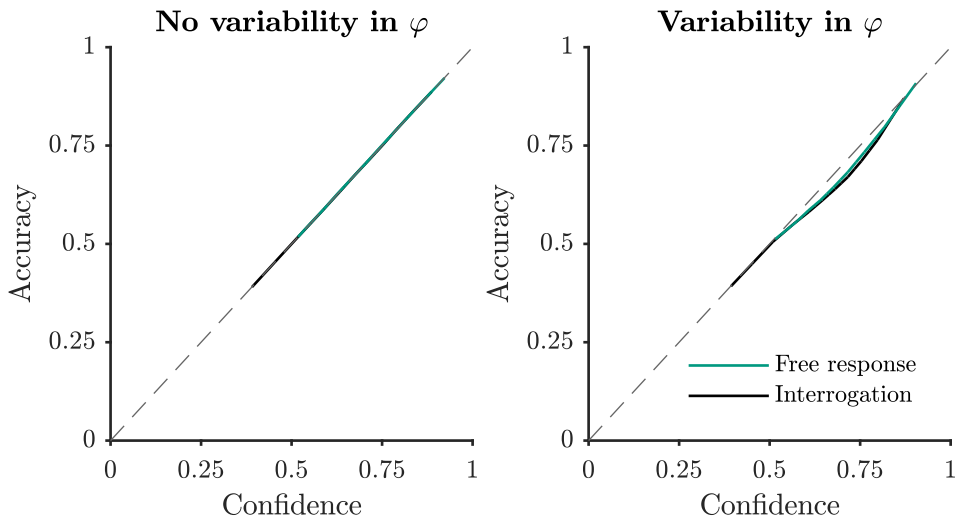

Figure 4.3: Accuracy as a function of confidence, where confidence is calculated using the observer's generative model of their evidence measurements. This model only approximates the true generative model. When there is variability in standardised drift rate, the approximations cause the observer's confidence to show some deviations from calibration.

separately from accumulator variability in the computational models applied (e.g. Kiani et al., 2014; Ratcliff and McKoon, 2008; van den Berg, Anandalingam et al., 2016, although see Zylberberg et al., 2016). Fig. 4.3 suggests this approximation does not cause large issues.

4.2.5 Confidence reporting model

Having described the task, observer beliefs, tested those beliefs, and found the confidence of a Bayesian observer, we turn to the question of what exact quantity confidence reports reflect. While the log-posterior ratio, x_{ll} is monotonically related to the probability of being correct, we do not assume that confidence is a direct readout. Instead, we allow the possibility that metacognitive noise corrupts this estimate (De Martino et al., 2013; Maniscalco & Lau, 2012, 2016), and hence that confidence is based on a noisy representation of x_{ll}, denoted x_c,

$$p(x_c|x_{ll}) = N(x_c; x_{ll}, \sigma_m^2)\,, \tag{4.28}$$

where σ_m is the standard deviation of metacognitive noise.

We would also like to minimise the number of assumptions we make about how x_c is transformed into a confidence report. There is evidence that different people use confidence scales in different ways (Ais et al., 2016; Navajas et al., 2017). To make minimal assumptions about how people view and treat confidence scales, here and throughout the thesis, we treat confidence reports, C, as ordinal data only (Aitchison et al., 2015). Note that we do not believe only the ordering of confidence reports is useful or interesting, just that this ordering is most likely to reflect the perceptual aspect of confidence that we wish to investigate (for further discussion of this choice see Section 1.3.2). Confidence reports on a continuous scale can be analysed by binning them first.

If people report greater confidence when x_c favours their decision to a greater extent, then all confidence reports falling into a higher confidence bin will have come from further along the x_c scale (in the direction that favours the choice made). Using d_i we denote the boundary on x_c that separates the confidence reports that fall into confidence category $C = i - 1$ from $C = i$, when the observer reports $R = 2$. When $R = 1$, the boundary applies to $-x_c$, or equivalently, a boundary of $-d_i$ is applied to x_c.

4.3 Results

We now have a complete description of the model, and everything we need to derive the probability distribution over confidence reports in both the interrogation and free response conditions. We would like to find the probability distribution over confidence, given the evidence presented, $\boldsymbol{E}$, the response, R, and, in the free response condition, the amount of time the observer monitors the stimulus before making a response, t_r. ($\boldsymbol{E}$ is a vector containing every $E_i = E_{2i} - E_{1i}$.) A key variable is the observer's log posterior ratio after they have seen all evidence, x_{ll}. Our general strategy will be to find a probability distribution over this variable. From this distribution we will be able to infer a distribution over the noisy readout of the log-posterior ratio, x_c. As described in the previous section, on a trial with

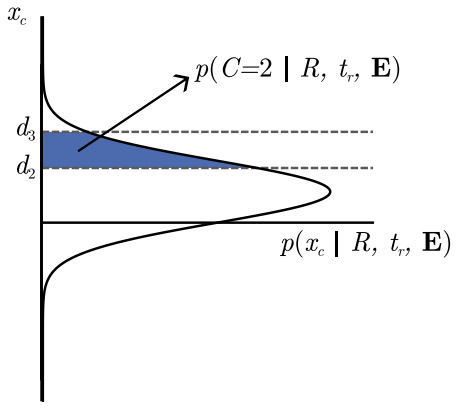

Figure 4.4: Probability of a confidence report, based on a distribution over x_c. x_c is a noisy representation of the log-posterior ratio. We make no specific assumptions about how observers use confidence scales apart from assuming that, if the observer reports higher confidence C, then the underlying variable x_c has a greater absolute value, in the direction corresponding to the response made. We use d_i to denote the boundaries between values of x_c that lead to different confidence reports. If we know the distribution over x_c, then the probability of a specific confidence report can be found by integrating x_c or $-x_c$ between the corresponding boundaries. Whether we integrate over x_c or $-x_c$ depends on the response made.

a response $R = 2$, if x_c falls between d_i and d_{i+1} the observer reports confidence $C = i$. If $R = 1$, the boundaries are $-d_{i+1}$ and $-d_i$. The probability of a confidence report $C = i$ will be given by the probability that x_c falls between the corresponding boundaries (Fig. 4.4).

Interrogation condition

In a trial from the interrogation condition, the stimulus is presented for some amount of time, t_e. The observer can only respond after the end of the stimulus. We aim to find the probability of confidence reports, given the response and evidence presented. Assuming that the response occurs at some fixed amount of time following t_e, the response time provides us with no information. This is because t_e is set by the researcher and, hence, unaffected by processes internal to the observer. A summary of the generative model for interrogation condition confidence reports, from the perspective of the researcher, is shown in Fig. 4.5.

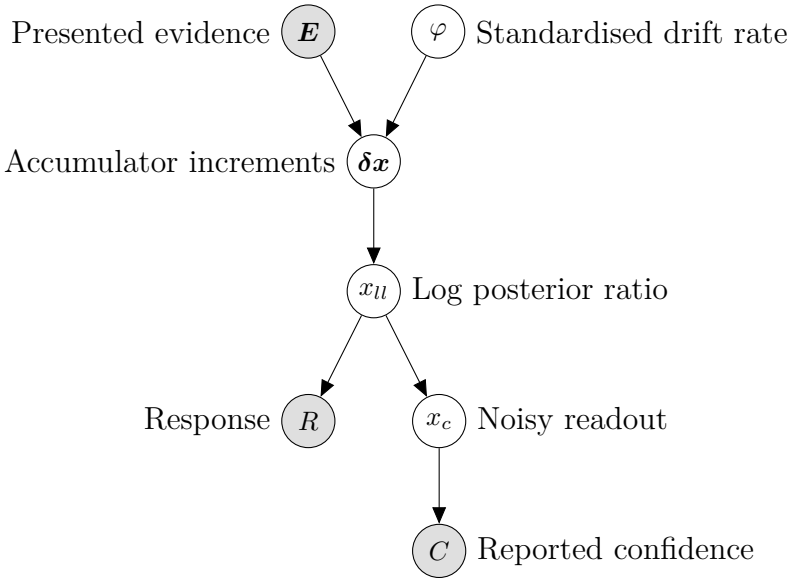

Figure 4.5: Representation of the generative model for confidence reports in the interrogation condition, from the perspective of the researcher. We want to infer the probability of reported confidence, using our knowledge of the evidence presented and the response given.

We start by integrating x_c over the region which leads to a confidence report $C = i$. When the response is $R = 2$, this is the region between d_i and d_{i+1} (Fig. 4.4). The case where $R = 1$ is identical except the limits become $-d_{i+1}$ and $-d_i$. Also marginalising over, x_{ll} and using Bayes rule,

$$p(C = i|R, \boldsymbol{E}) = \int_{d_i}^{d_{i+1}} dx_c \int dx_{ll}\; p(x_c|x_{ll})p(x_{ll}|R, \boldsymbol{E}) \tag{4.29}$$

$$= \int_{d_i}^{d_{i+1}} dx_c \int dx_{ll}\; \frac{p(x_c|x_{ll})p(R|x_{ll})p(x_{ll}|\boldsymbol{E})}{p(R|\boldsymbol{E})}\,. \tag{4.30}$$

We want to find,

$$I(x_c) = \int dx_{ll}\; p(x_c|x_{ll})p(R|x_{ll})p(x_{ll}|\boldsymbol{E})\,, \tag{4.31}$$

and then to use,

$$p(C = i|R, \boldsymbol{E}) = \int_{d_i}^{d_{i+1}} dx_c\; \frac{I(x_c)}{p(R|\boldsymbol{E})}\,. \tag{4.32}$$

For now just consider $p(x_{ll}|\boldsymbol{E})$. Let's marginalise over φ,

$$p(x_{ll}|\boldsymbol{E}) = \int d\varphi\; p(x_{ll}|\varphi, \boldsymbol{E})p(\varphi|\boldsymbol{E})\,. \tag{4.33}$$

$\boldsymbol{E}$ and φ are independent of each other (when not conditioned on other variables; see Fig. 4.5). Hence,

$$p(\varphi|\boldsymbol{E}) = p(\varphi) = N(\varphi; 1, \sigma_\varphi^2)\,, \tag{4.34}$$

where we have used (4.7).

Recall from (4.6) that,

$$p(\delta x_{ij}|\delta E_{ij}, \varphi) \;=\; N(\delta x_{ij}; \delta E_{ij}\varphi, \frac{\sigma_{acc}^2}{2}\delta t)\,. \tag{4.35}$$

Using standard formulae for the sum of normally distributed random variables, we have that $x = \sum_i \delta x_{2i} - \delta x_{1i}$ will be given by,

$$p(x|\varphi, \boldsymbol{E}) \;=\; N(x; E\varphi, t_e\sigma_{acc}^2)\,. \tag{4.36}$$

where $E = \sum_i \delta E_{2i} - \delta E_{1i}$, and $t_e = \sum_i \delta t$. Using (4.26) we have,

$$p(x_{ll}|\varphi, \boldsymbol{E}) = N(x_{ll}; \frac{E\varphi}{\theta(t_e)}, \frac{t_e\sigma_{acc}^2}{\theta^2(t_e)})\,. \tag{4.37}$$

Substituting these results into (4.33) gives us,

$$p(x_{ll}|\boldsymbol{E}) \;=\; \int d\varphi\; N(x_{ll}; \frac{E\varphi}{\theta(t_e)}, \frac{t_e\sigma_{acc}^2}{\theta^2(t_e)})N(\varphi; 1, \sigma_\varphi^2) \tag{4.38}$$

$$\;=\; \frac{\theta(t_e)}{E}\int d\varphi\; N(\varphi; \frac{x_{ll}\theta(t_e)}{E}, \frac{t_e\sigma_{acc}^2}{E^2})N(\varphi; 1, \sigma_\varphi^2)\,. \tag{4.39}$$

We can apply the result in Appendix B.1. Simplifying, we find,

$$p(x_{ll}|\boldsymbol{E}) \;=\; N(x_{ll}; \frac{E}{\theta(t_e)}, \frac{E^2\sigma_\varphi^2 + t_e\sigma_{acc}^2}{\theta^2(t_e)})\,. \tag{4.40}$$

Returning to (4.31), we now have,

$$I(x_c) = \int dx_{ll}\; p(x_c|x_{ll})p(R|x_{ll})N(x_{ll}; \frac{E}{\theta(t_e)}, \frac{E^2\sigma_\varphi^2 + t_e\sigma_{acc}^2}{\theta^2(t_e)})\,. \tag{4.41}$$

In the case in which we have no metacognitive noise, $x_c = x_{ll}$. We can express this using the Dirac delta function as $p(x_c|x_{ll}) = \delta(x_c - x_{ll})$. Then we have,

$$I(x_c) \;=\; \int dx_{ll}\; \delta(x_c - x_{ll})p(R|x_{ll})N(x_{ll}; \frac{E}{\theta(t_e)}, \frac{E^2\sigma_\varphi^2 + t_e\sigma_{acc}^2}{\theta^2(t_e)}) \tag{4.42}$$

$$\;=\; p(R|x_{ll} = x_c)N(x_c; \frac{E}{\theta(t_e)}, \frac{E^2\sigma_\varphi^2 + t_e\sigma_{acc}^2}{\theta^2(t_e)})\,, \tag{4.43}$$

Consider the term $p(R|x_{ll} = x_c)$, which describes the probability of a response, given a final log-posterior x_{ll}. The Bayesian observer's decision rule is deterministic, and was described in Section 4.2. If $x_{ll} < 0$ the observer reports that $R = 1$, and reports $R = 2$ if $x_{ll} > 0$. In all cases, the observer makes the response that is most likely to be correct. Hence, the log-posterior ratio at the time of the decision, x_{ll}, always favours the response made, and the probability of x_{ll} and R being inconsistent is zero. Due to the absence of metacognitive noise, $x_{ll} = x_c$. Hence, $p(R|x_{ll} = x_c)$ will also be one when x_c and R are consistent, and zero when x_c is inconsistent with R, i.e. when x_c suggests a different response should have been made. Because x_c and R are always consistent, the observer will never report a confidence of less than 50%.

In the no metacognitive noise case we have, using (4.32) and our expression for $I(x_c)$,

$$
\begin{aligned}
&p(C = i|R = 2, \boldsymbol{E}) \\
&\quad = \frac{1}{p(R = 2|\mathbf{E})} \int_{d_i}^{d_{i+1}} dx_c \; p(R = 2|x_{ll} = x_c) N(x_c; \frac{E}{\theta(t_e)}, \frac{E^2\sigma_\varphi^2 + t_e\sigma_{acc}^2}{\theta^2(t_e)}) \quad (4.44) \\
&\quad = \frac{1}{p(R = 2|\mathbf{E})} \int_{\max\{0,d_i\}}^{\max\{0,d_{i+1}\}} dx_c \; N(x_c; \frac{E}{\theta(t_e)}, \frac{E^2\sigma_\varphi^2 + t_e\sigma_{acc}^2}{\theta^2(t_e)}) \,, \quad (4.45)
\end{aligned}
$$

where we have used that $p(R = 2|x_{ll} = x_c)$ is zero anywhere x_{ll} is inconsistent with the response $R = 2$ (i.e. $x_{ll} < 0$), and one elsewhere. In the case where $R = 1$, an identical expression holds, except the limits become $\min\{0, -d_{i+1}\}$ and $\min\{0, -d_i\}$.

We can compute the probability of the response, $p(R|\mathbf{E})$, using,

$$
\begin{aligned}
p(R|\boldsymbol{E}) &= \int_{-\infty}^{\infty} dx_{ll} \; p(R|x_{ll}) p(x_{ll}|\boldsymbol{E}) \quad (4.46) \\
&= \int_a^b dx_{ll} \; N(x_{ll}; \frac{E}{\theta(t_e)}, \frac{E^2\sigma_\varphi^2 + t_e\sigma_{acc}^2}{\theta^2(t_e)}) \quad (4.47)
\end{aligned}
$$

Where $a = -\infty, b = 0$ for $R = 1$ or $a = 0, b = \infty$ for $R = 2$. These limits again come from using that $p(R|x_{ll}) = 1$ when R and x_{ll} are consistent, and $p(R|x_{ll}) = 0$ otherwise.

Let's now consider what happens in the presence of metacognitive noise, when x_c no longer equals x_{ll}, but is a noisy version of it as in (4.28). Returning to

(4.41), in this case $I(x_c)$ becomes,

$$I(x_c) = \int dx_{ll}\ N(x_c; x_{ll}, \sigma_m^2) p(R|x_{ll}) N(x_{ll}; \frac{E}{\theta(t_e)}, \frac{E^2\sigma_\varphi^2 + t_e\sigma_{acc}^2}{\theta^2(t_e)}) \,. \tag{4.48}$$

Because $p(R|x_{ll})$ is 1 when R is consistent with the log-posterior ratio, and 0 otherwise, the product of this term and the normal distribution over x_{ll} is a normal distribution truncated to the region where R and x_{ll} are consistent. Note that the product of these two terms is not a probability distribution over x_{ll} because it is not normalised. However, we can write this product as a scaled truncated normal distribution,

$$\begin{aligned} J(x_{ll}) &= p(R|x_{ll}) N(x_{ll}; \frac{E}{\theta(t_e)}, \frac{E^2\sigma_\varphi^2 + t_e\sigma_{acc}^2}{\theta^2(t_e)}) & (4.49)\\ &= TN(x_{ll}; \mu_{ll}, \sigma_{ll}^2, a, b) \int_a^b dx'_{ll}\ N(x'_{ll}; \frac{E}{\theta(t_e)}, \frac{E^2\sigma_\varphi^2 + t_e\sigma_{acc}^2}{\theta^2(t_e)}) & (4.50)\\ &= TN(x_{ll}; \mu_{ll}, \sigma_{ll}^2, a, b) p(R|\boldsymbol{E}) \,. & (4.51) \end{aligned}$$

We have used (4.47) to get the final line. As before $a = -\infty, b = 0$ or $a = 0, b = \infty$ depending on the response made, and TN indicates a truncated normal distribution, truncated at a and b. The first and second parameters, μ_{ll} and σ_{ll}^2, are the mean and variance of the distribution prior to truncation, and therefore are,

$$\begin{aligned} \mu_{ll} &= \frac{E}{\theta(t_e)} & (4.52)\\ \sigma_{ll}^2 &= \frac{E^2\sigma_\varphi^2 + t_e\sigma_{acc}^2}{\theta^2(t_e)} & (4.53) \end{aligned}$$

We have then,

$$I(x_c) = p(R|\boldsymbol{E}) \int_{-\infty}^{\infty} dx_{ll}\ N(x_c; x_{ll}, \sigma_m^2) TN(x_{ll}; \mu_{ll}, \sigma_{ll}^2, a, b) \,. \tag{4.54}$$

Consider a change of variables $y = x_c - x_{ll}$,

$$\begin{aligned} I(x_c) &= -p(R|\boldsymbol{E}) \int_{\infty}^{-\infty} dy\ N(y; 0, \sigma_m^2) TN(x_c - y; \mu_{ll}, \sigma_{ll}^2, a, b) & (4.55)\\ &= p(R|\boldsymbol{E}) \int_{-\infty}^{\infty} dy\ N(y; 0, \sigma_m^2) TN(x_c - y; \mu_{ll}, \sigma_{ll}^2, a, b) \,. & (4.56) \end{aligned}$$

This expression is the convolution of a normal distribution, and a truncated normal distribution. We can perform the convolution by writing the distributions out in

full, collecting all terms containing y into a single exponential, and integrating (S. Turban, personal communication, December, 2019):

$$I(x_c) = p(R|\boldsymbol{E}) \int_{-\infty}^{\infty} dy\ N(y; 0, \sigma_m^2) TN(x_c - y; \mu_{ll}, \sigma_{ll}^2, a, b) \tag{4.57}$$

$$= p(R|\boldsymbol{E}) \int_{-\infty}^{\infty} dy \frac{1}{2\pi\sqrt{\sigma_m^2 \sigma_{ll}^2}} \frac{1}{\Phi(\frac{b-\mu_{ll}}{\sigma_{ll}}) - \Phi(\frac{a-\mu_{ll}}{\sigma_{ll}})} e^{-\frac{y^2}{2\sigma_m^2}} e^{-\frac{(x_c-y-\mu_{ll})^2}{2\sigma_{ll}^2}} \mathbb{1}_{y\in[x_c-b, x_c-a]} \tag{4.58}$$

$$= p(R|\boldsymbol{E}) \frac{1}{\Phi(\frac{b-\mu_{ll}}{\sigma_{ll}}) - \Phi(\frac{a-\mu_{ll}}{\sigma_{ll}})} N(x_c; \mu_{ll}, \sigma_m^2 + \sigma_{ll}^2) \left[\Phi(\frac{x_c - a - \alpha}{\beta}) - \Phi(\frac{x_c - b - \alpha}{\beta})\right] \tag{4.59}$$

where $\Phi()$ indicates the standard cumulative normal distribution, $\mathbb{1}$ is an indicator function (which is one when its argument is true, and zero otherwise), $\alpha = \frac{\beta^2(x_c-\mu_{ll})}{\sigma_{ll}^2}$ and $\beta^2 = \frac{\sigma_m^2\sigma_{ll}^2}{\sigma_m^2+\sigma_{ll}^2}$.

In our case, a and b take very specific values. Let,

$$L = \begin{cases} -1 & \text{if } a = -\infty, b = 0; \text{ i.e. R=1} \\ 1 & \text{if } a = 0, b = \infty; \text{ i.e. R=2}\,, \end{cases}$$

then we can write,

$$I(x_c) = p(R|\boldsymbol{E}) \frac{1}{\Phi(L\frac{\mu_{ll}}{\sigma_{ll}})} N(x_c; \mu_{ll}, \sigma_m^2 + \sigma_{ll}^2) \Phi\left(L \frac{x_c\sigma_{ll}^2 + \mu_{ll}\sigma_m^2}{\sigma_m\sigma_{ll}\sqrt{\sigma_m^2 + \sigma_{ll}^2}}\right), \tag{4.60}$$

and using (4.32) our final result for confidence is,

$$p(C = i|R, t_e, \boldsymbol{E}) = L \int_{Ld_i}^{Ld_{i+1}} dx_c\ \frac{1}{\Phi(L\frac{\mu_{ll}}{\sigma_{ll}})} N(x_c; \mu_{ll}, \sigma_m^2 + \sigma_{ll}^2) \Phi\left(L \frac{x_c\sigma_{ll}^2 + \mu_{ll}\sigma_m^2}{\sigma_m\sigma_{ll}\sqrt{\sigma_m^2 + \sigma_{ll}^2}}\right). \tag{4.61}$$

We have also used L in the above expression to account for the different limits of integration when $R = 1$ and $R = 2$. We can solve the integral in (4.61) via numerical integration. It is also possible to rearrange the expression to a form which is faster to evaluate numerically. We rearrange the expression in Appendix

B.4, and state the result here. The expression in (4.61) becomes,

$$p(C=i|R,t_e,\boldsymbol{E}) = \frac{L}{\Phi(L\frac{\mu_{ll}}{\sigma_{ll}})}\Bigg[\quad BvN\Big(\frac{g}{\sqrt{1+h^2}}, e_{i+1}; \frac{-h}{\sqrt{1+h^2}}\Big) \tag{4.62}$$

$$- \quad BvN\Big(\frac{g}{\sqrt{1+h^2}}, e_i; \frac{-h}{\sqrt{1+h^2}}\Big)\Bigg], \tag{4.63}$$

where,

$$BvN(l_1,l_2,\rho) = \int_{-\infty}^{l_1}\int_{-\infty}^{l_2} \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}(x^2+y^2-2xy\rho)} dxdy\,, \tag{4.64}$$

is the bivariate cumulative normal distribution, corresponding to a distribution with mean and covariance,

$$\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{4.65}$$

$$\Sigma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}. \tag{4.66}$$

Additionally, g, h, and e_i denote

$$g = L\frac{\mu_{ll}\sigma_{ll}^2 + \mu_{ll}\sigma_m^2}{\sigma_m\sigma_{ll}\sqrt{\sigma_m^2+\sigma_{ll}^2}} \tag{4.67}$$

$$h = \frac{L\sigma_{ll}^2}{\sigma_m\sigma_{ll}} \tag{4.68}$$

$$e_i = \frac{Ld_i - \mu_{ll}}{\sqrt{\sigma_m^2+\sigma_{ll}^2}}\,. \tag{4.69}$$

This integral can be numerically evaluated using standard functions.

Free response condition

In the free response condition, we again want to find the probability distribution over confidence reports, but now have an additional piece of information to incorporate into our predictions, the time of the response t_r. Response time will be determined by the evolution of the accumulator, and specifically, by the first time the accumulator reaches a decision threshold (Fig. 4.6).

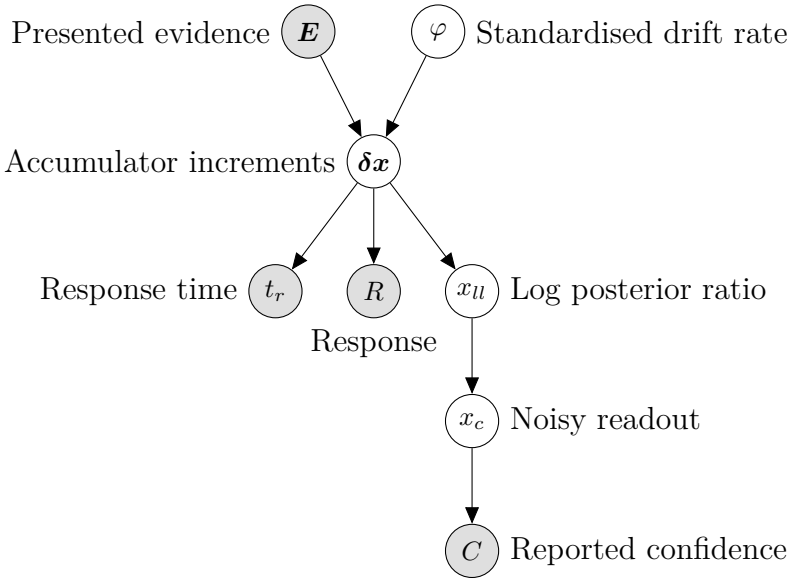

Figure 4.6: Representation of the generative model for confidence reports in the free response condition, from the perspective of the researcher. We want to infer the probability of reported confidence, using our knowledge of the evidence presented, the response given, and the response time.

Integrating x_c over the region which leads to a confidence report $C = i$ after $R = 2$, and marginalising over x_{ll},

$$p(C = i|R, t_r, \boldsymbol{E}) \quad = \quad \int_{d_i}^{d_{i+1}} dx_c \int dx_{ll}\; p(x_c|x_{ll}) p(x_{ll}|R, t_r, \boldsymbol{E})\,. \tag{4.70}$$

(As before, for $R = 1$ we use integration limits $-d_{i+1}$ and $-d_i$ instead.) The second distribution in this expression can be obtained by marginalising over φ,

$$p(x_{ll}|R, t_r, \boldsymbol{E}) = \int d\varphi\; p(x_{ll}|\varphi, R, t_r, \boldsymbol{E}) p(\varphi|R, t_r, \boldsymbol{E})\,. \tag{4.71}$$

The final term is a distribution over φ, the standardised drift rate. Observers often need to calculate the distribution over drift rate, given the evidence received (Drugowitsch et al., 2012; Moran, 2015; Moreno-Bote, 2010). As researchers, we can infer the value of φ in a similar way.

Evidence presented immediately prior to a response does not contribute to the response itself, as it is still being processed (Resulaj et al., 2009). If, for an observer, the interval of stimulus in this processing pipeline is I, we can infer that the amount of time they spent accumulating evidence prior to a decision, t_d, is

$t_d = t_e - I = t_r - I$. Additionally, if an observer uses a decision threshold, $f(t)$, to trigger responses $R = 2$, and $-f(t)$ to trigger responses $R = 1$, then we know that at the time they made their decision t_d,

$$x = \begin{cases} f(t_d) & \text{if } R = 2 \\ -f(t_d) & \text{if } R = 1\,, \end{cases} \tag{4.72}$$

because the time they made their decision was the time the accumulator hit the decision threshold (Moreno-Bote, 2010). Denote the value of x at t_d by x_d. Hence for known (or hypothesised) values of I and $f(t)$, we can infer t_d and x_d from R and t_r. Note that in this case we can also infer R and t_r from t_d and x_d. Hence, these quantities are interchangeable, i.e.

$$p(\varphi|R, t_r, \boldsymbol{E}) = p(\varphi|x_d, t_d, \boldsymbol{E}) \tag{4.73}$$

The probability distribution over φ will depend on the entire stream of evidence up to the time of the decision (i.e. all elements of $\boldsymbol{E}$ that correspond to evidence received before a decision). This is because measurements of all evidence prior to a decision, in conjunction with φ, determine changes in the accumulator, which in turn determines the time of the response, and response itself (see Fig. 4.6). However, for the purpose of inferring φ, we approximate the evidence stream by its average prior to the time of the decision, $\overline{E}$,

$$p(\varphi|R, t_r, \boldsymbol{E}) = p(\varphi|x_d, t_d, \overline{E})\,. \tag{4.74}$$

Precisely, $\overline{E}$ is given by $\overline{E} = \sum_i (\delta E_{2i} - \delta E_{1i})/t_d$, where the summation is taken over time steps prior to the decision. We test this approximation, in conjunction with other approximations, once the derivation is complete. Using Bayes rule,

$$\begin{aligned} p(\varphi|R, t_r, \boldsymbol{E}) &= p(\varphi|x_d, t_d, \overline{E}) && (4.75)\\ &\propto p(x_d, t_d|\varphi, \overline{E}) p(\varphi|\overline{E}) && (4.76)\\ &\propto p(\varphi|\overline{E}) \sum_{\omega\in\Omega} p(\boldsymbol{\delta x}^{(\omega)}|\varphi, \overline{E}) && (4.77)\\ &\propto p(\varphi|\overline{E}) \sum_{\omega\in\Omega} \prod_i p(\delta x_i^{(\omega)}|\varphi, \overline{E})\,. && (4.78) \end{aligned}$$

In this equation, Ω is the set of all paths that do not cross the decision threshold prior to t_d, and arrive at the threshold at t_d (Moreno-Bote, 2010). $\boldsymbol{\delta x}^{(\omega)}$ is the $\boldsymbol{\delta x}$ vector for the particular path ω.

We approximate $p(\varphi|\overline{E}) \approx p(\varphi)$. φ is independent of $\boldsymbol{E}$, but could depend on $\overline{E}$, because the average evidence is related to the time of the response, which φ also affects (Fig. 4.6). Again, this approximation will be tested once we have derived predictions for confidence. The approximation gives,

$$p(\varphi|R, t_r, \boldsymbol{E}) \propto p(\varphi) \sum_{\omega\in\Omega} \prod_i p(\delta x_i^{(\omega)}|\varphi, \overline{E})\,. \tag{4.79}$$

Using (4.6) we have that $\delta x_i = \delta x_{2i} - \delta x_{1i}$ is given by

$$\begin{aligned} p(\delta x_i|\varphi, \overline{E}) &= N(\delta x_i; (\delta E_{2i} - \delta E_{1i})\varphi, \sigma_{acc}^2\delta t) && (4.80)\\ &= N(\delta x_i; \varphi\overline{E}\delta t, \sigma_{acc}^2\delta t) && (4.81)\\ &\propto N(\varphi; \frac{\delta x_i}{\overline{E}\delta t}, \frac{\sigma_{acc}^2}{\overline{E}^2\delta t}) && (4.82)\end{aligned}$$

where we have again used our approximation replacing the full evidence stream with the average evidence. The proportionality is with respect to φ.

Using standard results for the product of normal distributions we find

$$\begin{aligned} \prod_i p(\delta x_i|\varphi, \overline{E}) &\propto \prod_i N(\varphi; \frac{\delta x_i}{\overline{E}\delta t}, \frac{\sigma_{acc}^2}{\overline{E}^2\delta t}) && (4.83)\\ &\propto N(\varphi; \frac{x_d}{\overline{E}t_d}, \frac{\sigma_{acc}^2}{\overline{E}^2 t_d}) && (4.84)\\ &= K^{(\omega)} N(\varphi; \frac{x_d}{\overline{E}t_d}, \frac{\sigma_{acc}^2}{\overline{E}^2 t_d})\,, && (4.85)\end{aligned}$$

where we have used $t_d = \sum \delta t$, $x_d = \sum \delta x_i$, and $K^{(\omega)}$ is the constant of proportionality. This constant will be different for different paths, so we explicitly indicate the path, ω. Hence, we have in (4.79)

$$\begin{aligned} p(\varphi|R, t_r, \boldsymbol{E}) &\propto p(\varphi) \sum_{\omega\in\Omega} K^{(\omega)} N(\varphi; \frac{x_d}{\overline{E}t_d}, \frac{\sigma_{acc}^2}{\overline{E}^2 t_d}) && (4.86)\\ &\propto p(\varphi) N(\varphi; \frac{x_d}{\overline{E}t_d}, \frac{\sigma_{acc}^2}{\overline{E}^2 t_d}) && (4.87)\\ & && (4.88)\end{aligned}$$

Recalling (4.7), φ is distributed as,

$$p(\varphi) = N(\varphi; 1, \sigma_\varphi^2)\,, \tag{4.89}$$

giving

$$\begin{aligned} p(\varphi|R, t_r, \boldsymbol{E}) &\propto N(\varphi; 1, \sigma_\varphi^2) N(\varphi; \frac{x_d}{\overline{E} t_d}, \frac{\sigma_{acc}^2}{\overline{E}^2 t_d}) & (4.90) \\ &= N\Big(\varphi; \frac{\sigma_{acc}^2 + x_d \sigma_\varphi^2 \overline{E}}{\sigma_{acc}^2 + t_d \sigma_\varphi^2 \overline{E}^2}, \frac{\sigma_{acc}^2 \sigma_\varphi^2}{\sigma_{acc}^2 + t_d \sigma_\varphi^2 \overline{E}^2}\Big)\,, & (4.91) \end{aligned}$$

where we have used standard formulae for the product of normal distributions.

Returning to (4.71), we see that in addition to an expression for $p(\varphi|R, t_r, \boldsymbol{E})$, we need an expression for $p(x_{ll}|\varphi, R, t_r, \boldsymbol{E})$. As discussed, from the response and response time, we can infer the time spent accumulating evidence prior to the decision, and the state of the accumulator at the time a decision was made (Fig. 4.2). Following a decision, evidence accumulation continues in the same manner as in the interrogation case (Pleskac & Busemeyer, 2010), until all evidence measurements have been processed. Therefore, the expression in (4.36) is valid for predicting the accumulation between the time of the decision and the end of stimulus processing (t_d, t_e). Denote the accumulation in this time Δx, and the sum over evidence presented in this interval $\Delta E = \sum_i \delta E_{2i} - \delta E_{1i}$ (the summation is taken over all i that correspond to time steps following a decision). Using (4.36),

$$p(\Delta x|\varphi, R, t_r, \boldsymbol{E}) = N(\Delta x; \varphi \Delta E, \sigma_{acc}^2 I)\,, \tag{4.92}$$

where I denotes the duration of the pipeline, $t_e - t_d$. Using our knowledge of the location of x at t_d, denoted x_d, and that the final state of x is given by $x_d + \Delta x$, the distribution over the final state is given by,

$$p(x|\varphi, R, t_r, \boldsymbol{E}) = N(x; x_d + \varphi \Delta E, \sigma_{acc}^2 I) \tag{4.93}$$

Using (4.26) we have,

$$p(x_{ll}|\varphi, R, t_r, \boldsymbol{E}) = N(x_{ll}; \frac{x_d + \varphi \Delta E}{\theta(t_e)}, \frac{\sigma_{acc}^2 I}{\theta^2(t_e)}) \tag{4.94}$$

We have now derived expressions for both the distributions in (4.71). Returning to this equation we have,

$$p(x_{ll}|R,t_r,\boldsymbol{E}) = \int d\varphi\ p(x_{ll}|\varphi,R,t_r,\boldsymbol{E})p(\varphi|R,t_r,\boldsymbol{E}) \tag{4.95}$$

$$= \int d\varphi\ N(x_{ll};\frac{x_d+\varphi\Delta E}{\theta(t_e)},\frac{\sigma_{acc}^2 I}{\theta^2(t_e)}) \\ N\Big(\varphi;\frac{\sigma_{acc}^2+x_d\sigma_\varphi^2\overline{E}}{\sigma_{acc}^2+t_d\sigma_\varphi^2\overline{E}^2},\frac{\sigma_{acc}^2\sigma_\varphi^2}{\sigma_{acc}^2+t_d\sigma_\varphi^2\overline{E}^2}\Big) \tag{4.96}$$

$$= \frac{\theta(t_e)}{\Delta E}\int d\varphi\ N(\varphi;\frac{\theta(t_e)x_{ll}-x_d}{\Delta E},\frac{\sigma_{acc}^2 I}{\Delta E^2}) \\ N\Big(\varphi;\frac{\sigma_{acc}^2+x_d\sigma_\varphi^2\overline{E}}{\sigma_{acc}^2+t_d\sigma_\varphi^2\overline{E}^2},\frac{\sigma_{acc}^2\sigma_\varphi^2}{\sigma_{acc}^2+t_d\sigma_\varphi^2\overline{E}^2}\Big)\,. \tag{4.97}$$

Applying the result in Appendix B.1 and rearranging we find,

$$p(x_{ll}|R,t_r,\boldsymbol{E}) = N(x_{ll};\mu_{ll},\sigma_{ll}^2) \tag{4.98}$$

Where,

$$\mu_{ll} = \frac{1}{\theta(t_e)}\left(x_d+\Delta E\frac{\sigma_{acc}^2+x_d\sigma_\varphi^2\overline{E}}{\sigma_{acc}^2+t_d\sigma_\varphi^2\overline{E}^2}\right) \tag{4.99}$$

$$\sigma_{ll}^2 = \frac{\sigma_{acc}^2}{\theta^2(t_e)}\left(\frac{\Delta E^2\sigma_\varphi^2}{\sigma_{acc}^2+t_d\sigma_\varphi^2\overline{E}^2}+I\right) \tag{4.100}$$

Using this result in (4.70), and allowing for normally distributed metacognitive noise as in (4.28),

$$p(C=i|R,t_r,\boldsymbol{E}) = \int_{d_i}^{d_{i+1}} dx_c \int dx_{ll}\ p(x_c|x_{ll})p(x_{ll}|R,t_r,\boldsymbol{E}) \tag{4.101}$$

$$= \int_{d_i}^{d_{i+1}} dx_c \int dx_{ll}\ N(x_c;x_{ll},\sigma_m^2)N(x_{ll};\mu_{ll},\sigma_{ll}^2) \tag{4.102}$$

$$= \int_{d_n}^{d_{n+1}} dx_c\ N(x_c;\mu_{ll},\sigma_{ll}^2+\sigma_m^2)\,. \tag{4.103}$$

This expression applies for $R = 2$. For the case of $R = 1$ the limits change to $-d_{i+1}$ and $-d_i$.

Testing the approximations

We made several approximations in the derivations above, so it is important to check that our predictions for confidence closely match confidence reports, when

these are simulated. We simulated the diffusion process using small time steps, and produced confidence reports in accordance with the model (see (4.27), (4.28) and Fig. 4.2; for details of the simulations see Appendix B.3). We then took each trial and computed predictions for the probability of each confidence report using the derived expressions, before randomly drawing a confidence report in accordance with the probability assigned to it. This allowed us to plot confidence simulated from the model, and confidence reports that match the derived predictions.

Fig. 4.7 shows simulations of confidence using the model (error bars), and confidence based on the derived predictions (error shading). No approximations were made in deriving confidence predictions in the interrogation condition. Consistent with this, simulated confidence, and the variance of simulated confidence, closely match predicted confidence and predicted confidence variance, as functions of response time and average evidence, both with and without variability in standardised drift rate.

For the free response derivations, we approximated evidence prior to a decision by its average, for the purpose of estimating standardised drift rate, and approximated standardised drift rate as independent of the average evidence prior to a decision. These approximations are only relevant when standardised drift rate is variable. Consistent with this, simulated and predicted confidence closely match in plots corresponding to no variability in standardised drift rate φ (Fig. 4.7). When variability in standardised drift rate is present, we can see that the approximations introduce some small discrepancies between simulations and predictions. For example, the predictions appear to overestimate the variability in confidence reports, in trials with long response times.

4.4 Discussion

Combining the normative framework of the DDM with the idea of a Bayesian readout for confidence, we derived predictions for the probability distribution over confidence reports, given the response, response time, and stimulus presented on a

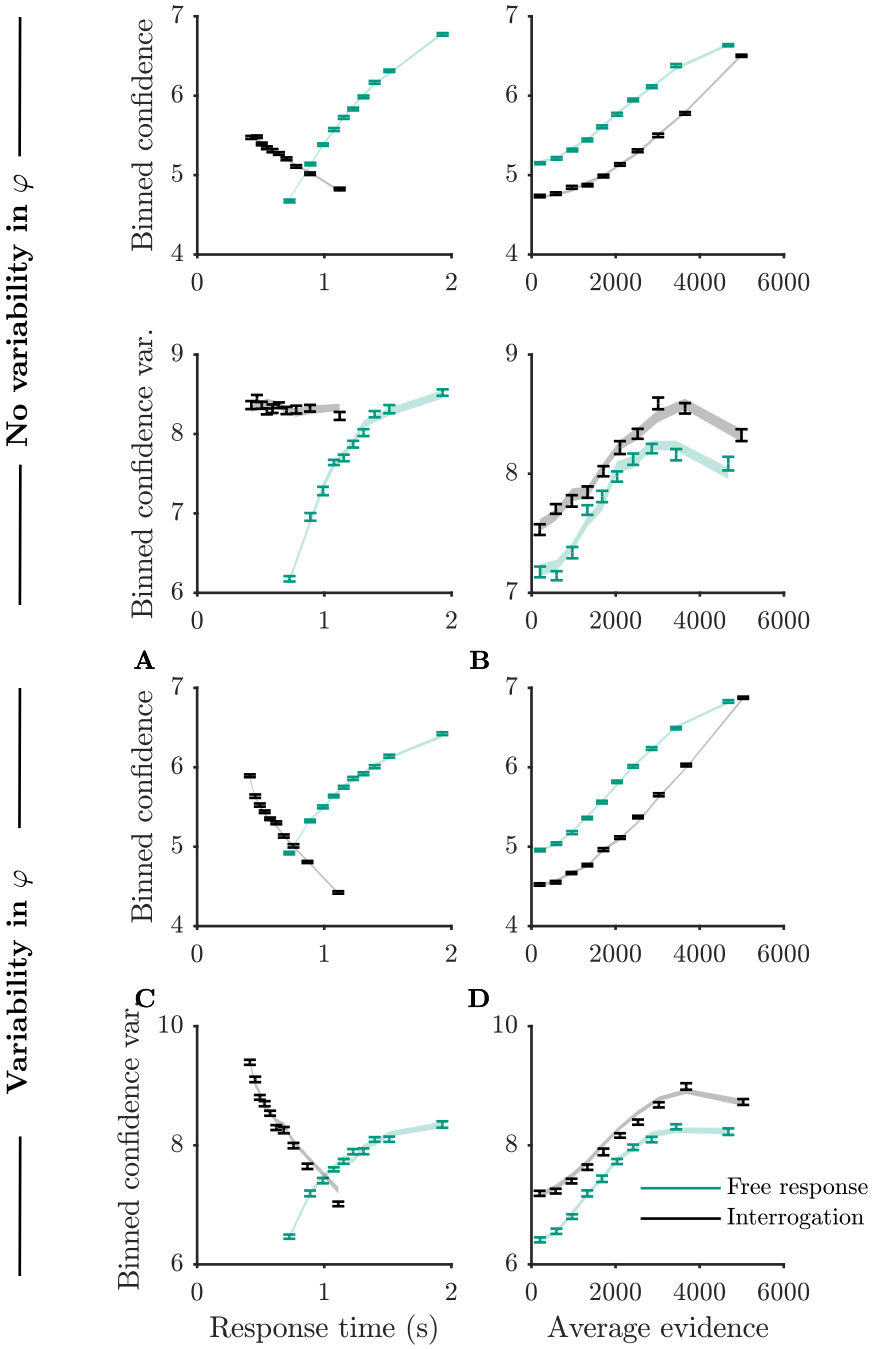

Figure 4.7: Mean and variance of binned confidence, produced via simulation of the model (error bars), and through the derived predictions (shading). Details of the simulation and plotting are provided in Appendix B.3. Predictions matched the simulation closely. There were some signs that approximations used to derive the predictions lead to a small overestimation of variability in confidence, when there is variability in standardised drift rate.

trial. We considered both the typical case, where response time is under the control of the participant (free response), and the less common case in which the observer has to respond at a particular time (interrogation). In the free response case, where the observer must set decision thresholds to trigger a response, we allow for the use of decision thresholds of arbitrary shape. These results build on the work of Moreno-Bote (2010), by including features known to be important in the construction of confidence. Specifically, the derivations account for accumulation of pipeline evidence (Moran et al., 2015), the effect of drift rate variability on pipeline evidence (Pleskac & Busemeyer, 2010), and metacognitive noise (Maniscalco & Lau, 2012, 2016).

Importantly, the derivations cover both static stimuli and dynamic stimuli that generate normally distributed fluctuations in evidence signal each frame. The derived expressions only require one evaluation per trial, in contrast to previous approaches that could handle dynamic stimuli, which required the evaluation of some function at every time step prior to a decision (Chang & Cooper, 1970; Smith, 2000; Zylberberg et al., 2018). Reducing computational cost is crucial for making trial-by-trial modelling of dynamic stimuli feasible. Trial-by-trial modelling may provide a stronger constraint when fitting models, than predictions made for large groups of trials at once, which has until now been the standard approach (see Chapter 1). Computationally cheap predictions may also allow us to use techniques that require predictions to be evaluated many times, such as cross-validation and Markov chain Monte Carlo (MCMC; Bishop, 2006).

Simple expressions may also provide additional insight into the mechanisms responsible for confidence. It is instructive to consider a situation in which observers can report their confidence on a very fine-grained scale. In this case, the most likely confidence report corresponds to the most likely value of x_c, the observer's noisy, continuous readout of confidence. In the interrogation condition without metacognitive noise, the most likely value of x_c is given by the mean of the distribution over x_c in (4.45). Substituting in all abbreviations, we have that

the most likely confidence report is

$$\frac{E}{\theta(t_e)} = \frac{2\nu_0 E}{s^2 + t_e\sigma_\nu^2} = \frac{2t_f\Delta\mu E}{t_f^2\sigma_{acc}^2 + t_f\sigma_E^2 + t_e\Delta\mu^2\sigma_\varphi^2}\,. \tag{4.104}$$

As might be expected, the most likely confidence report is proportional to the total evidence presented, E. We also see that any form of noise, whether internal accumulator noise, variability in the stimulus, or drift rate variability, reduces the most likely confidence. Note that these effects apply once the observer has learned the statistics of the task. Hence, this result does not conflict with the finding that when stimulus variability is uncued, greater variability causes increases in confidence (Zylberberg et al., 2016). It is simple to extend the derivations to account for the case where observers incorrectly estimate variance in the stimulus (see Appendix A.2.2). We only need to change the generative model used by the observer, that determines how the observer transforms the time and accumulator state, x, into an estimate of the log posterior ratio, x_{ll}. Such changes would affect the function $\theta()$ but would leave the dependence on total evidence, E, unchanged.

We can take the same approach for the free response condition, and consider the most likely confidence report if a fine grained confidence scale is used. In the absence of metacognitive noise, the most likely confidence report is given by (4.99),

$$\begin{aligned}\frac{1}{\theta(t_e)}&\left(x_d+\Delta E\frac{\sigma_{acc}^2 + x_d\sigma_\varphi^2\overline{E}}{\sigma_{acc}^2 + t_d\sigma_\varphi^2\overline{E}^2}\right)\\ &= \frac{2t_f\Delta\mu}{t_f^2\sigma_{acc}^2 + t_f\sigma_E^2 + t_e\Delta\mu^2\sigma_\varphi^2}\left(x_d + \Delta E\frac{\sigma_{acc}^2 + x_d\sigma_\varphi^2\overline{E}}{\sigma_{acc}^2 + t_d\sigma_\varphi^2\overline{E}^2}\right).\end{aligned} \tag{4.105}$$

Breaking this expression down, we see that the prefactor which multiplies the term in brackets is the same as the prefactor that multiplies E in the interrogation case, hence the same dependencies from this prefactor will carry over. It is interesting to note that if there is no variability in standardised drift rate ($\sigma_\varphi = 0$), then the only evidence used for predicting confidence is evidence from the processing pipeline, ΔE. It seems counter-intuitive that the evidence on which the decision was based adds nothing to our prediction for confidence. This occurs because, at the time of the decision, we know that the state of the accumulator is at the decision boundary

corresponding to the response made (Section 2.1.2; Fig. 4.2; Pleskac and Busemeyer, 2010). Given knowledge of the state of the accumulator at the time of the decision, we do not need to know the evidence presented up to this point. This is the same line of reasoning that led us to prediction A and B in Chapter 2.

Aside from the prefactor, the most likely level of confidence is determined by the state of the accumulator at decision time x_d, plus an additional term that captures the path of the accumulator following the decision,

$$\Delta E \frac{\sigma_{acc}^2 + x_d \sigma_\varphi^2 \overline{E}}{\sigma_{acc}^2 + t_d \sigma_\varphi^2 \overline{E}^2} \,. \tag{4.106}$$

The fraction multiplying ΔE, which is 1 when $\sigma_\varphi = 0$, is a term that accounts for the effect of drift-rate variability on confidence. This term provides a mathematical description of the idea discussed in Chapter 1 that, if strong evidence has been gathered by the time of the decision, relative to the time spent deliberating, the standardised drift rate is likely to be high (Moreno-Bote, 2010), and pipeline evidence will have a big impact on decision confidence (Pleskac & Busemeyer, 2010). On the other hand, if at the time of decision, little evidence has been gathered relative to the time spent deliberating, evidence is accumulating slowly, suggesting a low standardised drift rate. In turn, this suggests that pipeline evidence will be processed poorly and will have a small effect on confidence. This is why the fraction contains decision time in the denominator, reducing the effect of pipeline evidence, and the height of the threshold at decision time is in the numerator, increasing the effect of pipeline evidence. We are not the first to describe this effect of drift rate variability; it was a central idea in the model for confidence proposed by Pleskac and Busemeyer (2010). Our contribution is to derive an expression for this effect.

If we assume that the decision threshold, x_d, does not depend on time, then decision time and stimulus duration only appear in denominators in (4.105). Hence, in the absence of other changes, decision time will decrease confidence. This is consistent with the findings and conclusions drawn by Kiani et al. (2014), and the discussion in Chapters 1 and 2. Kiani et al. (2014) increased decision time by

using stimuli which, for 360 ms, provided no net evidence for either choice. This manipulation decreased confidence, without clear changes in accuracy, suggesting that the effect of decision time on confidence is direct, and not mediated by changes in evidence.

Expressions for confidence may also help us understand confidence findings that previously appeared difficult to explain. In Section 2.1.1 we noted that one confusing feature of the literature has been inconsistency in the observed relationship between confidence and signal strength, on error trials. Sanders et al. (2016) found confidence on error trials decreased as signal strength increased. However, Kiani et al. (2014) found that confidence on error trials increased with signal strength. In Section 2.1.1 we provided an intuitive argument suggesting that DDM models of confidence can reconcile these conflicting results, by taking into account the timing of confidence reports (Desender et al., 2020; Kiani et al., 2014). Sanders et al. (2016) collected confidence reports following a decision, suggesting that pipeline evidence would have contributed to confidence. In contrast, Kiani et al. (2014) asked participants to simultaneously report decisions and confidence using a saccade. The stimulus turned off as soon as a saccade began, minimising the chance that confidence was based on any more information than the decision, i.e. $\Delta E = 0$.

To test intuitions used in Section 2.1.1, we simulated response times, decisions, and confidence reports, under two conditions. In one condition, we simulated the task designed by Kiani et al. (2014), setting pipeline evidence to zero. In the other condition, we simulated a processing pipeline containing (a conservative) 100ms of the stimulus prior to response (see Appendix B.3 for details of the simulations). Consistent with our previous conclusions, confidence in errors increased with signal strength when decisions and confidence reports were simultaneous, while confidence in errors decreased with signal strength when observers received pipeline evidence (Fig. 4.8).

An additional strength of the derivations is the treatment of metacognitive noise. If we fit the predictions to data, we will get an estimate for metacognitive noise

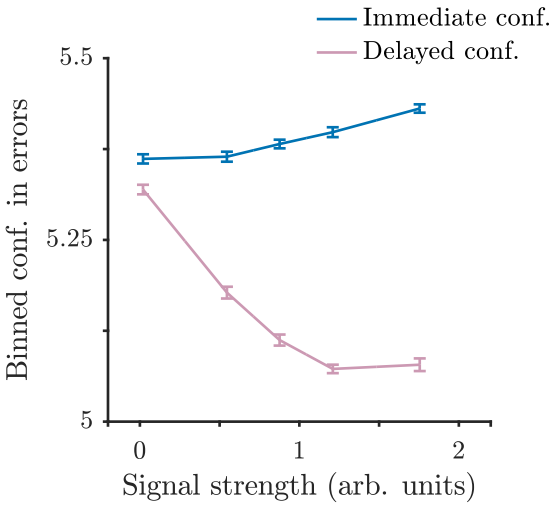

Figure 4.8: Confidence on error trials as a function of signal strength. When confidence is reported immediately, and therefore does not reflect evidence measurements in the processing pipeline, confidence on error trials increases with signal strength. On the other hand, when confidence reports are made after a decision, and include 100ms of evidence measurements from the processing pipeline, confidence on error trials decreases with signal strength.

from the value of the fitted metacognitive noise parameter. Existing methods for inferring the level of metacognitive noise are based on signal detection frameworks (Maniscalco & Lau, 2012). Therefore, these methods implicitly assume that signal detection theory is an adequate description of decision making mechanisms. The DDM can be viewed as an extension of signal detection theory to the case where response time, and hence the amount of information obtained, is under the control of the observer (Smith, 2000). This is a common situation in empirical studies. Therefore, more accurate estimates of metacognitive noise may be obtainable using derivations in which the DDM is the assumed framework, such as here.

There are a number of limitations to this work, which can be divided into those that affect the scope of the derivations themselves, and those that concern the practical application of the predictions in model fitting. Regarding the former, limits to the derivations themselves, we made two particularly important assumptions. Namely, we used a DDM framework, and assumed stimuli of a specific kind. A family of alternatives to the DDM assume not just one accumulator, but two

(Chapter 1; Bogacz et al., 2006; Moreno-Bote, 2010). Each accumulator corresponds to one of the choices, and the accumulators may be more or less correlated or anti-correlated with each other. It may be possible to use recently derived expressions for the state of the second accumulator at decision time, to extend the approach developed here, although it may be necessary to approximate stimuli as static prior to a decision (Shan et al., 2019). Notwithstanding this consideration, we believe the DDM to be a particularly interesting case due to the normative properties of the diffusion mechanism (Chapter 1).

The second limitation to the scope of the derivations is that we assumed stimuli of a specific kind. It may be possible to extend the derivations to more general dynamic stimuli, rather than just those with normally distributed evidence fluctuations. Properties of the stimuli are only used to derive confidence of the optimal observer, given an accumulator state and deliberation time. Properties of the stimuli are not important for deriving predictions for confidence, once the observer's internal mapping from accumulator state and time, to confidence, is known. Therefore, any stimulus for which we can find the optimal observer's decision and confidence rules could be modelled using the approach we have described. One of the most common stimuli, the random dot motion stimulus, contains dots that move randomly over the course of a trial, creating random fluctuations in evidence for the prevailing motion direction (Kiani et al., 2008; Pilly & Seitz, 2009). If it was possible to characterise the nature of these fluctuations, and derive the optimal observer's decision and confidence rule, a very large quantity of data could be analysed on a trial-by-trial basis using the approach set out here.

Turning to limitations that affect the applicability of these derivations to modelling, the fact that we only make predictions for confidence, not for response times and decisions may be a concern, along with the fact that we have not yet accounted for lapses. It is certainly the case that we can only fit to confidence given decisions, and response times, rather than fitting to decisions and response times themselves. However, only predicting confidence is key for avoiding difficulties in

deriving computationally cheap expressions for responses and response times. As discussed in Section 4.1, there is no simple expression for the probability distribution over the state of the accumulator prior to the decision, because of the effect of the decision thresholds (Cox & Miller, 1965; Smith, 2000). Additionally, when fitting confidence reports, we will still model all aspects of the decision mechanism, in the sense that we will generate estimates for all parameters of the decision mechanism. Using these parameters we can make predictions for decisions and response times, providing an additional check of the model and its assumptions. This is exactly the approach we took in Section 3.4.

Regarding lapses, for the modelling in Chapter 3 we include a confidence lapse rate parameter that describes some probability that a random confidence report is given (Appendix A.2.3). It would be trickier to model lapses that also affect the response given (Adler & Ma, 2018a). The probability that a given response and response time resulted from a lapse would be difficult to calculate because, whilst we might know the probability distribution over responses and response times generated by lapses, we do not have expressions for the probability distribution over decisions and response times generated by the non-lapse diffusion process. For similar reasons, it would be difficult to incorporate the idea of variability in the start point of the accumulator, and variability in the duration of the pipeline (Ratcliff & McKoon, 2008; Ratcliff & Tuerlinckx, 2002).

Notwithstanding the limitations discussed, we believe these derivations will prove useful, both for gaining insight into the model of confidence implied by our current knowledge, and for model fitting. We have already seen in this discussion that the derived expressions offer the potential to directly explore the relationship between confidence and other variables, without the need to run simulations using a wide range of parameter values. The expressions also offer explanatory power difficult to gain otherwise. For example, in (4.106) we saw that the effect of evidence accumulated following a decision is modulated by a complicated factor. We were

able to understand this factor as a term which describes the effect of drift rate variability, and to make sense of the way this term was affected by evidence and time.

Regarding model fitting, we believe the derivations will make it more feasible to conduct trial-by-trial modelling of confidence data. Trial-by-trial modelling allows us to capitalise on variability in dynamic stimuli, rather than ignoring it. From a model fitting and comparison point of view, models must now explain far more detail. For example, models will need to account for the difference between confidence on trials in which little evidence was presented but the response was quick, and confidence on trials in which a large amount of evidence was presented but the response was slow. They will need to explain the precise quantitative effect of response and evidence, rather than simply the effect of fast vs. slow trials, or high vs. low evidence. Additionally, the models will need to account for the time course of the effect of evidence. We explored whether DDM variants could account for all these features using trial-by-trial modelling in Chapter 3. This was made possible by the computationally cheap expressions developed here.

5

Challenges to Bayesian confidence

Contents

5.1 Introduction

We have seen that the DDM with a miscalibrated Bayesian readout for confidence can provide an excellent qualitative and quantitative account of patterns in confidence data. In this chapter we consider findings, reviewed in Section 1.6, which appear to challenge the idea of Bayesian confidence. We consider two such phenomena,

but we discuss further findings in Section 7.2.2. The two phenomena which we investigate are congruent-evidence over-weighting and the positive evidence effect (Koizumi et al., 2015; Peters et al., 2017; Rollwage et al., 2020; Samaha et al., 2016; Zylberberg et al., 2012). We outline these phenomena in turn, and consider whether they can be explained using Bayesian models of confidence.

5.1.1 Congruent-evidence over-weighting

Congruent-evidence over-weighting refers to the idea that, while two-alternative perceptual decisions are determined by the balance of evidence between the options, perceptual confidence is more influenced by decision-congruent information than decision-incongruent information (Zylberberg et al., 2012). "Decision-congruent" information refers to information presented in the selected stimulus item or feature (e.g. selected array of dots or motion direction), and "decision incongruent" refers to information presented in the unselected stimulus item (Peters et al., 2017; Zylberberg et al., 2012). There is substantial evidence for congruent evidence over-weighting (Koizumi et al., 2015; Peters et al., 2017; Rollwage et al., 2020; Samaha et al., 2016; Zylberberg et al., 2012; although see Charles and Yeung, 2019; note we analyse the data collected by Charles and Yeung, 2019 again below).

At first sight this phenomenon points to a separate process for confidence readouts as compared to the decision itself, one that does not conform to Bayesian principles (Peters et al., 2017): Intuitively, if decision-congruent and decision-incongruent evidence is equally reliable, it should be weighted equally by a Bayesian observer. It follows from this line of reasoning that observers should use the difference in evidence between the two alternatives (i.e. the balance of evidence), to determine both their choices and confidence (Bogacz et al., 2006; Vickers & Packer, 1982). We checked this reasoning using the model developed in previous chapters that features a miscalibrated Bayesian readout for confidence. We simulated stimuli with properties that matched those used in the experiment for dataset B (Section 3.1; simulation details in Appendix C.2). Behaviour was then simulated using the

best fitting Bayesian observer model from Chapter 3, with parameters set to the median of the values obtained when fitting to the data (Appendix C.2).

At each 50ms stimulus frame relative to stimulus onset and response, we looked at the correlation between decision-congruent evidence (evidence presented in the selected array), decision-incongruent evidence (evidence presented in the unselected array), and confidence (Fig. 5.1). Fig. 5.1 only reflects data from the free response condition. Crucially, while the effects of evidence fluctuations on confidence varied over time, the effect of congruent and incongruent evidence was matched: In no frame was there a clear difference between the correlation coefficients describing the effects of congruent and incongruent evidence, regardless of whether the data were analysed time-locked to the stimulus or to the response. When time-locked to the stimulus, the effect of evidence varied little as a function of time. When time-locked to the response, a clear effect of time was observed (and this effect was very similar for decision-congruent and decision-incongruent evidence). Consistent with the reasoning in Chapter 2, evidence just before a response had a large effect on confidence. This evidence is in sensory processing pipelines at the time of the decision, or is received between internal commitment to the decision and overt response, and so is processed after a decision threshold is crossed (Resulaj et al., 2009). Evidence processed before the threshold crossing is irrelevant when the decision threshold is flat because, whatever evidence is presented, at the time of the decision the evidence accumulator will be at the height of the decision threshold. Regardless of the strong dependence of the effect of evidence on time, the effect of congruent and incongruent evidence remained matched.

Although this result matches intuition and strongly suggests that a Bayesian observer equally weights decision-congruent and incongruent evidence, this turns out not to be the case in general. That is, there are circumstances in which a Bayesian observer does not behave like an observer who uses the balance of evidence alone to determine their confidence (Aitchison et al., 2015). To clarify the discussion we work, for now, from an SDT framing of perceptual decisions in which there is a static

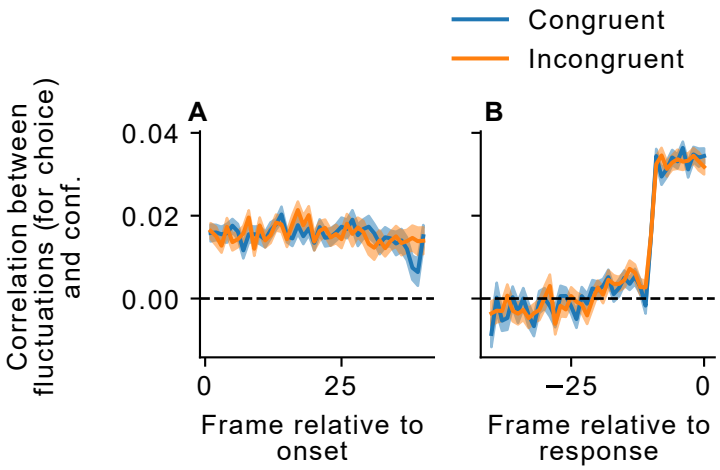

Figure 5.1: The effect of congruent and incongruent evidence in the free response condition of a simulated dataset, featuring Bayesian observers. Data were simulated using the best fitting model from Chapter 3, see Appendix C.2 for details. For this plot we excluded frames which were more than 2 seconds after onset, when looking at frames relative to onset, and frames more than 2 seconds before or after the response, when looking at frames relative to the response.

representation of evidence which does not change over time (Green & Swets, 1966). It we accept that observers do in fact separately represent evidence pertaining to the two options in a two-alternative perceptual task, then we can represent the observer's measurement of evidence for the two options on a graph (Aitchison et al., 2015). The x-axis represents measured evidence for option A, and the y-axis represents measured evidence for option B (Fig. 5.2). For an observer who uses the balance of evidence for both decisions and confidence, the contours that trace out constant levels of confidence run parallel to the diagonal, reflecting the fact that this observer ignores the sum of evidence for the two alternatives (Fig. 5.2, A). For an observer who uses the balance of evidence to select their choice, but then only uses decision congruent information to determine their confidence, contours of constant confidence run parallel to the axis for the unselected option, reflecting the fact that this observer ignores incongruent evidence when determining their confidence (Fig. 5.2, B).

We can now consider the confidence of an observer who reports the probability they are correct, given their evidence measurements (Kiani et al., 2014; Pew, 1969; Vickers & Packer, 1982). We plot the confidence of this Bayesian observer

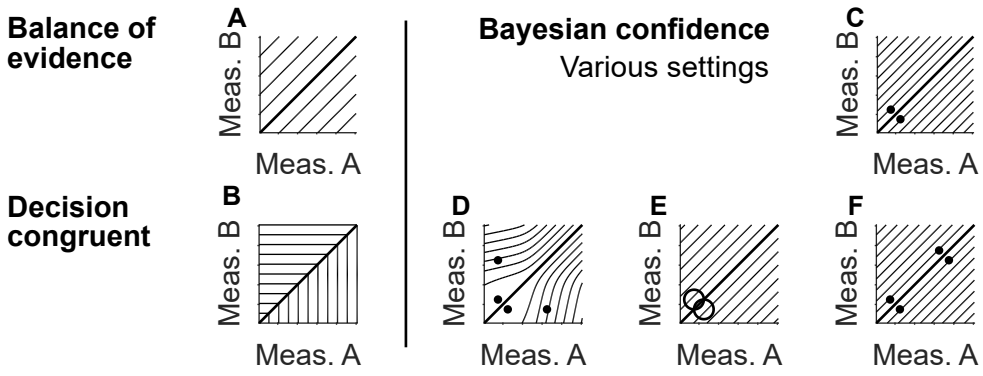

Figure 5.2: Contours of constant confidence in different models and experimental settings. The axes in the plots represent the noisy measurements of evidence for option A and B. Working in an SDT framework, we assume the observer receives a single measurement for each option. (A) Contours of constant confidence for an observer who uses the balance of evidence to determine their confidence. (B) Confidence contours for an observer who only uses decision-congruent evidence to determine their confidence. (C - F) Contours of constant confidence for an observer who uses a Bayesian readout of the probability of being correct for their confidence. (C, D, F) One of a small number of stimulus combinations is presented. The most probable measurement of these stimulus combinations is represented by a dot. (E) Stimuli are drawn from uncorrelated normal distributions. This situation is represented using circles to indicate the range of typical evidence measurements. See Aitchison et al. (2015) for similar plots to A, B, and D.

in four different settings (Fig. 5.2, C-F). As usual, the observer's task is to determine whether there is more evidence for option A or option B. In one setting considered, there are only two possible stimulus combinations (Fig. 5.2, C). Either there is slightly more evidence presented for option A, or there is slightly more evidence presented for option B. The observer's measurement is noisy, but we plot the most likely measurement for the two stimulus combinations as a point. In this context, we see that the probability of being correct only depends on the balance of evidence (Maniscalco et al., 2016; Vickers & Packer, 1982). The sum of evidence for the two options provides no extra information. Hence, in this case, our intuition that the Bayesian observer relies on the balance of evidence to determine their confidence is correct.

This experimental situation is surprisingly rare. Typically, the difficulty of the task is manipulated (e.g. Koizumi et al., 2015; Peters et al., 2017). A common way to vary the difficulty of the task is to present more or less evidence in the correct stimulus

item, while leaving the evidence in the incorrect stimulus item unchanged (e.g. Kiani et al., 2014; Peters et al., 2017). Miyoshi and Lau (2020) showed that an optimal observer does not use the balance of evidence for their confidence reports when more variability is associated with evidence for the correct option than the incorrect option (there referred to as target and nontarget). This is exactly the situation that is constructed when we vary difficulty by providing more or less evidence for the correct option, while keeping evidence for the incorrect option unchanged.

In Fig. 5.2, D, we plot Bayesian confidence in a situation where there are the same two stimulus combinations as in Fig. 5.2, C, but now two additional possibilities, representing cases in which the task is made easier (Aitchison et al., 2015). We see that the balance of evidence remains important for confidence, but that the sum of evidence for the two alternatives also becomes important (because the contours describing constant confidence are no longer parallel to the diagonal; Aitchison et al., 2015). The sum of the evidence now provides a cue as to the difficulty of the task: More measured evidence for either option suggests that more evidence was presented, and that we are in the easier condition. As a result, confidence increases. In the absence of this effect, decision-congruent evidence increases confidence, while decision-incongruent evidence decreases confidence. However, taking into account the new effect that leads to an increase in confidence with either congruent or incongruent evidence, the positive effect of decision-congruent evidence will be boosted, while the negative effect of decision-incongruent evidence will be diminished. As a result, in this context, a Bayesian observer will weight decision-congruent evidence more heavily in their confidence judgement. Note that the boundary between choices for option A and option B is unaffected and still reflects the balance of evidence.

Key studies that have found decision-congruent evidence to have more of an effect on confidence than decision-incongruent evidence have varied the evidence provided for the correct stimulus item or feature. Peters et al. (2017) varied the level of contrast used to present a house or a face in a discrimination between these.

In other discrimination tasks, Zylberberg et al. (2012) varied the proportion of dots moving left or right, and the luminance of a target patch, to maintain a target level of performance. Given the stimulus sets used, it is not at all clear these results are in conflict with the idea of Bayesian confidence.

Aitchison et al. (2015) directly compared a model in which confidence reflected a Bayesian readout with models in which confidence was determined by the balance of evidence, or by decision-congruent evidence alone. Importantly, difficulty was varied by increasing the evidence presented for the correct option. When confidence reports were collected following a decision there was clear evidence for a Bayesian readout. When confidence was collected simultaneously with a decision, the Bayesian readout and decision-congruent model were more closely matched, but both performed better than the balance of evidence model. These results support the idea of Bayesian readout for confidence, and also support the claim that when difficulty is varied, a Bayesian readout differs markedly from a readout of the balance of evidence.

We aimed to provide an empirical test of the claim that, for confidence, decision-congruent evidence is generally weighted more heavily than decision-incongruent evidence, even when this is inappropriate from a Bayesian point of view. We used three datasets with a clear face-valid measure of evidence for the two options. All three datasets used the dot discrimination task described in Chapter 2 and 3, in which observers aimed to decide which of two arrays contained the most dots on average. The number of dots in the chosen (unchosen) array provides a simple measure of decision-congruent (incongruent) evidence. In contrast, previous studies have often relied on derived measures of evidence. For example, Peters et al. (2017) trained a decoder on electrocorticography (ECoG) signals, and then used the resulting weights to estimate decision-congruent and incongruent evidence. Zylberberg et al. (2012) derived a measure for decision-congruent and incongruent evidence by applying motion filters to a stimulus containing moving dots. Although Zylberberg et al. (2012) used a face-valid measure of evidence in their second experiment.

A crucial feature of two of the three datasets was that, due to the stimuli used, a Bayesian observer would use the balance of evidence to determine their confidence. In these datasets, difficulty varied trial-to-trial, but due to the structure of this variation, the difference in evidence remained the only relevant variable to confidence. For example, in one dataset, evidence values for both the correct and incorrect options were manipulated, with the same variance, and independently of each other. We represent this situation in Fig. 5.2, E, using circles to reflect the fact that the stimulus may take a variety of values, and that this distribution follows a bivariate independent Gaussian. In a third dataset, difficulty was manipulated in complex ways following a decision. Nevertheless, this dataset provides a valuable additional data point in this debate.

5.1.2 Positive evidence effect

The second phenomenon we investigate in this chapter, in addition to congruent-evidence over-weighting, is the positive evidence effect. The positive evidence effect refers to the finding that by boosting evidence for both alternatives, confidence can be increased without affecting performance (Koizumi et al., 2015; Rollwage et al., 2020; Samaha et al., 2016). The positive evidence effect has been linked to congruent-evidence over-weighting. Specifically, it has been suggested that the positive evidence effect arises because of the over-weighting of congruent-evidence (Koizumi et al., 2015; Rollwage et al., 2020). The thought is that perceptual decisions, reflecting the balance of evidence, will not be affected by an increase in evidence for both alternatives. On the other hand, if decision-congruent evidence is over-weighted in confidence then increasing evidence for both alternatives should increase confidence.

Bayesian confidence does not predict the positive evidence effect when a new condition is created by adding the same amount of evidence to both alternatives (Fig. 5.2, F). In this situation it remains optimal to continue to use the balance of evidence for confidence. Whilst it is never desirable to postulate additional mechanisms, it may be possible to account for the positive evidence effect within

our framework by incorporating extra, empirically supported, detail into our models. Specifically, we suggest this effect may be explained by an increase in noise with increasing evidence (Pardo-Vazquez et al., 2019; Teodorescu et al., 2016). Higher levels of noise can generate higher levels of confidence because noise can increase the probability of reaching very high levels of accumulated evidence (Zylberberg et al., 2016). Noise will of course increase the chance of reaching high levels of evidence for both the correct and the incorrect options, but in both these circumstances confidence will be high. Hence, higher evidence for either option, if this generates higher levels of noise, may generate higher confidence.

A number of mechanisms could generate increasing noise with increasing evidence. First, and most simply, noise in the encoding of stimulus intensity may be proportional to that intensity (Pardo-Vazquez et al., 2019; Teodorescu et al., 2016). Evidence for this claim comes from the finding that increasing the intensity of two stimuli, without changing their ratio, does not change accuracy but does make decisions faster, and through a specific transformation of the response time distribution (Pardo-Vazquez et al., 2019). This finding can be accounted for by a model that assumes the noise in a representation is proportional to the value of the quantity represented. Highly relevant to the present discussion, Teodorescu et al. (2016) found that the DDM provided a good account of the effects of boosting evidence for both options, but only when the model featured noise that was proportional to the evidence accumulation inputs.

An alternative explanation for increasing noise with increasing evidence comes from the idea that there are fluctuations in the precision with which different items in a single display are encoded (Mazyar et al., 2012; although see Shen and Ma, 2019). Consider a model in which drift-rate varies from trial-to-trial (Ratcliff & McKoon, 2008; Ratcliff et al., 1999) and has a multiplicative effect on evidence measurements as considered in previous chapters (Section 2.1). However, unlike in previous chapters, now consider the case where drift rate varies from trial-to-trial independently for the two stimuli, so that on one trial the drift rate might be high

for one stimulus but low for the other. In this case, because the effect of drift-rate variability is multiplicative, the level of noise generated by drift-rate variability increases if more evidence is presented for item 1, but this is also true for item 2 (Appendix C.1.3). As a result, the noise generated by variability in drift-rate increases as the average evidence increases.

In the dataset collected for Chapter 3, we varied the average evidence without changing the difference in evidence. Hence, this dataset provides an excellent opportunity to examine whether stimulus-dependent noise can account for effects observed on confidence, when increasing evidence for both options simultaneously.

To summarise, here we combine empirical data analysis with modelling to explore two closely related questions. First, we explore whether decision-congruent evidence is over-weighted in confidence reports, even when this would not be the policy of a Bayesian observer. Second, we look for the positive evidence effect in our data, and examine whether it can be explained using the idea of stimulus dependent noise.

5.2 Methods

5.2.1 Experimental methods

Datasets. We used three previously collected datasets for the analysis.

Dataset A was collected by Charles and Yeung (2019) and was the dataset used in Chapter 2. Unlike in Chapter 2, here we analyse data from both conditions in this experiment, one in which response speed was emphasised and one in which response accuracy was emphasised. Dataset A contained 23 participants, and a total of 11040 trials.

Dataset A used an experimental setting similar to that considered in Fig. 5.2, C: Task difficulty was not varied, nor was the evidence presented in the two alternatives simultaneously increased or decreased. This suggests that, as in Fig.

5.2, Bayesian confidence only depends on the balance of evidence, and decision-congruent and decision-incongruent evidence should be weighted equally. However, Fig. 5.2 represents Bayesian confidence within an SDT framework, where the observer receives a single static representation of evidence (Chapter 1; Green and Swets, 1966). This seems an inappropriate model for an experiment where evidence was presented over an extended period of time. We made predictions for confidence using a dynamic representation of evidence in Chapter 4. There we allowed for the integration of all information presented during a trial. In Chapter 4 we considered a case in which the midpoint between the mean evidence presented for the correct and incorrect options varied from trial to trial (this value was denoted $\overline{\mu}$). However, the derivations apply equally well to the special case in which the evidence midpoint is fixed. The predictions for Bayesian confidence made in Chapter 4 only contained evidence variables that reflected the difference in evidence (e.g. they contained the difference in evidence following a decision, ΔE). Hence, if observers are reporting Bayesian confidence in the task used for dataset A, we expect decision-congruent and decision-incongruent information to be weighted equally, just as the SDT modelling suggested.

Dataset B was the dataset collected for Chapter 3. We analysed the data from the free condition (participants could respond when they wanted) and the interrogation condition (participants cued when to respond) separately (Bogacz et al., 2006). Dataset B contained 48 participants, and a total of 30720 trials.

As described in Section 3.1, the midpoint between the means of the two distributions for the number of dots in the correct and incorrect arrays itself varied from trial to trial. Hence, on each trial, the evidence for both options was increased or decreased simultaneously. This is a situation analogous to the SDT situation depicted in Fig. 5.2, F. While SDT can provide us with intuitions about the effects of dynamic stimuli (which turn out to be correct), the derivations in Chapter 4 apply directly. Therefore, we again expect a Bayesian observer would report confidence which weighted decision-congruent and decision-incongruent information equally.

Dataset C was collected by Carlebach and Yeung (2020), and included the data from the five experiments Carlebach and Yeung (2020) reported. These experiments were very similar to the experiments used for datasets A and B. Participants reported whether there were on average more dots in an array on the left of the display, or in an array on the right of the display. As in the other datasets, the number of dots in each array was updated every 50ms by sampling from two distributions, one with a higher mean (distribution for correct array), than the other (distribution for incorrect array). The distributions used were similar to those used by Charles and Yeung (2019). (For details see Carlebach and Yeung, 2020.) Dataset C contained 118 participants, and a total of 37760 trials.

In the studies conducted by Carlebach and Yeung (2020), the display continued to be presented and updated for a short period following a response. For dataset C we do not analyse data from beyond the response time because, at this point, the distributions describing the evidence presented for the two alternatives changed in a complex manner, that depended on the experiment, accuracy, and a previous choice made by the participant. Note that, because of this, it is unclear to us what the policy of a Bayesian observer would be, with regard to converting their evidence measurements into a confidence report. However, this dataset remains useful for examining whether the over-weighting of decision-congruent information in confidence is a general phenomenon, or whether it applies to very specific stimulus sets, as would be expected if confidence is Bayesian.

Testing the effect of congruent and incongruent evidence. Following Zylberberg et al. (2012) we analysed fluctuations in evidence values around their mean, not raw evidence values. That is, for each trial, array, and stimulus frame, we subtracted from the number of dots presented, the mean of the distribution from which the number of dots was sampled. We use "stimulus frame" to refer to the periods during which the stimuli were constant, before the number of dots in the two arrays was again resampled. In all datasets, stimulus frames were 50ms long. The advantage of analysing evidence fluctuations is that they will not correlate

with accuracy, nor with evidence fluctuations at other time points. The evidence fluctuations will also not be affected by trial-by-trial changes in the mid-point of the means from which evidence is sampled, as in dataset B.

We examined the effect of evidence fluctuations on response, response time, and confidence on a frame-by-frame basis (Resulaj et al., 2009; Zylberberg et al., 2012). We considered the effect of evidence both at frames relative to stimulus onset, and at frames relative to response time. Separately for each participant, we looked at the evidence (i.e. dots) presented in the chosen (decision-congruent) and unchosen (decision-incongruent) array in a particular frame, and correlated this with the outcome variable being considered (response, response time or confidence). We then computed the mean value of this correlation coefficient across participants. This generates a time series describing the average correlation between evidence in each frame, and the outcome variable.

Specific correlations were performed for each outcome variable. Within each experiment, one of the arrays was labelled as item 1, and one as item 2, with corresponding responses, 1 and 2. When response was the outcome variable, we computed the point-biserial correlation coefficient between the response given and the evidence fluctuations, signed so that a more positive value indicated more evidence for item 2. When confidence or response time was the outcome variable, we computed the Kendall rank correlation coefficient between the outcome variable and evidence fluctuations, signed so that a more positive value indicated more dots were presented in the array currently under consideration (chosen or unchosen). For the case of confidence and response times, following the computation of correlations, we signed them so that they indicate the effect of evidence fluctuations in the direction which favours the chosen option. Hence, a positive correlation between decision-incongruent evidence and confidence means that, for evidence presented in the unchosen array, evidence favouring the choice made (i.e. below average numbers of dots) was related to higher confidence.

We only computed the correlation for a participant and frame if there were more than 10 relevant trials. For each correlation time-series considered, we excluded frames for which there was not a measure of the correlation for all participants.

We used a permutation-based approach (Ross, 2014) to test for significant differences between the two correlation time series for the effect of congruent and incongruent evidence. Specifically, we repeatedly shuffled the data to estimate the distribution over differences between the two series, under the null hypothesis of there being no difference. In each of 10,000 repetitions, for each participant, we randomly shuffled the labels for the congruent and incongruent correlation time series, before recomputing the across-participant mean difference between the congruent and incongruent series. After all repetitions, at each time point, we used percentiles of the 10,000 sampled points to determine what difference between the congruent and incongruent series would justify rejecting the null hypothesis of no difference between the series. Looking at the 97.5 percentile, and the 0.025 percentile corresponds to a 5% type I error rate. However, we perform a comparison for each stimulus frame. We used the Bonferroni correction for multiple comparisons, taking into account the number of frames in the time series being compared. This process was repeated for each pair of congruent and incongruent time series.

Testing the effect of evidence midpoint. For plots of the effect of evidence midpoint, the plotting procedure was the same as in Chapters 2 and 3, and as described in Appendix A.4. Binned confidence again refers to confidence divided into 4 bins on a participant-by-participant basis. However, as in many analyses we only used trials from one condition of dataset B (the interrogation condition), confidence was binned separately for the free response and interrogation conditions in this dataset. Where needed, we compute the one-sample variant of the Cohen's d effect size measure (Cohen, 1988),

$$d = \frac{\mu}{\sigma},$$

using the mean of the values being compared to zero, μ, and the estimated population standard deviation σ.

Plots. Unless noted, error bars and error shading correspond to $\pm$ 1 standard error of the mean.

5.2.2 Modelling methods

Simulations. Where simulations were used, we simulated stimuli together with the full evidence accumulation process, from the onset of the trial to the end of stimulus processing. Simulated stimuli had properties closely matched to those used in dataset B. See Appendix C.2 for details.

Three computational models. We consider three computational models to explore the effect of evidence midpoint on accuracy, confidence and response time. To bypass the complexities associated with the effects of decision thresholds (Section 4.1), we only model interrogation condition data in which the observer's response time was cued. In this case the observer does not need to apply a decision threshold, because the time at which they make a response is determined by the researcher (Bogacz et al., 2006).

As in the DDM (Ratcliff & McKoon, 2008), in all three models we assume that the observer simply sums all evidence measurements for both options, and takes the difference to compute the balance of evidence (i.e. the difference in evidence between the two options; Bogacz et al., 2006; Vickers and Packer, 1982). The observer then responds on this basis, although there is some chance of a lapse, which leads to a random response. In all models considered, the observer's confidence is solely determined by the balance of evidence, in the direction of the choice made (Vickers & Packer, 1982). As in previous chapters, we assume that the observer's evidence measurements at a time point are determined by the number of array dots in the stimulus frame currently being processed (Smith, 2000). These

measurements are noisy and are corrupted by a constant level of "accumulator noise", σ_{acc}, which affects each measurement that is added to the accumulator tracking the evidence (Bogacz et al., 2006). The evidence accumulation continues until all stimulus information has been processed. Apart from model 3, we do not consider drift-rate variability (Chapter 2; Ratcliff et al., 1999).

Up to this point, we have described the base model (model 1; Table 5.1). The other models can be viewed as extensions of this model. Model 1 has two free parameters, the level of accumulator noise, and the lapse rate. Note that model 1 does not include any form of noise that scales with the stimulus. Model 2 retains the features of model 1, but includes an additional source of noise (Table 5.1). This additional source of noise, σ_s, behaves similarly to accumulator noise, except that it scales with stimulus intensity (Pardo-Vazquez et al., 2019; Teodorescu et al., 2016). Specifically, the noise generated from this source, when measuring evidence in a stimulus array, is proportional to the number of dots presented in that stimulus array. Model 3 includes noise that scales with the stimulus through a different mechanism (Table 5.1). Unlike the other models, model 3 includes drift-rate variability (σ_φ). This is operationalised in the same way as in previous chapters, as "standardised drift rate", which has a multiplicative effect on the strength of the signal from the stimulus. Unlike in previous chapters, we consider drift-rate variability that independently affects the encoding of the two arrays. Hence, in a single trial, standardised drift-rate could be high for the processing of information from one array, but low for the other. As discussed in the introduction, this leads to increasing noise with increasing average evidence.

For the mathematical details of the models, see Appendix C.1

5.2.3 Model fitting

Separately for each participant, and model, we fit the relevant free parameters (Table 5.1). We fit the models using maximum likelihood fitting. In Appendix C.1 we derive expressions for the probability of each response. Using these expressions,

Model number		1	2	3
Constant accumulator noise	σ_{acc}	✓	✓	✓
Intensity dependent noise	σ_s	-	✓	-
Ind. drift rate variability	σ_φ	-	-	✓
Lapse rate	λ	✓	✓	✓
Total parameters		2	3	3

Table 5.1: Free parameters in the three models considered.

we can calculate the likelihood for any set of parameter values, θ. The likelihood is given by the probability of the responses made, given the stimuli presented, according to the model with parameters set to θ. We assume the probability of responses on each trial are conditionally independent given the stimuli presented, and the parameters. As a result we can compute the likelihood as the product of the probabilities of all the responses made by a participant,

$$L(\theta) = p(R^{(1)}, R^{(2)}, ...|\theta, \mathbf{E}^{(1)}, \mathbf{E}^{(2)}, ...) \tag{5.1}$$

$$= \prod_i p(R^{(i)}|\theta, \mathbf{E}^{(i)})\,. \tag{5.2}$$

Here $R^{(i)}$ and $\mathbf{E}^{(i)}$ are, respectively, the participant's response and the stimulus presented, on the ith trial.

Maximum likelihood fitting involved two stages. First, the likelihood was evaluated at 200 randomly drawn sets of parameters (for details on how the parameters were drawn see Appendix C.6). The parameter set with the highest likelihood was then used as the start point for the MATLAB fmincon optimiser ('Matlab Optimization Toolbox', 2017). For details on the limits applied to parameters during optimisation, see Appendix C.6. For each participant and model, this fitting procedure was repeated 16 times. Restarting the optimiser multiple times allowed us to perform a heuristic assessment of the optimiser's performance in this context (Acerbi et al., 2018). This assessment suggested the

optimiser was performing very well, and didn't highlight any difficulties with local maxima of the log likelihood function (Appendix C.5).

5.2.4 Model comparison

We compared the models by evaluating the Akaike information criterion (AIC), and the Bayesian information criterion (BIC). These provide measures of how well a model fits the data, while penalising the extra flexibility that additional parameters can provide. We computed the mean AIC and BIC across participants, to find the best fitting model according to the AIC, and according to the BIC. On a participant-by-participant basis we computed the difference in AIC and BIC between each model and the best fitting model. This allowed us to compute the mean difference between the best fitting model and the other models, along with 95% confidence intervals (found by bootstrapping 10000 times).

We visually inspected model fit by simulating data using the participant-by-participant fitted parameter values (Appendix C.2).

5.3 Results

5.3.1 Congruent-evidence over-weighting

We first looked at the relative weighting of decision-congruent and decision-incongruent evidence on responses and confidence. We were particularly interested in the two datasets where an over-weighting of decision-congruent evidence would clearly contradict the idea that confidence reflects a Bayesian readout (dataset A and dataset B). We additionally looked at the effect of evidence on response and response time. In cases where confidence has reflected an over-weighting of decision-congruent information, responses have reflected an equal weighting of congruent and incongruent information (Peters et al., 2017; Zylberberg et al., 2012). If the over-weighting is specific to confidence, we would also not expect to see any

differential weighting of congruent and incongruent evidence on response time, which is a product of the decision-making mechanism.

While not the main analysis we first looked to see if responses reflected decision-congruent and incongruent evidence equally. We determined whether there were significant differences between the effects of congruent and incongruent evidence using the procedure described in Section 5.2. Time points at which there was a significant difference are indicated with grey shading in the plots. Similar patterns were observed for the three cases in which participants had a degree of control over the duration of stimulus presentation: Dataset A, dataset B in the free condition, and dataset C (Fig. 5.3). In the experiments for dataset A and C, participants could respond when they wanted, although relatively short deadlines were imposed, within which participants were required to respond. Nevertheless, the weighting of evidence appeared qualitatively similar in the three cases, with correlation coefficients for the effect of evidence on response larger early in a trial, when looking relative to trial onset (Fig. 5.3, A, B, D), and peaking prior to response when looking relative to response (note that dataset A contains frames following a response, so the response time is shifted along the x-axis relative to other plots; Fig. 5.3, E, F, H). Although evidence at different time points was not weighted equally, the patterns were consistent for decision-congruent and incongruent evidence such that congruent and incongruent evidence appeared to be weighted equally at each individual time point in a trial. This was also true in the interrogation condition of dataset B (Fig. 5.3, C, G). Only for a couple of time points across the datasets was a significant difference found (shown with grey shading). Note that while we corrected for multiple comparisons within each pair of congruent-incongruent correlation time series, the chance of a false positive across all analyses would have been above 5%.

We next turned to the main analysis, the effect of congruent and incongruent evidence on confidence. Apart from in the interrogation condition of dataset B, evidence correlation coefficients appeared to increase up to the time of response, when looking relative to response time (Fig. 5.4, E, F, H). Relative to stimulus onset,

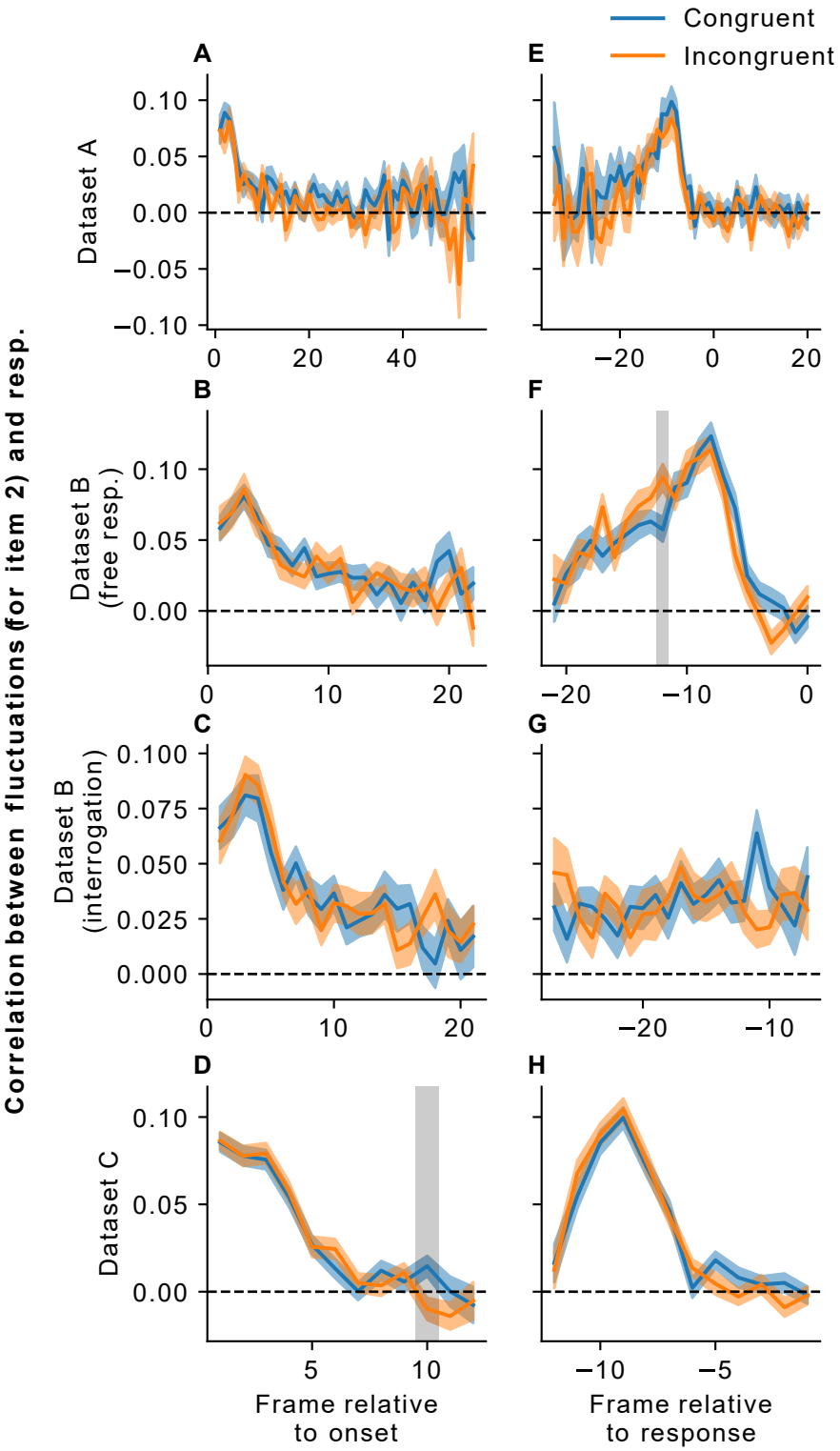

Figure 5.3: The relationship between decision-congruent evidence, decision-incongruent evidence, and the response given, in the three datasets (interrogation condition and free response condition from dataset B plotted separately). Grey shading indicates a frame in which there was a significant difference between the two series.

different qualitative patterns were present in different datasets (Fig. 5.4, A, B, C, D). Crucially, one qualitative feature was consistent over all the datasets considered: The weighting of decision-congruent and decision-incongruent evidence in confidence was similar across all time points. While the weighting was similar, there was a numerical trend in the data, such that correlation coefficients for the effect of evidence on confidence were greater for congruent than incongruent evidence. This trend is visible in Fig. 5.4, B, D, F, H. Indeed, for a period of time approximately 8 frames prior to response this difference is significant in dataset C (Fig. 5.4, H).

For completeness we looked at the effect of decision-congruent and incongruent evidence on response time. For datasets in which participants had a degree of control over the duration of stimulus presentation, there was a distinctive negative peak prior to response in the coefficients for the effect of evidence on response time (Fig. 5.5, E, F, H). Correlation coefficients for the effect of evidence on response time were particularly large and negative at the onset of the stimulus, suggesting more early evidence was related to faster response times (Fig. 5.5, A, B, D). The qualitative patterns were different when response time was determined by the researcher (interrogation condition, dataset B; Fig. 5.5, C, G). However, in all cases, decision-congruent and incongruent evidence appeared to show a similar relationship to response time throughout a trial.

In summary, the results suggest that decision-congruent and incongruent evidence is weighted approximately equally in both perceptual decisions, and in confidence. There was some evidence for a slightly greater weighting of decision-congruent evidence in confidence reports. It is interesting to note that this effect appeared strongest in dataset C, for which we are unclear on the policy of a Bayesian observer. Dataset C also contained the most trials, and would therefore have the most power to detect an effect. Hence, no strong conclusions can be drawn regarding the cause of this slight over-weighting.

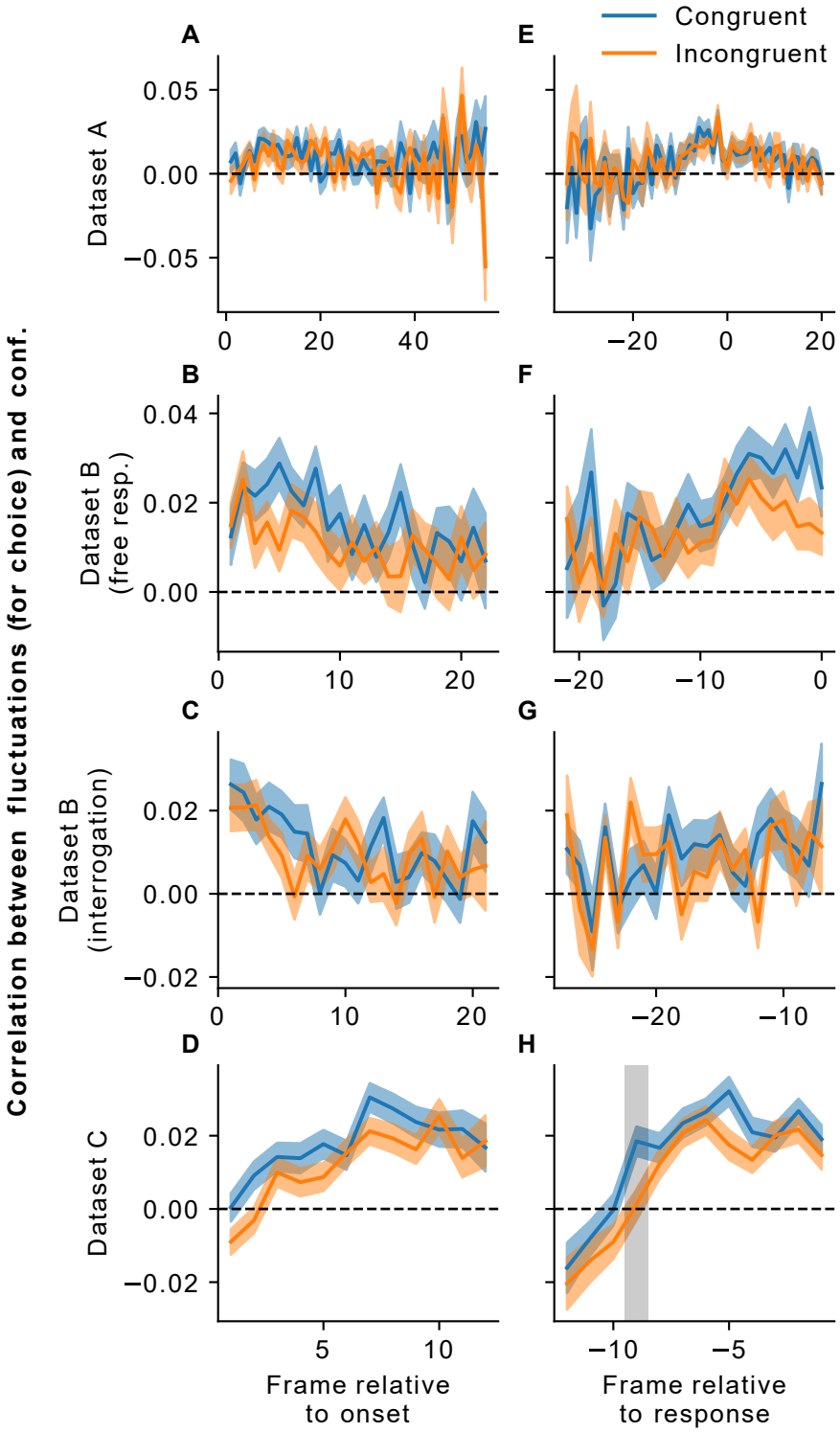

Figure 5.4: The relationship between decision-congruent evidence, decision-incongruent evidence, and confidence in the three datasets (interrogation condition and free response condition from dataset B plotted separately). Grey shading indicates a frame in which there was a significant difference between the two series.

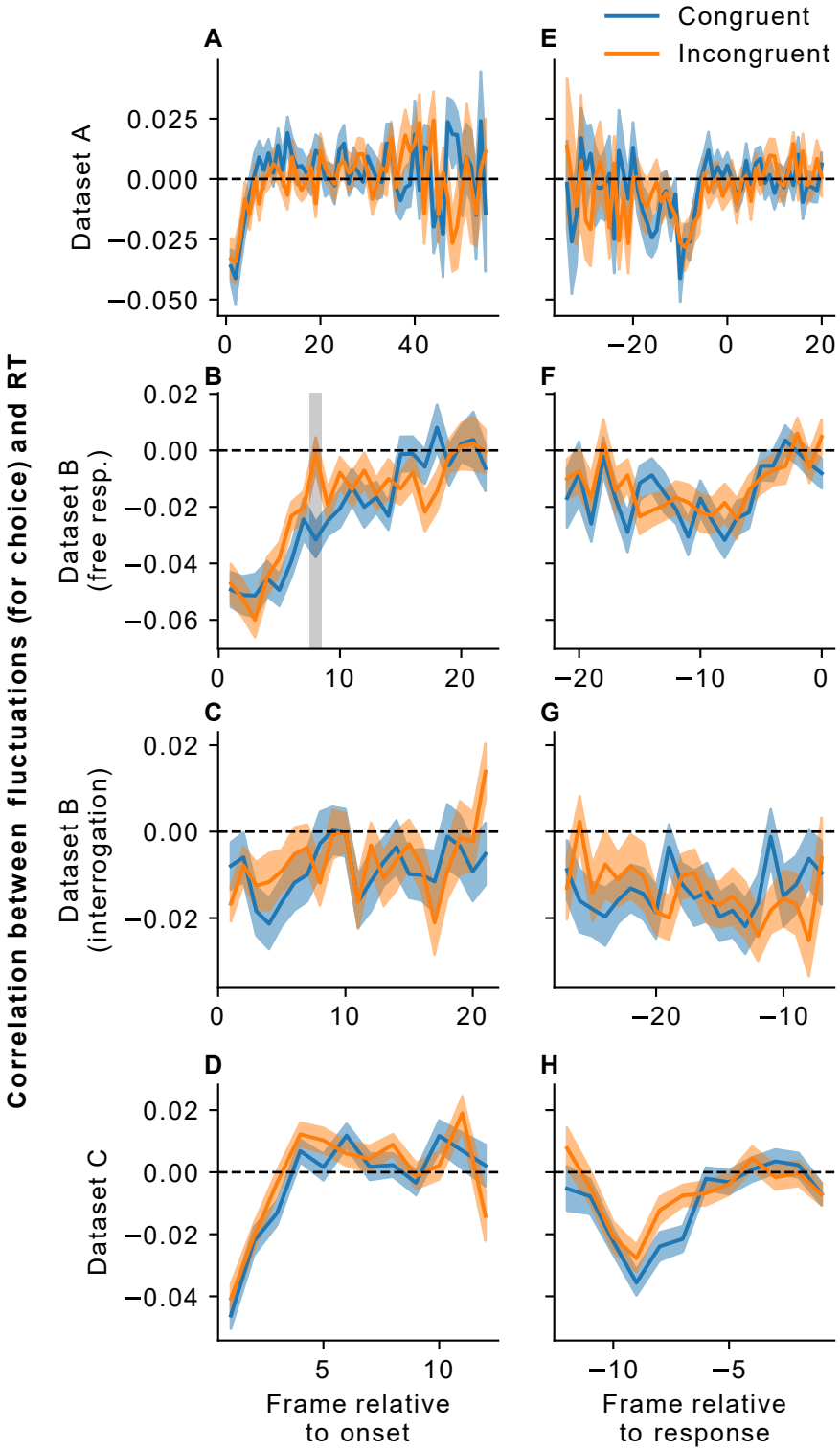

Figure 5.5: The relationship between decision-congruent evidence, decision-incongruent evidence, and response time in the three datasets (interrogation condition and free response condition from dataset B plotted separately). Grey shading indicates a frame in which there was a significant difference between the two series.

5.3.2 Positive evidence effect

Having found little evidence for a strong over-weighting of decision-congruent evidence when using simple face-valid measures of evidence, in contexts where Bayesian confidence should reflect the balance of evidence, we turn to the second phenomenon discussed in Section 5.1. It has been reported that if evidence for both options is boosted in a certain way, then confidence can be increased without a corresponding increase in performance (Koizumi et al., 2015; Rollwage et al., 2020; Samaha et al., 2016). This "positive evidence effect" has been attributed to the hypothesised over-weighting of decision-congruent evidence. However, having argued that this over-weighting is not a general phenomenon, and is only observed in situations where a Bayesian observer would over-weight decision-congruent evidence, we must provide an alternate explanation of these effects.

In dataset B (collected for Chapter 3), we varied the mean level of evidence in the two boxes on a trial-to-trial basis. More precisely, we varied the midpoint between the means of the two evidence distributions (we refer to this quantity as the evidence midpoint). The number of dots in the correct and incorrect array were sampled from these two evidence distributions. Hence, a higher evidence midpoint corresponds to boosting the evidence for both alternatives by an equal amount.

We first looked at what effect simultaneously increasing and decreasing the evidence midpoint had in dataset B. Accuracy decreased with evidence midpoint in both the free response ($t(47) = -2.8$, $p = 0.0078$, $d = -0.40$) and interrogation conditions (Fig. 5.6, A; $t(47) = -4.0$, $p = 0.000\,23$, $d = -0.58$). In contrast, confidence increased in both the free response ($t(47) = 3.9$, $p = 0.000\,31$, $d = 0.56$) and interrogation conditions (Fig. 5.6, B; $t(47) = 3.1$, $p = 0.0037$, $d = 0.44$). In the interrogation condition, response time is set by the experimenter, and independently of evidence midpoint. In the free response condition, response time decreased with evidence midpoint (Fig. 5.6, C; $t(47) = -3.3$, $p = 0.0018$, $d = -0.48$).

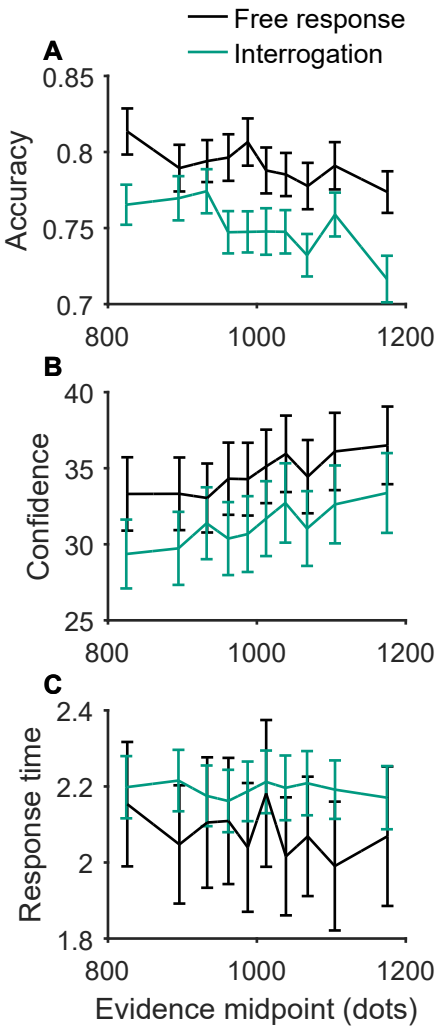

Figure 5.6: The effect of evidence midpoint on accuracy, confidence, and response time in dataset B.

These effects cannot be explained by standard DDM variants (like those set out in Chapter 2) because, according to the DDM, only the difference in evidence is tracked by the observer, and the absolute level of evidence has no effect on the evidence accumulation mechanism. If we held that decisions were determined by the balance of evidence, but that confidence was primarily informed by decision congruent evidence (Peters et al., 2017), we could explain why confidence increases as more evidence is presented for both options (Rollwage et al., 2020). However, note, this account cannot explain why simultaneously increasing congruent and incongruent evidence has effects on accuracy and response time.

We may be able to account for these effects using DDM variants that feature noise that scales with the average evidence (Section 5.1; Teodorescu et al., 2016). We consider three DDM models (Section 5.2, Table 5.1). Model 1 is a standard DDM, model 2 features noise that scales with stimulus intensity, while model 3 features drift-rate variability that independently affects the two arrays presented on a trial. In this section we perform modelling only using data from the interrogation condition of dataset B. This condition allows for relatively simple computational modelling, as we do not need to take into account decision thresholds. We verified that the three models were in principle distinguishable, by simulating data using each of the three models, and then fitting each of the three models to each simulated dataset. In each case, the true data generating model was correctly recovered (Appendix C.3 for model recovery; Appendix C.2 for simulation details).

We fit all three models to the data, and compared the fits using the AIC and BIC (Fig. 5.7; unlike elsewhere, error bars represent 95% confidence intervals). Model 2 fit reliably worse according to both the AIC and BIC. According to the BIC, which more heavily penalises model complexity, model 1 fit reliably better. However, according to the AIC, model 3 fit best. The difference in fit between model 3 and model 1, according to the AIC, was not significant. To gain further insight into these results we performed another round of model recovery. We simulated datasets of the same size as the real dataset, using the fitted parameter values for each participant.

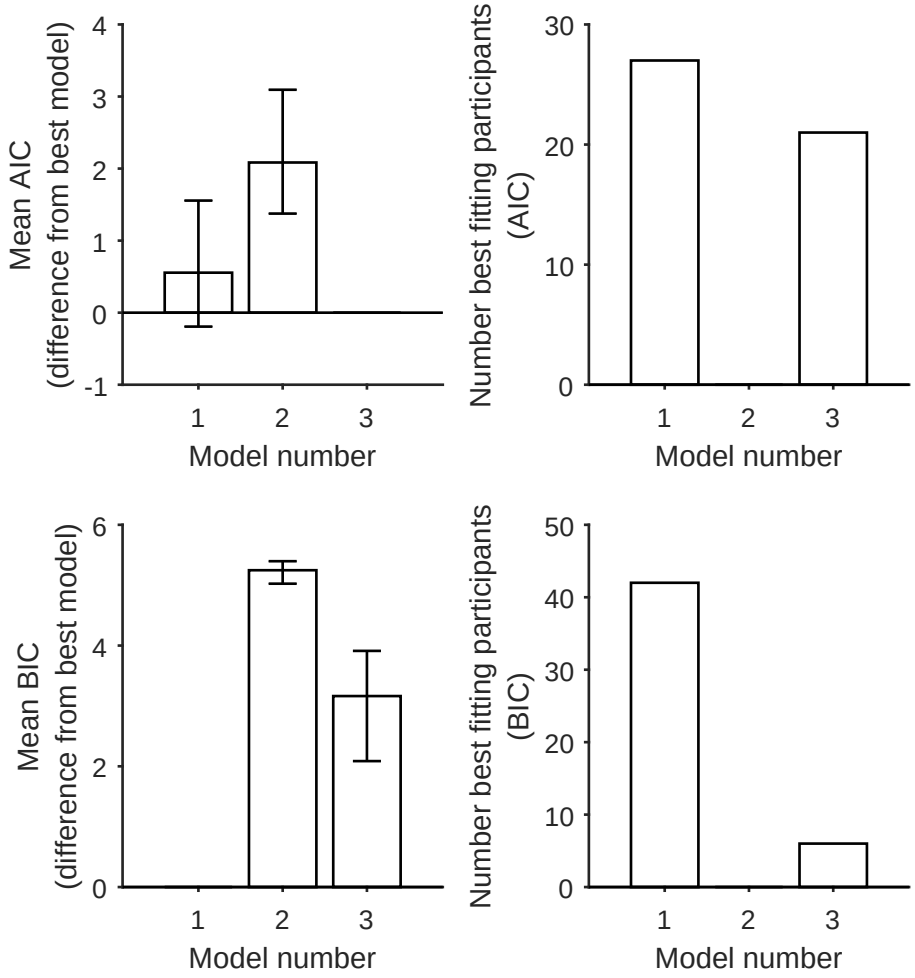

Figure 5.7: AIC and BIC relative to the overall best fitting model (according to the AIC or BIC), and number of participants for which each model provided the best fit. Unlike in other plots, error bars represent 95% bootstrapped confidence intervals.

We then fit all three models to each simulated dataset. According to both the AIC and BIC, the model with fewest parameters, model 1, fit best regardless of the model used to generate the data (Appendix C.4 for model recovery; Appendix C.2 for simulation details). This suggests that we may not have enough data to successfully distinguish the models, when these are given plausible parameter values.

While the quantitative model comparison was inconclusive, we also wanted to look at whether the models could account for the qualitative patterns we had observed. To do this, we simulated data using the models and the fitted parameters

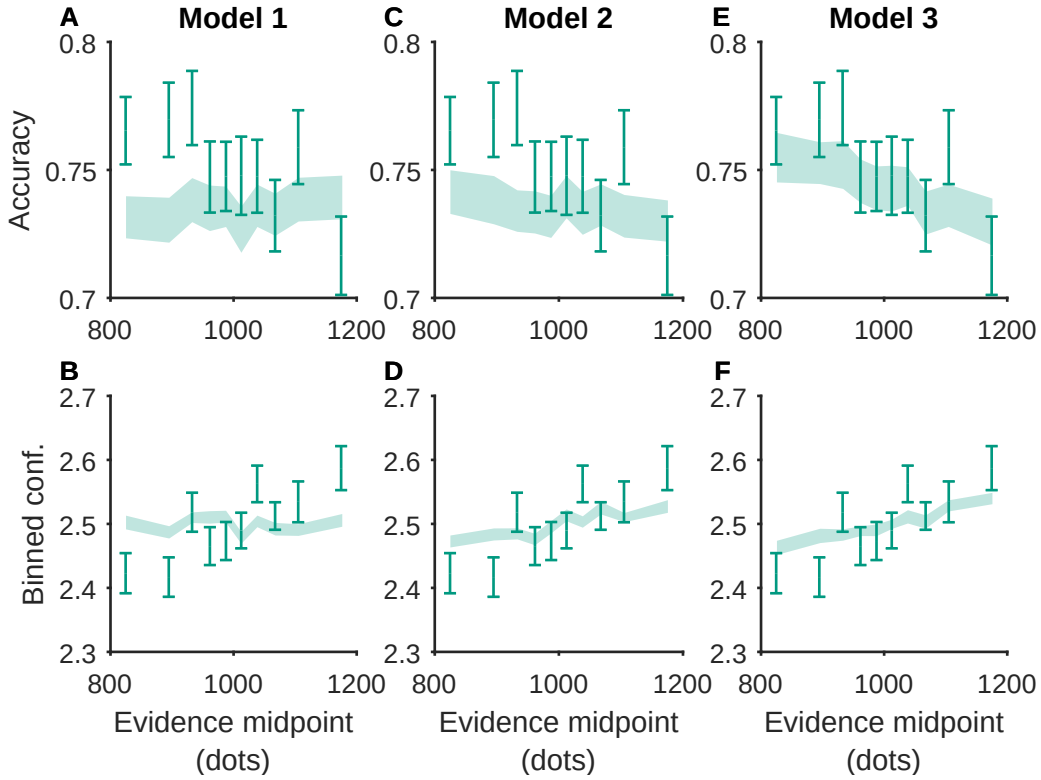

Figure 5.8: The effect of evidence midpoint on accuracy and binned confidence, in interrogation trials from dataset B (error bars), and the fit for model 1, 2, and 3 (shading).

for each participant (Appendix C.2). As expected, a standard DDM without stimulus-dependent noise (model 1) could not account for a change in accuracy ($t(47) = 0.027$, $p = 0.98$, $d = 0.0039$), or binned confidence ($t(47) = 0.60$, $p = 0.55$, $d = 0.087$), with evidence midpoint (Fig. 5.8, A and B).

Model 2 could account for the qualitative patterns of accuracy decreasing with evidence midpoint ($t(47) = -3.1$, $p = 0.0029$, $d = -0.45$), and confidence increasing with evidence midpoint ($t(47) = 4.0$, $p = 0.000\,23$, $d = 0.58$). However, the fits were clearly worse than those for model 3, which appeared to capture the observed patterns well (Fig. 5.8). Confirming the patterns in the figure, model 3 captured the decrease in accuracy with evidence midpoint $t(47) = -7.5$, $p = 1.6 \times 10^{-9}$, $d = -1.1$), and the increase in binned confidence with evidence midpoint ($t(47) = 5.3$, $p = 3.5 \times 10^{-6}$, $d = 0.76$).

Fitted parameter values are provided in Appendix C.7.

5.4 Discussion

Using computational modelling and three previously collected datasets we explored two phenomena which might appear to contradict Bayesian confidence. Namely, congruent-evidence over-weighting, and the positive evidence effect.

5.4.1 Congruent-evidence over-weighting

We examined the effect of decision-congruent and incongruent evidence in three previously-collected datasets, with the aim of testing the claim that over-weighting of decision-congruent information is a general phenomenon in perceptual confidence (Peters et al., 2017). Whereas the stimulus sets used in previous studies may have made it Bayes optimal to over-weight decision-congruent evidence, in two of the three datasets we used, a Bayesian observer would equally weight congruent and incongruent evidence in their confidence reports.

In these two datasets, and in a third dataset that used a related but more complicated stimulus set, we found that decision-congruent and incongruent evidence were approximately equally weighted, over the entire duration of stimulus presentation. There was a trend towards slight over-weighting of decision-congruent information, and evidence that this over-weighting was significant in one dataset. Nevertheless, these results support the view that the strong over-weighting of decision congruent evidence, which has previously been observed (Peters et al., 2017; Zylberberg et al., 2012), may not be a general phenomenon but a consequence of Bayesian confidence and the stimuli used.

If it is real, there are a number of possible explanations for the slight over-weighting of decision-congruent evidence. One explanation is closely related to an idea put forward by Peters et al. (2017), that it is sensible to use decision-congruent evidence for confidence in ecological settings because in such settings there may in fact be no stimulus present. In this case, decision-congruent evidence is crucial because it tells the observer that there is in fact an object present, as oppose to no

object, and hence will be an important determinant of confidence. This situation is very similar to a situation we considered in Section 5.1. In particular, in Fig. 5.2, D, we saw that when there are two conditions, one that is difficult, and one that is easier because there is a strong signal for the correct option, decision-congruent evidence is important for confidence. If the difficult condition was made impossible, by providing no evidence for either option, the importance of decision-congruent information for confidence, would likely increase. In the lab, observers may believe that there is a very small but non-zero chance that no stimulus is present, and instead the researcher has simply presented noise. (This is a prior that highly experienced psychophysics participants may have even learnt from previous studies.)

Although the link made by Peters et al. (2017) to the kind of decisions faced by animals and humans in ecological settings is insightful, we suggest that there is an additional and important effect in the data collected by Peters et al. (2017), and the data collected by Zylberberg et al. (2012), which may explain why these studies found a strong over-weighting of confidence. As discussed in Section 5.1, in these studies evidence for the correct option was varied, and it may have been Bayes optimal to strongly over-weight decision congruent evidence in confidence. (Although we concede we have not demonstrated this, and the policy of a Bayesian observer may depend highly on parameter values such as the observer's perceptual noise.) A limitation of our study is that we did not demonstrate both an equal weighting where it is Bayesian to do so, and an over-weighting of congruent evidence when this is Bayesian. We only demonstrated the former. The idea of Bayesian confidence leads to the strong prediction that the very same type of stimuli could generate both patterns, through manipulation of the stimulus sets used, a prediction that would be very feasible to test in future research.

5.4.2 Positive evidence effect

In the second half of this chapter we explored possible explanations for the positive evidence effect, the finding that boosting evidence for both alternatives can lead to

increased performance without changes in accuracy (Koizumi et al., 2015; Rollwage et al., 2020; Samaha et al., 2016). This phenomenon has previously been explained with reference to the idea of a general over-weighting of decision-congruent evidence (Koizumi et al., 2015; Rollwage et al., 2020). In a dataset where evidence for both options was simultaneously increased or decreased from trial-to-trial, we found effects of the evidence midpoint. Consistent with previous findings (Rollwage et al., 2020; Teodorescu et al., 2016), increasing evidence for both options increased confidence, but decreased accuracy and response times. A standard DDM, which ascribes relevance only to the difference in evidence, cannot explain these effects (Teodorescu et al., 2016). However, the idea that decision-congruent evidence is over-weighted in confidence but not decisions, does not account for the observed effects on accuracy and response times. A different explanation is needed. Teodorescu et al. (2016) found that the DDM can account well for the effects of evidence midpoint on responses and response times when the model is adapted to include noise that depends on input to the evidence accumulator. A model in which drift-rate variability independently affected the two stimuli (model 3), generated noise which scaled with evidence midpoint, and could capture the qualitative effects of evidence midpoint on accuracy and confidence. However, quantitative model comparison was ambiguous about whether this model provided a better fit than a standard DDM (model 1).

One puzzle in these results is why we found, as others have found, that evidence midpoint decreased performance (Teodorescu et al., 2016), when it has also been reported that confidence can be increased without changes in performance (Koizumi et al., 2015; Rollwage et al., 2020; Samaha et al., 2016). The answer may be methodological. We simultaneously increased or decreased evidence for both alternatives, and by the same amount, whereas other studies have used different procedures. For example, Rollwage et al. (2020) boosted evidence for both alternatives by first increasing evidence for the incorrect option, and then staircasing the increase in evidence in the correct alternative, so that a desired level of performance was met. Hence the apparent conflict may simply be due to a methodological difference.

There are a number of limitations to the modelling work we undertook. Firstly, for computational simplicity we limited modelling work to interrogation condition trials. It will be both important and potentially beneficial to include free response condition data. Both stimulus dependent noise (Teodorescu et al., 2016), and other sources of variability (Zylberberg et al., 2016), have been found to decrease response times, but it will be important to see if models can simultaneously capture patterns in response times and confidence. Including free response data may also generate a stronger test for the models, and hence lead to a more conclusive quantitative model comparison.

Another important limitation is that we included noise that scales linearly with stimulus intensity in one of our models, but this may be an oversimplification: Internally represented intensity may differ from stimulus intensity (Geisler, 1989; Pardo-Vazquez et al., 2019). Where evidence has been found for noise which scales with intensity, evidence has also been found for a non-linear transformation of physical intensity into represented intensity (Pardo-Vazquez et al., 2019; Teodorescu et al., 2016). It is then represented intensity which determines the level of noise. This may explain why a model in which one source of noise scales directly with stimulus intensity (model 2), performed poorly here. In previous modelling work, the inclusion of a source of noise which scales linearly with transformed stimulus intensity has proved very successful (Pardo-Vazquez et al., 2019; Teodorescu et al., 2016), but exploration of these detailed properties was beyond the scope of the present work.

A fascinating and open question is what effect such nonlinear stimulus transformation would have on Bayesian confidence. This question is related to another limitation of the modelling work in this study: For modelling the effect of stimulus-dependent noise we did not use Bayesian confidence, but instead simply assumed that observers report the balance of evidence. This was a reasonable choice when investigating whether over-weighting of decision-congruent evidence is the explanation for increased confidence with increasing evidence midpoint. However, our hypothesis is that confidence reflects a Bayesian readout, not the balance of

evidence. Therefore, it will be important in future to investigate what effect stimulus dependent noise has on a Bayesian observer's confidence.

Tentatively, we suggest that a Bayesian readout for confidence will lead to a weaker dependence of confidence on evidence midpoint, but will not eliminate it. Our reasoning is that a Bayesian observer will take into account the fact that noise increases with evidence midpoint, and will compensate for this effect. However, a Bayesian observer will never know the true value of evidence midpoint on a trial, and their estimate will be shifted toward the mean of the prior over evidence midpoints (Tassinari et al., 2006). Hence, when evidence midpoint is high, the Bayesian observer will underestimate it, but when it is low, the Bayesian observer will overestimate it. As a result, a Bayesian observer will not fully compensate for the effect of stimulus-dependent noise. This pattern is similar in some ways to one explanation for the hard-easy effect. The hard-easy effect refers to the finding that people tend to be over-confident on easy trials, but under-confident on difficult trials (Drugowitsch et al., 2014; Lichtenstein & Fischhoff, 1977). This pattern can be explained from a Bayesian point of view, because the Bayesian observer will try to estimate the difficulty of the task, but their estimates will be shifted towards the mean of the prior over difficulty (Drugowitsch et al., 2014). Hence they will estimate hard trials to only be moderately-hard, and the reverse for easy trials.

We have seen that the over-weighting of decision-congruent evidence is not a general perceptual phenomenon. Instead, we have observed cases where it does not apply, and should not, if observers are Bayesian. We have also seen that stimulus-dependent noise may be able to explain why confidence increases when evidence values for both options are simultaneously increased. Taken together, these results show that effects which might at first sight appear to contradict Bayes, are in fact consistent with a Bayesian readout for confidence, or may be consistent with plausible additional mechanisms regarding encoding noise. We have seen no strong reason to abandon the claim that confidence is Bayesian in favour of the idea that

it is built from heuristic computations. Nevertheless, we will return to this point, and discuss further challenges to the idea of Bayesian confidence in Chapter 7.

6

DDM in settings with greater ecological validity

Contents

6.1 Introduction

In previous chapters we explored whether qualitative and quantitative patterns in confidence reports can be accounted for from within the normative framework of the DDM, by extending the DDM with a miscalibrated Bayesian readout for confidence. In Chapter 5 we examined challenges to the idea that confidence reflects a Bayesian

readout. Here we turn to focus on perceptual decisions themselves, and look at a challenge regarding our characterisation of the optimality of the DDM.

The DDM describes a "stopping rule", a rule for terminating observation, and taking an action (Ferguson, 2011). The DDM also prescribes the action that should be taken when observation is stopped. As discussed in Chapter 1, the DDM with time-dependent thresholds can be viewed as a version of the stopping rule which maximises reward rate when faced with multiple alternatives, "scaled down" for the case of only 2 alternatives. In Section 1.5 we considered challenges to the claim that the DDM is optimal in any relevant sense. In particular, we noted that in naturalistic settings the maximum reward rate may never be obtained, because of the complexity of the optimal stopping rule, and because the "tasks" faced by animals and humans are constantly changing. Instead, what may be important is rapid learning in new situations.

Tajima et al. (2019) provided a result which may have important implications when considering how animals and humans learn to maximise reward rate. Tajima et al. (2019) showed that approximately optimal stopping rules may be considerably simpler than optimal stopping rules. Consider the problem that animals and humans face: In the standard framework that we have used throughout this thesis, when deciding between N options each of the N options generate a noisy signal in the brain representing the evidence for that option (Bogacz et al., 2006; Usher & McClelland, 2001). The observer aims to determine which evidence signal has the higher mean. The optimal policy involves accumulating the evidence for each of the options (Tajima et al., 2019). In the N dimensional space defined by the N evidence accumulators, the optimal time to make a decision is when the point in this space representing the N accumulators reaches a time-dependent curved bound (Tajima et al., 2019). Importantly, Tajima et al. (2019) showed that if the state of the accumulators is mapped onto a curved manifold, approximately optimal stopping can be achieved by stopping when the value of the maximum

mapped accumulator crosses a threshold (a procedure which is very similar to the race model; Bogacz et al., 2006, Section 1.2.2).

The finding that approximately optimal stopping rules may be far simpler than optimal stopping rules suggests that animals and humans may be able to simplify the learning problem they face, while compromising little in terms of reward rate. An approximation that reduced the number of relevant variables would clearly simplify the learning problem. Instead of needing to learn a mapping from time and the state of all evidence accumulators, as would be required for true optimality, the observer only needs to learn a mapping from a reduced number of variables to waiting vs. stopping (Fig. 6.1). (One detail is that the dimensionality of the problem can be reduced by 1 without any loss in performance because the sum of all evidence accumulators is irrelevant in the optimal rule; Tajima et al., 2019.) If animals and humans learn the parameters of this mapping through associative learning, the use of an approximate stopping rule which reduces the number of relevant variables may speed up learning, leading to greater reward rate even before reward rate has been maximised.

We take the idea of an approximately optimal stopping rule, and consider a rule that reduces the number of variables used by the observer for deciding between stopping vs. waiting. Specifically we consider the possibility that observers use a criterion on time and the maximum posterior probability (Fig. 6.1). (Note the similarity between this stopping rule and the multihypothesis sequential probability ratio test discussed in Section 1.5.1; Baum and Veeravalli, 1994; Bogacz and Gurney, 2007.) Importantly, this approximately optimal stopping rule reduces to the DDM with time-dependent thresholds in the situations considered in previous chapters (e.g. two alternatives, uncorrelated evidence streams of equal variance). We have seen that, for the contexts considered in previous chapters, at any specific time point during observation, the log-posterior ratio between the two alternatives is proportional to the accumulated difference in evidence (see Eq. 4.26 in Chapter 4). Hence, a criterion that is a function of time and posterior probability can be

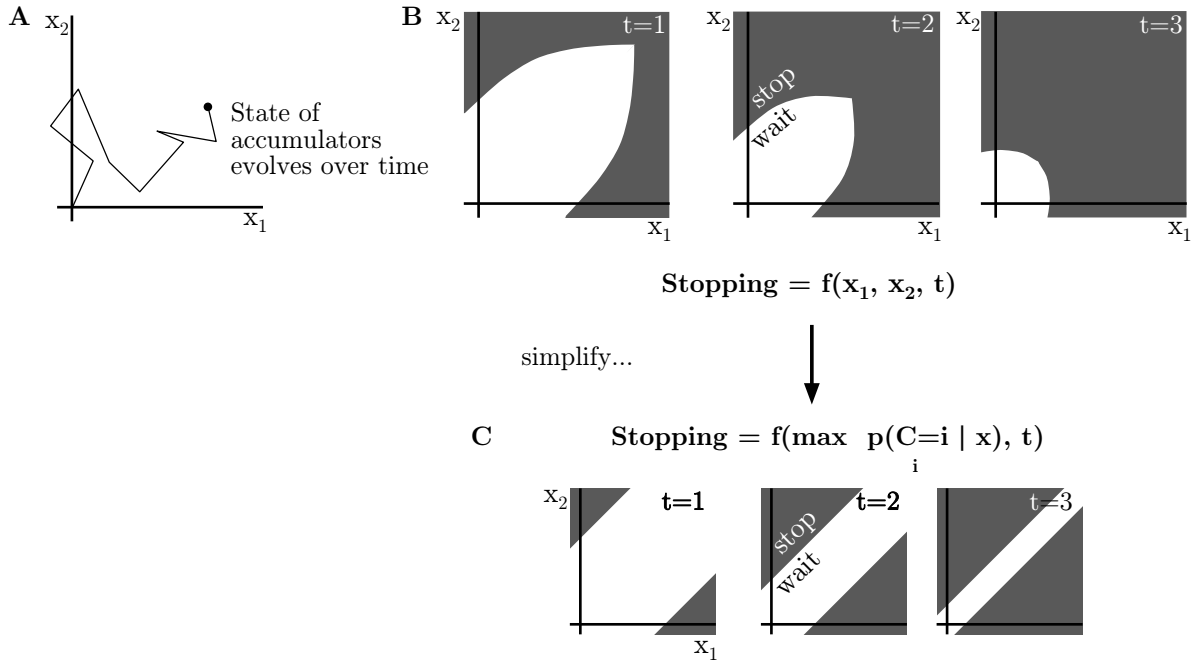

Figure 6.1: Animals may use N accumulators to track the cumulative evidence for the N choices. The case of $N = 2$ is shown here. (A) The state of the accumulators can be represented as a point in N dimensional space that moves over time. Each dimension represents the value of one accumulator x_i. (B) Observers could learn a criterion on time, t, and the state of the N accumulators, x_i, to waiting vs. stopping. The criterion shown is a function of both evidence accumulators and decreases with time. (C) Alternatively, observers could simplify the problem by learning a mapping from a smaller number of variables. For example, the observer could use a time-dependent criterion on time and the posterior probability of the most probable option.

converted to a criterion that is a function of time and accumulated difference in evidence (Moran, 2015). The accumulated difference in evidence is precisely the variable that is tracked in the DDM. This line of reasoning implies that the DDM is a special case of a policy in which time and maximum posterior probability are used to decide between stopping vs. waiting.

Considering a generic perceptual decision making task we show that, under a wide range of conditions, a criterion on time and the maximum posterior probability is sufficient for approximately optimal stopping. Using neural networks as model learners, we find that this approximation facilitates rapid learning. Taken together, these results suggest that a stopping rule which reduces to the DDM leads to higher reward rate, even if reward rate is never optimised. Thus, the DDM may be normative even in ecological settings where the "task" is constantly changing.

In Section 6.2 we discuss in detail the mathematical formalism. Specifically, we outline how one may approximate optimal stopping using a criterion on time and maximum posterior probability. Apart from the first paragraph, and equations 6.1 and 6.2, Section 6.2 may be safely skipped by a reader uninterested in the derivation of this result. In Section 6.4 we explore whether this approximation leads to advantages in terms of faster learning, and hence, greater reward.

6.2 Mathematical results

6.2.1 Decision problem and posterior

We consider a more general mathematical framing of perceptual decisions than in previous chapters. We retain the standard idea that, for each option, the brain computes a noisy evidence signal (Bogacz et al., 2006). The goal of an observer is to determine which object or property, C, is the cause of the observed evidence samples $\delta\boldsymbol{x}(1), \delta\boldsymbol{x}(2), ...\boldsymbol{x}(M)$. $\delta\boldsymbol{x}(t)$ denotes a vector containing the values of the evidence signals for each of the N options, at time step t. The prior probability of object $C = i$ is $p(C = i)$, and $C = i$ generates average evidence samples given

by $\boldsymbol{\mu}(i)$. $\boldsymbol{\mu}(i)$ indicates a vector, the same length as the number of possible values of C, where each element is given by

$$\mu_j(i) = \begin{cases} \mu_0 + \mu & \text{if } i = j \\ \mu_0 & \text{otherwise}\,. \end{cases} \tag{6.1}$$

At each time step, the object generates associated evidence samples, $\delta\boldsymbol{x}(t)$, according to

$$p(\delta\boldsymbol{x}(t)|C = i) = N(\delta\boldsymbol{x}(t); \boldsymbol{\mu}(i)\delta T, \Sigma\delta T)\,, \tag{6.2}$$

where t indicates the current time step, δT is the duration of a time step, $N()$ indicates the multivariate normal distribution, and Σ is the covariance of evidence samples (Tajima et al., 2019). In Appendix D.1 we find the posterior over C using Bayes rule to be

$$\begin{aligned} &p(C = i|\delta\boldsymbol{x}(1), \delta\boldsymbol{x}(2), ..., \delta\boldsymbol{x}(M)) \\ &\qquad \propto p(\delta\boldsymbol{x}(1), \delta\boldsymbol{x}(2), ..., \delta\boldsymbol{x}(M)|C = i)p(C = i) &(6.3)\\ &\qquad \propto N\left(\boldsymbol{\mu}(i); \frac{\boldsymbol{x}(t)}{t}, \frac{\Sigma}{t}\right) p(C = i)\,, &(6.4) \end{aligned}$$

where $\boldsymbol{x}(t) = \sum_i \delta\boldsymbol{x}(i)$, $t = \sum_i \delta T$ (Drugowitsch et al., 2012).

We consider only covariance matrices that can be written in the following form

$$\Sigma = a\boldsymbol{I} + b\boldsymbol{J}\,, \tag{6.5}$$

where $\boldsymbol{J}$ is a matrix of ones. While this covariance matrix allows for relationships between different accumulators, the covariance between any two accumulators will be the same. Hence, this covariance matrix does not cover a case in which we believe that some pairs of accumulators will show a stronger covariance than other pairs of accumulators.

In Appendix D.1 we find that in this case the posterior becomes

$$p(C = i|\boldsymbol{x}(t)) = \frac{1}{1 + \sum_{j\neq i} \frac{p(C=j)}{p(C=i)} e^{\frac{\mu(x_j - x_i)}{a}}}\,. \tag{6.6}$$

6.2.2 Optimal stopping

If a Bayesian observer stops, they will make the decision that maximises expected reward (Ma & Jazayeri, 2014). If the reward for all options is the same, they will pick the most likely option, i.e. they will have the response, $\hat{C}$,

$$\hat{C} = \arg\max_i p(C = i|\boldsymbol{x}(t)) \,. \tag{6.7}$$

If the observer receives reward r_c for a correct response and r_e for an incorrect response, then the expected reward from deciding, y, is

$$\begin{aligned} y(\boldsymbol{x}(t)) &= r_c p(C = \hat{C}|\boldsymbol{x}(t)) + r_e(1 - p(C = \hat{C}|\boldsymbol{x}(t))) + L(t) & (6.8)\\ &= (r_c - r_e)p(C = \hat{C}|\boldsymbol{x}(t)) + r_e + L(t) \,, & (6.9) \end{aligned}$$

where $L(t)$ is a function describing the cumulative cost of time. Note that we assume, consistent with previous work on the stopping rules used by animals and humans (Drugowitsch et al., 2012; Tajima et al., 2019), that the observer values temporally distant rewards the same as rewards obtained immediately.

Although we can calculate the reward from deciding, it is more difficult to calculate the expected reward from waiting. Suppose that we have a stopping rule, S, that describes for which values of $\boldsymbol{x}(t)$, and time steps t we will stop and decide. Then the expected reward from following S from time step t onward is

$$E_S[y|\boldsymbol{x}(t)] \,. \tag{6.10}$$

In words, this expression denotes the expected reward, following stopping rule S, given that we are currently at $\boldsymbol{x}(t)$. An optimal observer will only stop and decide now if there is no way of increasing the expected reward (Ferguson, 2011; Øksendal, 2013). That is, they will stop now if there is no stopping rule which, if followed from this point on, would lead to an expected reward that is greater than the reward for stopping now. The optimal observer will therefore only stop if

$$y(\boldsymbol{x}(t)) = \max_S E_S[y|\boldsymbol{x}(t)] \,. \tag{6.11}$$

It is difficult to find the set of points which satisfy this equation because is it difficult to calculate the maximum possible reward under any stopping rule.

One reason why it is difficult to calculate the maximum possible reward under any stopping rule is that the future evolution of $\boldsymbol{x}(t)$ is unknown, but described by

$$p(\delta\boldsymbol{x}|\boldsymbol{x}(t)) = \sum_j p(\delta\boldsymbol{x}|C=j)p(C=j|\boldsymbol{x}(t))\,, \tag{6.12}$$

where $\delta\boldsymbol{x}$ is the change in $\boldsymbol{x}$ at any future time step. Hence, we have to consider the expectation over all possible trajectories, and the time and place at which these paths are terminated according to the stopping rule, to determine the expected reward under that stopping rule.

One way to solve this problem is to work backwards (Drugowitsch et al., 2012), and capitalise on the fact that we can easily calculate the expected reward from deciding using (6.9). If we assume that after some large amount of time, T, the observer makes a decision regardless of the value of $\boldsymbol{x}(T)$, then we can compute the reward expected at this time for any value of $\boldsymbol{x}(T)$ using the expected reward for deciding. Taking one time-step backwards to $T-1$, we can compute the expected reward from deciding as usual. However, we can now also calculate the expected reward from waiting because we know the value associated with each state at time T, and know the probability distribution over the change in the accumulator from time $T-1$ to T, $\delta\boldsymbol{x}$, using 6.12.

While this offers a way for the researcher to calculate the optimal policy it seems less plausible to think that a strategy of working backwards, one time step at a time from some distant time, is used by animals and humans. Animals and humans may learn a mapping from time and the evidence accumulators to stopping vs. waiting, rather than solving for the optimal policy explicitly. As discussed, when learning a mapping, reducing the number of variables considered may speed up learning. We next look at an approximation which does just that: According to the approximate solution a smaller number of variables are relevant to the decision of whether to stop vs. wait.

6.2.3 Approximating the optimal solution

One approach to finding an approximate solution would be to use some form of myopic approximation and only look ahead a few steps. We make a different approximation here, looking as far ahead as needed, but at each time step only considering the most likely value for $\delta\boldsymbol{x}$. Considering only one value of $\boldsymbol{x}(t)$ at each time step reduces the problem of picking a stopping policy to the problem of picking a time to stop.

In Appendix D.1 we consider any two points $\boldsymbol{x}^{(d)}(t)$ and $\boldsymbol{x}^{(e)}(t)$ which share a specific property. Once the posterior probability over the N options has been computed at each of these two points, the greatest of the posterior probabilities at $\boldsymbol{x}^{(d)}(t)$ is the same as the greatest posterior probability at $\boldsymbol{x}^{(e)}(t)$. That is,

$$j = \arg\max_i p(C = i|\boldsymbol{x}^{(d)}(t)) \tag{6.13}$$

$$k = \arg\max_i p(C = i|\boldsymbol{x}^{(e)}(t)) \tag{6.14}$$

$$p(C = j|\boldsymbol{x}^{(d)}(t)) = p(C = k|\boldsymbol{x}^{(e)}(t))\,. \tag{6.15}$$

Under the approximation, the expected reward at any later time step $t + l$ is the same. As the expected future rewards are the same for all stopping times, and only stopping time matters (because $\boldsymbol{x}$ takes a single path), if it is optimal to stop (or continue) at $t, \boldsymbol{x}^{(d)}$, then it is also optimal to stop (or continue) at $t, \boldsymbol{x}^{(e)}$. As $\boldsymbol{x}^{(d)}$ and $\boldsymbol{x}^{(e)}$ can be any two points for which the posterior probability of the most probable option is the same, we can say that at a time step, t, for all points, $\boldsymbol{x}$, with same maximum posterior probability, it is either optimal to stop at all of them, or none of them. Hence, according to the approximate optimal stopping rule, only two variables are needed to determine whether or not to stop; time and the maximum posterior probability.

6.3 Modelling methods

Having found that an approximately optimal stopping rule exists that uses a stopping criterion on only time and the maximum posterior probability, we set out to test whether learning stopping rules with this reduced set of variables is faster. To do this we use neural networks as model learners. The networks were trained to give response times and choices on the basis of a representation of time and the accumulated evidence. We compared the performance of different networks that received different representations of the accumulated evidence. Our hypothesis was that a network that received time and the maximum posterior probability would learn fastest.

Task The evidence the networks received was determined by a simulated version of the perceptual decision making task described in the previous section (see equations 6.1 and 6.2). We used the settings shown in Table 6.1. We considered every combination of number of choices (2, 4, 10), uncorrelated and correlated accumulators ($b = 0$ and $b = 0.9$), and cost of time (constant, increasing). If constant, the cost of time was 1 per time step, if increasing, the cost of time was $\frac{2}{3}t$, where t is the time step. A cost of time was always included because without a cost of time it would be optimal to accumulate evidence forever. Deliberating time is always costly to animals and humans because it comes at the expense of other opportunities. For example, time can be spent finding food or avoiding predators (Drugowitsch et al., 2012). Each trial was made up of 40 time steps, and the networks received 3000 simulated trials (which were then split into training and validation sets).

For the case of 2 choices, constant cost of time, and no correlation between accumulators, we computed the expected reward for an optimal observer, allowing comparison of actual performance to optimal performance (derivation provided in Appendix D.2).

Variable	Expression	Simulation value
Number of choices	k	2, 4, 10
Correct mean	$\mu_0 + \mu$	1.3
Incorrect mean	μ_0	1
Covariance	b	0, 0.9
Variance	a + b	1
Correct reward	r_c	50
Error reward	r_e	-50
Prior	$p(C = i)$	$1/k$
Cumulative cost of time	L(t)	see text

Table 6.1: Properties of the simulated perceptual decision making task.

Networks We used deep neural networks implemented in `Keras` using the `TensorFlow` backend. The networks comprised of an input layer, an output layer, and 2 hidden layers (see Fig. 6.2). Each hidden layer had a number of units equal to the number of choices in the decision task, plus 1 ($k + 1$). The output layer contained a single unit.

The nature of the input data depended on the network. All networks received the current time step as an input (scaled and shifted, see below), but varied in their additional inputs. The "all accumulators" network received an additional input for each evidence accumulator, conveying its value (Fig. 6.2). The "max accumulator" only received one additional input, the value of the maximum accumulator. This network is related to the race model of stopping, according to which animals use a stopping criterion on the value of the maximum accumulator (Bogacz et al., 2006). The "all probabilities" network received an input for each choice, in addition to time. These inputs corresponded to the posterior probability of each choice. The "max probability" network only received one additional input, the value of maximum posterior probability. Hence, this network received only the variables that are needed for the approximately optimal stopping criterion, and was the network we hypothesised would perform best.

The values for time and evidence obviously change throughout the course of a trial. The input to the networks corresponds to a snapshot of the value for time and evidence at one particular point. The output corresponds to the networks behaviour at that particular point. To see how the network behaves over the course of a trial we input the data for each time point, in turn, and look at the corresponding outputs. More precisely, in each training epoch we presented the networks with the input data for every trial, and every time step. The activation of the output unit in response to the input data for trial i and time step t was interpreted as the conditional probability that the network stopped and decided at time step t in trial i, given that it had not already stopped and decided. The tanh activation of the output unit lead to activations in the range $(-1, 1)$, so we mapped linearly to the range $(0, 1)$ to provide a number that could be interpreted as a probability. On the final time step, step 40, the networks were forced to stop and decide. The result of this procedure was a set of 40 probabilities, corresponding to the probability that the network stopped at each time point, conditional on reaching that time point in the trial (i.e. without stopping prior to that point).

The networks were trained to minimise the average loss across all trials and time steps. The loss is the negative reward earned by the network. Once the network "stopped", we assumed that the network then made the Bayes optimal decision, and the network received the corresponding reward. From this reward, we subtracted the cost of time used to accumulate until the time step of the decision. As we interpreted the output of the network probabilistically, we did not enforce that the network stop at a specific time. Instead, our interpretation of the output gives us a probability distribution over stopping times. The network earned the reward associated with a particular stopping time in proportion to the probability mass assigned by the network to that stopping time. Hence, the reward earned by the network represents the expected reward under a probabilistic stopping rule. An alternative approach would have been to force the network to make a single choice about when to respond. We discuss the choice to interpret the network output probabilistically in Section 6.5.

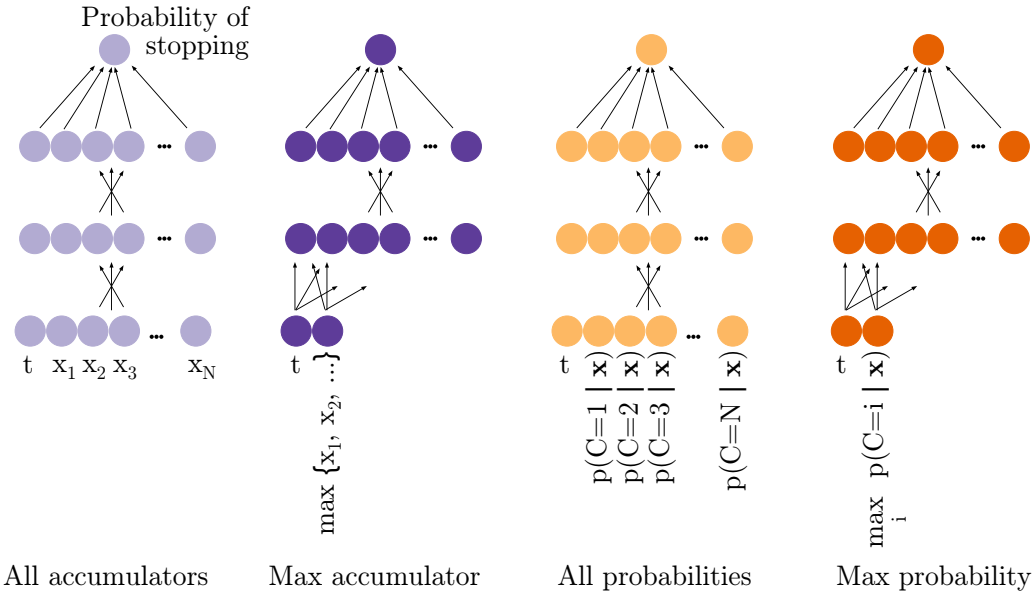

Figure 6.2: The networks used.

We prepared the training data for the constant cost of time condition by z-scoring the evidence values for all trials, all time steps and all choices together, and separately z-scoring the time step value for all trials and all time steps together (LeCun et al., 1998). The cost of time does not affect the distribution from which training data is drawn. In preparing data for the increasing cost of time condition, we did not z-score but applied the same centring and scaling values used for the constant cost of time condition. (Each centring and scaling from the constant cost of time condition was applied once in the increasing condition.)

When assessing the performance of the networks, we used the reward obtained per trial in the validation datasets (the part of the simulated datasets not used for training). All plots show this reward. To obtain an estimate of the uncertainty associated with the performance of the networks, we trained each network 20 times (from random initial weights).

Hyperparameter optimisation An issue with using neural networks as model learners is that they introduce an element of arbitrariness. One must select the parameters of the networks, and the concern is that this may lead to conclusions

which depend on those parameters. To alleviate this, we used the Hyperopt Python library to perform hyperparameter optimisation (Bergstra et al., 2013). Hyperparameter optimisation involves scanning over the network hyperparameters and choosing the configuration that maximises some objective function. Thus the choice of network properties is, to some extent, taken out of the hands of the experimenter.

We optimised over the parameters that plausibly have a large impact on the results; for example, a high learning rate may impair a network's ability to find the global minimum, whilst a very low learning rate may artificially slow down a network. This becomes a particular problem if different networks perform best at different learning rates. By optimising over learning rate, we can reduce the probability that there are such effects in our results. For the hyperparameter optimisation options considered, see Table 6.2. Every possible combination of parameters was considered at least once.

Our objective function was the reward obtained in 3000 unseen trials, after 500 epochs of training on 2250 trials from the 4 choice condition, with a constant cost of time, and no correlation between accumulators. Using previously unseen trials allows us to avoid parameter combinations that produce overfitting. Importantly, we performed hyperparameter optimisation using the "all accumulators" network. That is, network properties were selected as those that allowed this network to perform best. Hence, if any network had an unfair advantage due to the specific network settings used, it was this one, and any increased performance we see in our favoured model, "max probability", is a conservative estimate of the improved performance compared to "all accumulators". Hyperparameter optimisation produced the choice of network parameters shown in Table 6.2.

Network variable	Parameter used	Hyperparameter optimisation choices
Optimiser	RMSprop	Adam, RMSprop, SGD, Adadelta
Learning rate	0.0005	0.0001, 0.0005, 0.001
Dropout rate	0.5	0.2, 0.5
Number of layers	4	3, 4, 5, 6, 7, 8, 11
Network type	Dense, fully connected	-
Training-validation split	75:25	-
Activation function	tanh	-

Table 6.2: Network parameters used for the model learner. For parameters that were optimised, we indicate the possible choices that were scanned over.

6.4 Modelling results

We trained neural networks on a simulated perceptual decision making task, in which the networks needed to learn a policy for stopping evidence accumulation and triggering a decision based on various possible representations of accumulated evidence. Here we discuss the results for the the simplest (2-way choice) and most difficult (10-way choice) cases. Results were similar for the intermediate difficulty (4-way choice), and we provide these in Appendix D.3.

6.4.1 2 options

After training the networks, we explored the central question of whether learning an appropriate mapping from time and state of the evidence accumulators, to stopping vs. waiting, can be sped up through approximation. Figs. 6.3 shows the time course, over rounds (epochs) of training, of the performance of the networks. The figure shows this time course in environments with increasing and constant cost functions, and with uncorrelated and correlated evidence accumulators. Lines represent the median reward obtained by the network over 20 attempts to train the

network (from random weights), and shading represents 95% confidence intervals computed using bootstrapping.

Under no condition tested was the "max probability" network outperformed; in all cases it learned as fast or faster than all other networks. In the case of uncorrelated evidence with a constant cost of time (Fig. 6.3, A), it is the "max probability" network that is able to achieve the optimal reward within 1500 training epochs. These results suggest that approximating the optimal stopping criterion on time and the state of all accumulators with a stopping criterion on time and the maximum posterior probability, leads to faster learning.

Interestingly, in most cases, the "all probabilities" network learned faster than the "all accumulators" network, suggesting that using probabilities enables faster learning than when using the raw accumulated evidence samples. The dimensionality of input for both networks is the same, although the probabilities are constrained to a line such that $p(C = 1|\boldsymbol{x}) + p(C = 2|\boldsymbol{x}) = 1$. The raw evidence samples are not so constrained, despite the fact that the sum of evidence samples, $\sum_i x_i(t)$, is irrelevant to the question of when to stop (Tajima et al., 2019). (This sum is irrelevant because adding the same amount of extra evidence to all accumulators does not change the posterior probabilities, and because of this it does not change the future behaviour of the accumulators relative to their current point, nor the value of deciding at those future points; see equations 6.6, 6.12, 6.9.) This additional unnecessary variability may explain why it is harder to learn a mapping from the raw evidence samples, than from probabilities.

We note that the approximation embodied by the "max probability" network does not lead to fast learning simply because the approximation reduces the dimensionality of the input for the mapping that must be learned. The "max accumulator" network also learns a mapping from two dimensions (time and the maximum accumulator) but this network learns far slower than even the "all probabilities" network, which learns a mapping from a three dimensional space.

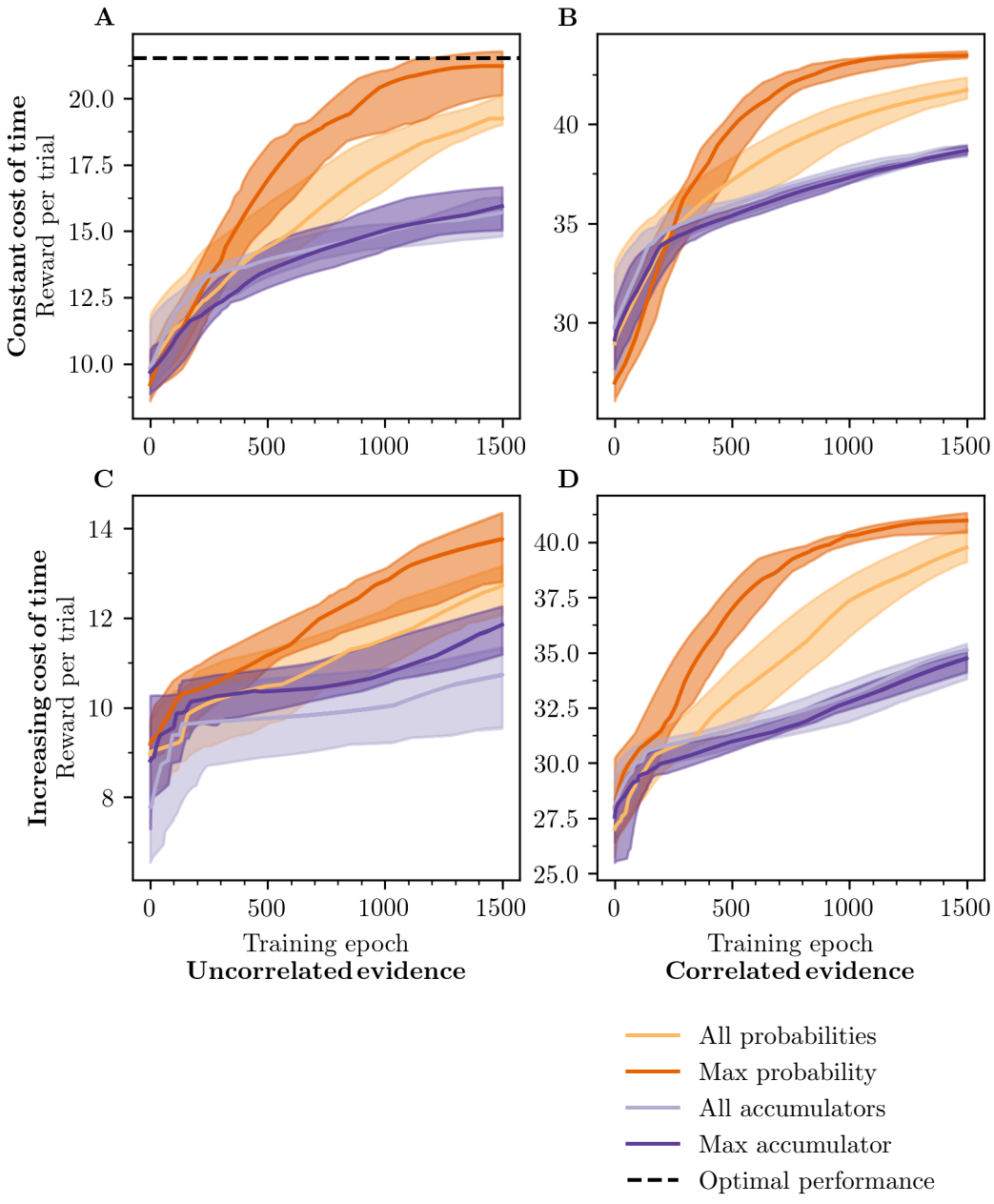

Figure 6.3: Average reward gained per trial for the case of 2 choices. Results are shown for cases with and without correlations between evidence accumulations, and with a constant and an increasing cost of time. The optimal performance for the uncorrelated constant cost of time case is calculated according to the discussion in Appendix D.2.

One may wonder why the networks achieve a greater reward in the correlated environment, rather than in the uncorrelated environment. This effect arises because, when the evidence accumulators are correlated, a greater proportion of variability is along the direction $[1, 1]^T$. The value of the state of the accumulators along this dimension gives the value of the sum of the accumulators and, as discussed, this sum is irrelevant to the posterior probabilities for the two options. Hence, in the correlated case, a larger proportion of variability is along a dimension that does not affect how discriminable the two options are, making the task easier.

Taken together these results suggest that animals, if they are aiming to rapidly learn optimal stopping, should learn a mapping from time and maximum posterior probability. Using the full posterior, while it offers no dimension reduction is also a sensible strategy when faced with 2-choices. In either case, animals should use the products of Bayesian inference in their criteria for stopping.

We additionally explored the decision rule learned by each of the networks. To do this we simulated new trials, and looked at how the networks behaved in these new trials. For each trial we found the networks' probabilistic policy by interpreting the output in the usual way (see Section 6.3), and simulated the behaviour of an observer who followed this policy. Hence, for each trial we found one stopping time, drawn according to the probabilistic policy. For each trial, and each time step at which the simulated agent waited, we plotted the difference in evidence for the two choices (accumulator difference), against time. We also plotted these values at the time step when the agent made a decision. We show the results of this analysis in Fig. 6.4, for the "max probability" and "all accumulator" networks, in the case where there is correlation between accumulators.

When the cost of time is constant, the optimal policy does not depend on time (Ferguson, 2011). This is because, whether you are at time step 5 or time step 10, if the current posterior is the same, the probability distribution over the evolution of the accumulator state in future time steps, and the cost of time in future time steps is identical. Certainly at time step 10 you will already have incurred a greater cost

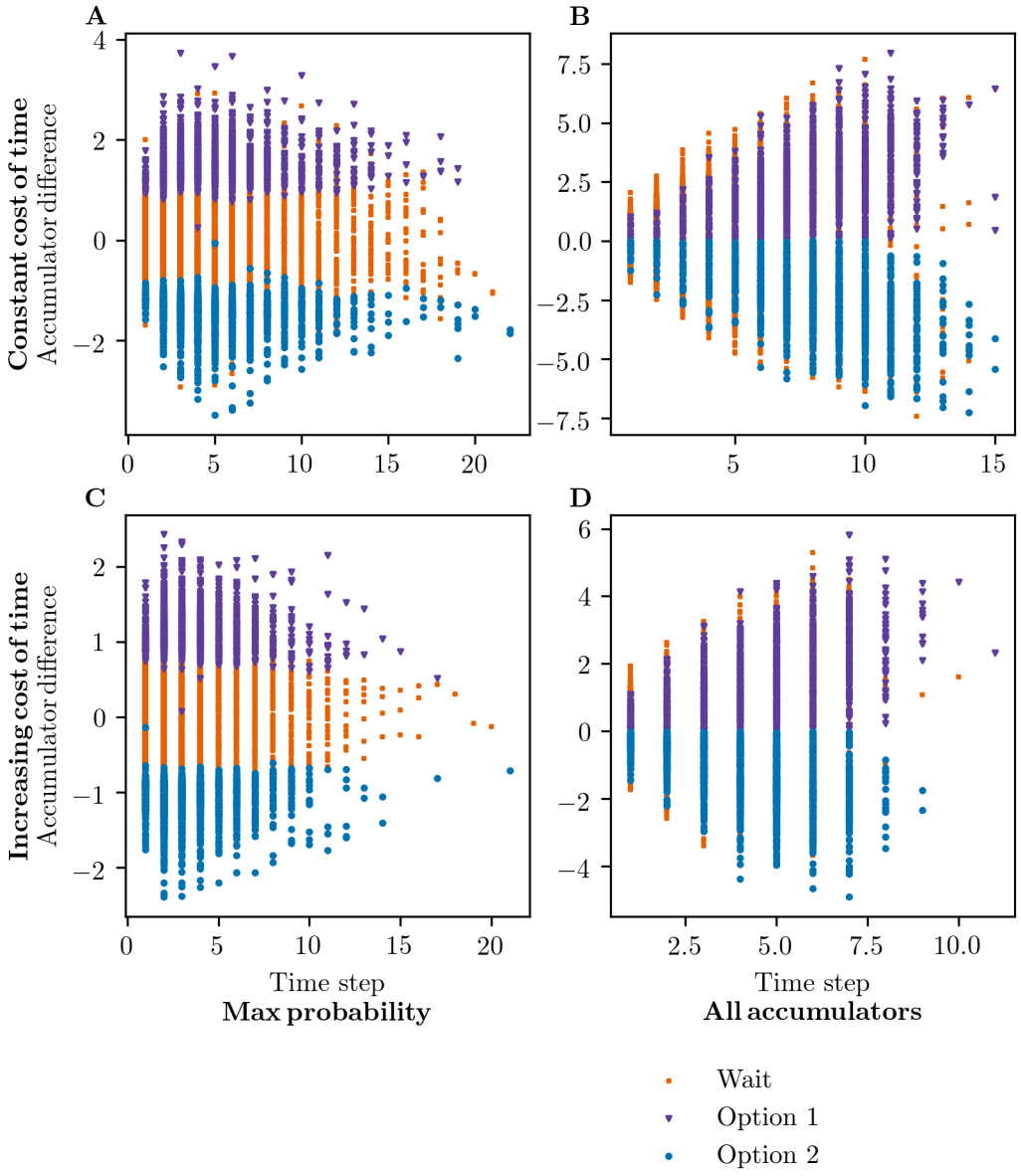

Figure 6.4: Learned decision rule for "max probability" (left) and "all accumulators" (right) networks, for correlated evidence accumulations with a constant cost of time (top) or increasing cost of time (bottom) for the case of 2 choices. Accumulator difference refers to the difference in the values of the two evidence accumulators.

of waiting than at time step 5, but these are sunk costs you will have to pay whether you wait or stop, and hence, do not affect whether waiting or stopping will provide greater reward. In the case we consider, the difference between accumulators is monotonically related to the posterior probabilities (see equation 6.6; Moran, 2015). Therefore, the networks should require the same difference between accumulators before they stop, regardless of how long they have been deliberating.

For the "max probability" model, we observe that, as expected, the difference in evidence required to make a decision is approximately constant with time (Fig. 6.4; A). The "all accumulator" network learned a very different rule, reflecting that fact that even after 1500 training epochs, this network was still performing well below the level of other networks (Fig. 6.4; B). The network had not learned a clear policy, and the network stopped accumulation and made decisions even when the difference in evidence between the two alternatives was zero; i.e. it was deciding despite having no evidence to favour either option.

Under conditions in which the cost of time is increasing, we expect the networks to require less evidence as time increases, as it becomes increasingly detrimental for the network to wait. Again the "all accumulator" network did not learn a clear policy (Fig. 6.4; D). For the "max probability" network it is difficult to see from the plot whether the correct stopping rule, with a decreasing criterion on the difference in evidence, has been learned (Fig. 6.4; C). The policy learned by the network becomes clearer if we plot it in a different way. Instead of plotting particular simulated behaviours, we plot the probability the network stops, as a function of the difference between the two accumulators, and the time step (Fig. 6.5). From these plots it is clear that the "max probability" network has learned a stopping policy that features a decreasing stopping criterion on the difference in accumulated evidence (Fig. 6.5; B). It is interesting to note that the network even learns a sensible policy for time steps which it rarely encounters (compare with the time steps reached in Fig. 6.4; C).

These results suggest that, in the case of two choices, learning an approximation to the full mapping from time and state of the accumulators, to stopping vs. waiting,

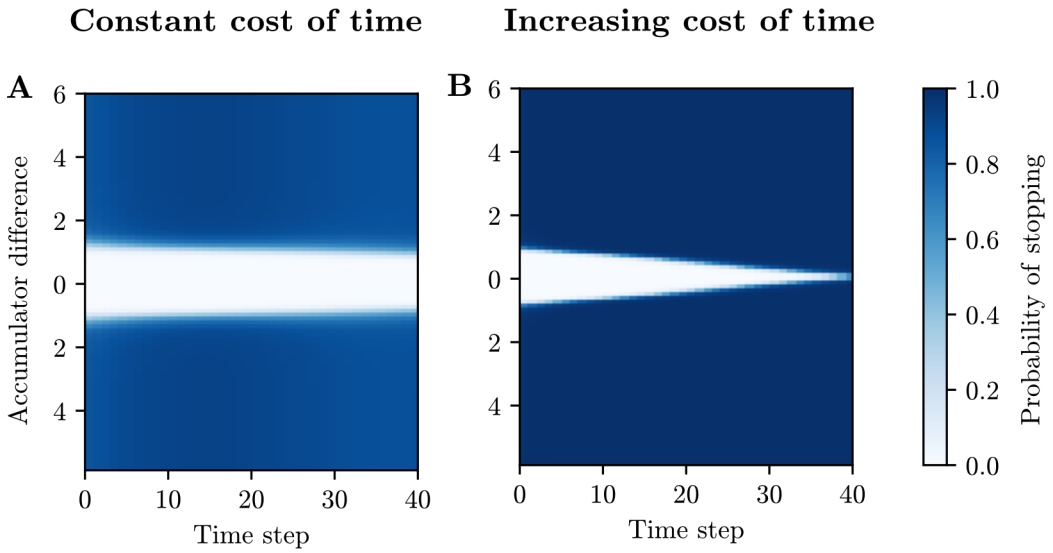

Figure 6.5: Probability the "max probability" network stops as a function of the difference between evidence accumulators, and time. Results shown for 2 choice options, correlated evidence accumulators and with a constant (A) or increasing (B) cost of time.

can speed up learning. Learning the mapping from time and maximum posterior probability is particularly fast. Furthermore, learning such an approximate mapping leads to a sensible stopping policy, suited to the current conditions.

6.4.2 10 options

The "max probability" network outperforms the other networks when choosing between 2 options. We also considered the much more challenging situation in which the network had to choose between 10 options. In Fig. 6.6 we plot the average reward per trial achieved by the different networks, over the course of training. Again the "max probability" network learned as fast as, or faster, than the other networks. In some cases the advantage was very large (Fig. 6.6, D). The only case in which the "max probability" network did not learn faster was with uncorrelated evidence samples and an increasing cost of time. As discussed in the previous section, the task with uncorrelated evidence samples is considerably harder than the task with correlated evidence samples. Coupled with the increasing cost of time, it may simply be that there is little for any network to learn in this case.

With 10 accumulators, one concern might be that the benefit of the approximation embodied by the "max probability" model is largely down to the fact that this approximation reduces the dimensionality of the input space from 11 (time and the 10 accumulators) to 2. However, we again see that the "max accumulator" network, which also learns a mapping from 2 dimensions (time and the maximum accumulator), performs far worse than the "max probability" network and, in many cases, also worse than the "all probabilities" network. Hence, the specific information represented by the maximum posterior probability seems to be the key to rapid learning.

As before, we are also able to check whether the networks learn a sensible stopping policy. It is more difficult to visualise the learned behaviour with 10 choices because of the dimensionality of the evidence space. We have to reduce the dimensionality of the cumulative evidence (10 values) in order to visualise the decision rule. To this end, we pick two options at random and plot the behaviour against the difference in the evidence accumulators for these two options. Naturally a lot of information is lost by doing this, but nevertheless the outcomes of the analysis are clear.

In Fig. 6.7 we show plots of the difference between evidence accumulators for two randomly selected accumulators, and behaviour at that difference, against time, as we did in the previous section. The "max probability" network learns to behave sensibly; if there is no evidence to distinguish between the two options, the network does not pick one of them (Fig. 6.7, A). In contrast, the "all accumulators" network stops and decides even when there is no difference in evidence between the chosen option and the other option used for plotting (Fig. 6.7, B, D). This might be a sensible thing to do if the cost of time is increasing, however, the network also behaves this way when the cost of time is constant. When the cost of time is constant, the network should use the same stopping rule at all time steps, as discussed. However, the "all accumulators" network clearly violates this feature of the optimal solution (Fig. 6.7, B).

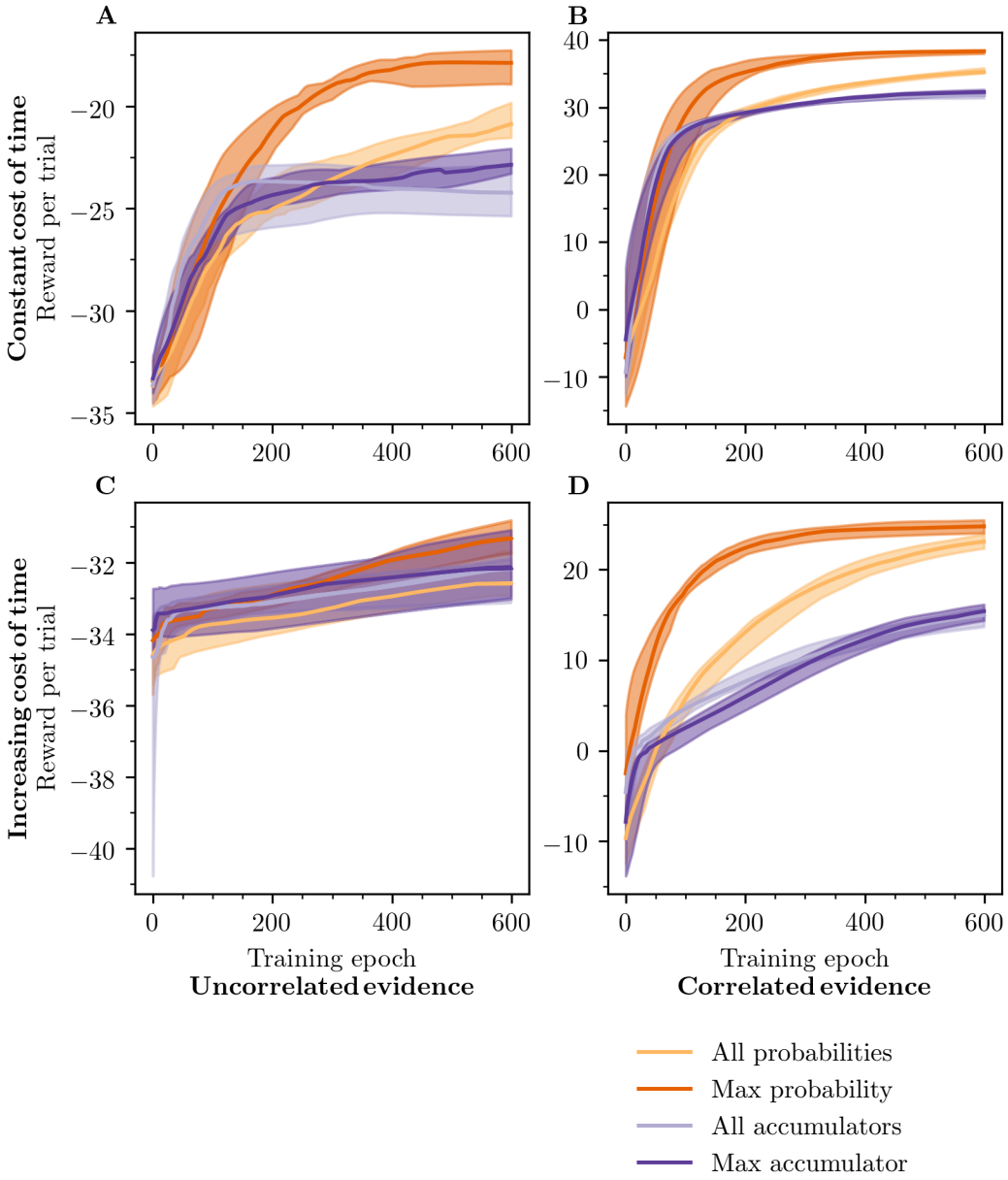

Figure 6.6: Average reward gained per trial for the case of 10 choices. Results are shown for cases with and without correlations between evidence accumulators, and with a constant and an increasing cost of time.

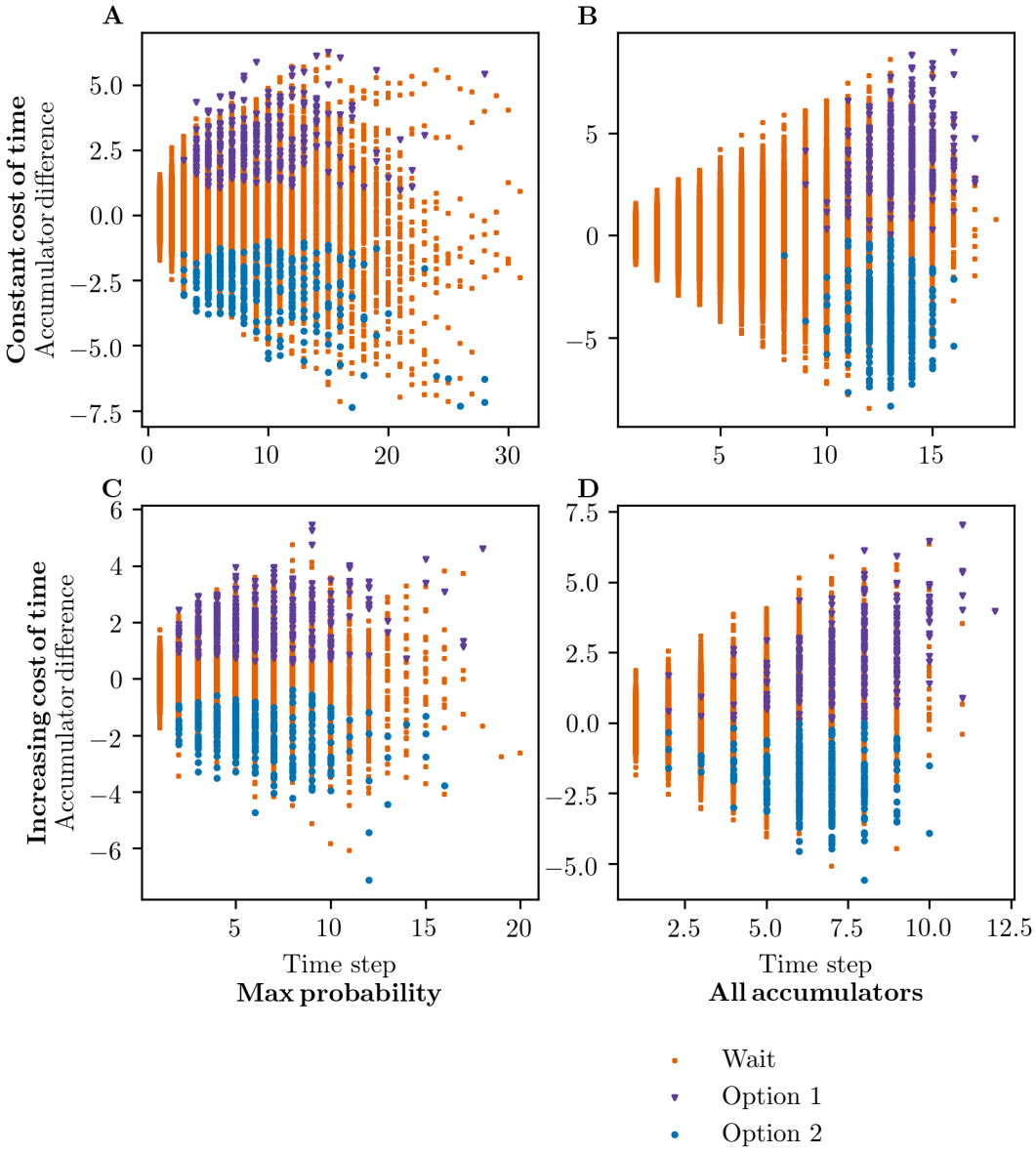

Figure 6.7: Learned decision rule for "max probability" (left) and "all accumulators" (right) networks, for correlated evidence accumulators with a constant cost of time (top) or increasing cost of time (bottom) for the case of 10 choices.

In conjunction with the results from the 2 options case, these findings suggest that learning an approximate stopping rule, by using a criterion on time and maximum posterior probability, leads both to (a) sensible stopping policies and (b) considerably faster learning than when learning a criterion on other variables.

6.5 Discussion

The aim of this chapter was to explore the idea that the DDM is not optimal in a sense that is relevant in ecological settings. Specifically, we noted that in ecological settings observers may never reach optimal reward rate, so the fact that a model can achieve this may be irrelevant. Instead, what may be important is rapid learning. We considered the problem of optimal stopping in a generic perceptual decision making task. We found that under a certain mathematical approximation – considering only the most probable events at future time steps – only time and the maximum posterior probability are relevant to the question of when to stop deliberation. We went on to test this result using neural networks as model learners. Crucially, a criterion on time and maximum posterior probability reduces to the DDM in the settings considered in previous chapters (Section 6.1).

We explored whether the simplification afforded by using a criterion on only two variables (time and maximum posterior probability) sped up learning. We compared several networks, including a network which learned a stopping criterion on time and the full description of the accumulation state (the value of all accumulators). The network that used a criterion on only time and maximum posterior probability, and hence used the structure of the approximate optimal stopping rule, learned fastest. We found evidence that it is the maximum posterior probability specifically, and not simply the dimensionality reduction, that supports the rapid learning of sensible and near optimal stopping rules. Specifically, a network which learned a mapping from time and the maximum accumulator performed poorly.

The success of a stopping rule based on time and maximum posterior probability has implications for the DDM. As discussed in Section 6.1, a stopping criterion based on time and maximum posterior probability is equivalent to a specific DDM. Assume, for argument's sake, we believe that animals and humans will use stopping rules suited to multi-alternative decisions, and that facilitate efficient learning in new situations. Then, the success of a criterion based on time and maximum posterior probability in these regards, is evidence for such a criterion. In turn, this is evidence that in the context of two-alternative decisions, the DDM will provide a good description of perceptual decision-making mechanisms.

The race model is also related to one of the networks used in this study: In the race model, it is the maximum accumulator that triggers a response (Bogacz et al., 2006). We tested a network that used time and the maximum accumulator in its stopping rule, however this network learned far more slowly than networks which used probabilities in their stopping rules. These results suggest that an animal that used a time-dependent race model would suffer from slower learning, and hence, a decrease in efficiency and/or quality of decisions.

Conceptually, our work is largely consistent and inspired by the work of Tajima et al. (2019), who showed that animals could approximate the optimal stopping rule by mapping the state of the accumulators onto a manifold. Our contribution is, first, to consider how approximation might affect the speed of acquisition, and second, to tie optimal stopping to Bayesian inference, by showing how conversion of accumulators to posterior probability is a specific mapping that can both be theoretically motivated, and convey practical benefits to animals. Moreover, we have focused on the implications of the results for the DDM.

The idea that animals might use stopping criteria involving posterior probabilities has been proposed before, for example, in the multihypothesis sequential probability ratio test (MSPRT; Section 1.5.1; Bogacz and Gurney, 2007). Using this rule, animals would continue accumulating evidence until the maximum posterior probability reaches a threshold that is constant over time. This policy is only optimal

in specific situations, such as when the observer makes very few errors (Dragalin et al., 1999). In general, a time-dependent threshold is required for optimal stopping (Tajima et al., 2019). Our work builds on the idea that the brain implements the MSPRT, by suggesting there are time-dependent thresholds on the maximum posterior probability, and by showing that these provide approximately optimal stopping in a wide range of conditions. Taking a wider perspective, our work adds to findings suggesting that the probability of being correct, which is the same as the maximum posterior probability and may be the variable reflected in subjective confidence (Kiani et al., 2014; Pew, 1969; Vickers & Packer, 1982), is a value of very general importance. The results in this chapter provide another example of a process where a representation of the probability of being correct may be important (Section 1.1; Bahrami et al., 2010; Balsdon et al., 2020; Drugowitsch et al., 2019).

The study has two main limitations. First, the conclusions drawn only directly apply to perceptual decisions of the form studied. We considered a decision in which the observer could have any prior belief about the options, we allowed correlations between the accumulators, and allowed any function for how the cost of continued deliberation changes with time. When creating any model, a large number of assumptions will have to be made. Perhaps the most important ones in our case are that we did not consider arbitrary covariance structures, different rewards associated with different options, or variability in the signal-to-noise ratio (μ). We anticipate that in these cases similar principles to those identified here will apply; with a sensible approximation it should be possible to reduce high dimensional problems down to a few dimensions, and posterior probabilities over the state of the world should have a key role to play in this.

The second main limitation is the nature of the neural networks. We worked to ensure maximum generalisability of our neural network results by selecting neural network properties as those which most helped a competitor to our favoured network (see "Hyperparameter optimisation" above). In any case, there are features of the network training that are unlike the process animals and humans engage

in. Specifically, the networks did not have to commit to a single decision, but provided a probability of stopping at each time step, and received feedback on this probabilistic policy. Hence, the networks received more information about good and bad behaviours than an animal that stops at a single time step, and can only learn about the value of stopping at this one step. We were not aiming to provide realistic models of the mechanism by which animals and humans learn. Instead we were aiming to test the idea that approximation could speed up learning in general. Nevertheless, in future work it would be valuable to consider more realistic models of the learning algorithm used by humans and animals.

Our findings provide evidence that using the maximum posterior probability, animals may be able to simplify stopping problems and speed up learning, providing evolutionary advantages, especially in changing environments. In the situations considered in previous chapters, this approach reduces to the DDM, suggesting the DDM is indeed normative in a sense that is relevant to naturalistic settings.

7

General discussion

Contents

7.1 Summary of findings

In Chapter 1 we explored current models for perceptual decisions and confidence, and set out the motivation for the thesis. We saw that a wide range of models have been proposed for perceptual decisions, but that the DDM has both strong empirical support and is a normative model of decision making (Bogacz et al., 2006; Ratcliff & McKoon, 2008). We explored the reasons why the DDM struggles to account for confidence reports, and saw that empirically satisfactory models of confidence can be built if one assumes a non-normative perceptual decision making mechanism

(Kiani et al., 2014). A central aim of the thesis was to establish whether, and how, the normative DDM could be reconciled with observed patterns in confidence data. We discussed how some combination of decision-thresholds that decrease with time (Drugowitsch et al., 2012; Malhotra et al., 2017), drift-rate variability (Pleskac & Busemeyer, 2010), and a Bayesian readout for confidence (Kiani et al., 2014; Pew, 1969; Vickers & Packer, 1982), or a miscalibrated Bayesian readout (Drugowitsch et al., 2014), might allow the DDM to account for the strength of the relationship between confidence and time. This relationship is not adequately captured by one of the most successful previous DDM accounts of confidence, two stage dynamic signal detection theory (2DSD; Pleskac and Busemeyer, 2010).

In Chapter 2 we set out a core DDM and 9 additional DDM variants. We considered models with the features identified in Chapter 1: Decreasing decision thresholds, drift-rate variability, a Bayesian readout for confidence, and a miscalibrated Bayesian readout. Focusing on the DDM variant with a miscalibrated Bayesian readout, we saw that this model can account for a very wide range of qualitative patterns that have previously been observed, and can also explain results that at first sight might appear to conflict (Desender et al., 2020). We made three qualitative predictions from the set of DDM models, and verified one of these in a previously collected dataset (Charles & Yeung, 2019). Specifically, we found that when observers are free to set the time of response, and once the effect of response time has been accounted for, pre-decision evidence has a smaller effect on confidence than evidence gathered between the time of internal commitment to a decision and overt response. This is expected if decisions and confidence are the result of the same process because, in this case, pre-decision evidence can only affect confidence by changing the time that the decision threshold is reached (Section 2.1.2). The results of Chapter 2 suggest that the DDM, coupled with some form of confidence readout from the evidence accumulation which generates the decision, is consistent with patterns observed in confidence data.

Directly building on the conclusion that the DDM may be able to account for patterns in confidence, in Chapter 3 we collected a new dataset and applied additional qualitative, and quantitative tests to the DDM variants. A key feature of the new dataset was that it contained trials in which participants could respond when they wished (free response), and trials in which participants were cued when to respond (interrogation; Bogacz et al., 2006). In a preregistered set of tests, we found all three qualitative predictions from Chapter 2 held in the new dataset. In addition to the prediction tested in Chapter 2, we also confirmed predicted differences between the interrogation and free response conditions: The effect of pre-decision evidence was stronger in the interrogation condition than in the free response condition, and confidence was more negatively related to response time in the free response condition than in the interrogation condition. These findings further supported the assumptions on which the predictions were made, assumptions that are shared by all the models developed in Chapter 2. Specifically, the DDM variants considered all assumed that the same evidence accumulation which generates the decision also leads to confidence. The models also assumed that decision thresholds are used in the free response, but not in the interrogation condition.

Having found qualitative support for the 10 DDM variants developed in Chapter 2, for the remainder of Chapter 3 we explored whether any of the variants could account for the precise quantitative patterns in confidence data. Patterns in confidence data differed between the free response and interrogation conditions. For example, confidence decreased with response time in the free response condition (Pleskac & Busemeyer, 2010), but remained constant with response time in the interrogation condition. These different patterns allowed us to generate a strong test for the models by requiring them to fit data from both conditions, with shared values for parameters relevant to both conditions (Ratcliff, 2006). DDM variants in which confidence reflected a miscalibrated Bayesian readout provided an excellent fit to the effects of time and evidence on confidence. For example, the DDM with a miscalibrated Bayesian readout could account for the fact that evidence and time have effects on confidence which are independent of each other, and it could

account for the precise nature of these effects even though they differed in the free response and interrogation conditions (Fig. 3.8). DDM variants with a miscalibrated Bayesian readout outperformed other models in a formal model comparison that included a model closely related to 2DSD. The results of Chapter 3 further support the conclusions of Chapter 2, that DDM variants with a miscalibrated Bayesian readout are consistent with, and provide a unified explanation for, a wide range of effects observed in confidence data.

In Chapter 4 we derived simple mathematical expressions for Bayesian confidence of DDM observers. These expressions are simple in the sense that they contain a limited number of standard terms and, as a result, can be rapidly evaluated. These results made it possible to model the effects of dynamic stimuli on a trial-by-trial basis in Chapter 3. They also allowed us to gain direct insight into patterns in the confidence of Bayesian observers.

In Chapters 5 and 6, we addressed arguments suggesting that the DDM with Bayesian confidence is not a sensible model to pursue. In Chapter 5 we explored challenges to the idea of Bayesian confidence (Peters et al., 2017; Zylberberg et al., 2012). Specifically, we looked for evidence of an over-weighting of decision-congruent evidence in settings where a Bayesian observer would equally weight decision-congruent and incongruent evidence. We found no evidence for a strong general over-weighting. In contrast, across three previously collected datasets containing data from 189 participants (Carlebach & Yeung, 2020; Charles & Yeung, 2019), we only found tentative evidence for a slight over-weighting of decision-congruent evidence. We suggested that previous studies may have found an over-weighting of decision-congruent evidence because of the stimuli used. Specifically, participants may have learned that there was greater variability in the signals associated with the correct option. In this situation, a Bayesian observer does not equally weight decision-congruent and incongruent evidence in their confidence reports (Miyoshi & Lau, 2020).

In Chapter 5 we also investigated a second phenomenon. It has been found that it is possible to increase confidence without corresponding changes in performance (Koizumi et al., 2015; Rollwage et al., 2020; Samaha et al., 2016). We explored this effect in a dataset where we simultaneously varied evidence for both alternatives. We found that the DDM, with an additional mechanism which generates stimulus dependent noise (Teodorescu et al., 2016), could account for the observed effects. The findings of Chapter 5, build on the results in previous chapters by providing further evidence that the DDM with a miscalibrated Bayesian readout provides an account that is consistent with a wide range of choice, response time and confidence data. Although, the results do suggest that the idea of stimulus dependent noise may need to be included to account for some effects.

Chapter 6 moved away from the focus on explicit confidence reports, instead returning to look at perceptual decision mechanisms, and a specific challenge to our use of the DDM. In focusing on models built on the DDM we were motivated by both the empirical success of the DDM, but also its normative properties (Section 1.4.2). However, a potential critique is that the DDM is not optimal in a way that is relevant in ecological settings (Section 1.5). Instead of a "stopping rule" for terminating deliberation that could generate maximal reward rate after extensive training, fast learning is likely to be crucial in ecological settings, where the "task" faced by animals and humans is constantly changing. We considered an approximate stopping rule that involved terminating deliberation on the basis of only two values: The time spent deliberating so far, and the maximum posterior probability across the posterior probabilities for all the options. This rule is closely related to the multihypothesis probability ratio test (Bogacz & Gurney, 2007). Importantly, the rule reduces to the DDM with time-dependent thresholds in the case of two-alternatives (Moran, 2015). First, we showed mathematically why this rule can be viewed as an approximation of the optimal solution. Second, using neural networks as model learners, we found that use of this rule leads to faster learning and consequently a greater reward rate, even before the theoretical maximum for reward rate has been reached. These

results suggest that the DDM does not just offer the theoretical possibility of maximal reward rate; it also offers a quicker route to get there.

As a whole, the findings suggest that there is no need to abandon the DDM, and in fact, the DDM and miscalibrated Bayesian confidence provide a powerful explanation for patterns observed in confidence data. For example, this model correctly makes the counter-intuitive prediction of pre-decision evidence having a reduced effect on confidence (Section 3.2), and accounts for the precise effects of response time and evidence in different experimental conditions, even when these variables are considered simultaneously (Fig. 3.8). Retaining the DDM is highly desirable because the DDM is likely to be normative not just in the lab, but in ecological settings, as it can generate faster learning than other models (Section 6.4). Furthermore, we explored how the idea of a miscalibrated Bayesian readout for confidence might be used to explain some findings which, on the surface, challenge this account (Section 5.3.1 and 5.3.2).

7.2 Results in a wider context

7.2.1 Relation to value-based decision making

In this thesis we have focused on perceptual decisions, categorical choices made on the basis of sensory information. We have been interested in perceptual decisions in their own right, as a basic cognitive function, and therefore a topic of fundamental importance for psychology. Perceptual decisions are not the only class of decisions that animals and humans face. Another important class of decisions are value-based decisions. In perceptual decisions, the observer makes a choice on the basis of sensory information, whereas in value-based decisions the observer makes a choice using internal value signals (Milosavljevic et al., 2010; Tajima et al., 2019; Tajima et al., 2016). At first sight, it would appear that perceptual and value-based decisions should be very different because of the source of noise: In perceptual decisions there is uncertainty about an external state of the world because of ambiguous

or noisy sensory information, but in value-based decisions the value of an item is defined by the observer's own representation of that value. Some form of noise in value-based decisions seems necessary to account for variability in choices, and changes in performance with time pressure and difficulty (Milosavljevic et al., 2010). One suggestion is that noise arises because people consider different features of stimuli in a stochastic manner, and each of the features are associated with a different value (Milosavljevic et al., 2010; Tajima et al., 2016). Intuitively though, such a process would not lead to the moment-to-moment Gaussian noise that is a central component of DDM models (Ratcliff & McKoon, 2008).

In some perceptual decisions a major source of noise does not appear to be sensory, but rather introduced during computation (Drugowitsch et al., 2016). (The proportion of noise which is computational does appear to depend on the task used, as might be expected: Computational noise appears to have a greater effect, relative to sensory noise, when a perceptual task is set up in such a way that sensory noise is likely to be low; Stengård and van den Berg, 2019.) The finding that computation itself may be a major source of noise suggests that value-based and perceptual decisions may be more similar than they appear, as they will both be corrupted by noise from this source. Consistent with this view, choice and response times in value-based decisions have been successfully modelled using the DDM (Milosavljevic et al., 2010), and patterns in confidence in value-based decisions have been accounted for using the race model, one of the models of perceptual decision making discussed in Section 1.2.2 (Bogacz et al., 2006; De Martino et al., 2013). Further support for the idea that perceptual and value-based decisions rely on similar mechanisms comes from the finding that the optimal algorithm for making decisions on the basis of noisy value signals is very similar to the optimal algorithm for making decisions on the basis of noisy sensory information (Tajima et al., 2016). While we hope the work in this thesis might be relevant to value-based decisions because of these similarities, it does not seem reasonable to make any stronger a claim without empirical support.

One interesting question – that could be studied using a modified version of the experiment in Chapter 3 – is how perceptual and value information are combined. Value presumably enters into all perceptual decisions, with observers aiming to make responses that, on the basis of sensory information, will maximise reward (Ma & Jazayeri, 2014). We artificially remove any interesting value-based effects through experimental design, by providing no complex or probabilistic rewards, or even no explicit reward at all, and instead relying on the value people place on making correct responses. However, an owl's decision about whether to fly after a possible movement in the grass depends on both perceptual uncertainty (what did the owl see and how big was it), and value (how much is prey of that type and size worth).

The brain could take different approaches to combining this information. A simple approach would be to estimate a single value for prey size and then combine this information with the associated value (Lak et al., 2017). A more sophisticated strategy would be to estimate a probability distribution over prey size, and combine this with the corresponding values (Lak et al., 2017). Ideally, the observer would separately maintain and update a representation of the states of the world associated with sensory signals, and a representation of value associated with states of the world (Ma and Jazayeri, 2014; Marr, 2010/1982, p.102). This would be advantageous because the link between states and sensory signals changes independently of the link between state and value, and vice versa. For example, different sensory states are associated with different distributions over prey size, when the owl is hunting in the day and when hunting in the night, but the value of a medium-sized mouse is unchanged. Conversely, an owl that has eaten well will attach less value to a medium-sized mouse, but the link between sensory signals and world states is unchanged. Separately maintaining and updating these perceptual and value representations would allow rapid changes in behaviour as environmental conditions change (Ma & Jazayeri, 2014).

To study the combination of value and perceptual information using the experiment in Chapter 3, we could modify the experiment by including explicit rewards.

Previous work has manipulated rewards so that they are not equal for both options and must, therefore, be taken into account (Ratcliff & McKoon, 2008; Summerfield & Koechlin, 2010; Whiteley & Sahani, 2008). It would be interesting to use more complicated manipulations of value, beyond manipulating the value associated with different options. For example, we could manipulate the value associated with stimuli of different signal strengths. With the appropriate manipulation, we may be able to distinguish behaviour resulting from the different approaches for combining value and perceptual information (e.g. combining a perceptual point estimate with value vs. combining a probability distribution with a full representation of value).

Bringing together the question of how perceptual and value information are combined, and the finding that evidence accumulation can be used to model value-based decisions as well as perceptual decisions, an intriguing question is whether the "evidence" accumulation mechanism studied by researchers in perceptual decision-making is the very same "evidence" accumulation mechanism studied by researchers in value-based decision-making. If the evidence accumulation mechanism is downstream of the combination of perceptual and value information, then what might be accumulated is a signal that in some way relates to expected reward. Alternatively, evidence accumulation could be the point at which this information is combined (Ratcliff & McKoon, 2008; Summerfield & Koechlin, 2010). A fascinating question is what confidence reflects in combined perceptual and value tasks. In Chapter 2 and Chapter 3 we saw evidence that confidence is read out from the same process that determines the decision. If the evidence accumulation reflects both perceptual and value information then we would expect value information to affect confidence.

7.2.2 More challenges to Bayesian confidence

In Chapter 5 we saw that over-weighting of decision-congruent evidence does not appear to be a general phenomenon in perceptual confidence (Peters et al., 2017).

However, in Chapter 1 we mentioned a number of other results which appear to conflict with the idea of Bayesian confidence.

Adler and Ma (2018a) provided strong evidence against the idea of Bayesian confidence. The case was especially strong because Adler and Ma (2018a) carefully ruled out a very large number of explanations that might be given in defence of Bayesian confidence. The task involved determining which of two categories a stimulus belonged to. The categories were overlapping, so even in the absence of perceptual noise the correct response would have been ambiguous. Additionally, the level of reliability of the stimuli was varied (e.g. by changing contrast). As a Bayesian observer would, people took into account the reliability of the stimuli when reporting their confidence. However, confidence did not change in the precise manner expected from a Bayesian observer: When the stimulus was very reliable the Bayesian model underestimated or overestimated participants' confidence, depending on the task and response. Instead, the data were better explained by the idea that people took perceptual uncertainty into account in a heuristic manner, by comparing their measurement of the stimulus to a simple function of perceptual uncertainty when determining their response and confidence. This pattern persisted even when participants were provided with trial-to-trial feedback. The findings were also unaffected when Adler and Ma (2018a) removed the assumption that observers knew the standard deviation of variability in the stimuli and perceptual noise.

One possible explanation for these findings, that was not (one of the very many) ruled out, is that these effects arise due to the mechanics of a dynamic representation of evidence. The Bayesian model considered by Adler and Ma (2018a) used a static representation of evidence (Section 1.2.1). We have seen that complex dependencies can arise between confidence, presented evidence, decision time, and noise, when observers have a dynamic representation of evidence (e.g. equation 4.100 in Chapter 4). Of particular relevance is that a Bayesian readout, when based on a dynamic representation of evidence, takes into account time spent accumulating evidence (an effect we have referred to as difficulty discounting; Moran, 2015; Moreno-Bote, 2010).

However, by definition, time is not a factor in a Bayesian readout from a static representation of evidence. Therefore, it could be that Bayesian confidence based on a dynamic representation of evidence provides a good account of confidence, even when Bayesian confidence based on a static representation does not. An objection to this suggestion is that Adler and Ma (2018a) used stimuli that were presented only very briefly. While it is intuitive that briefly presented stimuli would generate a static representation of evidence, there is empirical support for the view that briefly presented stimuli generate an enduring representation, which then drives the typical evidence accumulation process (Ratcliff & Rouder, 2000; Ratcliff et al., 2016). Hence, even for briefly presented stimuli, a dynamic representation of evidence may provide a more accurate model.

Another set of findings suggest that people's confidence reflects probabilistic quantities, but not solely the probability of being correct. In two-alternative choices, there is evidence that observers may report a quantity that at least partially reflects their uncertainty about the relevant stimulus feature (e.g. uncertainty about stimulus orientation), rather than simply uncertainty in their choice (e.g. whether the stimulus orientation is to the left or right of vertical; Navajas et al., 2017). However, Adler and Ma (2018b) noted that the model used by Navajas et al. (2017) was not Bayesian. In particular, the model involved an estimate of average stimulus orientation that was continuously updated as new information arrived. It was not possible in this model for the observer to equally weight all information, as a Bayesian would. The observer's estimate of the probability of being correct was based on this non-Bayesian estimate of average stimulus orientation. Such a model could not account for qualitative patterns in the data, without confidence also reflecting the additional probabilistic quantity of stimulus uncertainty. However, as Adler and Ma (2018b) point out, because a Bayesian observer was not used, these results do not speak against the idea that confidence solely reflects the probability of being correct, according to a Bayesian observer.

In the context of three-alternative choices, Li and Ma (2020) recently provided evidence that confidence does not reflect the probability of being correct, but rather the difference between the probability of the best and second best options. In the task, participants had to classify a dot as belonging to one of three other clusters of dots (simultaneously presented) on the basis of its spatial position. This is an interesting finding, especially because in the context of two-alternative decisions, the idea that people report the probability of being correct and the idea that people report the difference between the best and second best options lead to identical predictions for the ordering of confidence reports (Li & Ma, 2020). (Throughout this thesis it is the ordering of confidence reports, only, which we have studied.) It will be important to see how far the result generalises; in the task used by Li and Ma (2020) stimuli were presented indefinitely, at high contrast, so there was presumably little perceptual uncertainty. In this situation the observer's understanding of the meaning of category membership may be crucial. Well-known heuristics associated with explicit probabilistic reasoning may be used, such as the representativeness heuristic (Tversky & Kahneman, 1974). Therefore, it will be very interesting to see if confidence reflects the same or a different quantity, in three-alternative choices made under high levels of perceptual uncertainty.

We have discussed empirical challenges to the idea of Bayesian confidence in Section 1.6, we explored two challenges in depth in Chapter 5, and we have discussed some outstanding challenges to this idea here. Throughout these discussions, one key theme to emerge is that a readout of the probability of being correct strongly depends on the mechanism being read from. Whether the mechanism features a static (Adler & Ma, 2018a) or dynamic representation of evidence, an estimate that is continuously updated (Navajas et al., 2017), or stimulus dependent noise (Chapter 5), are all crucial. A criticism of work that focuses on optimality has been that it has tended to under-emphasise mechanism (Jones & Love, 2011; Rahnev & Denison, 2018). This is a criticism we return to below (Section 7.3.3), but for now we note that in debates about whether confidence is Bayesian, focusing on underlying mechanism may be key to making progress.

7.3 Potential criticisms

While we have addressed specific limitations relevant to each chapter, in those chapters, certain criticisms apply to the motivation, approach and conclusions of the thesis as a whole.

7.3.1 Is the DDM better than alternatives?

We have argued that the DDM has many theoretical strengths (Chapter 1 and 6), and found that DDMs can provide excellent qualitative and quantitative accounts of confidence (Chapter 2 and 3). Nevertheless, one critique of the thesis could be that we have not shown the DDM to provide a better model for decisions and confidence than alternatives because no direct comparison has been conducted. For example, it could be that models in which partially correlated accumulators support the decision (Section 1.2.2; Kiani et al., 2014), can provide a better account of response, response time and confidence data.

We accept both that alternative models could provide a better description of the data, and that if a model provided better description and prediction of data then it would be superior to the DDM. Our aim at the outset was more limited: We aimed to show just that the DDM, so successful at describing decisions and response times, could also account for confidence reports. Including aiming to show that the DDM could account for apparently challenging findings such as the strong relationship between confidence and decision time. In doing so, we showed that we do not need to abandon the DDM in order to account for an important feature of behaviour. This concern may have been one of the motivations for the use of models featuring non-normative decision mechanisms in confidence research (De Martino et al., 2013; Kiani et al., 2014; van den Berg, Anandalingam et al., 2016; Zylberberg et al., 2012).

While we have argued the DDM is superior on theoretical grounds, the comparison of the DDM to alternatives is clearly ultimately an empirical issue. We had hoped to be able to expand the scope of the thesis, and run a direct comparison

between DDM models of confidence, and accounts which use partially correlated accumulators (e.g. Kiani et al., 2014). Such an extension may present extensive technical challenges and, as a result, this direct comparison lay beyond the scope of the present thesis. In Chapter 4 we derived predictions for confidence in DDM models. It would be a substantial undertaking to expand these derivations to other cases (for relevant previous results see Moreno-Bote, 2010; Shan et al., 2019). A key feature of the DDM, on which we relied heavily, is that at the time of the decision, we know the state of the evidence accumulator. It is at the decision threshold. In models with two partially correlated accumulators, we only know that the winning accumulator is at the decision threshold, whereas the losing accumulator could be anywhere below the threshold (Kiani et al., 2014). Because of this feature, we cannot apply the same trick we used in Chapter 4 and infer the state of the accumulator at the time of the decision, using the height of the decision threshold at that time. As a result we cannot largely bypass consideration of the path of the accumulators prior to a decision. Instead we have to deal with the complicated relationship that exists between presented evidence and the state of the accumulators, when decision thresholds are present (Section 4.1; Cox and Miller, 1965; Smith, 2000).

There are at least two other approaches which could be used with the aim of testing the claim that the DDM is a better model than alternatives. First, there is the brute-force approach: With enough computational time, we can use numerical approaches to solving the equations which govern the evolution of the evidence accumulators (Chang & Cooper, 1970; Shinn et al., 2020; Smith, 2000; Zylberberg et al., 2018). In this case, there are few constraints and one could compare different models with any combination of features (such as drift-rate variability, decision threshold shapes, and correlation among accumulators), using the trial-by-trial properties of any stimuli. How much time such a model comparison would take on a computer cluster is an open question. A second, more sophisticated approach would be to first explore the conditions under which the distinctive property of the DDM – a single accumulation of the difference in evidence – leads to clearly different predictions from other models. If such situations could be identified, there

would be no need for computationally expensive trial-by-trial modelling to draw out very subtle differences between models.

7.3.2 Are the models too flexible, or too constrained?

Throughout, we have focused heavily on Bayesian models. There are well-known and very reasonable critiques of the entire Bayesian approach to psychology and cognitive science (Bowers & Davis, 2012; Jones & Love, 2011). Perhaps the strongest reason to worry about the Bayesian approach is the idea that Bayesian models may be too flexible, and for this reason not falsifiable. The specific charge is that the Bayesian modeller can change the underlying assumptions to account for any pattern in the data (Bowers & Davis, 2012; Jones & Love, 2011; Rahnev & Denison, 2018). For example, the modeller could change the form of the noise corrupting evidence measurements, the type of evidence measurements made, the beliefs of the observer, or the observer's preferences. With any scientific theory, whether computational or verbal, it is always possible to change auxiliary assumptions, with the result that the predictions of the theory change (Quine, 1951, p.39-40). What matters, for a Bayesian account or any other, is whether those changes can be justified. We would agree that, frequently, substantial changes are made to the assumptions underlying Bayesian models without sufficient justification, or without those modified assumptions being verified in the future (Rahnev & Denison, 2018). Perhaps Bayesian models have been particularly vulnerable to this charge because the details of the observer model – which apart from the rule for making decisions on the basis of measurements are neither Bayesian nor non-Bayesian – are often under-emphasised (Rahnev & Denison, 2018).

Throughout, we have tried to emphasise the details of the observer model used, specifically, the DDM with its dynamic representation of the difference in evidence. We have tried to make clear that the Bayesian readout is a readout with respect to the evidence measurements collected by a DDM observer. In the absence of claims about the measurements that an observer receives, the term "Bayesian

confidence" is meaningless (Rahnev & Denison, 2018). Regarding flexibility in the choice of assumptions, in two places we have diverged from standard assumptions. We allowed the possibility that the observer's Bayesian readout for confidence is miscalibrated because they incorrectly estimate the magnitude of different sources of variability (Chapter 2 and 3; Drugowitsch et al., 2014), and we considered the possibility of stimulus-dependent noise (Chapter 5; Teodorescu et al., 2016). In both cases we tried to make the departure from usual assumptions explicit by, for example, frequently referring to the "miscalibrated Bayesian" readout. Both changes were justified with reference to previous findings (de Gardelle & Mamassian, 2015; Herce Castañón et al., 2019; Pardo-Vazquez et al., 2019; Teodorescu et al., 2016; Zylberberg et al., 2016; Zylberberg et al., 2014). A miscalibrated readout is especially plausible in the context of the experiment of Chapter 3, in which participants did not receive trial-by-trial feedback.

Nevertheless, we acknowledge there is more to be done in terms of justifying these claims. For example, regarding DDM accounts with a miscalibrated Bayesian readout, new preregistered studies with and without trial-by-trial feedback could provide a strong test of the assumptions made. Without trial-by-trial feedback, we expect similar results to those found here. With trial-by-trial feedback we expect the observer to become calibrated over time. Below we discuss another approach to testing the beliefs held by observers (Section 7.4).

While one potential criticism is that the models used have too much flexibility, another is that the models are too constrained. For example, we made particular assumptions about the form of metacognitive noise (Maniscalco & Lau, 2012). In particular, we assumed that normally distributed noise corrupts the variable that determines confidence (accumulated difference in evidence, or estimated log-posterior ratio) as it is read out (Chapter 2). However, there is evidence that the noise corrupting this variable increases with the variable itself (Shekhar & Rahnev, 2020). We saw another important example in Chapter 5: Throughout we have ignored the possibility that internally represented intensity is a nonlinear function of

stimulus intensity (Geisler, 1989; Pardo-Vazquez et al., 2019). It is always possible to add more complexity and detail to computational models, so there is a balance to strike between the insight which can be gained from simpler models, and the precision that can be gained from more complex models (Usher & McClelland, 2001). For the phenomena on which we focused, the approximations made appear to have been reasonable, but we consider ways to include further detail below (Section 7.4).

7.3.3 Is optimality a useful notion?

Another challenge that applies to the thesis as a whole is the idea that optimality is not a useful notion (Rahnev & Denison, 2018). This has been argued on the grounds that optimality is impossible to define in the absence of specific assumptions about the observer, such as the information measured by the observer, the form of noise which corrupts this information, and the value they attribute to different possible outcomes. A major motivation for this thesis has been the optimality of the DDM (Bogacz et al., 2006; Gold & Shadlen, 2007; Moran, 2015; Tajima et al., 2019; Wald & Wolfowitz, 1948). The DDM is certainly only optimal relative to specific assumptions about the information encoded, and the variance of this information (as discussed in Section 1.2.2). However, we can say that DDM observers make the best use of the information available to them. On the other hand, an observer who uses the race model or partially-correlated evidence accumulators is not optimal, even relative to the assumptions made by these models. That is, these observers do not make best use of the information that is available to them (Chapter 1). It is this wastefulness which we label as suboptimal, and hypothesise would have been removed through evolutionary processes.

We agree that focus on optimality has sometimes had the unhelpful effect of obscuring or reducing focus on debates about the mechanisms underlying behaviour (Jones & Love, 2011; Rahnev & Denison, 2018). As optimality can only be defined relative to hypothesised mechanism, questions about mechanism are at least as important as questions about optimality (Rahnev & Denison, 2018). In relation to

this thesis, a key investigation for the future will be one we mentioned very briefly in Chapter 1, regarding the mechanism responsible for Bayesian confidence. Bayesian confidence could arise either because brains actually represent and compute with probability distributions, or because people gradually learn a mapping from sensory signals to confidence (Kiani & Shadlen, 2009; Ma & Jazayeri, 2014). If we propose that people actually compute with probability distributions, then the nature of the representations used will be of central importance (for candidate representations see e.g. Jazayeri and Movshon, 2006; Ma et al., 2006; Sahani and Dayan, 2003). For example, an interesting question will be whether people represent probability distributions of arbitrary complexity, or if they are limited to distributions like the normal, which can be characterised simply by their mean and variance (Chetverikov et al., 2017). If we believe that observers learn a mapping, then there will be important questions regarding the learning mechanism used. For example, we would want to know how flexible the mapping is, the parameters of the mapping, and how rapidly these parameters can be estimated in a new task or context.

7.4 Future directions

We have seen in the discussion above that the models we have used can be criticised both for being too flexible and too constrained. One approach we would like to try in the future is increasing the flexibility of models, but in a way which makes this flexibility completely explicit (Stocker & Simoncelli, 2006). The behaviour of a Bayesian observer is dependent on the sources of noise corrupting their measurements, their prior beliefs, and the value they attach to the various possible outcomes (Ma & Jazayeri, 2014). In situations where we have multiple hypotheses about the form of the noise corrupting measurements, or the beliefs of the observer, these parts of the model should be flexible and have associated parameters which are also fit (Stocker & Simoncelli, 2006).

There are arguably several advantages, and some difficulties with this approach. In terms of advantages, such an approach would allow us to test claims about

whether behaviour follows Bayesian principles, regardless of the beliefs and precise form of the measurements feeding into the decision stage. While such models would have far more flexibility, this flexibility would be explicit, and hence penalised by model comparison measures. Furthermore, to the extent that such a model outperforms other models and is viewed as a good description of behaviour, the fitted values of the parameters for the model would give us information about the noise corrupting measurements, and the prior beliefs held by the observer. In the context of speed perception, using this approach Stocker and Simoncelli (2006) found evidence for non-Gaussian priors, and noise that scaled with speed. In this thesis, we have considered some flexibility for priors (Chapter 2 and 3), and noise (Chapter 5), but it would be fascinating to explore a much more flexible parameterisation.

With more parameters, much more data is likely to be required just to adequately fit such flexible models, but also to ensure that model comparison is reliable. However, larger and larger amounts of data are becoming publicly available thanks to the open science movement (Rahnev et al., 2020). With databases containing large numbers of studies it will be far easier to assess the extent to which particular models provide a consistent account across a wide range of contexts, and the extent to which models can predict completely unseen data. Regardless of how flexible a model is, if it can predict data that it has not been trained on better than its competitors, then we can be assured that the additional flexibility is not simply fitting noise (Yarkoni & Westfall, 2017).

7.5 Conclusion

Regarding the mechanism of perceptual decisions, we focused on the DDM, providing both theoretical and empirical evidence justifying its arguable status as the most successful model of response times and choices. We saw that the idea of a Bayesian readout of the probability of being correct, using the measurements gathered by a DDM observer, offers intuitive explanations for a diverse and complex pattern of effects in confidence data. While a greater focus on the mechanism underlying

confidence reports is required, we suggest that a combination of the DDM and miscalibrated Bayesian confidence provides great promise for accurately describing choices, response times and confidence. We look forward to seeing how much more data, and how many more effects, this combination of ideas can explain.

References

Acerbi, L., Dokka, K., Angelaki, D. E. & Ma, W. J. (2018). Bayesian comparison of explicit and implicit causal inference strategies in multisensory heading perception. *PLOS Computational Biology*, *14*(7), e1006110. https://doi.org/10.1371/journal.pcbi.1006110

Adler, W. T. & Ma, W. J. (2018a). Comparing bayesian and non-bayesian accounts of human confidence reports (S. J. Gershman, Ed.). *PLOS Computational Biology*, *14*(11), e1006572. https://doi.org/10.1371/journal.pcbi.1006572

Adler, W. T. & Ma, W. J. (2018b). Limitations of proposed signatures of bayesian confidence. *Neural Computation*, *30*(12), 3327–3354. https://doi.org/10.1162/neco_a_01141

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

Aitchison, L., Bang, D., Bahrami, B. & Latham, P. E. (2015). Doubly bayesian analysis of confidence in perceptual decision-making. *PLoS Computational Biology*, *11*(10), 1–23. https://doi.org/10.1371/journal.pcbi.1004519

Babchishin, K. M. & Helmus, L.-M. (2016). The influence of base rates on correlations: An evaluation of proposed alternative effect sizes with real-world data. *Behavior Research Methods*, *48*(3), 1021–1031. https://doi.org/10.3758/s13428-015-0627-7

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

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

Bang, J. W., Shekhar, M. & Rahnev, D. (2019). Sensory noise increases metacognitive efficiency. *Journal of experimental psychology. General*, *148*(3). https://doi.org/10.1037/xge0000511

Baranski, J. V. & Petrusic, W. M. (1998). Probing the locus of confidence judgments: Experiments on the time to determine confidence. *Journal of experimental psychology. Human perception and performance*, *24*(3), 929–945. https://doi.org/10.1037/0096-1523.24.3.929

Baum, C. & Veeravalli, V. (1994). A sequential procedure for multihypothesis testing. *IEEE Transactions on Information Theory*, *40*(6), 1994–2007. https://doi.org/10.1109/18.340472

Bergstra, J., Yamins, D. & Cox, D. D. (2013). Making a science of model search: Hyperparameter optimization in hundreds of dimensions for vision architectures. *Proceedings of the 30th International Conference on International Conference on Machine Learning - Volume 28*, I–115–I–123.

Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer.

Bitzer, S., Park, H., Blankenburg, F. & Kiebel, S. J. (2014). Perceptual decision making: Drift-diffusion model is equivalent to a bayesian model. *Frontiers in Human Neuroscience*, *8*, 1–17. https://doi.org/10.3389/fnhum.2014.00102

Bogacz, R., Brown, E., Moehlis, J., Holmes, P. & Cohen, J. D. (2006). The physics of optimal decision making: A formal analysis of models of performance in two-alternative forced-choice tasks. *Psychological Review*, *113*(4), 700–765. https://doi.org/10.1037/0033-295X.113.4.700

Bogacz, R. & Gurney, K. (2007). The basal ganglia and cortex implement optimal decision making between alternative actions. *Neural Computation*, *19*(2), 442–477. https://doi.org/10.1162/neco.2007.19.2.442

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

Bowers, J. S. & Davis, C. J. (2012). Bayesian just-so stories in psychology and neuroscience. *Psychological Bulletin*, *138*(3), 389–414. https://doi.org/10.1037/a0026450

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*(4), 433–436. https://doi.org/10.1163/156856897X00357

Bromiley, P. A. (2018). *Products and convolutions of gaussian probability density functions*. Retrieved 2020, from http://www.tina-vision.net/docs/memos/2003-003.pdf

Brown, S. D., Ratcliff, R. & Smith, P. L. (2006). Evaluating methods for approximating stochastic differential equations. *Journal of Mathematical Psychology*, *50*(4), 402–410. https://doi.org/10.1016/j.jmp.2006.03.004

Calder-Travis, J., Bogacz, R. & Yeung, N. (2020). Bayesian confidence for drift diffusion observers in dynamic stimuli tasks. *bioRxiv*. https://doi.org/10.1101/2020.02.25.965384

Calder-Travis, J., Charles, L., Bogacz, R. & Yeung, N. (2020). *Bayesian confidence in optimal decisions* (preprint). PsyArXiv. https://doi.org/10.31234/osf.io/j8sxz

Calder-Travis, J. & Ma, W. J. (2020). Explaining the effects of distractor statistics in visual search. *bioRxiv*. https://doi.org/10.1101/2020.01.03.893057

Calder-Travis, J. & Slade, E. (2020). Rapid acquisition of near optimal stopping using bayesian by-products [Publisher: Cold Spring Harbor Laboratory Section: New Results]. *bioRxiv*, 844969. https://doi.org/10.1101/844969

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*. https://doi.org/10.1037/xhp0000747

Chang, J. & Cooper, G. (1970). A practical difference scheme for fokker-planck equations. *Journal of Computational Physics*, *6*(1), 1–16. https://doi.org/10.1016/0021-9991(70)90001-X

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

Cheadle, S., Wyart, V., Tsetsos, K., Myers, N., de Gardelle, V., Herce Castañón, S. & Summerfield, C. (2014). Adaptive gain control during human perceptual choice. *Neuron*, *81*(6), 1429–1441. https://doi.org/10.1016/j.neuron.2014.01.020

Chetverikov, A., Campana, G. & Kristjánsson, Á. (2017). Rapid learning of visual ensembles. *Journal of Vision*, *17*(2). https://doi.org/10.1167/17.2.21

Cisek, P., Puskas, A., El-Murr, S., Puskas, G. A., El-Murr, S., Puskas, A. & El-Murr, S. (2009). Decisions in changing conditions: The urgency-gating model. *Journal of Neuroscience*, *29*(37), 11560–11571. https://doi.org/10.1523/JNEUROSCI.1844-09.2009

Cobb, P. W. (1932). Weber's law and the fechnerian muddle. *Psychological Review*, *39*(6), 533–551.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Erlbaum Associates.

Cox, D. R. & Miller, H. D. (1965). *The theory of stochastic processes*. Methuen.

De Martino, B., Fleming, S. M., Garrett, N. & Dolan, R. J. (2013). Confidence in value-based choice. *Nature Neuroscience*, *16*, 105–110. https://doi.org/10.1038/nn.3279

de Gardelle, V. & Mamassian, P. (2015). Weighting mean and variability during confidence judgments. *PLoS ONE*, *10*(3), 1–11. https://doi.org/10.1371/journal.pone.0120870

Desender, K., Boldt, A., Verguts, T. & Donner, T. H. (2019). Confidence predicts speed-accuracy tradeoff for subsequent decisions (R. Kiani, M. J. Frank & S. P. Kelly, Eds.). *eLife*, *8*, e43499. https://doi.org/10.7554/eLife.43499

Desender, K., Boldt, A. & Yeung, N. (2018). Subjective confidence predicts information seeking in decision making. *Psychological Science*, *29*(5), 761–778. https://doi.org/10.1177/0956797617744771

Desender, K., Donner, T. H. & Verguts, T. (2020). Dynamic expressions of confidence within an evidence accumulation framework. *bioRxiv*. https://doi.org/10.1101/2020.02.18.953778

Diederich, A. & Busemeyer, J. R. (2003). Simple matrix methods for analyzing diffusion models of choice probability, choice response time, and simple response time. *Journal of Mathematical Psychology*, *47*(3), 304–322. https://doi.org/10.1016/S0022-2496(03)00003-8

Dragalin, V., Tartakovsky, A. & Veeravalli, V. (1999). Multihypothesis sequential probability ratio tests .i. asymptotic optimality. *IEEE Transactions on Information Theory*, *45*(7), 2448–2461. https://doi.org/10.1109/18.796383

Drugowitsch, J. (2016). Fast and accurate monte carlo sampling of first-passage times from wiener diffusion models. *Scientific Reports*, *6*(1), 20490. https://doi.org/10.1038/srep20490

Drugowitsch, J., Mendonça, A. G., Mainen, Z. F. & Pouget, A. (2019). Learning optimal decisions with confidence. *Proceedings of the National Academy of Sciences*, *116*(49), 24872–24880. https://doi.org/10.1073/pnas.1906787116

Drugowitsch, J., Moreno-Bote, R., Churchland, A. K., Shadlen, M. N. & Pouget, A. (2012). The cost of accumulating evidence in perceptual decision making. *Journal of Neuroscience*, *32*(11), 3612–3628. https://doi.org/10.1523/JNEUROSCI.4010-11.2012

Drugowitsch, J., Moreno-Bote, R. & Pouget, A. (2014). Relation between belief and performance in perceptual decision making. *PLoS ONE*, *9*(5). https://doi.org/10.1371/journal.pone.0096511

Drugowitsch, J., Wyart, V., Devauchelle, A.-D. & Koechlin, E. (2016). Computational precision of mental inference as critical source of human choice suboptimality. *Neuron*, *92*(6), 1–14. https://doi.org/10.1016/j.neuron.2016.11.005

Ernst, M. O. & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*(6870), 429–433. https://doi.org/10.1038/415429a

Evans, N. J., Hawkins, G. E. & Brown, S. D. (2020). The role of passing time in decision-making. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *46*(2), 316–326. https://doi.org/10.1037/xlm0000725

Faul, F., Erdfelder, E., Lang, A.-G. & Buchner, A. (2007). G*power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

Ferguson, T. S. (2011). *Optimal stopping and applications*. Retrieved 2019, from https://www.math.ucla.edu/~tom/Stopping/Contents.html

Festinger, L. (1943). Studies in decision: I. decision-time, relative frequency of judgment and subjective confidence as related to physical stimulus difference. *Journal of Experimental Psychology*, *32*(4), 291–306. https://doi.org/10.1037/h0056685

Fleming, S. M. & Daw, N. D. (2017). Self-evaluation of decision-making: A general bayesian framework for metacognitive computation. *Psychological Review*, *124*(1), 91–114. https://doi.org/10.1037/rev0000045

Fleming, S. M., Maniscalco, B., Ko, Y., Amendi, N., Ro, T. & Lau, H. (2015). Action-specific disruption of perceptual confidence. *Psychological Science*, *26*(1), 89–98. https://doi.org/10.1177/0956797614557697

Gajdos, T., Fleming, S. M., Garcia, M. S., Weindel, G. & Davranche, K. (2019). Revealing subthreshold motor contributions to perceptual confidence [WOS:000462170200001]. *Neuroscience of Consciousness*, *5*(1), niz001. https://doi.org/10.1093/nc/niz001

Garrett, H. E. (1922). *A study of the relation of accuracy to speed,*

Geisler, W. S. (1989). Sequential ideal-observer analysis of visual discriminations. *Psychological Review*, *96*(2), 267–314. https://doi.org/10.1037/0033-295X.96.2.267

Gold, J. I. & Shadlen, M. N. (2007). The neural basis of decision making. *Annual Review of Neuroscience*, *30*(1), 535–574. https://doi.org/10.1146/annurev.neuro.29.051605.113038

Green, D. M. & Swets, J. A. (1966). *Signal detection theory and psychophysics*. Wiley.

Hauser, T. U., Allen, M., Rees, G. & Dolan, R. J. (2017). Metacognitive impairments extend perceptual decision making weaknesses in compulsivity. *Scientific Reports*, *7*(1). https://doi.org/10.1038/s41598-017-06116-z

Hawkins, G. E., Forstmann, B. U., Wagenmakers, E.-J., Ratcliff, R. & Brown, S. D. (2015). Revisiting the evidence for collapsing boundaries and urgency signals in perceptual decision-making. *Journal of Neuroscience*, *35*(6), 2476–2484. https://doi.org/10.1523/JNEUROSCI.2410-14.2015

Hawkins, G. E., Wagenmakers, E.-J., Ratcliff, R. & Brown, S. D. (2015). Discriminating evidence accumulation from urgency signals in speeded decision making. *Journal of Neurophysiology*, *114*(1), 40–47. https://doi.org/10.1152/jn.00088.2015

Heitz, R. P. (2014). The speed-accuracy tradeoff: History, physiology, methodology, and behavior. *Frontiers in Neuroscience*, *8*. https://doi.org/10.3389/fnins.2014.00150

Henmon, V. A. C. (1911). The relation of the time of a judgement to its accuracy [Institution: University of Wisconsin]. *Psychological Review*, *18*(3), 186–201.

Herce Castañón, S., Moran, R., Ding, J., Egner, T., Bang, D. & Summerfield, C. (2019). Human noise blindness drives suboptimal cognitive inference. *Nature Communications*, *10*(1). https://doi.org/10.1038/s41467-019-09330-7

Hertz, U., Palminteri, S., Brunetti, S., Olesen, C., Frith, C. D. & Bahrami, B. (2017). Neural computations underpinning the strategic management of influence in advice giving. *Nature Communications*, *8*(1), 1–12. https://doi.org/10.1038/s41467-017-02314-5

Honig, M., Ma, W. J. & Fougnie, D. (2020). Humans incorporate trial-to-trial working memory uncertainty into rewarded decisions. *Proceedings of the National Academy of Sciences*, *117*(15), 8391–8397. https://doi.org/10.1073/pnas.1918143117

Jang, Y., Wallsten, T. S. & Huber, D. E. (2012). A stochastic detection and retrieval model for the study of metacognition. *119*(1), 186–200. https://doi.org/10.1037/a0025960

Jazayeri, M. & Movshon, J. A. (2006). Optimal representation of sensory information by neural populations. *Nature Neuroscience*, *9*(5), 690–696. https://doi.org/10.1038/nn1691

Jones, M. & Love, B. C. (2011). Bayesian fundamentalism or enlightenment? on the explanatory status and theoretical contributions of bayesian models of cognition. *Behavioral and Brain Sciences*, *34*(4), 169–188. https://doi.org/10.1017/S0140525X10003134

Kelly, S. P. & O'Connell, R. G. (2013). Internal and external influences on the rate of sensory evidence accumulation in the human brain. *J Neurosci*, *33*(50), 19434–19441. https://doi.org/10.1523/JNEUROSCI.3355-13.2013

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

Kiani, R., Hanks, T. D. & Shadlen, M. N. (2008). Bounded integration in parietal cortex underlies decisions even when viewing duration is dictated by the environment. *J Neurosci*, *28*(12), 3017–3029. https://doi.org/10.1523/JNEUROSCI.4761-07.2008

Kiani, R. & Shadlen, M. N. (2009). Representation of confidence associated with a decision by neurons in the parietal cortex. *Science*, *324*(5928), 759–764. https://doi.org/10.1126/science.1169405

Kleiner, M., Brainard, D. & Pelli, D. (2007). What's new in psychtoolbox-3? [ECVP Abstract Supplement]. *Perception*, *36*.

Koizumi, A., Maniscalco, B. & Lau, H. (2015). Does perceptual confidence facilitate cognitive control? *Attention, Perception, & Psychophysics*, *77*(4), 1295–1306. https://doi.org/10.3758/s13414-015-0843-3

Kuhn, T. S. (1978). *The essential tension : Selected studies in scientific tradition and change*. University of Chicago Press.

Kuhn, T. S. (2012). *The structure of scientific revolutions: 50th anniversary edition*. University of Chicago Press.

Lak, A., Nomoto, K., Keramati, M., Sakagami, M. & Kepecs, A. (2017). Midbrain dopamine neurons signal belief in choice accuracy during a perceptual decision. *Current Biology*, *27*(6), 1–12. https://doi.org/10.1016/j.cub.2017.02.026

LeCun, Y., Bottou, L., Orr, G. B. & Müller, K.-R. (1998). Efficient BackProp. In G. B. Orr & K.-R. Müller (Eds.), *Neural networks: Tricks of the trade* (pp. 9–50). Springer Berlin Heidelberg. https://doi.org/10.1007/3-540-49430-8_2

Lee, M. S. Y., Jago, J. B., García-Bellido, D. C., Edgecombe, G. D., Gehling, J. G. & Paterson, J. R. (2011). Modern optics in exceptionally preserved eyes of early cambrian arthropods from australia. *Nature*, *474*(7353), 631–634. https://doi.org/10.1038/nature10097

Lennie, P. (2003). The cost of cortical computation. *Current Biology*, *13*(6), 493–497. https://doi.org/10.1016/S0960-9822(03)00135-0

Lewandowsky, S. & Farrell, S. (2011). *Computational modeling in cognition: Principles and practice*. SAGE Publications, Inc. https://doi.org/10.4135/9781483349428

Li, H.-H. & Ma, W. J. (2020). Confidence reports in decision-making with multiple alternatives violate the bayesian confidence hypothesis. *Nature Communications*, *11*. https://doi.org/10.1038/s41467-020-15581-6

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

Lieder, F. & Griffiths, T. L. (2020). Resource-rational analysis: Understanding human cognition as the optimal use of limited computational resources. *Behavioral and Brain Sciences*, *43*. https://doi.org/10.1017/S0140525X1900061X

Ljung, L. (1999). *System identification. theory for the user.* (2 nd.). Prentice Hall.

Long, J. S. (1997). *Regression models for categorical and limited dependent variables.* Sage.

Luce, R. D. (1986). *Response times: Their role in inferring elementary mental organization.* Oxford University Press.

Ma, W. J., Beck, J. M., Latham, P. E. & Pouget, A. (2006). Bayesian inference with probabilistic population codes. *Nature Neuroscience*, *9*(11), 1432–8. https://doi.org/10.1038/nn1790

Ma, W. J. & Jazayeri, M. (2014). Neural coding of uncertainty and probability. *Annual Review of Neuroscience*, *37*(1), 205–220. https://doi.org/10.1146/annurev-neuro-071013-014017

Ma, W. J., Navalpakkam, V., Beck, J. M., van den Berg, R. & Pouget, A. (2011). Behavior and neural basis of near-optimal visual search. *Nature Neuroscience*, *1*(6), 783–790. https://doi.org/10.1038/nn.2814

Malhotra, G., Leslie, D. S., Ludwig, C. J. H. & Bogacz, R. (2017). Overcoming indecision by changing the decision boundary. *Journal of Experimental Psychology: General*, *146*(6), 776–805. https://doi.org/10.1037/xge0000286

Malhotra, G., Leslie, D. S., Ludwig, C. J. H. & Bogacz, R. (2018). Time-varying decision boundaries : Insights from optimality analysis. *Psychonomic Bulletin & Review*, *25*, 971–996. https://doi.org/10.3758/s13423-017-1340-6

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

Maniscalco, B. & Lau, H. (2016). The signal processing architecture underlying subjective reports of sensory awareness. *Neuroscience of Consciousness*, *2016*. https://doi.org/10.1093/nc/niw002

Maniscalco, B., Peters, M. A. K. & Lau, H. (2016). Heuristic use of perceptual evidence leads to dissociation between performance and metacognitive sensitivity.

Attention, perception & psychophysics, *78*(3), 923–37. https://doi.org/10.3758/s13414-016-1059-x

Marr, D. (2010). *Vision: A computational investigation into the human representation and processing of visual information* [(Original work published 1982.)]. MIT Press.

Matlab optimization toolbox (Version 8.0). (2017). Natick, Massachusetts, United States, MathWorks.

Mazyar, H., van den Berg, R. & Ma, W. J. (2012). Does precision decrease with set size? *Journal of Vision*, *12*(6). https://doi.org/10.1167/12.6.10

McMillen, T. & Holmes, P. (2006). The dynamics of choice among multiple alternatives. *Journal of Mathematical Psychology*, *50*(1), 30–57. https://doi.org/10.1016/j.jmp.2005.10.003

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

Miller, J. & Ulrich, R. (2003). Simple reaction time and statistical facilitation: A parallel grains model. *Cognitive Psychology*, *46*(2), 101–151. https://doi.org/10.1016/S0010-0285(02)00517-0

Milosavljevic, M., Malmaud, J., Huth, A., Koch, C. & Rangel, A. (2010). The drift diffusion model can account for the accuracy and reaction time of value-based choices under high and low time pressure. *Judgment and Decision Making*, *5*(6), 437–449. https://doi.org/10.2139/ssrn.1901533

Miyoshi, K. & Lau, H. (2020). A decision-congruent heuristic gives superior metacognitive sensitivity under realistic variance assumptions. *Psychological Review*. https://doi.org/10.1037/rev0000184

Moran, R. (2015). Optimal decision making in heterogeneous and biased environments. *Psychonomic Bulletin & Review*, *22*(1), 38–53. https://doi.org/10.3758/s13423-014-0669-3

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

Moreno-Bote, R. (2010). Decision confidence and uncertainty in diffusion models with partially correlated neuronal integrators. *Neural computation*, *22*(7), 1786–1811. https://doi.org/10.1162/neco.2010.12-08-930

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*, 1–23. https://doi.org/10.7554/eLife.11946

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

Navarro, D. J. & Fuss, I. G. (2009). Fast and accurate calculations for first-passage times in wiener diffusion models. *Journal of Mathematical Psychology*, *53*(4), 222–230. https://doi.org/10.1016/j.jmp.2009.02.003

Norris, D. (2006). The bayesian reader: Explaining word recognition as an optimal bayesian decision process. *Psychological Review*, *113*(2), 327–357. https://doi.org/10.1037/0033-295X.113.2.327

O'Connell, R. G., Dockree, P. M. & Kelly, S. P. (2012). A supramodal accumulation-to-bound signal that determines perceptual decisions in humans. *Nature Neuroscience*, *15*(12), 1729–1735. https://doi.org/10.1038/nn.3248

Okazawa, G., Sha, L., Purcell, B. A. & Kiani, R. (2018). Psychophysical reverse correlation reflects both sensory and decision-making processes. *Nature Communications*, *9*. https://doi.org/10.1038/s41467-018-05797-y

Øksendal, B. K. (2013). *Stochastic differential equations: An introduction with applications* (Sixth edition, sixth corrected printing). Springer.

Owen, D. B. (1980). A table of normal integrals: A table. *Communications in Statistics - Simulation and Computation*, *9*(4), 389–419. https://doi.org/10.1080/03610918008812164

Palestro, J. J., Weichart, E., Sederberg, P. B. & Turner, B. M. (2018). Some task demands induce collapsing bounds: Evidence from a behavioral analysis. *Psychonomic Bulletin and Review*, *25*, 1225–1248. https://doi.org/10.3758/s13423-018-1479-9

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

Palmer, J., Ames, C. T. & Lindsey, D. T. (1993). Measuring the effect of attention on simple visual search. *Journal of Experimental Psychology*, *19*(1), 108–130.

Palmer, J., Verghese, P. & Pavel, M. (2000). The psychophysics of visual search. *Vision Research*, *40*(10), 1227–1268. https://doi.org/10.1016/S0042-6989(99)00244-8

Pardo-Vazquez, J. L., Castiñeiras-de Saa, J. R., Valente, M., Damião, I., Costa, T., Vicente, M. I., Mendonça, A. G., Mainen, Z. F. & Renart, A. (2019). The mechanistic foundation of weber's law. *Nature Neuroscience*, *22*(9), 1493–1502. https://doi.org/10.1038/s41593-019-0439-7

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*(4), 437–442. https://doi.org/10.1163/156856897X00366

Peters, M. A. K., Thesen, T., Ko, Y. D., Maniscalco, B., Carlson, C., Davidson, M., Doyle, W., Kuzniecky, R., Devinsky, O., Halgren, E. & Lau, H. (2017). Perceptual confidence neglects decision-incongruent evidence in the brain. *Nature Human Behaviour*, *1*(7). https://doi.org/10.1038/s41562-017-0139

Peterson, W., Birdsall, T. & Fox, W. (1954). The theory of signal detectability. *Transactions of the IRE Professional Group on Information Theory*, *4*(4), 171–212. https://doi.org/10.1109/TIT.1954.1057460

Petrusic, W. M. & Baranski, J. V. (2009). Probability assessment with response times and confidence in perception and knowledge. *Acta Psychologica*, *130*(2), 103–114. https://doi.org/10.1016/j.actpsy.2008.10.008

Pew, R. W. (1969). The speed-accuracy operating characteristic [Publisher: Elsevier BV]. *Acta Psychologica*, *30*, 16–26. https://doi.org/10.1016/0001-6918(69)90035-3

Pike, R. (1973). Response latency models for signal detection [Institution: U. Queensland, St. Lucia, Australia]. *Psychological Review*, *80*(1), 53–68.

Pilly, P. K. & Seitz, A. R. (2009). What a difference a parameter makes: A psychophysical comparison of random dot motion algorithms. *Vision Research*, *49*(13), 1599–1612. https://doi.org/10.1016/j.visres.2009.03.019

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

Quine, W. V. (1951). Two dogmas of empiricism. *The Philosophical Review*, *60*(1), 20–43. https://doi.org/10.2307/2181906

Rafiei, F. & Rahnev, D. (2020). Does the diffusion model account for the effects of speed-accuracy tradeoff on response times? [Publisher: PsyArXiv]. https://doi.org/10.31234/osf.io/bhj85

Rahnev, D. & Denison, R. N. (2018). Suboptimality in perceptual decision making. *Behavioral and Brain Sciences*, *41*. https://doi.org/10.1017/S0140525X18000936

Rahnev, D., Desender, K., Lee, A. L. F., Adler, W. T., Aguilar-Lleyda, D., Akdoğan, B., Arbuzova, P., Atlas, L. Y., Balcı, F., Bang, J. W., Bègue, I., Birney, D. P., Brady, T. F., Calder-Travis, J., Chetverikov, A., Clark, T. K., Davranche, K., Denison, R. N., Dildine, T. C., … Zylberberg, A. (2020). The confidence database [Number: 3 Publisher: Nature Publishing Group]. *Nature Human Behaviour*, *4*(3), 317–325. https://doi.org/10.1038/s41562-019-0813-1

Rahnev, D., Koizumi, A., McCurdy, L. Y., D'Esposito, M. & Lau, H. (2015). Confidence leak in perceptual decision making. *Psychological Science*, *26*(11), 1664–1680. https://doi.org/10.1177/0956797615595037

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*(2), 59–108.

Ratcliff, R. (2006). Modeling response signal and response time data. *Cognitive Psychology*, *53*(3), 195–237. https://doi.org/10.1016/j.cogpsych.2005.10.002

Ratcliff, R., Gomez, P. & McKoon, G. (2004). A diffusion model account of the lexical decision task. *Psychological Review*, *111*(1), 159–182. https://doi.org/10.1037/0033-295X.111.1.159

Ratcliff, R. & McKoon, G. (2008). The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Computation*, *20*(4), 873–922. https://doi.org/10.1162/neco.2008.12-06-420

Ratcliff, R. & Rouder, J. N. (1998). Modeling response times for two-choice decisions. *Psychological Science*, *9*(5), 347–356. https://doi.org/10.1111/1467-9280.00067

Ratcliff, R. & Rouder, J. N. (2000). A diffusion model account of masking in two-choice letter identification [Institution: (1)Department of Psychology, Northwestern University.]. *Journal of Experimental Psychology*, *26*(1), 127–140.

Ratcliff, R. & Smith, P. L. (2004). A comparison of sequential sampling models for two-choice reaction time. *Psychological Review*, *111*(2), 333–367. https://doi.org/10.1037/0033-295X.111.2.333

Ratcliff, R., Smith, P. L., Brown, S. D. & McKoon, G. (2016). Diffusion decision model: Current issues and history. *Trends in Cognitive Sciences*, *20*(4), 260–281. https://doi.org/10.1016/j.tics.2016.01.007

Ratcliff, R. & Tuerlinckx, F. (2002). Estimating parameters of the diffusion model: Approaches to dealing with contaminant reaction times and parameter variability. *Psychonomic Bulletin & Review*, *9*(3), 438–481. https://doi.org/10.3758/BF03196302

Ratcliff, R., Van Zandt, T. & Mckoon, G. (1999). Connectionist and diffusion models of reaction time. *Psychological review*, *106*(2), 261–300. https://doi.org/10.1037/0033-295X.106.2.261

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

Rollwage, M., Loosen, A., Hauser, T. U., Moran, R., Dolan, R. J. & Fleming, S. M. (2020). Confidence drives a neural confirmation bias [Number: 1 Publisher:

Nature Publishing Group]. *Nature Communications*, *11*(1), 1–11. https://doi.org/10.1038/s41467-020-16278-6

Rosenholtz, R. (2001). Visual search for orientation among heterogeneous distractors: Experimental results and implications for signal-detection theory models of search. *Journal of Experimental Psychology: Human Perception and Performance*, *27*(4), 985–999. https://doi.org/10.1037/0096-1523.27.4.985

Ross, S. M. (2014). Simulation, bootstrap statistical methods, and permutation tests. In S. M. Ross (Ed.), *Introduction to probability and statistics for engineers and scientists (fifth edition)* (pp. 619–646). Academic Press. https://doi.org/10.1016/B978-0-12-394811-3.50015-0

Rouault, M., Seow, T., Gillan, C. M. & Fleming, S. M. (2018). Psychiatric symptom dimensions are associated with dissociable shifts in metacognition but not task performance. *Biological Psychiatry*, *84*(6), 443–451. https://doi.org/10.1016/j.biopsych.2017.12.017

Sahani, M. & Dayan, P. (2003). Doubly distributional population codes: Simultaneous representation of uncertainty and multiplicity. *Neural Computation*, *15*, 2255–2279.

Samaha, J., Barrett, J. J., Sheldon, A. D., LaRocque, J. J. & Postle, B. R. (2016). Dissociating perceptual confidence from discrimination accuracy reveals no influence of metacognitive awareness on working memory [Publisher: Frontiers]. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.00851

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

Shadlen, M. N., Hanks, T. D., Churchland, A. K., Kiani, R. & Yang, T. (2006). The speed and accuracy of a simple perceptual decision: A mathematical primer. In K. Doya, S. Ishii, A. Pouget & R. P. Rao (Eds.), *Bayesian brain* (pp. 208–237). The MIT Press. https://doi.org/10.7551/mitpress/9780262042383.003.0010

Shadlen, M. N. & Kiani, R. (2013). Decision making as a window on cognition. *Neuron*, *80*(3), 791–806. https://doi.org/10.1016/j.neuron.2013.10.047

Shan, H., Moreno-Bote, R. & Drugowitsch, J. (2019). Family of closed-form solutions for two-dimensional correlated diffusion processes. *Phys. Rev. E*, *100*(3). https://doi.org/10.1103/PhysRevE.100.032132

Shekhar, M. & Rahnev, D. (2020). *The nature of metacognitive inefficiency in perceptual decision making* (preprint). PsyArXiv. https://doi.org/10.31234/osf.io/g9qzh

Shen, S. & Ma, W. J. (2019). Variable precision in visual perception. *Psychological Review*, *126*(1), 89–132. https://doi.org/10.1037/rev0000128

Shinn, M., Lam, N. H. & Murray, J. D. (2020). *A flexible framework for simulating and fitting generalized drift-diffusion models* (preprint). Neuroscience. https://doi.org/10.1101/2020.03.14.992065

Shorrocks, B. (2016). *The giraffe: Biology, ecology, evolution and behaviour*. John Wiley & Sons.

Siedlecka, M., Paulewicz, B. & Koculak, M. (2020). Task-related motor response inflates confidence [Publisher: Cold Spring Harbor Laboratory Section: New Results]. *bioRxiv*, 2020.03.26.010306. https://doi.org/10.1101/2020.03.26.010306

Smith, P. L. (2000). Stochastic dynamic models of response time and accuracy: A foundational primer. *Journal of mathematical psychology*, *44*(3), 408–463. https://doi.org/10.1006/jmps.1999.1260

Smith, P. L. & Vickers, D. (1988). The accumulator model of two-choice discrimination. *Journal of Mathematical Psychology*, *32*(2), 135–168. https://doi.org/10.1016/0022-2496(88)90043-0

Stengård, E. & van den Berg, R. (2019). Imperfect bayesian inference in visual perception (W. Einhäuser, Ed.). *PLOS Computational Biology*, *15*(4), e1006465. https://doi.org/10.1371/journal.pcbi.1006465

Stine, G. M., Zylberberg, A., Ditterich, J. & Shadlen, M. N. (2020). Differentiating between integration and non-integration strategies in perceptual decision making (V. Wyart, M. J. Frank, V. Wyart & M. Usher, Eds.) [Publisher: eLife Sciences Publications, Ltd]. *eLife*, *9*, e55365. https://doi.org/10.7554/eLife.55365

Stocker, A. A. & Simoncelli, E. P. (2006). Noise characteristics and prior expectations in human visual speed perception. *Nature Neuroscience*, *9*(4), 578–585. https://doi.org/10.1038/nn1669

Summerfield, C. & Koechlin, E. (2010). Economic value biases uncertain perceptual choices in the parietal and prefrontal cortices [Publisher: Frontiers]. *Frontiers in Human Neuroscience*, *4*. https://doi.org/10.3389/fnhum.2010.00208

Tajima, S., Drugowitsch, J., Patel, N. & Pouget, A. (2019). Optimal policy for multi-alternative decisions. *Nature Neuroscience*, *22*, 1503–1511. https://doi.org/10.1038/s41593-019-0453-9

Tajima, S., Drugowitsch, J. & Pouget, A. (2016). Optimal policy for value-based decision-making. *Nature communications*, *7*, 12400. https://doi.org/10.1038/ncomms12400

Tassinari, H., Hudson, T. E. & Landy, M. S. (2006). Combining priors and noisy visual cues in a rapid pointing task. *Journal of Neuroscience*, *26*(40), 10154–10163. https://doi.org/10.1523/JNEUROSCI.2779-06.2006

Teodorescu, A. R., Moran, R. & Usher, M. (2016). Absolutely relative or relatively absolute: Violations of value invariance in human decision making. *Psychonomic Bulletin & Review*, *23*, 22–38. https://doi.org/10.3758/s13423-015-0858-8

Teodorescu, A. R. & Usher, M. (2013). Disentangling decision models: From independence to competition. *Psychological review*, *120*(1), 1–38. https://doi.org/10.1037/a0030776

Thura, D. (2016). How to discriminate conclusively among different models of decision making? *Journal of Neurophysiology*, *115*(5), 2251–2254. https://doi.org/10.1152/jn.00911.2015

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

Tsetsos, K., Gao, J., McClelland, J. L. & Usher, M. (2012). Using time-varying evidence to test models of decision dynamics: Bounded diffusion vs. the leaky competing accumulator model. *Frontiers in Neuroscience*, *6*. https://doi.org/10.3389/fnins.2012.00079

Tuerlinckx, F. (2004). The efficient computation of the cumulative distribution and probability density functions in the diffusion model. *Behavior Research Methods, Instruments, & Computers*, *36*(4), 702–716. https://doi.org/10.3758/BF03206552

Tuerlinckx, F., Maris, E., Ratcliff, R. & De Boeck, P. (2001). A comparison of four methods for simulating the diffusion process. *Behavior Research Methods, Instruments, & Computers*, *33*(4), 443–456. https://doi.org/10.3758/BF03195402

Tversky, A. & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases [Publisher: American Association for the Advancement of Science Section: Articles]. *Science*, *185*(4157), 1124–1131. https://doi.org/10.1126/science.185.4157.1124

Usher, M. & McClelland, J. L. (2001). The time course of perceptual choice: The leaky, competing accumulator model. *108*. https://doi.org/10.1037/0033-295X.108.3.550

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*, 1–21. https://doi.org/10.7554/eLife.12192

van den Berg, R. & Ma, W. J. (2018). A resource-rational theory of set size effects in human visual working memory. *eLife*, *7*. https://doi.org/10.7554/eLife.34963

van den Berg, R., Yoo, A. H. & Ma, W. J. (2017). Fechner's law in metacognition: A quantitative model of visual working memory confidence. *Psychological Review*, *124*(2), 197–214. https://doi.org/10.1037/rev0000060

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

Vickers, D. (1970). Evidence for an accumulator model of psychophysical discrimination. *Ergonomics*, *13*(1), 37–58. https://doi.org/10.1080/00140137008931117

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

Vickers, D. & Packer, J. (1982). Effects of alternating set for speed or accuracy on response time, accuracy and confidence in a unidimensional discrimination task. *Acta Psychologica*, *50*(2), 179–197. https://doi.org/10.1016/0001-6918(82)90006-3

Voskuilen, C., Ratcliff, R. & Smith, P. L. (2016). Comparing fixed and collapsing boundary versions of the diffusion model. *Journal of Mathematical Psychology*, *73*, 59–79. https://doi.org/10.1016/j.jmp.2016.04.008

Voss, A. & Voss, J. (2008). A fast numerical algorithm for the estimation of diffusion model parameters. *Journal of Mathematical Psychology*, *52*(1), 1–9. https://doi.org/10.1016/j.jmp.2007.09.005

Wald, A. & Wolfowitz, J. (1948). Optimum character of the sequential probability ratio test. *The Annals of Mathematical Statistics*, *19*(3), 326–339. https://doi.org/10.1214/aoms/1177730197

Whiteley, L. & Sahani, M. (2008). Implicit knowledge of visual uncertainty guides decisions with asymmetric outcomes [Publisher: The Association for Research in Vision and Ophthalmology]. *Journal of Vision*, *8*(3), 2–2. https://doi.org/10.1167/8.3.2

Wickelgren, W. A. (1977). Speed-accuracy tradeoff and information processing dynamics. *Acta Psychologica*, *41*(1), 67–85. https://doi.org/10.1016/0001-6918(77)90012-9

Wixted, J. T. (2007). Dual-process theory and signal-detection theory of recognition memory. *Psychological Review*, *114*(1), 152–176. https://doi.org/10.1037/0033-295X.114.1.152

Yarkoni, T. & Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning [Publisher: SAGE Publications Inc]. *Perspectives on Psychological Science*, *12*(6), 1100–1122. https://doi.org/10.1177/1745691617693393

Yeung, N. & Summerfield, C. (2014). Shared mechanisms for confidence judgements and error detection in human decision making. In S. M. Fleming & C. D. Frith (Eds.),

The cognitive neuroscience of metacognition (pp. 147–167). Springer. https://doi.org/10.1007/978-3-642-45190-4_7

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

Zylberberg, A., Barttfeld, P. & Sigman, M. (2012). The construction of confidence in a perceptual decision. *Frontiers in Integrative Neuroscience*, *6*. https://doi.org/10.3389/fnint.2012.00079

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

Zylberberg, A., Roelfsema, P. R. & Sigman, M. (2014). Variance misperception explains illusions of confidence in simple perceptual decisions. *Consciousness and Cognition*, *27*, 246–253. https://doi.org/10.1016/j.concog.2014.05.012

Zylberberg, A., Wolpert, D. M. & Shadlen, M. N. (2018). Counterfactual reasoning underlies the learning of priors in decision making. *Neuron*, *99*(5), 1083–1097.e6. https://doi.org/10.1016/j.neuron.2018.07.035

Appendices

A

Appendix for "Models and qualitative tests" and "Quantitative tests"

Contents

A.1 Test of the race model

In the race model two accumulators separately accumulate evidence measurements for the two response options, until one reaches the decision threshold and triggers a response (Fig. 1.1; Bogacz et al., 2006; Vickers, 1970). As discussed in the main text, the race model makes a very strong prediction: The model predicts that presenting more evidence for either option will never slow down response times (Teodorescu & Usher, 2013).

We tested the effect of evidence for the incorrect option on response times. For trials with a confidence report, and in which the response time was greater than 200ms, we looked at the number of dots presented in the first four frames of the stimulus ($4 \times 50ms$), separately for the incorrect array (i.e. low dot mean), and the correct array (i.e. high dot mean). In the dataset used for the preliminary study (Section 2.2), and in the dataset used for the main analysis (Section 3.1), trials in which more evidence was presented for the incorrect option generated slower response times, contradicting the prediction made by the race model (Fig. A.1, A; Fig. A.2, A). This was confirmed by correlating response times with evidence (i.e. dots) presented in the first four frames in the incorrect array, on a participant-by-participant basis. We compared the resulting coefficients to zero across participants. Response times increased with evidence in both the preliminary study dataset ($t(22) = 4.9, p = 7.3 \times 10^{-5}, d = 1.0$) and the main study dataset ($t(47) = 2.9, p = 0.0050, d = 0.43$).

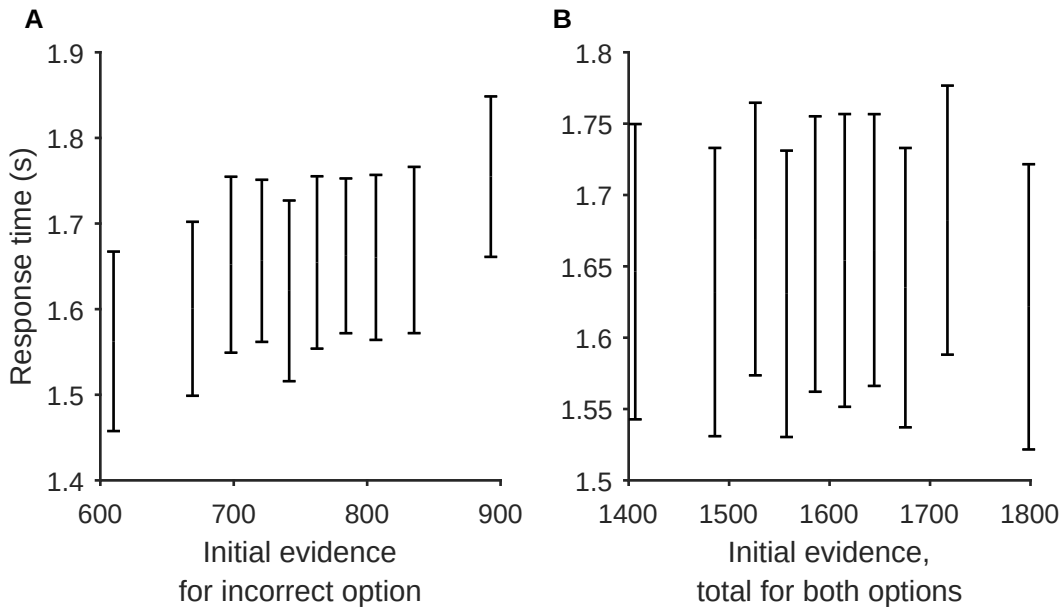

Figure A.1: The effect of initial evidence in the preliminary study. Error bars represent ±1 SEM of the mean. Plotting details in Appendix A.4. Only trials with a response time greater than 200ms, and on which a confidence report was obtained, are plotted.

The drift diffusion model (DDM) predicts that the sum of evidence for the two alternatives is irrelevant, as in the DDM only the difference in evidence is tracked (Bogacz et al., 2006). On free response trials with a confidence report, and in which the response time was greater than 200ms, we summed the evidence presented (i.e. dots) in the two arrays in the first four frames. We found no evidence for an effect of the sum of evidence in the previously collected dataset (Fig. A.1, B; $t(22) = -0.51, p = 0.62, d = -0.11$). However, we did find an effect in the newly collected dataset, such that an increased sum of evidence for the two options sped up response times (Fig. A.2, B; $t(47) = -3.2, p = 0.0023, d = -0.46$). We speculate that this effect may arise from sensory noise which scales with evidence, with greater variability in the evidence accumulation leading to faster responses (Pardo-Vazquez et al., 2019; Zylberberg et al., 2016). See Chapter 5 for a detailed investigation of this idea.

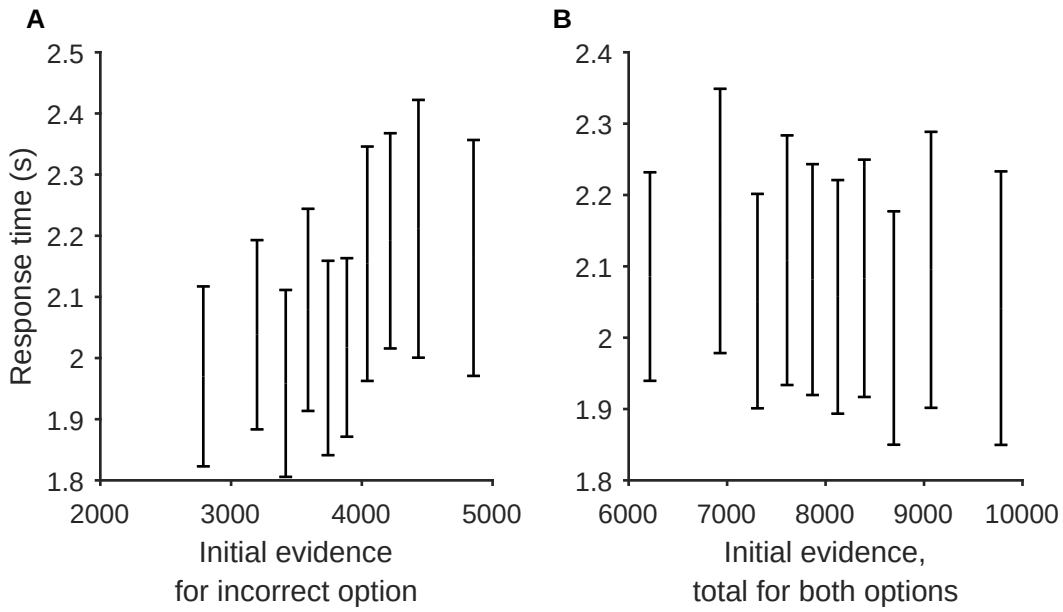

Figure A.2: The effect of initial evidence in the main study. Error bars represent ±1 SEM of the mean. Plotting details in Appendix A.4. Only trials in the free response condition, with a response time greater than 200ms, and on which a confidence report was obtained, are plotted.

A.2 Mathematical details of the models

Here we describe the mathematical framing of the models used in Chapter 3. We largely use the framework from Chapter 4, but with some extensions.

A.2.1 Evidence accumulation

Changes in the state of the accumulator are determined by the number of dots in the two stimuli arrays of the frame currently being processed. If E_{ij} is the number of dots presented in array i in stimulus frame j, then over a small time step δt, during which frame j is being processed, the probability distribution over the increment in the accumulator, δx, is given by (Drugowitsch et al., 2012),

$$p(\delta x|E_{1j}, E_{2j}, \varphi) \quad = \quad N(\delta x; \varphi(E_{2j} - E_{1j})\frac{\delta t}{t_f}, \sigma^2_{acc}\delta t)\,. \tag{A.1}$$

$N(y; a, b)$ indicates a normal distribution over y with mean, a, and variance, b. φ is the standardised drift rate. t_f is the duration of one stimulus frame. σ_{acc}, one of

the free parameters, determines the level of noise in accumulator increments.

In models in which standardised drift-rate is constant, $\varphi = 1$ on all trials. In models where standardised draft-rate varies it follows a normal distribution (Ratcliff & McKoon, 2008), with the level of variability set by the free parameter σ_φ:

$$p(\varphi) = N(\varphi; 1, \sigma_\varphi^2)\,. \tag{A.2}$$

The accumulator begins at 0 and a response is made when it reaches one of the two decision thresholds. The position of the upper and lower thresholds are given by $a - bt$ and $-a + bt$ where t is the time spent accumulating evidence, and a and b are free parameters. For models with flat decision thresholds, $b = 0$. The duration of sensory and motor processing pipelines is also a free parameter, I. Hence the response time is given by I plus the time spent accumulating evidence up to one of the decision thresholds.

A.2.2 Confidence readouts

In Chapter 4 we found expressions for the probability of a confidence report $C = i$, given a response R, time of response t_r, and evidence stream $\boldsymbol{E}$, in both the free response and interrogation conditions. The derivations covered variability in standardised drift-rate, time-dependent thresholds and metacognitive noise, but they only considered one form of readout for confidence. Specifically, they assumed a noisy but otherwise Bayesian, readout of confidence, based on a correct generative model of the environment. Here we consider how to extend the derivations to cover confidence which reflects a readout of the final state of the accumulator, and confidence which reflects a miscalibrated Bayesian readout.

In the expressions in Chapter 4, the observer maps the accumulator state, x, to a readout of the log-posterior ratio between the two options, x_{ll}. The mapping can be summarised by a single function, $\theta(t_e)$, of the time spent accumulating evidence, t_e,

$$x_{ll} = \frac{x}{\theta(t_e)}\,. \tag{A.3}$$

The division of accumulated evidence, x, by $\theta(t)$ is the effect which we refer to as difficulty discounting. Substituting in all abbreviations used in Chapter 4,

$$\theta(t_e) = \frac{t_f^2\sigma_{acc}^2 + t_f\sigma_E^2 + t_e\Delta\mu^2\sigma_\varphi^2}{2t_f\Delta\mu}\,, \tag{A.4}$$

where t_f is the duration of a "frame" over which evidence presented in the stimulus is constant (for the experiments above this was 50ms). Evidence presented in the stimulus for the two options is sampled from two distributions each frame. $\Delta\mu$ is the absolute value of the difference between the means of these two distributions. $\sigma_E/\sqrt{2}$ is the standard deviation of each of these two distributions. σ_{acc} is the standard deviation of noise which corrupts the incoming information as it is added to the accumulator, and σ_φ is the standard deviation in standardised drift-rate across trials.

Depending on the response, either the log-posterior ratio, x_{ll}, or its negative, $-x_{ll}$, are monotonically related to the probability of being correct. Hence, $1/\theta(t_e)$ converts the final state of the accumulator into a readout which determines the confidence of a calibrated Bayesian observer. We can change $\theta(t_e)$ to model observers who do not use a calibrated Bayesian readout. For an observer who's confidence reflects a noisy readout of the final state of the accumulator itself, we can just set $\theta(t_e) = 1$, so that,

$$x_{ll} = x\,. \tag{A.5}$$

In the case of an observer who uses a miscalibrated Bayesian readout of confidence, we allow the possibility that the observer incorrectly estimates the magnitude of difference sources of variability. The readout will have the same form as that in (A.4), except any variance terms will be replaced with their estimated, rather than their true value. For example, if the observer incorrectly estimates the magnitude of accumulator noise in (A.4) we will replace the true value of accumulator noise, σ_{acc}, with the observer's estimate of it, denoted $\widehat{\sigma_{acc}}$.

One complication is that multiplying $\theta(t_e)$ by a constant, K, doesn't change the predictions of the model if one also changes other parameters. (Specifically, the predictions do not change if one also multiplies by $1/K$, the variables d_i and σ_m, which are defined below.) To make the model identifiable we set $\theta(1) = 1$.

To see why this does not in any way constrain the predictions the model can make, we rearrange (A.4) to find,

$$\theta(t_e) = \frac{t_f}{2\Delta\mu}\Big(\sigma_{acc}^2 + \frac{\sigma_E^2}{t_f} + \Big[\frac{\Delta\mu}{t_f}\Big]^2\sigma_\varphi^2\Big)([1-\gamma] + \gamma t_e)\,, \tag{A.6}$$

where,

$$\gamma = \frac{\Big[\frac{\Delta\mu}{t_f}\Big]^2\sigma_\varphi^2}{\sigma_{acc}^2 + \frac{\sigma_E^2}{t_f} + \Big[\frac{\Delta\mu}{t_f}\Big]^2\sigma_\varphi^2}\,. \tag{A.7}$$

As discussed, we can use any $\theta'(t_e) = K\theta(t_e)$ in (A.3), without affecting the model's predictions. We chose K such that,

$$\theta'(t_e) = [1-\gamma] + \gamma t_e\,. \tag{A.8}$$

Then $\theta'(t_e) = 1$.

For convenience, we also use (A.8) for the calibrated Bayesian model, but of course in this case γ is computed using the true values of all variables. For the miscalibrated Bayesian observer we do not fit γ directly, but a transformed version:

$$\Gamma = -\log\Big(\frac{1}{\gamma} - 1\Big)\,. \tag{A.9}$$

Γ is not bounded at 0 and 1, but can take any value.

For all readouts we allow the possibility that the readout is corrupted by metacognitive noise (De Martino et al., 2013; Maniscalco & Lau, 2012). Hence, x_{ll} does not determine confidence directly, but a noisy version of it, x_c,

$$p(x_c|x_{ll}) = N(x_c; x_{ll}, \sigma_m^2)\,. \tag{A.10}$$

σ_m is a free parameter for the level of metacognitive noise. The value of x_c relative to confidence bin boundaries then determines the confidence report given. If x_c falls between d_i and d_{i+1} then a confidence report in bin i is given.

A.2.3 Lapses

In Chapter 4 we did not consider lapses. Here we include the possibility that on a certain proportion of trials observer's give a random confidence report.

We assume that the probability of a lapse in the confidence report given, does not depend on the response, response time, or evidence stream, with some exceptions. That is, generally,

$$p(L = 0|R, t_r, \boldsymbol{E}) = 1 - \lambda \quad \text{(A.11)}$$

$$p(L = 1|R, t_r, \boldsymbol{E}) = \lambda\,, \quad \text{(A.12)}$$

where $L = 1$ and $L = 0$ denote the presence or absence of a lapse. There is one exception in the free response condition, and one in the interrogation condition. In the free response condition, a free parameter of the model describes the total delay caused by sensory and motor processing pipelines (I). If any response in the free condition is faster than this estimated delay, then the corresponding trial is treated as a lapse. In the interrogation condition, if the standard deviation of metacognitive noise is far smaller than other sources of variability, then it is approximated as zero. If metacognitive noise is approximately zero we can compute confidence using a simpler equation (Chapter 4). The probability of the response made, $p(R|\boldsymbol{E})$, appears in the denominator of the relevant expression. To avoid numerical problems in this situation, we treat trials as certainly the result of a lapse when $p(R|\boldsymbol{E}) < 2.2 \times 10^{-307}$.

When a lapse occurs we assume it leads to all possible confidence reports with equal probability,

$$p(C = i|L = 1, R, t_r, \boldsymbol{E}) = M\,, \quad \text{(A.13)}$$

where M is a constant. Marginalising over whether or not a lapse has occurred

provides us with our final predictions for confidence,

$$p(C|R,t_r,\boldsymbol{E}) = p(L=0|R,t_r,\boldsymbol{E})p(C|L=0,R,t_r,\boldsymbol{E}) + p(L=1|R,t_r,\boldsymbol{E})p(C|L=1,R,t_r,\boldsymbol{E}) \quad \text{(A.14)}$$

$$= (1-\lambda)p(C|L=0,R,t_r,\boldsymbol{E}) + \lambda p(C|L=1,R,t_r,\boldsymbol{E})\,. \quad \text{(A.15)}$$

The probability of a confidence report in the absence of a lapse, $p(C=i|L=0,R,t_r,\boldsymbol{E})$, is given by the results of Chapter 4. These results apply directly in the case of a calibrated Bayesian readout, and they apply once adjusted according to the considerations in the previous subsection in the case of a direct readout of the state of the accumulator, or in the case of a miscalibrated Bayesian readout.

A.2.4 Changes for implementation

In the implementation of the derivations we made a number of low-level changes from the expressions in Chapter 4. Firstly, in the experiment the stimulus only cleared at the end of each 50ms frame, not immediately at the time of a response. We accounted for this additional $< 50ms$ of evidence presented to the participants. On a trial-by-trial basis, we treated this additional evidence in the same way as evidence from the processing pipeline.

Second, we precomputed values on a trial-by-trial basis for total pipeline evidence, and for mean pre-decision evidence. We precomputed these values assuming that the processing pipeline contained evidence presented in the $0ms$, or $50ms$, or $100ms$, ... prior to the end of the stimulus, and assuming that the initial $0ms$, or $50ms$, or $100ms$, ... of evidence presented in the stimulus was processed before a decision. We then linearly interpolated between these values to determine the pipeline or pre-decision evidence, given a hypothesised time at which a decision threshold was reached. Third, it turned out that the log-likelihood of the model varied in a discontinuous way as the parameter corresponding to the duration of the processing pipeline, I, changed. This was due to the fact that trials with response times shorter than the pipeline duration were classed as certain lapses. To avoid the difficulties

of finding the maximum of a non-smooth function, we evaluated the likelihood function at the two closest values to I that were multiples of 50ms, and linearly interpolated between these to estimate the likelihood function at I.

A.3 Simulation details

Confidence simulations

To plot data against model fits (i.e. to plot figures in the "Model comparison" and "Fits of the best model" subsections of Section 3.4), we simulated one confidence report for each real trial containing a valid confidence report, on the basis of the response, response time, and evidence presented on that trial, and in accordance with the model predictions for $p(C|R, t_r, \boldsymbol{E})$.

Diffusion simulations

For other simulations in chapters 2 and 3 we took a different approach and simulated completely new synthetic trials and stimuli, and then simulated the full diffusion process from the onset of the trial, to the end of all stimulus processing, including the decision threshold crossing in the free response case.

We simulated stimuli with properties which matched those used in the main experiment (Section 3.1). The evidence accumulation, decision mechanism and confidence reporting followed the models (Section 2.1). We simulated the internal evidence accumulation using timesteps of size 0.1ms, comparing the accumulator to the decision thresholds (in the free response condition) after each timestep. In accordance with the models, at each timestep the drift-rate was determined by the stimulus frame currently being processed. Specifically, the drift-rate was determined by the difference between the number of dots in the two arrays, multiplied by the standardised drift-rate.

"Simulating responses and response times" subsection of Section 3.4. For the plots of model predictions for response times and accuracy we ran the diffusion simulations as described, with some small differences. On each simulated trial, a random trial from the corresponding participant in the main experiment was selected. The distributions which were used on this trial to determine the number of dots in the high and low evidence arrays were then used in the simulated trial. Trial durations in the interrogation condition were sampled, on a participant-by-participant basis, from the trial durations used in the interrogation condition in the real experiment. We simulated 6400 trials per participant.

Model recovery. We simulated datasets for model recovery in almost the same way as we simulated data for the "Simulating responses and response times" subsection of Section 3.4. For model recovery we simulated the same number of trials as there are in the real dataset (640), because we wanted to test whether the analysis was successful with the number of trials available. For these simulations we also applied a time limit of 60 seconds to free response trials, to prevent the simulated dataset taking up too much memory on disk. After this time limit the simulated decision thresholds collapsed to almost zero.

Fig. 2.2. For the diffusion simulations used to generate Fig. 2.2, we simulated 64000 trials for 40 participants using model 7. Stimuli were generated in the same way as in the main experiment, except that all trials were from the free response condition. All simulated participants had the same parameters, shown in Table A.1. The were two "emphasis" conditions. In the "accuracy" condition, observers used a decision threshold that was set at a height of 2000. In the "speed" condition they used a threshold that was set at half this height.

Fig. 2.3 and 2.4. For the diffusion simulations used to generate Fig. 2.3 and those used to generate Fig. 2.4, we simulated 12800 trials for 20 participants using model 7 and the parameters in Table A.1. The decision threshold was set at a

height of 2000. Stimuli were generated in the same way as in the main experiment except all trials were from the free response condition, and discriminability was now manipulated. There were three discriminability levels, which we created by changing the number of dots between the means of the two distributions from which dots in the two arrays were sampled. In the main experiment the difference between the two means was 90 (prior to truncation; see Section 3.1). Here we used a difference of 90, 360, and 810 (prior to truncation). In order to simulate the situation where confidence reports are determined simultaneously with the decision (that is, simultaneously with the time of the threshold crossing), confidence was read out at the time of the decision using the state of the accumulator at that time. For these plots we manipulated whether there was variability in the time taken up by sensory and motor processing. In the simulations for Fig. 2.3 there was no variability in this "pipeline" duration. For Fig. 2.4 we added normally distributed variability in the pipeline duration with standard deviation of 0.1s. The normal distribution used was truncated at 0.

Parameter		Value
Accumulator noise	σ_{acc}	2900
Decision thresh. height	a	See text
Pipeline duration	I	0.35 s
Lapse rate	λ	0
Metacognitive noise	σ_m	1500
Estimated variability ratio	Γ	0.9

Table A.1: Parameters of the simulated observers. Details of the roles of the parameters in the computational model are provided in Appendix A.2.

A.4 Plotting procedure

To visualise the relationship between variables we binned variables on the x-axis of plots. Separately for each participant and each plotted series, we divided the x-axis variable into bins containing approximately equal numbers of trials. The

location of a bin on a plot was determined by the mean value across participants, of the mean value of the x-variable within each bin. The y-value of a bin represents the mean value of the y-variable across participants. Unless noted, error bars and error shading represent ±1 standard error of the mean (SEM) across participants or simulated participants. Unless noted, plots were based on the data from all participants, and all trials in which a confidence report was obtained.

For all plots which refer to "binned confidence", confidence reports were divided into 4 bins on a participant-by-participant basis.

A.5 Ordinal regression details

We repeatedly used a probit ordinal regression onto binned confidence. Predictors in these regressions were individually z-scored. Following the fitting of the probit regression model to the data, we applied a transformation to the resulting coefficients. Namely, we applied the reverse transformation to that which was generated by z-scoring. We did this so that coefficients associated with evidence predictors conveyed effects per unit of evidence (i.e. stimulus dots), rather than per standard deviation of predictor. For example, a greater coefficient for pre-decision evidence than pipeline evidence means that, if the evidence in one pre-decision frame was boosted by 10 dots, this would have a stronger effect on confidence than boosting one pipeline frame by 10 dots.

The transformation which, when applied to the coefficients, reverses the effect of z-scoring the predictors, requires some thought. Denote one of the predictors for a single trial x, and the z-scored version $\widetilde{x}$. Then,

$$\widetilde{x} = \frac{x - \mu}{\sigma}\,, \tag{A.16}$$

where μ and σ are the mean and standard deviation used for z-scoring.

In an ordinal regression, a coefficient, β, and the corresponding predictor for a single trial, $\widetilde{x}$, are multiplied (Long, 1997),

$$\boldsymbol{\beta}\widetilde{x}\,. \tag{A.17}$$

Substituting in our expression for $\widetilde{x}$,

$$\boldsymbol{\beta}\widetilde{x} = \boldsymbol{\beta}\frac{x-\mu}{\sigma} = \frac{\beta}{\sigma}x - \frac{\beta\mu}{\sigma}\,. \tag{A.18}$$

We see that the coefficient associated with the raw evidence predictor (i.e. evidence in units of dots), is given by β/σ. Hence, we divided the coefficients resulting from the regression by the standard deviation used for z-scoring of the corresponding predictors.

A.6 Preregistration details

We made one small deviation from the preregistration. For the ordinal regression analysis we excluded any trial in which the number of frames prior to response was less than, or equal to, the estimate of the pipeline duration for the participant. We preregistered "less than", not "less than or equal to". We did not run the analysis in the "less than" case.

A.7 Participant demographics

Gender	Female	36
	Male	13
	Non-binary	0
	Prefer not to say	0
Age	18-25	39
	26-35	10
	36-45	0
	45-50; 56-65; 65+	0
Handedness	Right	45
	Left	4
	Neither	0

Table A.2: Gender, age, and handedness of participants in the study.

A.8 Response distributions

Using the dataset collected for Chapter 3 we combined all trials from all participants and, by dividing the data into 200 equally spaced bins, looked at the distribution over response times and confidence reports (Fig. A.3). Efforts to match response times in the two conditions achieved reasonable results. Note that in the interrogation condition there is some time between the end of the stimulus and the response which the end of the stimulus triggers. Hence, in the interrogation condition the stimulus will be presented for a shorter amount of time than the response time, unlike in the free response condition. Confidence reports were recorded on a continuous scale from -90 to 90, with 0 corresponding to a report of "Don't know" (Section 3.1).

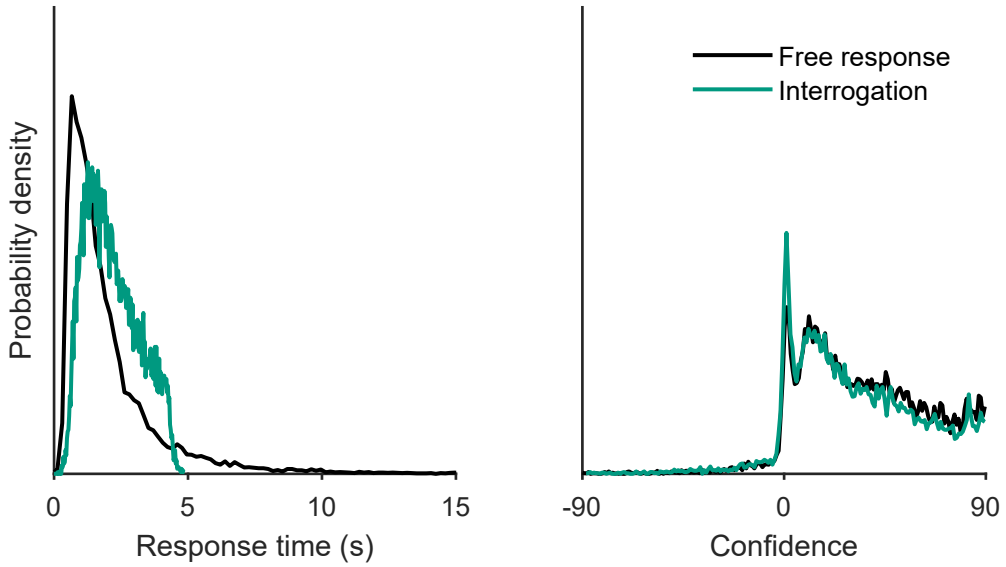

Figure A.3: The distributions over response times and confidence reports in the two conditions. All trials were included, including trials in which a confidence report was not obtained.

A.9 Assessment of optimisation performance

We ran every fit 40 times to minimise the chance of finding a local maximum, rather than the global maximum of the likelihood function. This also allowed us to look for indicators that the optimisation algorithm was struggling with local maxima.

For a heuristic assessment of problems with local maxima we followed Acerbi et al. (2018, Supplemental methods, Section 4.2). We estimated the number of fitting runs required for there to be a 99% chance of having at least one run end within 1 log-likelihood of the greatest value of the log-likelihood obtained over all 40 fitting runs. If many runs are required then the optimisation algorithm must be struggling, and repeatedly getting stuck in local maxima. We estimated the number of runs required by bootstrapping (Acerbi et al., 2018). For example, to estimate performance when using 30 fitting runs, we repeatedly sampled with replacement 30 values from the 40 values of the log-likelihood obtained from the 40 real fitting runs.

For this analysis we used the results of the fits performed on the entire dataset (that is, where the dataset was not divided into training and test data).

The estimated number of fitting runs required was generally low, indicating good performance of the optimiser on this problem (Fig. A.4). In Fig. A.4 NaN values represent cases where, even with a bootstrapped sample size of 40 runs, there was not a 99% chance of ending a run within 1 log-likelihood of the greatest value of the log-likelihood obtained. Hence, there were certainly some cases where the fitting performed poorly. For a couple of participants the fitting was poor across all models. However, in general, the optimisation algorithm appeared to perform well.

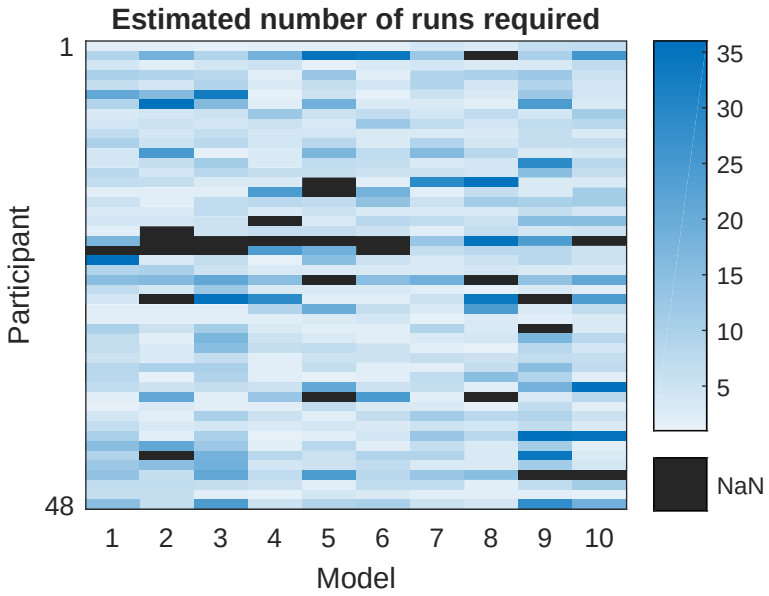

Figure A.4: Estimated number of fitting runs required to meet a heuristic performance criterion.

A.10 AIC and BIC results

In addition to comparing models using cross validation, we took another approach in which we fit the models to all data, without dividing into training and test sets. We then evaluated the performance of the models using the Akaike information criterion (AIC), and the Bayesian information criterion (BIC). These criteria take into account the maximum likelihood achieved by a model, but also apply a penalisation for the

number of free parameters, and the increased flexibility that additional parameters generally provide. A lower value on these criteria indicates better fit.

A similar picture emerged in the AIC and BIC results, to those found using cross validation. Specifically, it was clear that models which featured a miscalibrated Bayesian readout of confidence fit best, but it was not clear whether drift-rate variability or time-dependent thresholds were present (Fig. A.5). According to the BIC, the more conservative of the two criteria, the simplest model which featured a miscalibrated Bayesian readout fit best.

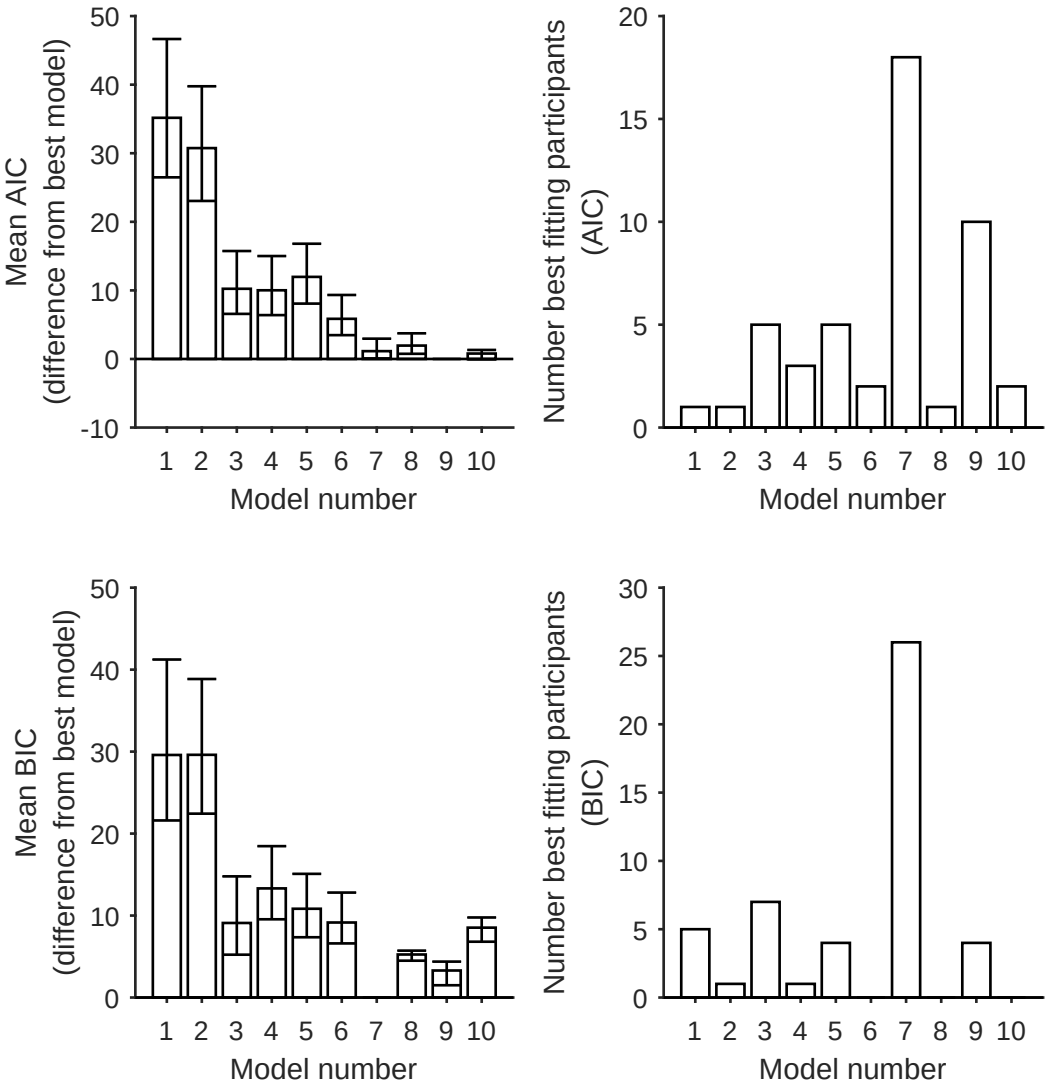

Figure A.5: AIC and BIC relative to the overall best fitting model (according to the AIC or BIC), and number of participants for which each model provided the best fit. Unlike in other plots, error bars represent 95% bootstrapped confidence intervals.

A.11 Model recovery

To test that the computational modelling, and model comparison was performing as expected, we performed a model recovery analysis. We simulated datasets from each of the 10 model variants, and then fit all 10 models to these 10 simulated

datasets. We looked to see if the true data generating model was recovered when the models were compared. To provide a particularly stringent test, we simulated datasets of the same size as the real dataset, using the fitted parameters for each participant (from the fits to the entire dataset without division into training and test trials). Using the fitted parameters generates a particularly strong test because these parameters have been tuned so that all models make similar predictions to the behavioural data, and hence also to each other. On each simulated trial we simulated the full diffusion process from the onset of evidence accumulation, through decision, to confidence report (Appendix A.3).

We fit all data (without partitioning into training and test data), and used AIC and BIC to compare models. Both with the AIC and with the BIC, the true generating model was often recovered (Fig. A.6). The type of confidence readout used was always correct in the recovered models, but the AIC and BIC were sometimes conservative regarding the presence of drift-rate variability and time-dependent thresholds: Sometimes the best fitting model was a simpler version of the data generating model.

Figure A.6: Mean AIC and BIC obtained after fitting data simulated using one of the 10 models.

A.12 Parameter bounds and start points

During optimisation using MATLAB's fmincon optimiser ('Matlab Optimization Toolbox', 2017), we specified bounds on the values the parameters could take

(Table A.3). At the start of each fit candidate parameter sets were drawn from uniform distributions between the "Lower initial" and "Upper initial" values in Table A.3. To support efficient fitting we offset and scaled some parameters, so that all parameters covered a similar range. The offset and scaling values used are also provided in the table. Details of the roles of the parameters in the computational model are provided in Appendix A.2.

Parameter		Lower bound	Lower initial	Upper bound	Upper initial	Offset	Scaling
Drift-rate variability	σ_φ	0	0	6	2	0	1/6
Accumulator noise	σ_{acc}	0	200	24000	8000	0	1/24000
Confidence bin bounds	d_i	-36000	-12000	72000	24000	36000	1/108000
Decision thresh. height	a	10	100	72000	24000	-10	1/72000
Decision thresh. slope	b	0	0	72000	24000	0	1/72000
Pipeline duration (s)	I	0.001	0.05	1	0.8	0	1
Lapse rate	λ	1/640	0.01	1	0.4	0	1
Metacognitive noise	σ_m	1	50	72000	24000	-1	1/72000
Estimated variability ratio	Γ	-20	-3	20	3	0	1/20

Table A.3: Parameter bounds, offset, and scaling used during fitting, and limits for initial parameter values.

A.13 Parameter values

Fitted parameter values for the three models on which we focused in Section 3.4 are provided below. Values are from the fits performed on the entire dataset (that is, where the dataset was not divided into training and test data). Details of the roles of the parameters in the computational model are provided in Appendix A.2.

Model	Parameter	Median	25th percentile	75th percentile
2	σ_φ	0.21	0.14	0.48
	σ_{acc}	940	230	1500
	d_i	3500	1900	7200
		3500	1100	8000
		3800	1600	7000
	a	3000	1600	6600
	I (s)	0.63	0.50	0.80
	λ	0.47	0.0017	0.70
	σ_m	2100	52	7300
5	σ_φ	0.77	0.52	1.5
	σ_{acc}	1600	750	2900
	d_i	4100	2600	6400
		3600	1200	7500
		4200	1400	7800
	a	4600	3000	10 000
	I (s)	0.35	0.20	0.65
	λ	0.14	0.0016	0.44
	σ_m	3300	1300	5700
7	σ_{acc}	2100	1400	3800
	d_i	4100	2300	5000
		2800	1300	4900
		2900	1400	5400
	a	3600	2600	6500
	I (s)	0.50	0.30	0.80
	λ	0.0017	0.0016	0.35
	σ_m	2400	960	3600
	Γ	0.26	−1.1	1.8

Table A.4: Fitted parameter values.

B

Appendix for "A new modelling method"

Contents

B.1 Product of normal distributions

We will repeatedly encounter a situation where we are interested in the probability distribution over z in the following equation,

$$p(z) \propto \int N(y; z, \sigma_A^2) N(y; \mu, \sigma_B^2) dy \,, \tag{B.1}$$

where N denotes the normal distribution. The product of two normal distributions is a scaled normal distribution (over y; Bromiley, 2018). The scaling is given by,

$$S = N(z; \mu, \sigma_A^2 + \sigma_B^2) \,. \tag{B.2}$$

Once we have integrated over y, only the scaling is left, as the normal distribution over y integrates to one. Hence,

$$p(z) = N(z; \mu, \sigma_A^2 + \sigma_B^2) \,. \tag{B.3}$$

B.2 Bayesian observer's policy

At some time t, the observer has gathered N sensory samples, $\delta x_1, \delta x_2, ..., \delta x_N$, collectively denoted $\boldsymbol{\delta x}$. The samples are generated according to (4.22), (4.23), and (4.24). Using Bayes rule the observer can infer the probability of the two possible values for S. Moran (2015), using a result from Drugowitsch et al. (2012), derived the posterior probability over the two options $S = 1$ and $S = 2$, for exactly the case we have. For completeness, we rederive the posterior here.

Given all the evidence measurements so far,

$$p(S|\boldsymbol{\delta x}) \propto p(\boldsymbol{\delta x}|S)p(S) \tag{B.4}$$

$$\propto \int \prod_{i=1}^{N} p(\delta x_i|\nu)p(\nu|S)d\nu\,, \tag{B.5}$$

where we have used the flat prior over S in (4.1). Looking at the product in (B.5), using (4.24), and using standard formulae for the product of normal distributions (e.g. see Bromiley, 2018) we have,

$$\prod_{i=1}^{N} p(\delta x_i|\nu) = \prod_{i=1}^{N} N(\delta x_i; \nu\delta t, s^2\delta t) \tag{B.6}$$

$$\propto N(\nu; \frac{x}{t}, \frac{s^2}{t})\,, \tag{B.7}$$

where the proportionality sign indicates proportionality with respect to ν, $x = \sum_{i=i}^{N} \delta x_i$, and we have used that $t = \sum_{i=1}^{N} \delta t$. Hence in (B.5) we have,

$$p(S|\boldsymbol{\delta x}) \propto \int N(\nu; \frac{x}{t}, \frac{s^2}{t}) N(\nu; \pm\nu_0, \sigma_\nu^2)d\nu\,, \tag{B.8}$$

where the sign on ν_0 is determined by S. Using the result in Appendix B.1,

$$p(S|\boldsymbol{\delta x}) \propto N(\pm\nu_0; \frac{x}{t}, \frac{s^2}{t} + \sigma_\nu^2)\,. \tag{B.9}$$

Rearranging, we find the log-posterior ratio is given by,

$$x_{ll} = \log \frac{p(S=2|\boldsymbol{\delta x})}{p(S=1|\boldsymbol{\delta x})} = \frac{2x\nu_0}{s^2 + t\sigma_\nu^2} = \frac{x}{\theta(t)}\,. \tag{B.10}$$

Where x_{ll}, and $\theta()$ provide abbreviations.

B.3 Simulation and plotting details

We simulated the task described in the main text with the parameters shown in Table B.1. The only difference between the simulation and the setup described in the main text is that evidence in a frame could not go below 0 or above 3096. Evidence within a frame was resampled until it met these constraints. The reference value, $\overline{\mu}$, the midpoint between the means of the two evidence signals, was sampled each trial. It was sampled from a normal distribution centred on 1000, with standard deviation of 100, that was truncated at 500 and 1500.

Variable	Expression	Simulation value
Duration of a frame	t_f	50ms
Max evidence in frame		3096
Min evidence in frame		0
Reference value	$p(\overline{\mu})$	See text
Mean difference between evidence signals	$\Delta\mu$	90
Standard deviation of evidence in a frame	σ_E	311

Table B.1: Parameters for the task used in the simulation.

The parameters used for the observer are shown in Table B.2. All participants were simulated using the same parameters. Half of the trials were simulated for the free response condition, and half for the interrogation condition. A linear decision threshold, symmetric for both options, was used in the free response condition. At the onset of evidence accumulation the threshold was at the value given in Table B.2 for threshold height, and it subsequently changed at the rate given by threshold slope. In the interrogation condition, stimulus presentation duration was drawn from a normal distribution of mean 0.8 s and standard deviation 0.3 s, truncated at 0.2 and 4 s.

The evolution of the observer's accumulator was simulated using small time steps of duration 0.1 ms. Each accumulator increment was determined by sampling

Variable		Fig. 4.3	Fig. 4.7	Fig. 4.8
Simulated participants		40	40	40
Trials per participant		128000	12800	128000
Pipeline duration	I	350 ms	350 ms	100ms
Standard deviation of accumulator noise	σ_{acc}	2860	2860	2860
Standard deviation of standardised drift rate	σ_{φ}	1 or 0	1 or 0	0.6
Standard deviation of metacognitive noise	σ_m	0	$1500/\theta(t=1)$	$1500/\theta(t=1)$
Threshold height		2000	2000	2000
Threshold slope		0	$-800\ s^{-1}$	0

Table B.2: Parameters of the simulated observer.

from the distribution in (4.6). Simulation of the accumulation continued until a decision threshold was crossed, and all remaining evidence in the pipeline had been processed. The only exception was the simulation for Fig. 4.8. For the "Immediate conf." condition in this plot, accumulation terminated as soon as the decision threshold was reached, consistent with the idea that the manipulation used by Kiani et al. (2014) ensured confidence was not based on pipeline evidence.

Confidence reports were produced in accordance with the model (see (4.27), (4.28) and Fig. 4.2). (The only difference to the model was that simulated confidence reports were scaled by multiplication with a constant, $\theta(t=1)$.) Where a plot refers to binned confidence, the confidence values produced by the model have been divided into 10 bins of equal numbers of data points on a participant-by-participant basis.

In order to make predictions for confidence using the derived expressions, we used the true parameters for σ_{acc}, σ_{φ}, σ_m, I, threshold height and threshold slope. For the bin edge parameters, d_i, we used the bin edges which were used for dividing the simulated confidence reports into 10 bins. (The code used the fact that simulated confidence reports were scaled by $\theta(t=1)$.)

Prior to plotting, data for the x-axis was binned. For each participant, and each bin, the average value of the variable on the x-axis was computed. Using these averages, the average value across participants was computed. 10 bins were used apart from in Fig. 4.8 where 5 bins were used. The y-variable was calculated separately for each participant and bin, before averaging across participants. Error bars, and the width of error shading, represent ±1 standard error of the mean, across participants.

B.4 Rearranging interrogation condition result

For the probability distribution over C in the forced condition, we found in (4.61) that

$$p(C=i|R,t_e,\boldsymbol{E}) = L\int_{Ld_i}^{Ld_{i+1}} dx_c \; \frac{1}{\Phi(L\frac{\mu_{ll}}{\sigma_{ll}})} N(x_c;\mu_{ll},\sigma_m^2+\sigma_{ll}^2)\Phi\left(L\frac{x_c\sigma_{ll}^2+\mu_{ll}\sigma_m^2}{\sigma_m\sigma_{ll}\sqrt{\sigma_m^2+\sigma_{ll}^2}}\right). \tag{B.11}$$

It is possible to rearrange this expression into a form which is easier to evaluate numerically.

Consider a change of variables,

$$z = \frac{x_c-\mu_{ll}}{\sqrt{\sigma_m^2+\sigma_{ll}^2}}\,. \tag{B.12}$$

Then we have,

$$dz = \frac{dx_c}{\sqrt{\sigma_m^2+\sigma_{ll}^2}} \tag{B.13}$$

and,

$$N(x_c;\mu_{ll},\sigma_m^2+\sigma_{ll}^2) = \frac{1}{\sqrt{\sigma_m^2+\sigma_{ll}^2}}N(z;0,1) \tag{B.14}$$

$$= \frac{1}{\sqrt{\sigma_m^2+\sigma_{ll}^2}}\phi(z)\,, \tag{B.15}$$

where $\phi()$ indicates the standard normal distribution. Also denote,

$$g = \frac{L\mu_{ll}\sqrt{\sigma_m^2+\sigma_{ll}^2}}{\sigma_m\sigma_{ll}} = L\frac{\mu_{ll}\sigma_{ll}^2+\mu_{ll}\sigma_m^2}{\sigma_m\sigma_{ll}\sqrt{\sigma_m^2+\sigma_{ll}^2}} \tag{B.16}$$

$$h = \frac{L\sigma_{ll}}{\sigma_m} = \frac{L\sigma_{ll}^2}{\sigma_m\sigma_{ll}}\,. \tag{B.17}$$

Then,

$$g+hz = L\frac{x_c\sigma_{ll}^2+\mu_{ll}\sigma_m^2}{\sigma_m\sigma_{ll}\sqrt{\sigma_m^2+\sigma_{ll}^2}}\,. \tag{B.18}$$

Hence, in (B.11) we have,

$$p(C=i|R,t_e,\boldsymbol{E}) = L\frac{1}{\Phi(L\frac{\mu_{ll}}{\sigma_{ll}})}\int_{e_i}^{e_{i+1}} dz\ \phi(z)\Phi(g+hz)\,, \tag{B.19}$$

where,

$$e_i = \frac{Ld_i-\mu_{ll}}{\sqrt{\sigma_m^2+\sigma_{ll}^2}}\,. \tag{B.20}$$

Using result 10010.1 from Owen (1980), gives us,

$$p(C=i|R,t_e,\boldsymbol{E}) = L\frac{1}{\Phi(L\frac{\mu_{ll}}{\sigma_{ll}})}\Bigg[\quad BvN\Big(\frac{g}{\sqrt{1+h^2}},e_{i+1};\frac{-h}{\sqrt{1+h^2}}\Big) \tag{B.21}$$

$$- \quad BvN\Big(\frac{g}{\sqrt{1+h^2}},e_i;\frac{-h}{\sqrt{1+h^2}}\Big)\Bigg]\,, \tag{B.22}$$

where,

$$BvN(l_1,l_2,\rho) = \int_{-\infty}^{l_1}\int_{-\infty}^{l_2}\frac{1}{2\pi\sqrt{1-\rho^2}}e^{-\frac{1}{2(1-\rho^2)}(x^2+y^2-2xy\rho)}dxdy\,, \tag{B.23}$$

is the bivariate cumulative normal distribution, corresponding to a distribution with mean and covariance,

$$\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{B.24}$$

$$\Sigma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}\,. \tag{B.25}$$

This integral can then be numerically evaluated using standard functions.

C

Appendix for "Challenges to Bayesian confidence"

Contents

C.1 Mathematical details of the models

In this section we begin by describing the three models in detail, while at the same time deriving predictions for the probability distribution over the observer's final balance of evidence, given the stimulus presented (which will be denoted $p(x|E)$). With this probability distribution we will be able to make trial-by-trial predictions for responses.

C.1.1 Model 1: Base model

We make many of the same basic assumptions as in Chapter 4 (see eq. 4.6). The evidence presented for option, i, in time step j, and denoted δE_{ij}, drives a noisy evidence measurement. The measurement of the evidence for this option in this time step is denoted δx_{ij}, and follows (Drugowitsch et al., 2012),

$$p(\delta x_{ij}|\delta E_{ij}) = N(\delta x_{ij}; \delta E_{ij}, \frac{\sigma_{acc}^2}{2}\delta t)\,. \tag{C.1}$$

Note that in this model we do not include drift-rate variability.

If the observer accumulates the evidence measurements, then over the course of a single frame, k, the change in the accumulator driven by option i, denoted z_{ik}, will be given by the sum of the measurements received in all the time steps occurring during frame k. Denote the set of these time steps Ω. Then,

$$z_{ik} = \sum_{j\in\Omega} \delta x_{ij} \tag{C.2}$$

$$p(z_{ik}|E_{ik}) = N(z_{ik}; E_{ik}, \frac{\sigma_{acc}^2}{2}t_f)\,, \tag{C.3}$$

where we have used that the duration of a frame t_f equals the number of time steps in that frame multiplied by the duration of a time step δt. We have also used that the sum over the evidence presented in each time step, j, of frame, k, is just the evidence presented in that frame, which we denote, E_{ik}. That is,

$$E_{ik} = \sum_j \delta E_{ij}\,. \tag{C.4}$$

The observer sums the evidence measurements over all frames, and takes the difference to compute the balance of evidence, x:

$$x = \sum_k z_{2k} - z_{1k} \tag{C.5}$$

$$p(x|E) = N(x; E, \sigma_{acc}^2 t)\,, \tag{C.6}$$

where t is the duration of evidence presentation, and $E = \sum_k E_{2k} - E_{1k}$.

C.1.2 Model 2: Stimulus intensity dependent noise

Model 2 is similar to the base model, but we now allow the possibility of noise which scales with stimulus intensity (Pardo-Vazquez et al., 2019; Teodorescu et al., 2016). Specifically, we allow the possibility of noise which scales with E_{ik}. To incorporate this stimulus dependent noise, σ_s, we modify (C.1), to become,

$$p(\delta x_{ij}|\delta E_{ij}) \quad = \quad N(\delta x_{ij}; \delta E_{ij}, \frac{\sigma^2_{acc} + \sigma^2_s E_{ik}}{2}\delta t)\,. \tag{C.7}$$

j is time step under consideration, and k is the stimulus frame active at time step j.

Taking the same steps as for the basic model, we sum over all time steps during frame k. The distribution over z_{ik} is given by,

$$p(z_{ik}|E_{ik}) \quad = \quad N(z_{ik}; E_{ik}, \frac{\sigma^2_{acc} + \sigma^2_s E_{ik}}{2}t_f)\,, \tag{C.8}$$

Again, the observer sums all evidence measurements received, and computes the balance of evidence, x,

$$x = \sum_k z_{2k} - z_{1k} \tag{C.9}$$

$$p(x|E) = N(x; E, \sigma^2_{acc}t + \frac{\widetilde{E_1} + \widetilde{E_2}}{2}\sigma^2_s t_f)\,, \tag{C.10}$$

where we have used an abbreviation, $\widetilde{E_i} = \sum_k E_{ik}$.

C.1.3 Model 3: Independent drift rate variability

Model 3 is again similar to the base model, but we now consider the possibility of trial-to-trial variability in drift rate (Ratcliff et al., 1999). Importantly, we assume that the drift rate which applies to evidence presented in one stimulus item, is independent of the drift rate which applies to evidence presented in the other stimulus.

We denote the drift rate which applies to evidence in the item for option 1, φ_1, and the drift rate which applies to evidence in the item for option 2, φ_2. As in Chapter 2 we operationalise drift-rate variability as a term which has a multiplicative

effect on the presented evidence (see Section 2.1 for discussion of this choice). We again assume that drift rate variability follows a normal distribution, but now independently for the two stimulus items,

$$p(\varphi_i) = N(\varphi_i; 1, \sigma_\varphi^2) \tag{C.11}$$

For this model we modify (C.1) to become,

$$p(\delta x_{ij}|\delta E_{ij}, \varphi_i) = N(\delta x_{ij}; \delta E_{ij}\varphi_i, \frac{\sigma_{acc}^2}{2}\delta t)\,. \tag{C.12}$$

Following the same steps as before, we sum over all the time steps in frame k,

$$z_{ik} = \sum_{j\in\Omega} \delta x_{ij} \tag{C.13}$$

$$p(z_{ik}|E_{ik}, \varphi_i) = N(z_{ik}; E_{ik}\varphi_i, \frac{\sigma_{acc}^2}{2}t_f)\,, \tag{C.14}$$

We next sum over all the frames to find the distribution over the total of the measurements for option i, denoted z_i,

$$z_i = \sum_k z_{ik} \tag{C.15}$$

$$p(z_i|\widetilde{E}_i, \varphi_i) = N(z_i; \widetilde{E}_i\varphi_i, \frac{\sigma_{acc}^2}{2}t)\,. \tag{C.16}$$

Using (C.11) we can marginalise over φ_i,

$$p(z_i|\widetilde{E}_i) = \int d\varphi_i\; p(z_i|\widetilde{E}_i, \varphi_i)p(\varphi_i|\widetilde{E}_i) \tag{C.17}$$

$$= \int d\varphi_i\; p(z_i|\widetilde{E}_i, \varphi_i)p(\varphi_i) \tag{C.18}$$

$$= \int d\varphi_i\; N(z_i; \widetilde{E}_i\varphi_i, \frac{\sigma_{acc}^2}{2}t)N(\varphi_i; 1, \sigma_\varphi^2) \tag{C.19}$$

$$= N(z_i; \widetilde{E}_i, \frac{\sigma_{acc}^2}{2}t + \widetilde{E}_i^{\,2}\sigma_\varphi^2)\,. \tag{C.20}$$

We can now calculate the distribution over the balance of evidence, x:

$$x = z_2 - z_1 \tag{C.21}$$

$$p(x|E) = N(x; E, \sigma_{acc}^2 t + (\widetilde{E_1}^{\,2} + \widetilde{E_2}^{\,2})\sigma_\varphi^2)\,. \tag{C.22}$$

Note that, $E = \widetilde{E_2} - \widetilde{E_1}$.

C.1.4 Response probabilities

We now have an expression for the probability distribution over the observer's balance of evidence, given the stimulus presented, according to each of the three models, $p(x|E)$. Assuming the response is not the result of a lapse, the observer uses the balance of evidence to select their response. They will select which ever option is favoured by the balance of evidence. As a result, the probability they select option 2, $R = 2$, is just given by the probability that the balance of evidence is positive,

$$p_{\text{no lapse}}(R = 2|E) = \int_0^\infty dx\ p(x|E) \tag{C.23}$$

However, we allow the possibility that on a certain proportion of trials, λ, the observer makes a random response. Hence the probability of response 2 is,

$$p(R = 2|E) = (1 - \lambda)p_{\text{no lapse}}(R = 2|E) + \frac{\lambda}{2} \tag{C.24}$$

C.2 Simulation details

For various analyses we simulated data. This involved simulating stimuli with properties closely matching those used in dataset B (from Chapter 3; Section 3.1), and simulating a diffusion process from the onset of each trial, until the stimulus had been completely processed. We simulated the evidence accumulation process using time steps of size 0.1ms, and following the assumptions of the relevant model.

Fig. 5.1 For simulations looking at the relative effect of decision-congruent and decision-incongruent evidence, we simulated 48 participants and 5120 trials each, with half of those trials being free response, and half being interrogation condition. We simulated data using the best fitting model from Chapter 3, which was referred to as "model 7" in that chapter. See Section 2.1 for a detailed description of the evidence accumulation, decision mechanism and confidence reporting in this model. For all simulated participants we used the same set of parameter values, the median of the fitted parameter values (Section A.13).

Simulated stimuli had the same properties as the stimuli used for dataset B (Section 3.1), with some small differences. On each trial, the pair of values used for the mean of the correct array dot distribution, and incorrect array dot distribution, were determined by the pair of values used on a trial in the main experiment. The trial was selected at random from the trials undertaken by the corresponding participant. On each simulated interrogation condition trial, the duration of stimulus presentation was determined by selecting a random interrogation condition trial, from the real data for the corresponding participant. The duration of stimulus scheduled to be presented in this real trial was used as the duration of stimulus presentation in the simulated trial.

"Positive evidence effect" and "Model recovery from fitted parameters" sections. We also ran simulations using the three models defined and fitted in Chapter 5, and as further described in Appendix C.1. For all these models, confidence was determined by the final balance of evidence in favour of the response.

For simulations of the models from fitted parameters, we simulated 48 participants. For each simulated participant we used the fitted parameters from the corresponding real participant. Simulated stimuli had the same properties as the stimuli used for dataset B, but with the same small differences regarding sampling of dot distribution means, and stimulus presentation durations, as in the previous subsection. When simulating data for plotting we simulated 5120 trials from the interrogation condition per participant. Simulations for model recovery used the same number of trials as in the real dataset.

"Model recovery from selected parameters" section. We simulated 48 participants using the three models defined in Chapter 5, with 1280 interrogation condition trials each. The duration of stimulus presentation was drawn from a truncated normal distribution. Prior to truncation it had a mean of 0.8s, standard deviation of 0.3, and it was truncated at 0.2s and 4s. The difference between the mean of the dot distribution for the correct array, and the incorrect array was the

same as used in dataset B; it was 90. However, we varied the midpoint between these distributions over a wider range to emphasise the effect of evidence midpoint. The midpoint was drawn from a normal distribution truncated at 400 and 2600 dots. Prior to truncation it had a mean of 1500, and standard deviation of 2000. The parameters for the models were set to the values in Table C.1.

Model number	1	2	3
σ_{acc}	2.86×10^3	0	716
σ_s	-	100	-
σ_φ	-	-	0.1
λ	0.1	0.1	0.1

Table C.1: Parameters used in the model recovery analysis to determine whether the models were, in principle, distinguishable.

C.3 Model recovery from selected parameters

To test whether the models were in principle distinguishable, we simulated data using each of the three models. We fit each of the resulting three datasets with each of the three models. The results of this exercise are shown in Fig. C.1, when evaluating fits with the AIC and BIC. Because the aim was to determine whether the models could be distinguished in principle, we simulated the models using stimuli and parameter values which we reasoned would lead the models to make distinctive predictions, that could not be recreated by the other models at any parameter values. For details of the settings used, see Section C.2.

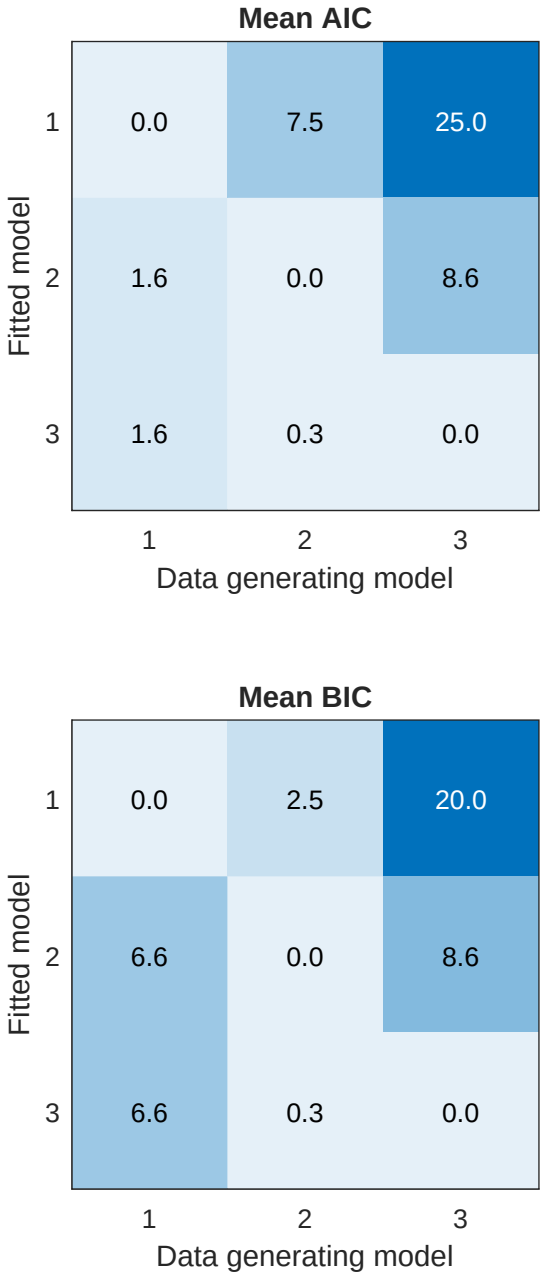

Figure C.1: Mean AIC and BIC obtained after fitting data simulated using one of the 3 models. Data was simulated with parameter values which we reasoned would lead the models to make distinctive predictions.

C.4 Model recovery from fitted parameters

We wanted to test whether the models, once fitted to produce predictions that closely matched the real data, could still be distinguished. We simulated data using each of the three models, and the fitted parameter values (Appendix C.7). We fit each of the resulting three datasets with each of the three models. The results of this exercise are shown in Fig. C.2, when evaluating fits with the AIC and BIC. For details of the simulations used, see Section C.2.

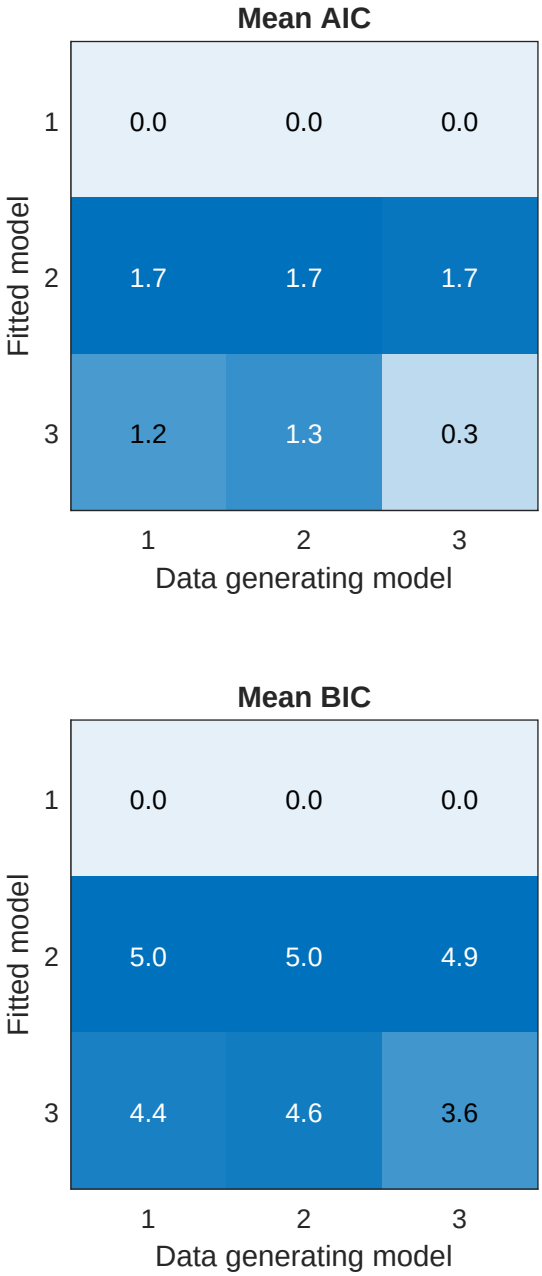

Figure C.2: Mean AIC and BIC obtained after fitting data simulated using one of the 3 models. Data was simulated using the parameter values from the fits to the real data.

C.5 Assessment of optimisation performance

To minimise the risk of only reaching local maxima of the likelihood function, we ran the fitting for each model and participant 16 times. This also allowed us to

assess the optimiser's performance. Using the same procedure described in Section A.9 (Acerbi et al., 2018, Supplemental methods, Section 4.2), we estimated the number of runs required for there to be a 99% chance of at least one run ending within 1 log-likelihood of a target value. For a participant and model, the target value was the greatest value of the log-likelihood obtained over all fitting runs. Results of this analysis are shown inf Fig. C.3. NaN values indicate that, even using a number of fitting runs equal to the number actually conducted (16), there was not a 99% chance of ending within 1 log-likelihood of the target value.

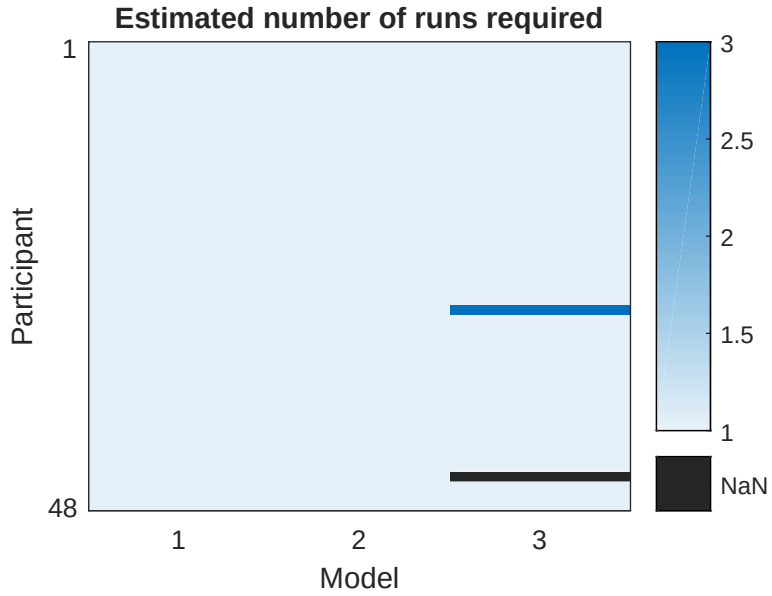

Figure C.3: Estimated number of fitting runs required to meet a heuristic performance criterion.

C.6 Parameter bounds and start points

During optimisation parameter values were bounded (Table C.2). At the start of each fit candidate parameter sets were drawn from uniform distributions between the "Lower initial" and "Upper initial" values in Table C.2. For details of the roles of the parameters see Appendix C.1.

Parameter	Lower bound	Lower initial	Upper bound	Upper initial
σ_{acc}	0	0	20000	4000
σ_s	0	5	500	100
σ_φ	0	0	8	4
λ	0	0.02	1	0.35

Table C.2: Parameter bounds used during fitting, and limits for initial parameter values.

C.7 Parameter values

Fitted parameter values for the three models are provided below. Details of the roles of the parameters in the computational model are provided in Appendix C.1.

Model	Parameter	Median	25th percentile	75th percentile
1	σ_{acc}	2300	1800	3600
	λ	0.054	0.	0.16
2	σ_{acc}	9.6	1.2	1800
	σ_s	58	2.9	81
	λ	0.048	0.	0.14
3	σ_{acc}	1600	250	2100
	σ_φ	0.052	0.026	0.078
	λ	0.	0.	0.096

Table C.3: Fitted parameter values for the three models.

D

Appendix for "DDM in settings with greater ecological validity"

Contents

D.1 Mathematical results

D.1.1 Decision problem and posterior

As discussed in the main text, every time step the observer receives evidence samples, $\delta\boldsymbol{x}(t)$, according to

$$p(\delta\boldsymbol{x}(t)|C=i) = N(\delta\boldsymbol{x}(t); \boldsymbol{\mu}(i)\delta T, \Sigma\delta T)\,, \tag{D.1}$$

where t indicates the current time step, δT is the duration of a time step, $N()$ indicates the multivariate normal distribution, and Σ is the covariance of evidence samples (Tajima et al., 2019).

We want to compute the posterior over C using Bayes rule. Before we can use Bayes rule we need to know the likelihood function of C, given evidence samples. After M time steps the observer will have gathered evidence samples $\delta\boldsymbol{x}(1), \delta\boldsymbol{x}(2), ..., \delta\boldsymbol{x}(M)$. Using the independence between the samples gathered at different time steps (conditioned on the value of C), the likelihood will be

$$\begin{aligned} p(\delta\boldsymbol{x}(1), \delta\boldsymbol{x}(2), ..., \delta\boldsymbol{x}(M)|C=i) &= \prod_{j=1}^{M} p(\delta\boldsymbol{x}(j)|C=i)\,, & \text{(D.2)} \\ &= \prod_{j=1}^{M} N\left(\delta\boldsymbol{x}(j); \boldsymbol{\mu}(i)\delta T, \Sigma\delta T\right)\,, & \text{(D.3)} \\ &= \prod_{j=1}^{M} N\left(\boldsymbol{\mu}(i)\delta T; \delta\boldsymbol{x}(j), \Sigma\delta T\right)\,, & \text{(D.4)} \end{aligned}$$

where we have used (D.1).

Using standard results for the product of multivariate normal distributions,

$$p(\delta\boldsymbol{x}(1), \delta\boldsymbol{x}(2), ..., \delta\boldsymbol{x}(M)|C=i) \quad \propto_{\boldsymbol{\mu}} \quad N\left(\boldsymbol{\mu}(i); \frac{\boldsymbol{x}(t)}{t}, \frac{\Sigma}{t}\right)\,, \qquad \text{(D.5)}$$

where $\boldsymbol{x}(t) = \sum_i \delta\boldsymbol{x}(i)$, $t = \sum_i \delta T$ and the proportionality is with respect to μ (Drugowitsch et al., 2012).

We can now calculate the posterior

$$\begin{aligned} &p(C=i|\delta\boldsymbol{x}(1), \delta\boldsymbol{x}(2), ..., \delta\boldsymbol{x}(M)) \\ &\qquad \propto p(\delta\boldsymbol{x}(1), \delta\boldsymbol{x}(2), ..., \delta\boldsymbol{x}(M)|C=i)p(C=i) & \text{(D.6)} \\ &\qquad \propto N\left(\boldsymbol{\mu}(i); \frac{\boldsymbol{x}(t)}{t}, \frac{\Sigma}{t}\right) p(C=i)\,. & \text{(D.7)} \end{aligned}$$

Note that

$$p(C=i|\delta\boldsymbol{x}(1), \delta\boldsymbol{x}(2), ..., \delta\boldsymbol{x}(M)) = p(C=i|\boldsymbol{x}(t))\,, \qquad \text{(D.8)}$$

consistent with the conclusions of Moreno-Bote (2010). We find a different form for the posterior by considering the log-posterior ratio

$$\log\frac{p(C=i|\boldsymbol{x}(t))}{p(C=j|\boldsymbol{x}(t))} \quad = \quad \log\frac{p(C=i)}{p(C=j)} + \log\frac{N(\boldsymbol{\mu}(i); \frac{\boldsymbol{x}(t)}{t}, \frac{\Sigma}{t})}{N(\boldsymbol{\mu}(j); \frac{\boldsymbol{x}(t)}{t}, \frac{\Sigma}{t})}\,. \qquad \text{(D.9)}$$

We consider only covariance matrices which can be written in the following form

$$\Sigma = a\boldsymbol{I} + b\boldsymbol{J}\,, \tag{D.10}$$

where $\boldsymbol{J}$ is a matrix of ones. We can write this covariance matrix as

$$\Sigma = a\boldsymbol{I} + b\begin{bmatrix}1\\1\\\vdots\\1\end{bmatrix}\begin{bmatrix}1\\1\\\vdots\\1\end{bmatrix}^T \tag{D.11}$$

and use the Sherman–Morrison formula to invert it, giving

$$\Sigma^{-1} = \frac{1}{a}\left(\boldsymbol{I} - \frac{b}{a+kb}\boldsymbol{J}\right), \tag{D.12}$$

where k is the dimensionality of $\boldsymbol{x}$ (and therefore, the number of possible values of C).

This allows us to rewrite the normal distribution

$$N\left(\boldsymbol{\mu}(i); \frac{\boldsymbol{x}(t)}{t}, \frac{\Sigma}{t}\right) = \quad (2\pi)^{-\frac{k}{2}} \det\left(\frac{\Sigma}{t}\right)^{-\frac{1}{2}} e^{-\frac{t}{2}\left(\boldsymbol{\mu}(i)-\frac{\boldsymbol{x}(t)}{t}\right)^T \Sigma^{-1}\left(\boldsymbol{\mu}(i)-\frac{\boldsymbol{x}(t)}{t}\right)}\,. \tag{D.13}$$

Consider the exponent

$$\Phi = -\frac{t}{2}\left(\boldsymbol{\mu}(i) - \frac{\boldsymbol{x}(t)}{t}\right)^T \Sigma^{-1}\left(\boldsymbol{\mu}(i) - \frac{\boldsymbol{x}(t)}{t}\right). \tag{D.14}$$

Using (D.12), we have

$$\begin{aligned}
\Phi &= -\frac{t}{2a}\left(\boldsymbol{\mu}(i) - \frac{\boldsymbol{x}(t)}{t}\right)^T \left(\boldsymbol{I} - \frac{b}{a+kb}\boldsymbol{J}\right)\left(\boldsymbol{\mu}(i) - \frac{\boldsymbol{x}(t)}{t}\right) && \text{(D.15)}\\
&= -\frac{t}{2a}\Bigg(\boldsymbol{\mu}^T(i)\boldsymbol{\mu}(i) - \frac{1}{t}\boldsymbol{x}^T\boldsymbol{\mu}(i) - \frac{1}{t}\boldsymbol{\mu}^T(i)\boldsymbol{x} + \frac{1}{t^2}\boldsymbol{x}^T\boldsymbol{x} - \\
&\quad \frac{b}{a+kb}\Big((k\mu_0+\mu)^2 - \frac{2(k\mu_0+\mu)}{t}\sum_j x_j + \frac{1}{t^2}(\sum_j x_j)^2\Big)\Bigg) && \text{(D.16)}\\
&\equiv \frac{\mu x_i}{a} + \varphi\,, && \text{(D.17)}
\end{aligned}$$

where x_i is the ith element of $\boldsymbol{x}$, and φ includes all terms which do not depend on i.

We can use this result in our expression for the log-posterior ratio (D.9),

$$\begin{aligned}
\log\frac{p(C=i|\boldsymbol{x}(t))}{p(C=j|\boldsymbol{x}(t))} &= \log\frac{p(C=i)}{p(C=j)} + \log\frac{e^{\frac{\mu x_i}{a}+\varphi}}{e^{\frac{\mu x_j}{a}+\varphi}} && \text{(D.18)}\\
&= \log\frac{p(C=i)}{p(C=j)} + \frac{\mu(x_i - x_j)}{a}\,. && \text{(D.19)}
\end{aligned}$$

Using the fact that $\sum_i p(C=i|\boldsymbol{x}(t)) = 1$, we have

$$p(C=i|\boldsymbol{x}(t)) = \frac{1}{1 + \sum_{j\neq i}\frac{p(C=j)}{p(C=i)}e^{\frac{\mu(x_j-x_i)}{a}}}\,. \tag{D.20}$$

D.1.2 Approximating the optimal solution

The future evolution of $\boldsymbol{x}(t)$ is unknown, but described by

$$p(\delta\boldsymbol{x}|\boldsymbol{x}(t)) = \sum_j p(\delta\boldsymbol{x}|C=j)p(C=j|\boldsymbol{x}(t))\,, \tag{D.21}$$

where $\delta\boldsymbol{x}$ is the change in $\boldsymbol{x}$ at any future time step. We approximate the future probability distribution over $\boldsymbol{x}$ by, at each time step, only considering the most likely value for $\delta\boldsymbol{x}$. This reduces the problem of picking a stopping policy to the problem of picking a time to stop.

The $p(\delta\boldsymbol{x}|C=j)$ terms in (D.21) are multivariate normal distributions given by (D.1). The maximum value reached by each is the same, because the covariance of these multivariate distributions does not depend on j. Hence, the most probable value for $\delta\boldsymbol{x}$ is determined by the multivariate normal distribution which is multiplied by the $p(C=j|\boldsymbol{x}(t))$ term which is greatest. The maximum value achieved by the corresponding multivariate normal distribution will be at its mean, $\boldsymbol{\mu}(j)\delta T$. That is,

$$\widehat{\delta\boldsymbol{x}} = \boldsymbol{\mu}(j)\delta T \text{ for } j = \arg\max_i p(C=i|\boldsymbol{x}(t))\,, \tag{D.22}$$

where $\widehat{\delta\boldsymbol{x}}$ is the most likely value of $\delta\boldsymbol{x}$.

Note that the most likely value of $\delta\boldsymbol{x}$ is such that $\delta\boldsymbol{x}$ will only provide additional evidence for the most probable option. Hence, if $C=j$ is currently the most probable option, under our approximation, it always will be.

Consider two points $\boldsymbol{x}^{(d)}(t)$ and $\boldsymbol{x}^{(e)}(t)$ for which the posterior probability of the most probable option is the same. That is,

$$j = \arg\max_i p(C=i|\boldsymbol{x}^{(d)}(t)) \tag{D.23}$$

$$k = \arg\max_i p(C=i|\boldsymbol{x}^{(e)}(t)) \tag{D.24}$$

$$p(C=j|\boldsymbol{x}^{(d)}(t)) = p(C=k|\boldsymbol{x}^{(e)}(t))\,. \tag{D.25}$$

Using (D.20) we have

$$\frac{1}{1+\sum_{i\neq j}\frac{p(C=i)}{p(C=j)}e^{\frac{\mu(x_i^{(d)}-x_j^{(d)})}{a}}} = \frac{1}{1+\sum_{i\neq k}\frac{p(C=i)}{p(C=k)}e^{\frac{\mu(x_i^{(e)}-x_k^{(e)})}{a}}} \tag{D.26}$$

$$\sum_{i\neq j}\frac{p(C=i)}{p(C=j)}e^{\frac{\mu(x_i^{(d)}-x_j^{(d)})}{a}} = \sum_{i\neq k}\frac{p(C=i)}{p(C=k)}e^{\frac{\mu(x_i^{(e)}-x_k^{(e)})}{a}}\,. \tag{D.27}$$

Consider some later time step $t+l$. Under our approximation,

$$\boldsymbol{x}^{(d)}(t+l) = \boldsymbol{x}^{(d)}(t)+\boldsymbol{\mu}(j)l \tag{D.28}$$

$$\boldsymbol{x}^{(e)}(t+l) = \boldsymbol{x}^{(e)}(t)+\boldsymbol{\mu}(k)l\,, \tag{D.29}$$

and hence considering components of the vector $\boldsymbol{x}$,

$$x_j^{(d)} \rightarrow x_j^{(d)}+l\mu \tag{D.30}$$

$$x_i^{(d)} \rightarrow x_i^{(d)} \text{ for } i\neq j \tag{D.31}$$

$$x_k^{(e)} \rightarrow x_k^{(e)}+l\mu \tag{D.32}$$

$$x_i^{(e)} \rightarrow x_i^{(e)} \text{ for } i\neq k\,. \tag{D.33}$$

Under the approximation, the expected reward for starting from $\boldsymbol{x}^{(d)}(t)$ and stopping at $t+l$ is, recalling (6.9), given by

$$E[y_{(t+l)}|\boldsymbol{x}^{(d)}(t)] = r_e+(r_c-r_e)p(C=j|\boldsymbol{x}^{(d)}(t+l))+L(t)\,, \tag{D.34}$$

because $C=j$ is the most probable option at time step t, and we know that under our approximation, the most likely option will not change. Again using, (D.20), and using our expressions for the components of $\boldsymbol{x}(t+l)$,

$$E[y_{(t+l)}|\boldsymbol{x}^{(d)}(t)] = r_e+(r_c-r_e)\frac{1}{1+\sum_{i\neq j}\frac{p(C=i)}{p(C=j)}e^{\frac{\mu(x_i^{(d)}-x_j^{(d)}-l\mu)}{a}}}+L(t) \tag{D.35}$$

$$= r_e+(r_c-r_e)\frac{1}{1+e^{-\frac{l\mu^2}{a}}\sum_{i\neq j}\frac{p(C=i)}{p(C=j)}e^{\frac{\mu(x_i^{(d)}-x_j^{(d)})}{a}}}+L(t) \tag{D.36}$$

$$= r_e+(r_c-r_e)\frac{1}{1+e^{-\frac{l\mu^2}{a}}\sum_{i\neq k}\frac{p(C=i)}{p(C=k)}e^{\frac{\mu(x_i^{(e)}-x_k^{(e)})}{a}}}+L(t) \tag{D.37}$$

$$= r_e+(r_c-r_e)\frac{1}{1+\sum_{i\neq k}\frac{p(C=i)}{p(C=k)}e^{\frac{\mu(x_i^{(e)}-x_k^{(e)}-l\mu)}{a}}}+L(t) \tag{D.38}$$

$$= E[y_{(t+l)}|\boldsymbol{x}^{(e)}(t)]\,, \tag{D.39}$$

where we have used equation (D.27) moving from line two to three. This result shows that if the maximum posterior probability at two points is equal at time step t, then the expected reward at any later time step $t+l$ is the same under our approximation.

D.2 Optimal rule for simple case

In the case of independent accumulators, 2 response options, and a constant cost of time, we can calculate performance under the optimal stopping rule.

As discussed in the main text (Sec. 6.4.1), when the cost of time is constant, the optimal stopping rule, when described as a criterion on the posterior probabilities, does not change with time (Ferguson, 2011). In our case, the posterior probabilities are monotonically related to the difference between the two accumulators (see equation (6.6)). Hence, the problem of finding the optimal stopping rule reduces to the problem of finding the height of a time-independent stopping threshold on the difference between the accumulators.

Denote $x = x_1 - x_2$, where x_i represents the ith element of $\boldsymbol{x}$, which is defined in the main text. We are looking for the optimal magnitude of x at which to stop and decide. That is, we will only make choices for $C = 1$ at $x = K$, and choices for $C = 2$ at $x = -K$. Using equation (6.6), an equal prior, and substituting in x we have

$$p(C = 1|x = K) = \frac{1}{1 + e^{-\frac{\mu K}{a}}}\,, \tag{D.40}$$

and a similar result holds for $C = 2$. If we are in a trial in which we reach $x = K$ and hence the response is $\widehat{C} = 1$, then the probability we are correct is $p(C = 1|x = K)$. In the perceptual decision on which the networks were trained, the reward for getting the wrong answer was the negative of the reward for getting the right answer. If we denote these rewards $-r$, and r, if we therefore reach $x = K$ the expected reward from deciding is

$$p(C = 1|x = K)r - (1 - p(C = 1|x = K))r\,. \tag{D.41}$$

A similar result holds for a response $\widehat{C} = 2$ at the boundary $x = -K$, and by symmetry $p(C = 1|x = K) = p(C = 2|x = -K)$. Denote this value $P = p(C = 1|x = K) = p(C = 2|x = -K)$, and note that P is the probability of making the correct decision. Hence, regardless of the boundary reached, the expected reward from deciding is

$$Pr - (1 - P)r\,. \tag{D.42}$$

In order to calculate the overall expected reward, we also need to know the cost of time, and hence the expected number of time steps spent accumulating evidence. From Shadlen et al. (2006) we have that the expected number of time steps equals the expected value of x at decision time, divided by the average change in x in the absence of decision boundaries. When the true value of C is $C = 1$, the average change in $x = x_1 - x_2$ in the absence of boundaries is just μ. Hence,

$$E[t] = \frac{E[x]}{\mu}\,. \tag{D.43}$$

When $C = 1$, the correct decision boundary is $x = K$ and the probability of reaching this boundary is P. There is a $1 - P$ probability of reaching the other boundary at $x = -K$. Hence,

$$E[x] = PK - (1 - P)K\,. \tag{D.44}$$

Which gives us

$$E[t] = \frac{PK - (1 - P)K}{\mu}\,. \tag{D.45}$$

An identical expression holds when $C = 2$.

The overall expected reward, $E[y]$, is the expected reward from deciding minus

the expected cost of time

$$\begin{aligned} E[y] &= Pr - (1-P)r - E[t]l && \text{(D.46)} \\ &= Pr - (1-P)r - \frac{l}{\mu}(PK - (1-P)K) && \text{(D.47)} \\ &= (r - \frac{lK}{\mu})(2P-1) && \text{(D.48)} \\ &= (r - \frac{lK}{\mu})(\frac{2}{1+e^{-\frac{\mu K}{a}}} - 1) && \text{(D.49)} \\ &= (r - \frac{lK}{\mu}) \tanh \frac{\mu K}{2a}\,, && \text{(D.50)} \end{aligned}$$

where l is the cost per time step.

To find the value of K which maximises the expected reward we take the derivative. Setting the derivative to zero and rearranging we find

$$\sinh \frac{\mu K}{a} - \frac{\mu}{a}(\frac{r\mu}{l} - K) = 0\,, \qquad \text{(D.51)}$$

which we can solve for K using standard solvers. Using K we can find the maximum obtainable reward using equation (D.50).

D.3 4 options

In addition to requiring the neural networks to choose between 2 and 10 options, we also looked at an intermediate case, where the networks had to choose between 4 options. In Fig. D.1 we again show the reward per trial achieved by the different networks. Unsurprisingly, the behaviour of the networks is somewhere between the behaviour observed for 2 and 10 locations. Again, in all cases, the "all probability" network is as good as, or better than all other networks; for all cases it is able to learn the optimal stopping rule the fastest.

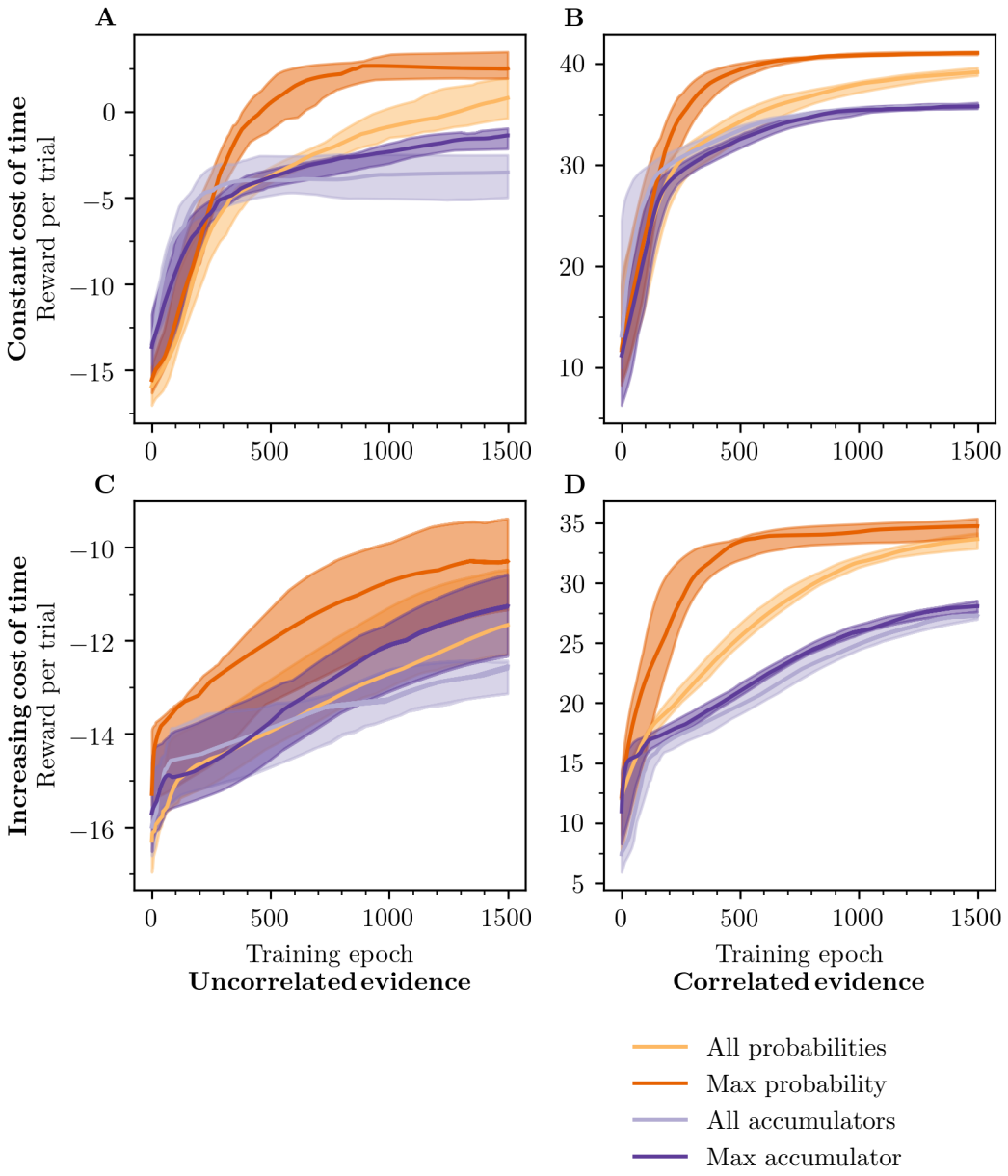

Figure D.1: Average reward gained per trial for the case of 4 choices. Results are shown for cases with and without correlations between evidence accumulations, and with a constant and an increasing cost of time.

In Fig. D.2 we show show the learned stopping rules for the "max probability" and "all accumulator" networks. In the case of a constant cost of time (Fig. D.2,

A, B), the "all accumulator" model is again not learning a sensible stopping rule, whilst the "max probability" network has learned a rule which is consistent with the features of optimal stopping discussed in the main text. There is a similar pattern of results in the increasing cost case, (Fig. D.2, C, D).

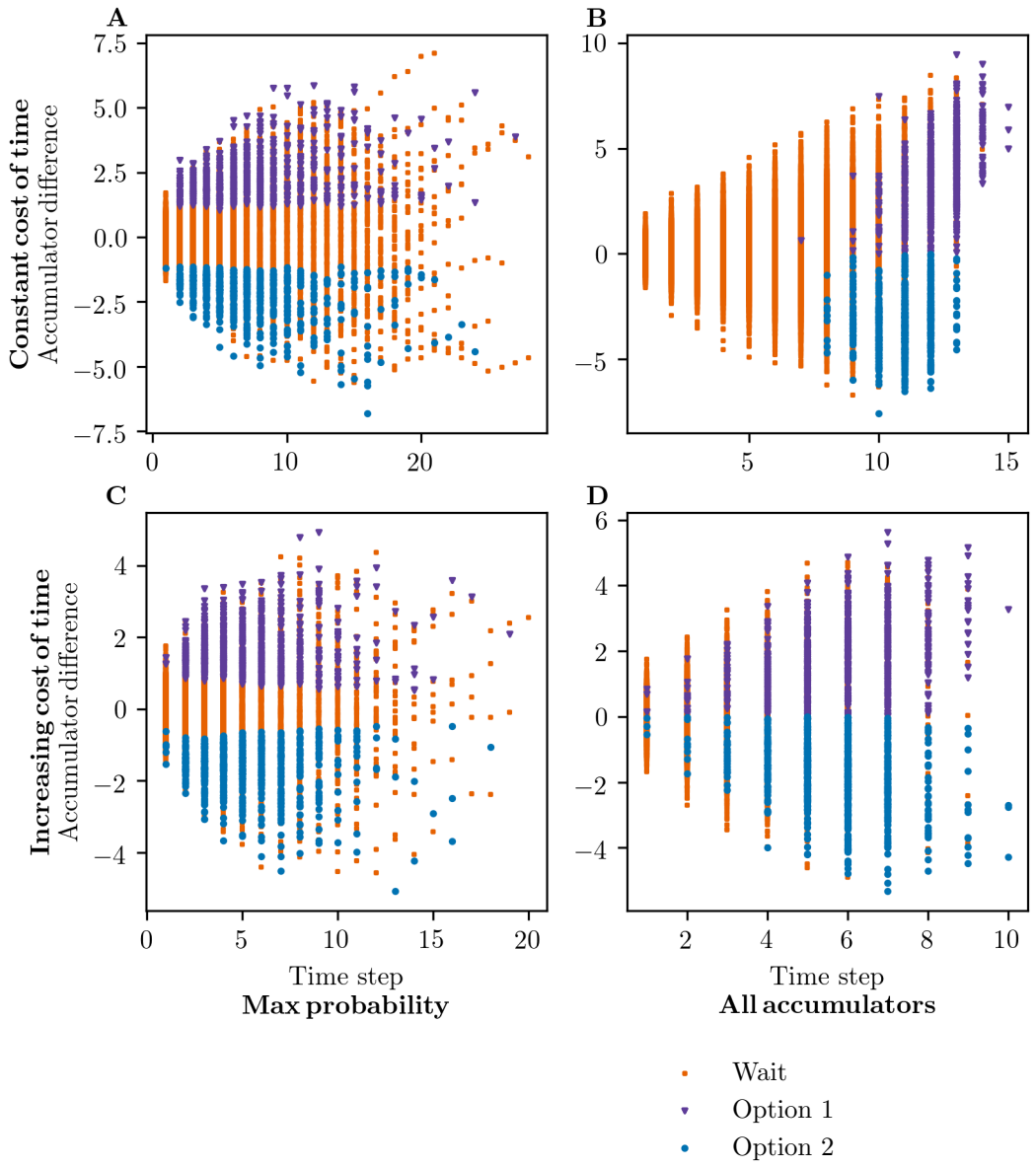

Figure D.2: Learned decision rule for "max probability" (left) and "all accumulator" (right) networks, for correlated evidence accumulations with a constant cost of time (top) or increasing cost of time (bottom) for the case of 4 choices.

E

Additional research: EEG correlates of evidence accumulation

Contents

E.1 Introduction

Throughout the thesis, we have explored perceptual decision making and confidence through computational modelling. In one of the first projects of my DPhil, we examined evidence accumulation from a different perspective, that of electroencephalographic (EEG) responses to evidence. O'Connell et al. (2012) identified a centro-parietal and positive (CPP) signal that behaves as if it indexes the accumulation of evidence in the brain, building up following onset of a target to a stereotyped height at the time of response. The signal builds up more slowly when weaker evidence is presented (Kelly & O'Connell, 2013), and is responsive to changes in evidence during decision formation (O'Connell et al., 2012). Importantly, the signal

is also present even when no immediate action is required (O'Connell et al., 2012). Collectively, these findings suggest that we may be able to effectively index the progress of evidence accumulation over time using EEG. If we can work out how the evidence presented in a stimulus frame feeds into the EEG signal over time – the impulse response function of a frame of evidence – we will be able to use this information to decode the sensory samples passed into the accumulator. Access to the information passed into the accumulator would allow us to directly determine the height of the threshold. In addition, we would be able to compare the evidence in the stimulus and the brain's representation of that evidence.

We have discussed that several findings support the view that evidence accumulation continues following a decision, and confidence is read out from the accumulator later than the time of the decision itself (Section 1.3.1; Yu et al., 2015). However, the CPP drops off following a response (O'Connell et al., 2012). In contrast to the account we have proposed (Chapter 1), this points to the idea that different processes may be responsible for decisions and confidence. It has been reported that there may be a second accumulation signal following decisions, when an error in the original decision is detected (Murphy et al., 2015). This accumulation signal may be the same signal as the positive central-parietal ERP component which follows errors, the error positivity (Pe; Murphy et al., 2015). The Pe varies with confidence, such that the ERP is more positive when participants are more sure they are wrong (Boldt & Yeung, 2015). These findings point to the idea that following a decision, a new accumulation – of evidence for error commission – takes place. This interpretation leads to a novel prediction that, following a perceptual decision, the sign of the evidence impulse response function should flip. A stimulus frame that provided evidence for the initial response in the pre-decision period (contributing to the CPP) should, after the response, be treated as evidence against the fact that an error occurred (reducing the Pe).

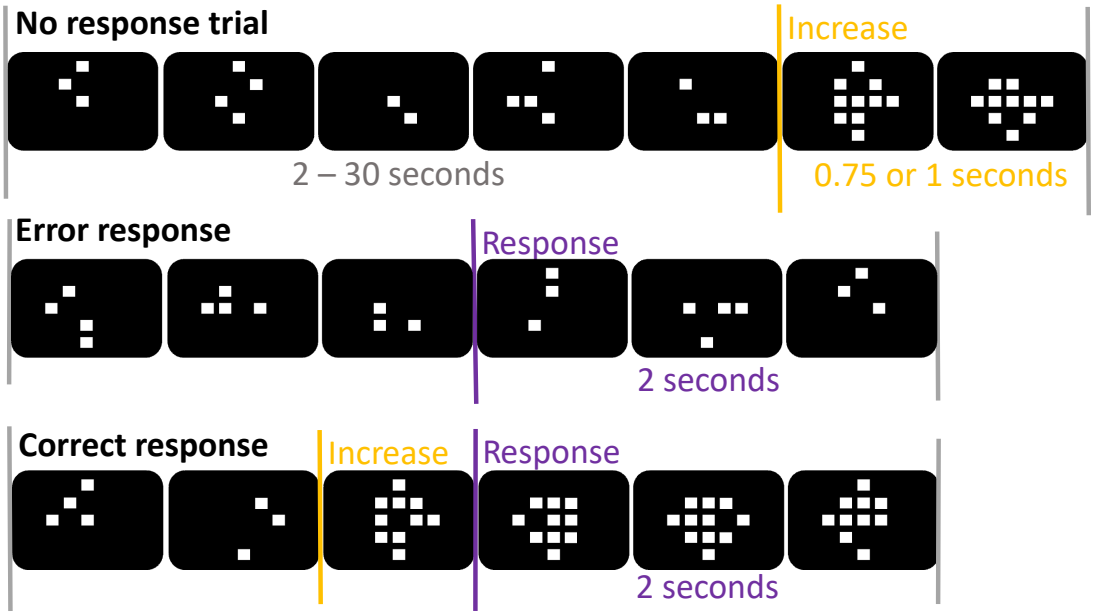

Figure E.1: The task used. Note that "screen shots" are schematics and for illustration only. Three types of trial are shown, those in which no response is made, those in which a correct response is made, and those in which an error response is made.

E.2 Method

Task. We used a task in which evidence fluctuated rapidly, so that we could examine the effect of evidence on the CPP. Additionally, we selected a task in which there was a single response, rather than a binary choice, so that the CPP would be unambiguously related to evidence for the single response. Human participants were presented with a circular patch of dots (Fig. E.1). Every 50 ms the number of dots in the patch was drawn from a Gaussian probability distribution centred on 184 (standard deviation 35). From 2 seconds after the onset of the stimulus, every frame there was a small chance that the mean of the Gaussian would increase to 229. Participants had 750 ms or 1000 ms to respond to this "target" (set based on performance). If participants responded at any time point, the mean of the distribution was frozen, and the stimulus continued for a further 2 seconds.

Participants. Excluding piloting, we collected EEG data from 20 participants.

Analysis. We focused on determining the impulse response function of a frame of evidence, on scalp potential. We took two approaches to answering this question. The first involved a simple ERP analysis, the second involved a regression based approach. The ERP analysis consisted of time locking to the onset of every stimulus frame. The frames were binned into quantiles based on the amount of evidence they provided that the target had appeared (more dots is more evidence). The analysis was performed separately for each bin. This produces an average time course of scalp potential associated with each of the different evidence quantiles.

In the regression based approach we assume that neural responses to evidence can be modelled by a finite impulse response function (Ljung, 1999). We assume the effect of a frame on scalp potential can simply be added to the effect of other frames. Specifically, we predict the value of the scalp potential at time, t, by looking at the value of the evidence in the stimulus n time lags back, and multiply this by the impulse response function at lag n. We do this for every time lag between time t, and $t - 2$ seconds.

Treating the values of the impulse response function as parameters to be inferred, we now have a linear regression, in which the values of stimulus evidence in previous frames are used to predict the scalp potential (Ljung, 1999). In spite of the fact that the scalp potential will in general exhibit strong autocorrelation, we can treat every sampled scalp potential as an independent case in the regression, and still produce consistent estimates of the impulse response function, provided noise in the scalp potential is suitably modelled (Ljung, 1999). The noise in EEG recordings is certainly not independent across time points, as EEG recordings exhibit long slow drifts. Instead we assume the noise can be adequately modelled as an autoregressive process of order one. This means that the noise at time, t, is given by the noise in the previous sample multiplied by some weight, plus noise drawn from a Gaussian. We can estimate this weight as a parameter in our linear regression (Ljung, 1999).

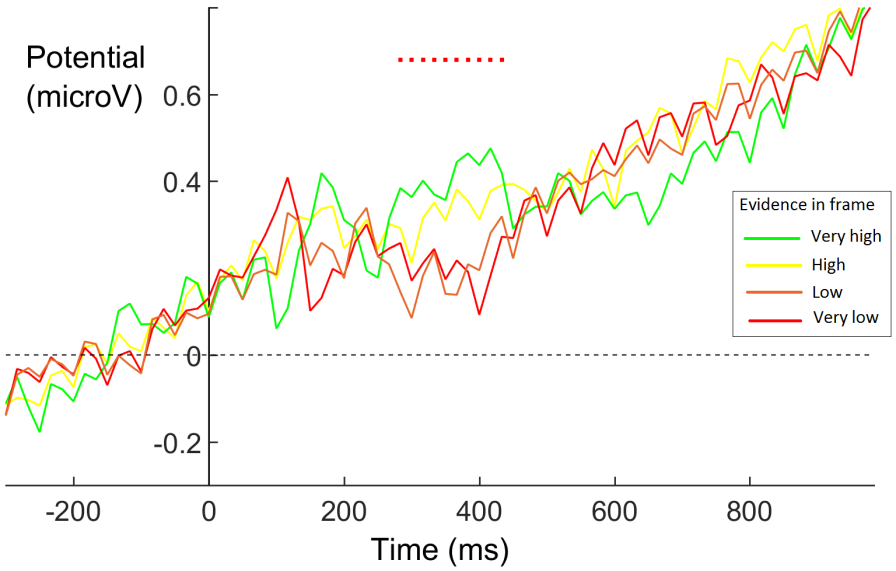

Figure E.2: Frame onset locked ERP at Pz for frames before the target and at least 2 seconds prior to response. Evidence samples are binned into four quantiles according to the amount of evidence they provide that the target has appeared. Dots represent a significant difference between the two highest and two lowest quantiles in a permutation test.

E.3 Results

We first considered frames of evidence that occur at least 2 seconds before any response, and which are prior to the target. The ERP analysis appeared to show an effect of number of stimulus dots between 300 and 500 ms following the presentation of a frame, at central parietal electrodes (Fig. E.2). Using permutation tests we looked for a difference between the two high evidence and two low evidence quantiles. The identified effect was significant, with more dots associated with a more positive potential. We suggest that this effect reflects evidence accumulation in the neural population that generates the CPP. Note that by 600 ms any effect of the number of dots in a frame on the signal has leaked away. This is expected given that the task involves change detection. In this case samples of evidence are only relevant for short periods of time before they become unreliable indicators of the state of the environment (Thura et al., 2012).

The regression based analysis of data from the same period appeared to show a

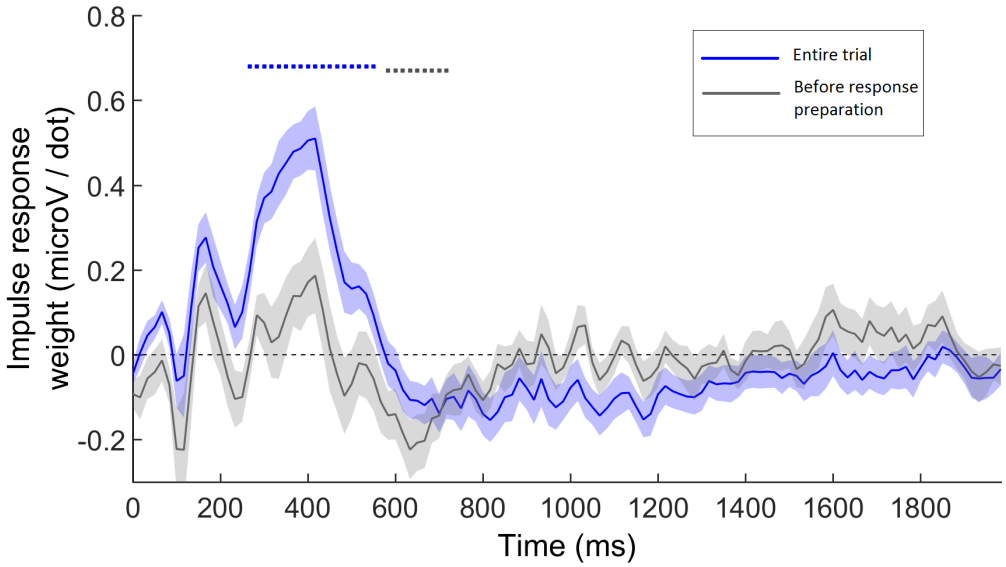

Figure E.3: The impulse response function of number of dots on scalp potential at Pz, over the entire duration of the trial (blue), and up to one second prior to response (grey). Dots represent significance in a permutation test. Error bars represent one standard error in a t-test against zero, uncorrected for multiple comparisons.

different pattern. When we considered data from the entire course of the trials, an effect in the same direction was present between 300 ms and 500 ms following the presentation of a frame (Fig E.3, blue line). However, this effect was not observed when we considered data occurring at least one second before any response, as we did in the ERP analysis (Fig E.3, grey line). Instead, we observed a negative effect of the number of dots at approximately 600 ms after frame onset (Fig E.3, grey line). A possible reason for this difference between the ERP and regression analysis is that the regression based approach assumes a linear relationship between the number of dots and scalp potential. This will increase power to detect linear effects, but obscure relationships which are not linear.

We next looked for the predicted flip in the impulse response function following decisions. For the scalp potential data between the response and the end of the trial, we see that the impulse response function is similar to the function before the response, and there is no evidence of a flip (Fig. E.4).

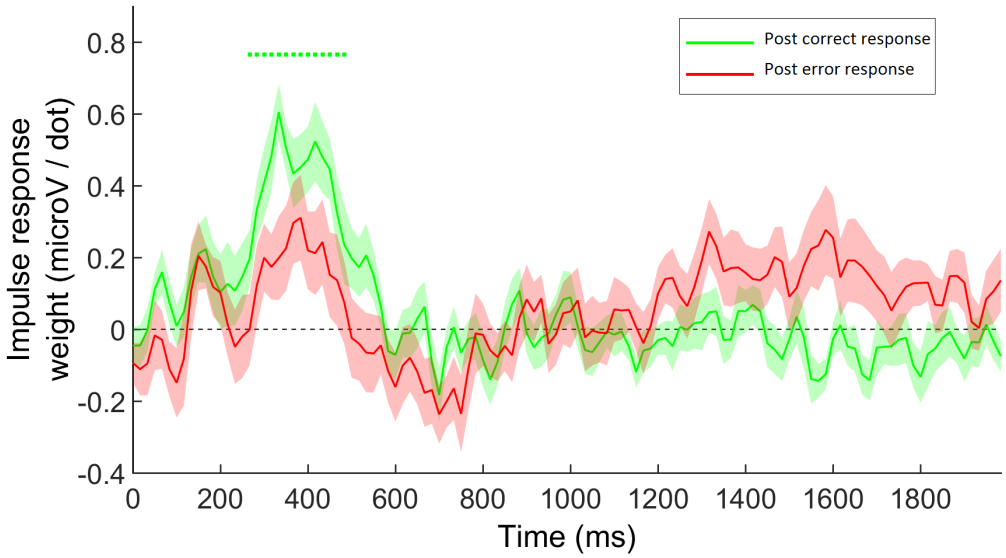

Figure E.4: The impulse response function of number of dots on scalp potential at Pz, after response. Correct (green) and error (red) trials were analysed separately. Dots represent significance in a permutation test. Error bars represent one standard error in a t-test against zero, uncorrected for multiple comparisons.

E.4 Discussion

The findings suggest that evidence in a stimulus frame has a positive effect on the CPP between 300 and 500 ms after the frame's onset, although the exact nature of the impulse response function is not yet clear. Contrary to our predictions, the impulse response function does not appear to flip following response. Hence, we did not find evidence for a confidence accumulation mechanism which is separate to the perceptual decision mechanism.

Before strong conclusions can be drawn from this work, it seems important to run a version of the study in which we dissociate the effects of evidence and stimulus luminosity, which are currently confounded. This will be trivial to do: In half the blocks the task will be to detect a decrease rather than an increase in the number of dots.

F

Additional research: A first attempt at a new modelling method

In our first attempt to model confidence we hoped to directly infer the shape of the decision thresholds which are used to trigger responses. We have seen how patterns in confidence are closely related to the shape of decision thresholds (e.g. Section 1.4.1), so directly estimating decision thresholds seemed like a sensible approach to take.

A recently developed method for inferring decision thresholds relies on stimuli that provide time-varying evidence for the two options (Malhotra et al., 2017). The total stimulus evidence presented in a given trial, up to a given time point, can be used as an estimate for the state of the observer's internal evidence accumulator at that time. If we have an estimate for the state of the accumulator at the time of the decision, we can infer the threshold used: The accumulator must be at the decision threshold at this time. However, we can also use the estimated state of the accumulator at time points prior to the decision to support our estimate of the threshold. At these time points the accumulator must be below the threshold. One way of combining this information is to use the estimated state of the accumulator at each time point, and the time point itself, as predictors in a logistic regression

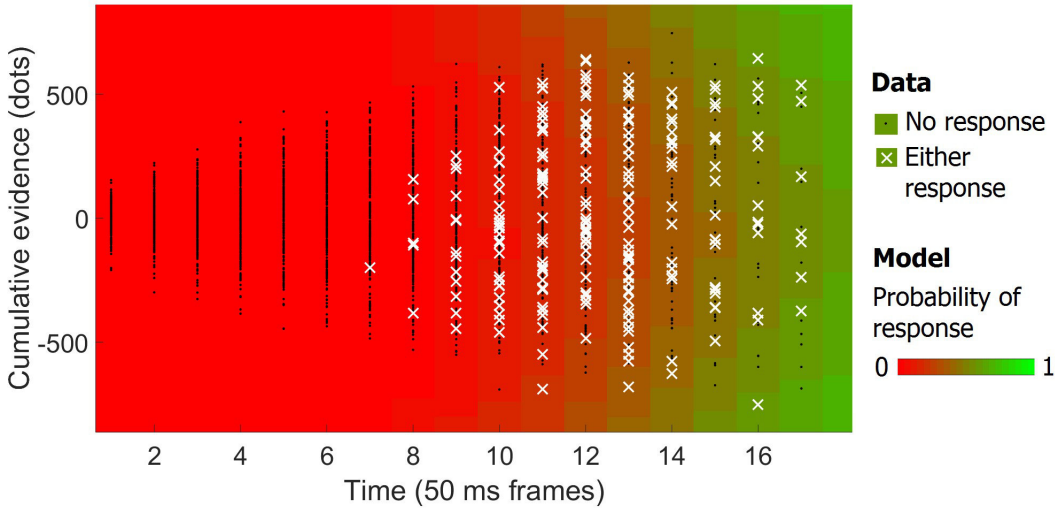

Figure F.1: Model estimated probability of responding as a function of time and evidence, and actual responses for one participant. If we assume the threshold follows the contour at which the probability of responding is 0.5, then this analysis suggests a threshold which decreases rapidly, so that at frame 16 almost no evidence is required for a response.

(Malhotra et al., 2017). The outcome in this regression is whether the participant responded for option A, responded for option B, or did not respond at the time point.

We applied this method to the dataset from Charles and Yeung (2019), which was also used in Chapter 2. Initial results were promising, and pointed to collapsing thresholds (Fig. F.1). However, when we performed model recovery analysis – in which datasets are simulated to test that the method generates accurate results – we found that the method consistently overestimated the complexity of the simulated data. In particular, whether the simulated threshold was flat or linear, the analysis consistently estimated a parameter associated with threshold curvature to be significant.

One of the main reasons for this effect may be that those trials in which a response is slow, are also trials in which the random effects of internal noise cancel out much of the evidence presented in the stimulus. Otherwise, the evidence presented in the stimulus would have driven the accumulator to cross the threshold. As a consequence, even when the decision threshold is flat, the total amount of evidence presented in the stimulus at response time is not constant. Instead, it is most likely a highly

complex function that can only be discovered through solution of the mathematical equations for diffusion towards absorbing boundaries (e.g. Shinn et al., 2020).

Malhotra et al. (2017) noted that the logistic regression model used, might not accurately describe the process through which responses and response times are generated. Through simulations they thoroughly tested and verified that their analysis approach provided valid results in the context of their experiment. This suggests the issues we identified are only important under certain conditions, and hence for certain stimuli. We were unlucky that the stimuli we used are ones for which differences between the data generating process and the logistic regression model are important.

G

Additional research: Single-trial SDT modelling of visual search

One project completed during the DPhil, but falling outside the scope of the thesis, involved trial-by-trial modelling of the effects of distractor statistics in visual search. We used SDT models, including a Bayesian SDT model.

The description of this project is omitted from this version of the thesis but will be available, open access, in the Journal of Vision. The version initially submitted to the Journal of Vision can be found at the following site (https://doi.org/10.1101/2020.01.03.893057). Upon publication in the Journal of Vision, this site will be updated with a link to the final open access version.