Expressions for Bayesian confidence of drift diffusion observers in dynamic stimuli tasks

Joshua Calder-Travis[1], Rafal Bogacz [2] & Nick Yeung [1]

[1] *Department of Experimental Psychology, University of Oxford*

[2] *MRC Brain Network Dynamics Unit, Nuffield Department of Clinical Neuroscience, University of Oxford*

Abstract

Much work has explored the possibility that the drift diffusion model, a model of response times and choices, could be extended to account for confidence reports. Many methods for making predictions from such models exist, although these methods either assume that stimuli are static over the course of a trial, or are computationally expensive, making it difficult to capitalise on trial-by-trial variability in dynamic stimuli. Using the framework of the drift diffusion model with time-dependent thresholds, and the idea of a Bayesian confidence readout, we derive expressions for the probability distribution over confidence reports. In line with current models of confidence, the derivations allow for the accumulation of "pipeline" evidence which has been received but not processed by the time of response, the effect of drift rate variability, and metacognitive noise. The expressions are valid for stimuli which change over the course of a trial with normally distributed fluctuations in the evidence they provide. A number of approximations are made to arrive at the final expressions, and we test all approximations via simulation. The derived expressions only contain a small number of standard functions, and only require evaluating once per trial, making trial-by-trial modelling of confidence data in dynamic stimuli tasks more feasible. We conclude by using the expressions to gain insight into the confidence of optimal observers, and empirically observed patterns.

1 Introduction

How animals and humans make perceptual decisions is of fundamental interest. It is increasingly recognised that decision confidence, an estimate of the probability a decision was correct, is both theoretically important and used by humans in a variety of ways (Bahrami et al., 2010; Desender, Boldt, Verguts & Donner, 2019; Desender, Boldt & Yeung, 2018; Drugowitsch, Mendonça, Mainen & Pouget, 2019; Sanders, Hangya & Kepecs, 2016; van den Berg, Zylberberg, Kiani, Shadlen & Wolpert, 2016). Confidence has also been linked to psychological disorder (Hauser, Allen, Rees & Dolan, 2017; Rouault, Seow, Gillan & Fleming, 2018). We aim to find expressions for confidence in a particular observer model, specifically the drift diffusion model (also known as the diffusion decision model; DDM; Ratcliff and McKoon, 2008), coupled with a Bayesian readout for confidence (Kiani & Shadlen, 2009; Moreno-Bote, 2010; Sanders et al., 2016). The two components of this model are of special interest for reasons set out below, and their combination has been found to account for distinctive patterns in confidence data (Desender, Donner & Verguts, 2020). Using this observer model we derive expressions for the probability distribution over confidence reports, given the response and response time on a trial. While we keep the dependence of these predictions on the model parameters implicit throughout (for the sake of readability) the predictions of course depend on the value of these parameters. Therefore we can use the derived expressions to fit the model, and variants of the model, to experimental data, providing parameter estimates and permitting model comparison.

It is important to ask at the outset why we would want to derive explicit mathematical expressions that are computationally cheap to evaluate, when computational modelling can often be performed in other ways. This exercise has two main purposes: We aim for computationally cheap predictions to make trial-by-trial modelling of dynamic stimuli feasible, and we aim for explicit mathematical expressions to gain deeper insight into the nature of confidence in an important model. Explicit mathematical expressions provide immediate knowledge of the relationships between different parameters and variables, and how they combine to produce confidence. Indeed, we will see that we can go further and interpret

such expressions to understand why different variables have the effect they do (Section 4). Regarding our other motivation, making trial-by-trial modelling feasible, such modelling allows us to capitalise on variability in stimuli, rather than ignoring it (Park, Lueckmann, von Kriegstein, Bitzer & Kiebel, 2016). When modelling on a trial-by-trial basis, a model that can capture the specific effects of each stimulus will outperform a model which can only capture general patterns in confidence across conditions. Hence, computationally cheap expressions may facilitate the development and testing of models for confidence that make increasingly precise predictions for behaviour.

The 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 (Bogacz, Brown, Moehlis, Holmes & Cohen, 2006; Green & Swets, 1966; Ratcliff & 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). When the accumulator tracking this difference reaches a fixed threshold (positive or negative), a response is triggered. The DDM has successfully been used to model decisions in a wide range of tasks (Ratcliff, Smith, Brown & McKoon, 2016).

The DDM is also a normative model of decision making. Under certain conditions, such as evidence measurements for the two options being equally reliable and signal strength known, the DDM is equivalent to tracking the posterior probability of each option until a fixed threshold on these probabilities is reached (Bitzer, Park, Blankenburg & Kiebel, 2014; Gold & Shadlen, 2007; Moran, 2015). In such a context, 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). When signal strength is unknown, a time dependent threshold is required, but under standard assumptions it nevertheless remains optimal to track the difference between the two accumulators (Drugowitsch, Moreno-Bote, Churchland, Shadlen & Pouget, 2012; Moran, 2015; Tajima, Drugowitsch, Patel & Pouget, 2019). For an intuition of why this policy is optimal, 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, so they should lower their decision threshold and make an immediate decision (Malhotra, Leslie, Ludwig & Bogacz, 2017).

While the DDM has optimal characteristics, and has been successfully applied to a wide range of decisions (Ratcliff et al., 2016), it is not clear the DDM provides an adequate account of confidence reports. Different ways of modelling confidence using the DDM have been proposed (Yeung & Summerfield, 2014). In one set of models, observers use some form of heuristic, based on variables which are directly accessible in the DDM, such as the state of the accumulator (Pleskac & Busemeyer, 2010), or the time taken to make a decision (Zylberberg, Barttfeld & Sigman, 2012). Another approach is to assume that observers map the state of the accumulator, and the time spent accumulating evidence, to the probability they are correct (Kiani, Corthell & Shadlen, 2014; Kiani & Shadlen, 2009; Moreno-Bote, 2010). A Bayesian readout of this kind could be learned over time, through the association of accumulator state and time with success or failure (Kiani & Shadlen, 2009). Alternatively, a Bayesian readout could reflect a probabilistic inference made using knowledge of the statistical structure of the task. One detail to consider is that the confidence readout could be based on a separate evidence accumulator to the one used for the decision (Balsdon, Wyart & Mamassian, 2020; Ganupuru, Goldring, Harun & Hanks, 2019; Jang, Wallsten & Huber, 2012). However, here we make the simplest assumption that decisions and confidence are based on the same evidence accumulator (Moreno-Bote, 2010).

There are several techniques which can be used for calculating the probability, according to the DDM, of different responses and response times (Brown, Ratcliff & Smith, 2006; Chang & Cooper, 1970; Cox & Miller, 1965; Diederich & Busemeyer, 2003; Drugowitsch, 2016; Navarro & Fuss, 2009; Shinn, Lam & Murray, 2020; Smith, 2000; Tuerlinckx, 2004; Tuerlinckx, Maris, Ratcliff & De Boeck, 2001; Voss & Voss, 2008). Importantly, approaches which can handle dynamic stimuli (stimuli which change over the course of a trial), and time-dependent thresholds, have been developed. One approach involves using finite difference methods to approximate the evolution of the probability distribution over accumulator state (which reflects the accumulated difference in evidence measurements; Chang and Cooper, 1970; Shinn et al., 2020; Voss and Voss, 2008; Zylberberg, Wolpert and Shadlen, 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 the trial-by-trial variability inherent in dynamic stochastic stimuli. (Practical solutions for trial-by-trial modelling already exist for static stimuli under certain conditions; e.g. Wiecki, Sofer and Frank, 2013). Often the computation time needed for calculating predictions is reduced by making predictions on a condition-by-condition basis (Park et al., 2016; e.g. Kiani et al., 2014; Ratcliff and McKoon, 2008; van den Berg, Anandalingam et al., 2016; Zylberberg, Fetsch and Shadlen, 2016). We design the experiment 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 the trial-by-trial variability of the dynamic stimuli which are often used (for model-free analyses that do capitalise on trial-by-trial variability see Charles and Yeung, 2019; Kiani, Hanks and Shadlen, 2008; Zylberberg et al., 2012). A method which reduced the computational cost of making trial-by-trial predictions for dynamic stimuli could allow us to perform model fitting which does not discard, but instead capitalises on the rich data produced from such stimuli.

One successful approach to the trial-by-trial modelling of evidence accumulation processes has avoided the usual computational costs by simulating evidence accumulation using large time steps (Park et al., 2016). While this is a promising approach that is particularly well suited to certain stimuli (such as those used by Park et al., 2016), we aim to model a process which evolves over very short times scales, driven by ongoing stimulus presentation and noise.

Another approach is to derive explicit mathematical expressions for model predictions, which could dramatically reduce computation time. Moreno-Bote (2010) derived expressions for confidence which take into account time-dependent thresholds, and which could be extended to account for dynamic stimuli of the kind we consider below. However, these derivations use two assumptions about the computation of confidence which are not in line with recent findings. First, Moreno-Bote (2010) made the intuitive assumption that decisions and confidence are based on the same information. However, as sensory and motor processing takes time, there will be stimulus information in these processing "pipelines" which does not contribute to the initial decision, but does inform confidence (Charles & Yeung, 2019; Moran, Teodorescu & Usher, 2015; Ratcliff & McKoon, 2008; Resulaj, Kiani, Wolpert & Shadlen, 2009; van den Berg, Anandalingam et al., 2016). Moreover, information in the processing pipeline is affected by trial-to-trial fluctuations in signal strength, and this may have important effects on confidence (Pleskac & Busemeyer, 2010). Second, there is now substantial evidence that the process which "reads out" confidence into a behavioural report is corrupted by "metacognitive noise" (Bang, Shekhar & Rahnev, 2019; De Martino, Fleming, Garrett & Dolan, 2013; Maniscalco & Lau, 2012, 2016; van den Berg, Yoo & Ma, 2017). We aim for mathematical expressions that take these important features into account.

A key idea that affects the scope of our derivations is that, while it may be very difficult or impossible to derive simple expressions for decisions and response times, it may be much simpler to find expressions for confidence. Prior to a decision, even if changes to the state of the accumulator are normally distributed over small intervals of time, 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. 1). Following a decision, decision boundaries are no longer relevant, hence normally distributed changes in the state of the accumulator lead to a normal distribution over this state (Fig. 1). Our contribution is not to provide new expressions for the probability distribution over response times and decisions in diffusion models. We aim to bypass the complexities associated with response times and decisions, and instead focus on confidence.

Using this strategy 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 (under standard assumptions) the optimal

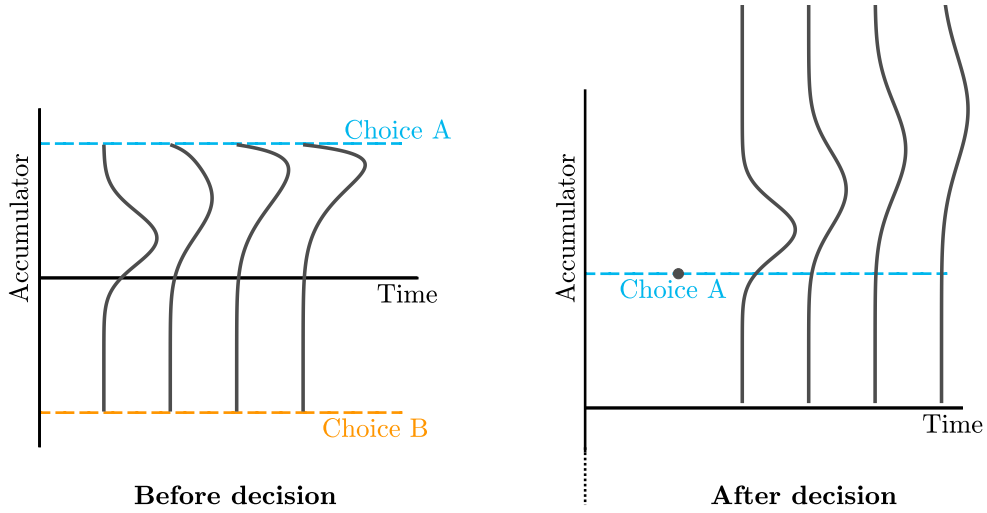

Figure 1: Probability distributions over accumulator state before and after a decision. Even if increments in the accumulator are normally distributed, prior to a decision the probability distribution over the state of the accumulator quickly becomes non-normal. This is because, if we get to time t without a decision, we know the accumulator has not been beyond either decision threshold prior to t. Following a decision (the time of which is represented by the dot in the right panel), normally distributed increments in the accumulator lead to a normal distribution over accumulator state, because there are no longer decision thresholds.

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. As discussed, making trial-by-trial modelling feasible is only one of our aims. Once we have derived the expressions we return to our second aim of gaining deeper insight into confidence from them.

2 Model

Overview

Our aim is to derive expressions for the probability distribution over confidence, given the response and response time on a trial. In order to derive these predictions we must first specify a model for decisions and confidence.

We consider a situation in which observers must make a choice between two alternatives. The presented stimulus provides two evidence signals, one for each option, and evidence provided by the stimulus can vary over time (Fig. 2; Bogacz et al., 2006; Moreno-Bote, 2010). For example, the observer might be presented with two clouds of dots, with the number of dots in each cloud constantly changing. Their 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 (Green & Swets, 1966; Ratcliff & McKoon, 2008). 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. 2).

Typically, following stimulus onset, a participant can respond whenever they wish. Some instruction

Symbol	Meaning
S	Stimulus (1 or 2)
R	Response (1 or 2)
C	Confidence
μ	Difference between mean stimulus evidence for the two options
$\Delta\mu$	Absolute value of the difference between the two means
E_i	Difference in evidence presented in stimulus frame i
δE_i	Difference in evidence presented in time step i.
	Note, $E_i = \sum_j \delta E_j$ (sum taken over time steps in a frame)
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
δx_i	Difference between the evidence measurements in time step i
$\boldsymbol{\delta x}$	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
x_{lp}	Scaled log-posterior ratio
x_c	Noisy measurement of x_{lp} which determines confidence
d_i	Boundaries on x_c which separate confidence reports falling into different bins
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 duration of the evidence pipeline
Symbol	**Abbreviates**
ν	$\varphi(\mu_2 - \mu_1)/t_f$
ν_0	$\Delta\mu/t_f$
σ_ν	$\nu_0\sigma_\varphi$
s^2	$\sigma_{acc}^2 + (\sigma_E^2/t_f)$
$\theta(t)$	$(s^2 + t\sigma_\nu^2)/(s^2 + \sigma_\nu^2)$
K	$(s^2 + \sigma_\nu^2)/2\nu_0$

Table 1: Symbols used in the derivations, along with key abbreviations.

or incentive may be given to respond in a certain way, such as fast and accurately, but beyond this the participant is free to set the time of response (e.g. Ratcliff & McKoon, 2008). The stimulus continues to be presented until a response is made. We call this condition "free response" (McMillen & Holmes, 2006). Following the DDM, 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. 2). As discussed in Section 1, measurements corresponding to evidence that has recently been presented will still be in sensory and motor processing pipelines at the time of response, and hence will not contribute to a decision (Resulaj et al., 2009). These measurements will be processed immediately following a response, and will be used to inform confidence (van den Berg, Anandalingam et al., 2016).

We also consider the "interrogation" condition (McMillen & Holmes, 2006), where the observer must respond at a time controlled by the researcher. In this case the stimulus is presented for a finite amount of time. Before the stimulus clears the participant cannot respond. Once the stimulus clears, the observer uses the final state of the accumulator (which reflects all evidence presented in the stimulus) to determine their response (Fig. 2; Bogacz et al., 2006). In all conditions, the observer uses a Bayesian readout of confidence which depends on the final state of the accumulator once all evidence has been processed, and the time spent accumulating evidence (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 mathematically, 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 noisy "read out" process which determines confidence (Fleming & Daw, 2017). In section 3 we turn to the main aim of the paper, which is to use the drift diffusion framework to derive simple expressions for the probability distribution over confidence reports, given a response and response time. A summary of symbols used in the derivations can be found in Table 1.

Task and observer beliefs

The observer's task is 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). As we are considering a DDM observer who tracks the difference in the evidence measurements for the two alternatives, we only need to consider the difference in evidence provided by the two evidence signals from the stimulus.

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, $\mu = \mu_2 - \mu_1$. We consider a situation in which the absolute value of the difference between the two means is a fixed value. However, we incorporate variability in signal strength below (see discussion of standardised drift rate variability, φ). Let $\Delta\mu$ denote a fixed positive value which determines the absolute value of the difference between the two means. This setup gives us,

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

$$p(\mu|S=1) = \delta(\mu + \Delta\mu) \tag{2}$$

$$p(\mu|S=2) = \delta(\mu - \Delta\mu)\,. \tag{3}$$

S denotes the stimulus (1 or 2) with the greater mean evidence, and hence the correct answer. (Note that a positive value for μ corresponds to the case in which $S = 2$, and a negative value to $S = 1$.)

In the interrogation condition the observer uses all information presented in the stimulus to determine both their response and confidence, however, in the free response condition we have to consider the effects of the decision thresholds (discussed above). 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 of the lag in sensory processing, and because of the 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). Hence, information presented in the stimulus immediately prior to the response will not contribute to the decision. Denote the duration

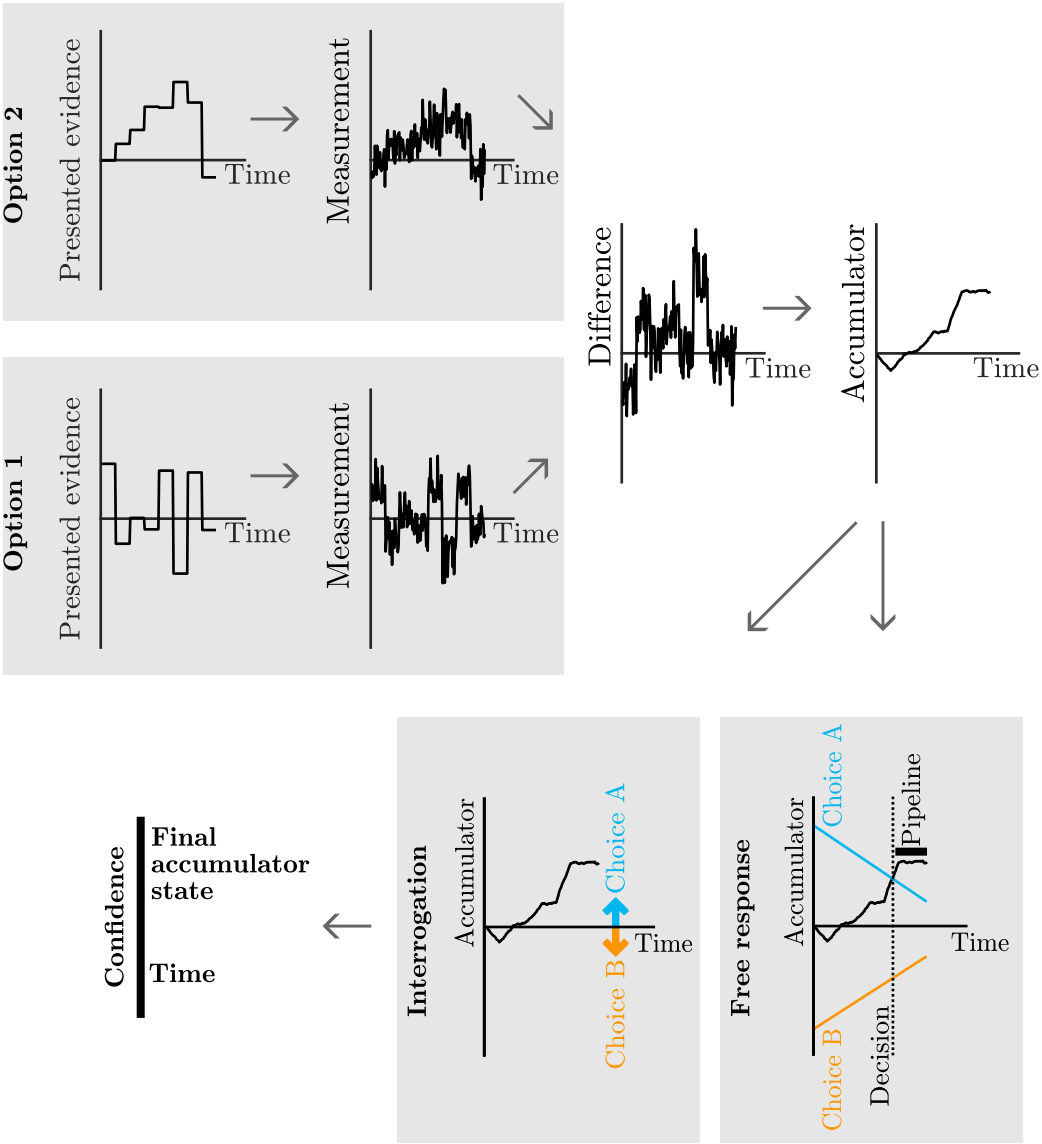

Figure 2: The model of confidence and decisions. Time varying evidence is presented for the two response options. The observer only receives noisy measurements of the evidence presented. The observer takes the difference between the evidence measurements and accumulates this difference. In the interrogation 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.

of the stimulus immediately prior to the response which does not contribute to the decision, because corresponding measurements are still in processing pipelines, I. (Note, the time spent accumulating evidence until the first decision threshold crossing and the duration of the processing pipeline together equal the time taken to respond, $t_d + I = t_r$.) 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.

We consider here the general case of a stimulus that provides time-varying evidence (Fig. 2). 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. We use E_i to denote the evidence presented for option 2 minus the evidence presented for option 1 (i.e. the difference in evidence) during frame i. If the difference in evidence presented in each frame is normally distributed around the underlying mean then we have,

$$p(E_i|\mu) = N(E_i; \mu, \sigma_E^2)\,, \tag{4}$$

where σ_E^2 is the variance over the presented difference-in-evidence. Note that μ does not correspond to anything the observer directly observes. μ is the underlying mean difference-in-evidence presented for the two options, and is constant throughout a trial. The actual presented evidence, E_i, is what varies over the course of a trial, and is drawn from a distribution centred on the underlying means.

While 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. 3, and equation 21). Consider a discretisation of time into very short time steps (much shorter than the duration of a frame) of duration δt. 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 short time step, j, as generated according to,

$$p(\delta E_j|\mu) = N(\delta E_j; \mu\frac{\delta t}{t_f}, \sigma_E^2\frac{\delta t}{t_f})\,, \tag{5}$$

where, δE_j indicates the difference-in-evidence presented in time step j only, not over the course of a frame ($E_i = \sum_j \delta E_j$ where the summation is taken over all the time steps in a stimulus frame). Importantly, the observer incorrectly assumes that the δE_j 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 a small time step, j, the incremental change in the state of the accumulator tracking the difference in evidence measurements, denoted δx_j, can be described by,

$$p(\delta x_j|\delta E_j, \varphi) \quad = \quad N(\delta x_j; \delta E_j\varphi, \sigma_{acc}^2\delta t)\,, \tag{6}$$

where σ_{acc}^2 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, Ratcliff & Smith, 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 which determines how well stimulus information is processed. To distinguish this operationalisation from the usual operationalisation, we sometimes refer to the multiplicative factor φ as standardised drift rate, because we set the mean value of this variable to be 1 regardless of the strength of the presented evidence. When standardised drift rate is high, the signal extracted from the stimulus is greater, and evidence 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,

$$p(\varphi) = N(\varphi; 1, \sigma_\varphi^2)\,. \tag{7}$$

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 accumulator are related to the underlying signal as follows,

$$\begin{aligned} p(\delta x_j|\mu,\varphi) &= \int d(\delta E_j)\; p(\delta x_j|\delta E_j,\varphi)p(\delta E_j|\mu) && (8)\\ &= \int d(\delta E_j)\; N(\delta x_j; \delta E_j\varphi, \sigma_{acc}^2\delta t)N(\delta E_j; \mu\frac{\delta t}{t_f}, \sigma_E^2\frac{\delta t}{t_f}) && (9)\\ &= N(\delta x_j; \mu\varphi\frac{\delta t}{t_f}, \delta t(\sigma_{acc}^2 + \varphi^2\frac{\sigma_E^2}{t_f}))\,, && (10)\end{aligned}$$

To reach the final line we rearranged, and applied the result in appendix A. 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:

$$p(\delta x_j|\mu,\varphi) = N(\delta x_j; \mu\varphi\frac{\delta t}{t_f}, \delta t(\sigma_{acc}^2 + \frac{\sigma_E^2}{t_f}))\,. \tag{11}$$

We test the effect of this assumption on the calibration of confidence below (Fig. 3).

To simplify these expressions we change variables to $\nu = \frac{1}{t_f}\varphi\mu$, and $s^2 = \sigma_{acc}^2 + \frac{\sigma_E^2}{t_f}$. This gives,

$$p(\delta x_i|\nu) = N(\delta x_i; \nu\delta t, s^2\delta t)\,. \tag{12}$$

We can compute the probability distribution over ν given S using (2), (3) and (7), giving,

$$\begin{aligned} p(\nu|S=1) &= N(\nu; -\frac{\Delta\mu}{t_f}, \frac{\Delta\mu^2}{t_f^2}\sigma_\varphi^2) = N(\nu; -\nu_0, \sigma_\nu^2) && (13)\\ p(\nu|S=2) &= N(\nu; \frac{\Delta\mu}{t_f}, \frac{\Delta\mu^2}{t_f^2}\sigma_\varphi^2) = N(\nu; \nu_0, \sigma_\nu^2)\,, && (14)\end{aligned}$$

where $\nu_0 = \frac{\Delta\mu}{t_f}$, and $\sigma_\nu = \frac{\Delta\mu}{t_f}\sigma_\varphi$.

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

$$\begin{aligned} p(\nu|S=1) &= N(\nu; -\nu_0, \sigma_\nu^2) && (15)\\ p(\nu|S=2) &= N(\nu; \nu_0, \sigma_\nu^2)\,. && (16)\end{aligned}$$

They receive two independent streams of evidence. They look at the difference in evidence in these streams at each time point, a quantity which they assume is related to ν via,

$$p(\delta x_i|\nu) = N(\delta x_i; \nu\delta t, s^2\delta t)\,. \tag{17}$$

The Bayesian observer

Having specified the properties of the task and the beliefs of the observer, we can infer the rule used by a Bayesian observer to translate all evidence measurements gathered so far into a decision and confidence.

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 (15), (16), and (17), 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, and simply state the result here. Using Bayes rule with all evidence samples the log-posterior ratio is given by (appendix B; Moran, 2015),

$$\log \frac{p(S = 2|\boldsymbol{\delta x})}{p(S = 1|\boldsymbol{\delta x})} = \frac{2x\nu_0}{s^2 + t\sigma_\nu^2}\,, \tag{18}$$

where $x = \sum_{i=1}^{N} \delta x_i$, and t is the time spent accumulating evidence. We will often work with a scaled version of the log-posterior ratio,

$$x_{lp} = K \log \frac{p(S = 2|\boldsymbol{\delta x})}{p(S = 1|\boldsymbol{\delta x})} = x\frac{s^2 + \sigma_\nu^2}{s^2 + t\sigma_\nu^2} = \frac{x}{\theta(t)}\,, \tag{19}$$

where x_{lp} denotes the scaled version of the log-posterior ratio. $\theta()$ provides an abbreviation for the purpose of making future equations less cluttered. K is the scaling constant and is equal to,

$$K = \frac{s^2 + \sigma_\nu^2}{2\nu_0}\,. \tag{20}$$

Note that the log-posterior ratio without any scaling is given by x_{lp}/K.

A Bayesian observer would report whichever option is more likely. Hence, they will report $S = 1$ when $x_{lp}/K < 0$, which is the same as $x_{lp} < 0$, or $x < 0$. 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 processed and accumulated all evidence measurements, the observer can respond according to the sign of the final accumulator state, x (Fig. 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 absolute value of the accumulator, $|x|$, 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. 2). Processing will continue until measurements from the full duration of stimulus presentation, t_e, have been processed.

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

$$p(\text{correct}|\boldsymbol{\delta x}) = p(S = 1|\boldsymbol{\delta x}) = \frac{1}{1 + e^{x_{lp}/K}} = \frac{1}{1 + e^{\frac{1}{K}\frac{x}{\theta(t_e)}}}\,, \tag{21}$$

if the observer reports $R = 1$. (This expression can be found by rearranging $p(S = 1|\boldsymbol{\delta x}) + p(S = 2|\boldsymbol{\delta x}) = 1$.) A similar expression holds for $R = 2$, the only difference being that x is replaced with $-x$, and x_{lp} with $-x_{lp}$.

Testing the observer's beliefs

Above, we assumed that observers approximate the true generative model for the evidence measurements. Specifically we assumed, (a) observers ignore the fact that evidence is constant within each frame, and (b) 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 (6), until

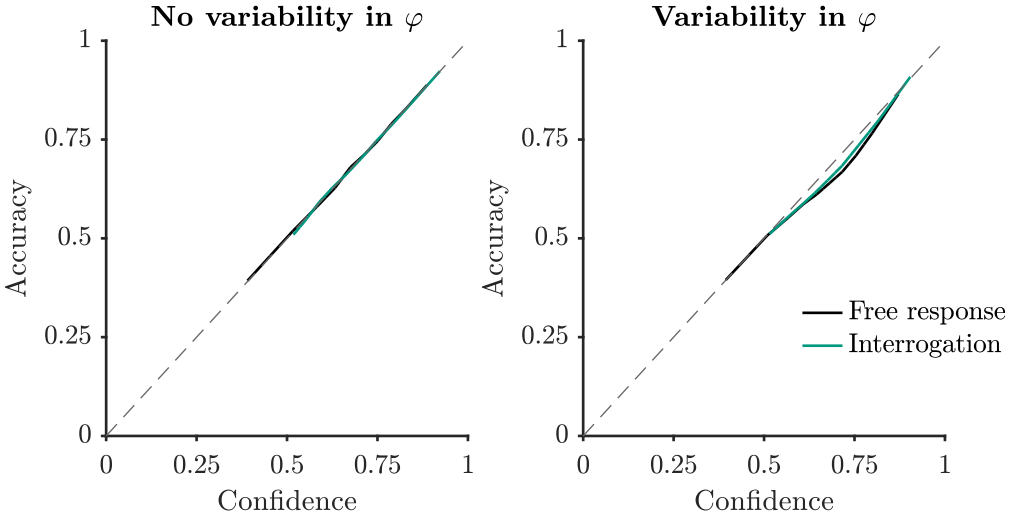

Figure 3: Accuracy as a function of confidence, where confidence is calculated using the observer's generative model of their evidence measurements. This model only approximates the true generative model. When there is variability in standardised drift rate, the approximations cause the observer's confidence to show some deviations from calibration. Simulation and plotting details in Appendix D.

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 (21). Full details of the simulation can be found in appendix D.

Fig. 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 D. 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. 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 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. 3 suggests this approximation does not cause large issues.

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 exactly what quantity confidence reports reflect. While the (scaled) log-posterior ratio, x_{lp} 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_{lp}, denoted x_c,

$$p(x_c|x_{lp}) = N(x_c; x_{lp}, \sigma_m^2)\,, \tag{22}$$

where σ_m is the standard deviation of metacognitive noise. (Note that this noisy readout of the scaled log-posterior ratio, is equivalent to a scaled version of a noisy readout of the unscaled log-posterior ratio; Appendix C.)

We would also like to minimise the number of assumptions we make about how x_c is transformed

Parameter / feature		Free response	Interrogation
Accumulator noise	σ_{acc}	✓	✓
Drift-rate variability	σ_{φ}	✓	✓
Metacognitive noise	σ_m	✓	✓
Confidence boundaries	d_i (for all i)	✓	✓
Duration of evidence pipeline	I	✓	-
Decision threshold shape	$f(t_d)$	✓	-

Table 2: Model parameters for fitting to data. The predictions for confidence depend on these parameters, but this dependence is kept implicit throughout for readability. The tables states "Parameter / feature" because $f(t_d)$ is not a parameter. $f(t_d)$ describes how the shape of the decision threshold changes over time. The modeller can parameterise this function as they wish. For example, they could use a flat threshold and simply fit threshold height, or they could use a complicated curved threshold with several parameters.

into a confidence report. There is evidence that different people use confidence scales in different ways (Ais, Zylberberg, Barttfeld & Sigman, 2016; Festinger, 1943; Navajas et al., 2017). To make minimal assumptions about how people view and treat confidence scales, we treat confidence reports, C, as ordinal data only (Aitchison, Bang, Bahrami & Latham, 2015). 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 which separates the confidence reports which 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.

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.) A key variable is the observer's (scaled) log-posterior ratio after they have seen all evidence, x_{lp}. 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 (scaled) log-posterior ratio, x_c. As described in the previous section, on a trial with 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).

Throughout we keep the dependence of the predictions on model parameters implicit: The probability distribution over confidence reports depends not just on the evidence presented, the response, and the time spent monitoring the stimulus, but also the parameters of the model. For the sake of readability this dependence is kept implicit in the derivations, however the parameters on which the predictions depend are of course of great practical importance. It will be these parameters that we adjust as we fit the model to data, and by constraining particular parameters to certain values we will be able to create different variants of the model for comparison. Parameters for fitting to data are listed in Table 2. Decision threshold is listed but this is not in itself a parameter. We will see that the modeller has freedom over what shape decision threshold to use, and how to parameterise this function.

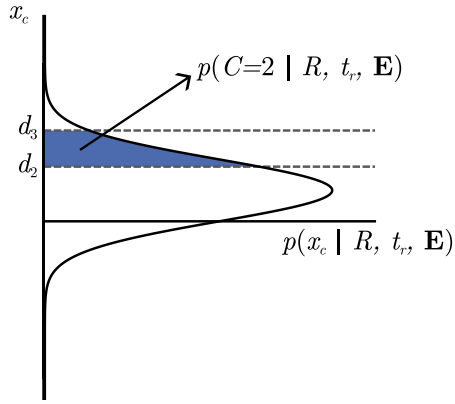

Figure 4: Probability of a confidence report, based on a distribution over x_c. x_c is a noisy representation of the (scaled) log-posterior ratio. We make no specific assumptions about how observers use confidence scales apart from assuming that, if the observer reports higher confidence C, then the 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.

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 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. 5.

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). The case where $R = 1$ is identical except the limits become $-d_{i+1}$ and $-d_i$. Also marginalising over, x_{lp} and using Bayes rule,

$$p(C = i|R = 2, \boldsymbol{E}) = \int_{d_i}^{d_{i+1}} dx_c \int dx_{lp}\ p(x_c|x_{lp})p(x_{lp}|R = 2, \boldsymbol{E}) \tag{23}$$

$$= \int_{d_i}^{d_{i+1}} dx_c \int dx_{lp}\ \frac{p(x_c|x_{lp})p(R = 2|x_{lp})p(x_{lp}|\boldsymbol{E})}{p(R = 2|\boldsymbol{E})}\,. \tag{24}$$

We want to find,

$$\Psi(x_c) = \int dx_{lp}\ p(x_c|x_{lp})p(R|x_{lp})p(x_{lp}|\boldsymbol{E})\,, \tag{25}$$

(where $\Psi(x_c)$ is an abbreviation), and then use,

$$p(C = i|R = 2, \boldsymbol{E}) = \int_{d_i}^{d_{i+1}} dx_c\ \frac{\Psi(x_c)}{p(R = 2|\boldsymbol{E})}\,. \tag{26}$$

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

$$p(x_{lp}|\boldsymbol{E}) = \int d\varphi\ p(x_{lp}|\varphi, \boldsymbol{E})p(\varphi|\boldsymbol{E})\,. \tag{27}$$

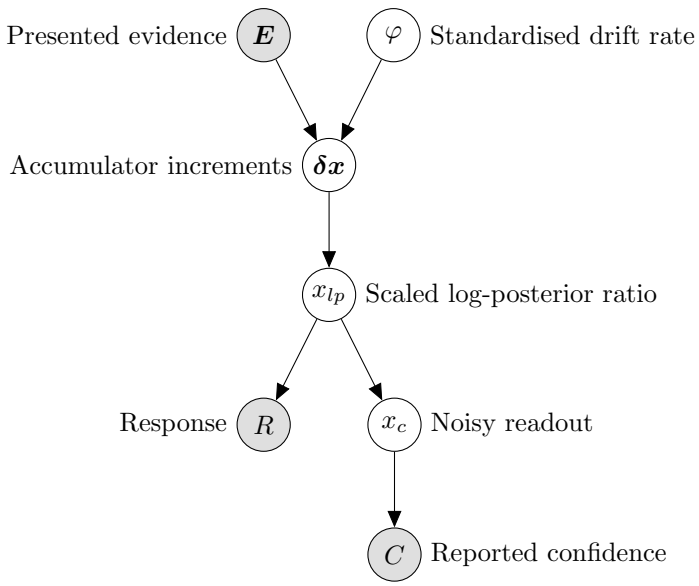

Figure 5: Representation of the generative model for confidence reports in the interrogation condition, from the perspective of the researcher. We want to infer the probability of reported confidence, using our knowledge of the evidence presented and the response given.

$\boldsymbol{E}$ and φ are independent of each other (when not conditioned on other variables; see Fig. 5). Hence,

$$p(\varphi|\boldsymbol{E}) = p(\varphi) = N(\varphi; 1, \sigma_\varphi^2)\,, \tag{28}$$

where we have used (7).

Recall from (6) that,

$$p(\delta x_i|\delta E_i, \varphi) \quad = \quad N(\delta x_i; \delta E_i \varphi, \sigma_{acc}^2 \delta t)\,. \tag{29}$$

Using standard formulae for the sum of normally distributed random variables, we have that $x = \sum_i \delta x_i$ will be given by,

$$p(x|\varphi, \boldsymbol{E}) \quad = \quad N(x; E\varphi, t_e \sigma_{acc}^2)\,. \tag{30}$$

where $E = \sum_i \delta E_i$ (sum taken over all time steps in all relevant frames), and $t_e = \sum_i \delta t$. Using (18) we have,

$$p(x_{lp}|\varphi, \boldsymbol{E}) = N(x_{lp}; \frac{E\varphi}{\theta(t_e)}, \frac{t_e \sigma_{acc}^2}{\theta^2(t_e)})\,. \tag{31}$$

Substituting these results into (27) gives us,

$$p(x_{lp}|\boldsymbol{E}) \quad = \quad \int d\varphi \; N(x_{lp}; \frac{E\varphi}{\theta(t_e)}, \frac{t_e \sigma_{acc}^2}{\theta^2(t_e)}) N(\varphi; 1, \sigma_\varphi^2) \tag{32}$$

$$\quad = \quad \frac{\theta(t_e)}{E} \int d\varphi \; N(\varphi; \frac{x_{lp}\theta(t_e)}{E}, \frac{t_e \sigma_{acc}^2}{E^2}) N(\varphi; 1, \sigma_\varphi^2)\,. \tag{33}$$

We can apply the result in appendix A. Simplifying, we find,

$$p(x_{lp}|\boldsymbol{E}) \quad = \quad N(x_{lp}; \frac{E}{\theta(t_e)}, \frac{E^2 \sigma_\varphi^2 + t_e \sigma_{acc}^2}{\theta^2(t_e)})\,. \tag{34}$$

Returning to (25), we now have,

$$\Psi(x_c) = \int dx_{lp} \; p(x_c|x_{lp}) p(R|x_{lp}) N(x_{lp}; \frac{E}{\theta(t_e)}, \frac{E^2 \sigma_\varphi^2 + t_e \sigma_{acc}^2}{\theta^2(t_e)})\,. \tag{35}$$

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

$$\Psi(x_c) = \int dx_{lp}\ \delta(x_c - x_{lp}) p(R|x_{lp}) N(x_{lp}; \frac{E}{\theta(t_e)}, \frac{E^2\sigma_\varphi^2 + t_e\sigma_{acc}^2}{\theta^2(t_e)}) \tag{36}$$

$$= p(R|x_{lp} = x_c) N(x_c; \frac{E}{\theta(t_e)}, \frac{E^2\sigma_\varphi^2 + t_e\sigma_{acc}^2}{\theta^2(t_e)}), \tag{37}$$

Consider the term $p(R|x_{lp} = x_c)$, which describes the probability of a response given a final (scaled) log-posterior x_{lp}. The Bayesian observer's decision rule is deterministic, and was described in section 2. If $x_{lp} < 0$ the observer reports that $R = 1$, and reports $R = 2$ if $x_{lp} > 0$. In all cases, the observer makes the response which is most likely to be correct. Hence, (scaled or not) the log-posterior ratio at the time of the decision always favours the response made, and the probability of x_{lp} and R being inconsistent is zero. Due to the absence of metacognitive noise, $x_{lp} = x_c$. Hence, $p(R|x_{lp} = x_c)$ 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 (26) and our expression for $\Psi(x_c)$,

$$p(C = i|R = 2, \boldsymbol{E}) = \frac{1}{p(R = 2|\mathbf{E})} \int_{d_i}^{d_{i+1}} dx_c\ p(R = 2|x_{lp} = x_c) N(x_c; \frac{E}{\theta(t_e)}, \frac{E^2\sigma_\varphi^2 + t_e\sigma_{acc}^2}{\theta^2(t_e)}) \tag{38}$$

$$= \frac{1}{p(R = 2|\mathbf{E})} \int_{\max\{0,d_i\}}^{\max\{0,d_{i+1}\}} dx_c\ N(x_c; \frac{E}{\theta(t_e)}, \frac{E^2\sigma_\varphi^2 + t_e\sigma_{acc}^2}{\theta^2(t_e)}), \tag{39}$$

where we have used that $p(R = 2|x_{lp} = x_c)$ is zero anywhere x_{lp} is inconsistent with the response $R = 2$ (i.e. $x_{lp} < 0$), and one elsewhere. In the case where $R = 1$ an identical expression holds, except the limits become $\min\{0, -d_{i+1}\}$ and $\min\{0, -d_i\}$, and we divide by $p(R = 1|\mathbf{E})$ not $p(R = 2|\mathbf{E})$.

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

$$p(R|\boldsymbol{E}) = \int_{-\infty}^{\infty} dx_{lp}\ p(R|x_{lp}) p(x_{lp}|\boldsymbol{E}) \tag{40}$$

$$= \int_a^b dx_{lp}\ N(x_{lp}; \frac{E}{\theta(t_e)}, \frac{E^2\sigma_\varphi^2 + t_e\sigma_{acc}^2}{\theta^2(t_e)}) \tag{41}$$

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

Let's now consider what happens in the presence of metacognitive noise, when x_c no longer equals x_{lp}, but is a noisy version of it as in (22). Returning to (35), in this case $\Psi(x_c)$ becomes,

$$\Psi(x_c) = \int dx_{lp}\ N(x_c; x_{lp}, \sigma_m^2) p(R|x_{lp}) N(x_{lp}; \frac{E}{\theta(t_e)}, \frac{E^2\sigma_\varphi^2 + t_e\sigma_{acc}^2}{\theta^2(t_e)}). \tag{42}$$

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

$$J(x_{lp}) = p(R|x_{lp}) N(x_{lp}; \frac{E}{\theta(t_e)}, \frac{E^2\sigma_\varphi^2 + t_e\sigma_{acc}^2}{\theta^2(t_e)}) \tag{43}$$

$$= TN(x_{lp}; \mu_{lp}, \sigma_{lp}^2, a, b) \int_a^b dx'_{lp}\ N(x'_{lp}; \frac{E}{\theta(t_e)}, \frac{E^2\sigma_\varphi^2 + t_e\sigma_{acc}^2}{\theta^2(t_e)}) \tag{44}$$

$$= TN(x_{lp}; \mu_{lp}, \sigma_{lp}^2, a, b) p(R|\boldsymbol{E}). \tag{45}$$

We have used (41) 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_{lp} and σ_{lp}^2, are the mean and variance of the distribution prior to truncation, and therefore are,

$$\mu_{lp} = \frac{E}{\theta(t_e)} \tag{46}$$

$$\sigma_{lp}^2 = \frac{E^2\sigma_\varphi^2 + t_e\sigma_{acc}^2}{\theta^2(t_e)} \tag{47}$$

Returning to (42) we now have,

$$\Psi(x_c) = p(R|\boldsymbol{E}) \int_{-\infty}^{\infty} dx_{lp}\ N(x_c; x_{lp}, \sigma_m^2) TN(x_{lp}; \mu_{lp}, \sigma_{lp}^2, a, b)\,. \tag{48}$$

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

$$\Psi(x_c) = -p(R|\boldsymbol{E}) \int_{\infty}^{-\infty} dy\ N(y; 0, \sigma_m^2) TN(x_c - y; \mu_{lp}, \sigma_{lp}^2, a, b) \tag{49}$$

$$= p(R|\boldsymbol{E}) \int_{-\infty}^{\infty} dy\ N(y; 0, \sigma_m^2) TN(x_c - y; \mu_{lp}, \sigma_{lp}^2, a, b)\,. \tag{50}$$

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):

$$\Psi(x_c) = p(R|\boldsymbol{E}) \int_{-\infty}^{\infty} dy\ N(y; 0, \sigma_m^2) TN(x_c - y; \mu_{lp}, \sigma_{lp}^2, a, b) \tag{51}$$

$$= p(R|\boldsymbol{E}) \int_{-\infty}^{\infty} dy \frac{1}{2\pi\sqrt{\sigma_m^2\sigma_{lp}^2}} \frac{1}{\Phi(\frac{b-\mu_{lp}}{\sigma_{lp}}) - \Phi(\frac{a-\mu_{lp}}{\sigma_{lp}})} e^{-\frac{y^2}{2\sigma_m^2}} e^{-\frac{(x_c-y-\mu_{lp})^2}{2\sigma_{lp}^2}} \mathbb{1}_{y\in[x_c-b, x_c-a]} \tag{52}$$

$$= p(R|\boldsymbol{E}) \frac{1}{\Phi(\frac{b-\mu_{lp}}{\sigma_{lp}}) - \Phi(\frac{a-\mu_{lp}}{\sigma_{lp}})} N(x_c; \mu_{lp}, \sigma_m^2 + \sigma_{lp}^2) \Big[\Phi(\frac{x_c - a - \alpha}{\beta}) - \Phi(\frac{x_c - b - \alpha}{\beta})\Big] \tag{53}$$

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), $\alpha = \frac{\beta^2(x_c-\mu_{lp})}{\sigma_{lp}^2}$ and $\beta^2 = \frac{\sigma_m^2\sigma_{lp}^2}{\sigma_m^2+\sigma_{lp}^2}$.

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,

$$\Psi(x_c) = p(R|\boldsymbol{E}) \frac{1}{\Phi(L\frac{\mu_{lp}}{\sigma_{lp}})} N(x_c; \mu_{lp}, \sigma_m^2 + \sigma_{lp}^2) \Phi\left(L \frac{x_c\sigma_{lp}^2 + \mu_{lp}\sigma_m^2}{\sigma_m\sigma_{lp}\sqrt{\sigma_m^2 + \sigma_{lp}^2}}\right), \tag{54}$$

and using (26) our final result for confidence is,

$$p(C = i|R, \boldsymbol{E}) = L \int_{Ld_i}^{Ld_{i+1}} dx_c\ \frac{1}{\Phi(L\frac{\mu_{lp}}{\sigma_{lp}})} N(x_c; \mu_{lp}, \sigma_m^2 + \sigma_{lp}^2) \Phi\left(L \frac{x_c\sigma_{lp}^2 + \mu_{lp}\sigma_m^2}{\sigma_m\sigma_{lp}\sqrt{\sigma_m^2 + \sigma_{lp}^2}}\right). \tag{55}$$

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 (55) 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

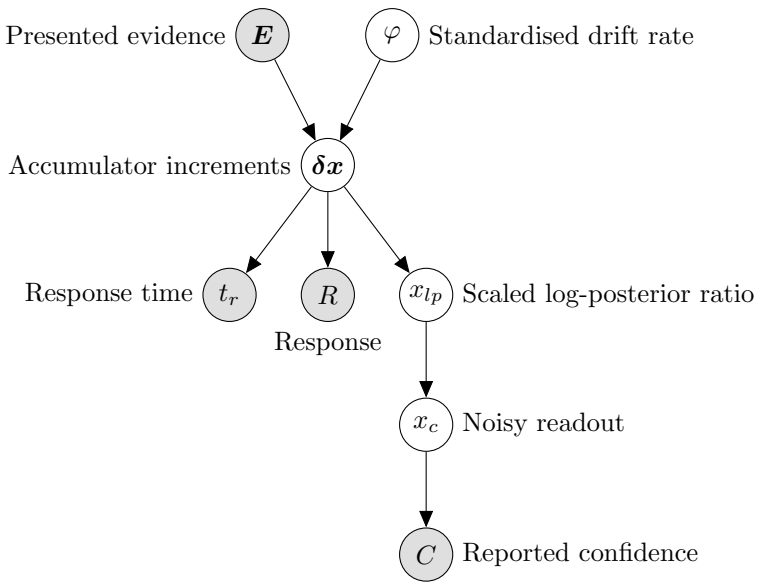

Figure 6: Representation of the generative model for confidence reports in the free response condition, from the perspective of the researcher. We want to infer the probability of reported confidence, using our knowledge of the evidence presented, the response given, and the response time.

E, and state the result here. The expression in (55) becomes,

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

where,

$$BvN(l_1,l_2,\rho) = \int_{-\infty}^{l_1}\int_{-\infty}^{l_2} \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}(x^2+y^2-2xy\rho)} dxdy\,, \tag{56}$$

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

$$\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \text{ and } \Sigma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}. \tag{57}$$

Additionally, g, h, and e_i denote $g = \frac{L\mu_{lp}\sqrt{\sigma_m^2+\sigma_{lp}^2}}{\sigma_m\sigma_{lp}}$, $h = \frac{L\sigma_{lp}}{\sigma_m}$ and $e_i = \frac{Ld_i-\mu_{lp}}{\sqrt{\sigma_m^2+\sigma_{lp}^2}}$. 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. 6).

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

$$p(C=i|R=2,t_r,\boldsymbol{E}) \quad = \quad \int_{d_i}^{d_{i+1}} dx_c \int dx_{lp}\; p(x_c|x_{lp})p(x_{lp}|R=2,t_r,\boldsymbol{E})\,. \tag{58}$$

(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,

$$p(x_{lp}|R, t_r, \boldsymbol{E}) = \int d\varphi \; p(x_{lp}|\varphi, R, t_r, \boldsymbol{E}) p(\varphi|R, t_r, \boldsymbol{E}) \,. \tag{59}$$

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,

$$x = \begin{cases} f(t_d) & \text{if } R = 2 \\ -f(t_d) & \text{if } R = 1 \,, \end{cases} \tag{60}$$

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 in

$$p(\varphi|R, t_r, \boldsymbol{E}) = p(\varphi|x_d, t_d, \boldsymbol{E}) \,. \tag{61}$$

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 $\boldsymbol{E}$ which 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. 6). However, for the purpose of inferring φ, we approximate the evidence stream by its average prior to the time of the decision, $\overline{E}$,

$$p(\varphi|R, t_r, \boldsymbol{E}) = p(\varphi|x_d, t_d, \overline{E}) \,. \tag{62}$$

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

$$\begin{aligned} p(\varphi|R, t_r, \boldsymbol{E}) &= p(\varphi|x_d, t_d, \overline{E}) && (63) \\ &\propto p(x_d, t_d|\varphi, \overline{E}) p(\varphi|\overline{E}) && (64) \\ &\propto p(\varphi|\overline{E}) \sum_{\omega \in \Omega} p(\boldsymbol{\delta x}^{(\omega)}|\varphi, \overline{E}) && (65) \\ &\propto p(\varphi|\overline{E}) \sum_{\omega \in \Omega} \prod_i p(\delta x_i^{(\omega)}|\varphi, \overline{E}) \,. && (66) \end{aligned}$$

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 may be related to the time of the response, which φ also affects (Fig. 6). Again, this approximation will be tested once we have derived predictions for confidence. The approximation gives,

$$p(\varphi|R, t_r, \boldsymbol{E}) \propto p(\varphi) \sum_{\omega \in \Omega} \prod_i p(\delta x_i^{(\omega)}|\varphi, \overline{E}) \,. \tag{67}$$

Returning to (6) but now using our approximation replacing the full evidence stream with the average evidence, we have that δx_i is given by

$$\begin{aligned} p(\delta x_j|\delta E_j, \varphi) &\approx p(\delta x_i|\overline{E}, \varphi) && (68) \\ &\approx N(\delta x_i; \varphi \overline{E} \delta t, \sigma_{acc}^2 \delta t) && (69) \\ &\propto N(\varphi; \frac{\delta x_i}{\overline{E} \delta t}, \frac{\sigma_{acc}^2}{\overline{E}^2 \delta t}) \,. && (70) \end{aligned}$$

The proportionality is with respect to φ.

Using standard results for the product of normal distributions we find

$$\prod_i p(\delta x_i|\overline{E},\varphi) \propto \prod_i N(\varphi;\frac{\delta x_i}{\overline{E}\delta t},\frac{\sigma_{acc}^2}{\overline{E}^2\delta t}) \tag{71}$$

$$\propto N(\varphi;\frac{x_d}{\overline{E}t_d},\frac{\sigma_{acc}^2}{\overline{E}^2 t_d}) \tag{72}$$

$$= K^{(\omega)}N(\varphi;\frac{x_d}{\overline{E}t_d},\frac{\sigma_{acc}^2}{\overline{E}^2 t_d}), \tag{73}$$

where we have used $t_d = \sum \delta t$, $x_d = \sum \delta x_i$, and $K^{(\omega)}$ is the constant of proportionality (see Drugowitsch et al., 2012 for closely related calculations). This constant will be different for different paths, so we explicitly indicate the path, ω. Hence, we have in (67)

$$p(\varphi|R,t_r,\boldsymbol{E}) \propto p(\varphi)\sum_{\omega\in\Omega} K^{(\omega)}N(\varphi;\frac{x_d}{\overline{E}t_d},\frac{\sigma_{acc}^2}{\overline{E}^2 t_d}) \tag{74}$$

$$\propto p(\varphi)N(\varphi;\frac{x_d}{\overline{E}t_d},\frac{\sigma_{acc}^2}{\overline{E}^2 t_d}). \tag{75}$$

Recalling (7), φ is distributed as,

$$p(\varphi) = N(\varphi;1,\sigma_\varphi^2), \tag{76}$$

giving

$$p(\varphi|R,t_r,\boldsymbol{E}) \propto N(\varphi;1,\sigma_\varphi^2)N(\varphi;\frac{x_d}{\overline{E}t_d},\frac{\sigma_{acc}^2}{\overline{E}^2 t_d}) \tag{77}$$

$$= N\Big(\varphi;\frac{\sigma_{acc}^2 + x_d\sigma_\varphi^2\overline{E}}{\sigma_{acc}^2 + t_d\sigma_\varphi^2\overline{E}^2},\frac{\sigma_{acc}^2\sigma_\varphi^2}{\sigma_{acc}^2 + t_d\sigma_\varphi^2\overline{E}^2}\Big), \tag{78}$$

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

Returning to (59), we see that in addition to an expression for $p(\varphi|R,t_r,\boldsymbol{E})$, we need an expression for $p(x_{lp}|\varphi,R,t_r,\boldsymbol{E})$. 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. 2). Following a decision, evidence accumulation continues in the same manner as in the interrogation condition (Pleskac & Busemeyer, 2010), until all evidence measurements have been processed. Therefore, the expression in (30) 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 $\Delta E = \sum_i \delta E_i$ (the summation is taken over all i which correspond to time steps following a decision). Using (30),

$$p(\Delta x|\varphi,\boldsymbol{E}) = N(\Delta x;\varphi\Delta E,\sigma_{acc}^2 I), \tag{79}$$

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,

$$p(x|\varphi,R,t_r,\boldsymbol{E}) = N(x;x_d+\varphi\Delta E,\sigma_{acc}^2 I) \tag{80}$$

Using (18) we have,

$$p(x_{lp}|\varphi,R,t_r,\boldsymbol{E}) = N(x_{lp};\frac{x_d+\varphi\Delta E}{\theta(t_e)},\frac{\sigma_{acc}^2 I}{\theta^2(t_e)}) \tag{81}$$

We have now derived expressions for both the distributions in (59). Returning to this equation we have,

$$p(x_{lp}|R,t_r,\boldsymbol{E}) = \int d\varphi\ p(x_{lp}|\varphi,R,t_r,\boldsymbol{E})p(\varphi|R,t_r,\boldsymbol{E}) \tag{82}$$

$$= \int d\varphi\ N(x_{lp};\frac{x_d+\varphi\Delta E}{\theta(t_e)},\frac{\sigma_{acc}^2 I}{\theta^2(t_e)})N\Big(\varphi;\frac{\sigma_{acc}^2+x_d\sigma_\varphi^2\overline{E}}{\sigma_{acc}^2+t_d\sigma_\varphi^2\overline{E}^2},\frac{\sigma_{acc}^2\sigma_\varphi^2}{\sigma_{acc}^2+t_d\sigma_\varphi^2\overline{E}^2}\Big) \tag{83}$$

$$= \frac{\theta(t_e)}{\Delta E}\int d\varphi\ N(\varphi;\frac{\theta(t_e)x_{lp}-x_d}{\Delta E},\frac{\sigma_{acc}^2 I}{\Delta E^2})N\Big(\varphi;\frac{\sigma_{acc}^2+x_d\sigma_\varphi^2\overline{E}}{\sigma_{acc}^2+t_d\sigma_\varphi^2\overline{E}^2},\frac{\sigma_{acc}^2\sigma_\varphi^2}{\sigma_{acc}^2+t_d\sigma_\varphi^2\overline{E}^2}\Big) \tag{84}$$

Applying the result in appendix A and rearranging we find,

$$p(x_{lp}|R,t_r,\boldsymbol{E}) = N(x_{lp};\mu_{lf},\sigma_{lf}^2) \tag{85}$$

Where,

$$\mu_{lf} = \frac{1}{\theta(t_e)}\left(x_d+\Delta E\frac{\sigma_{acc}^2+x_d\sigma_\varphi^2\overline{E}}{\sigma_{acc}^2+t_d\sigma_\varphi^2\overline{E}^2}\right) \tag{86}$$

$$\sigma_{lf}^2 = \frac{\sigma_{acc}^2}{\theta^2(t_e)}\left(\frac{\Delta E^2\sigma_\varphi^2}{\sigma_{acc}^2+t_d\sigma_\varphi^2\overline{E}^2}+I\right) \tag{87}$$

Using this result in (58), and allowing for normally distributed metacognitive noise as in (22),

$$p(C=i|R=2,t_r,\boldsymbol{E}) = \int_{d_i}^{d_{i+1}} dx_c \int dx_{lp}\ p(x_c|x_{lp})p(x_{lp}|R=2,t_r,\boldsymbol{E}) \tag{88}$$

$$= \int_{d_i}^{d_{i+1}} dx_c \int dx_{lp}\ N(x_c;x_{lp},\sigma_m^2)N(x_{lp};\mu_{lf},\sigma_{lf}^2) \tag{89}$$

$$= \int_{d_n}^{d_{n+1}} dx_c\ N(x_c;\mu_{lf},\sigma_{lf}^2+\sigma_m^2)\,. \tag{90}$$

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 (21), (22) and Fig. 2; for details of the simulations see appendix D). 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 which match the derived predictions.

Fig. 7 shows simulations of confidence using the model (error bars), and confidence based on the derived predictions (error shading). No additional 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. 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 over estimate the variability in

confidence reports, in trials with long response times. For completeness, we note that there were similarly small discrepancies in an observer who used an alternative to Bayesian confidence (Appendix F). Nevertheless, we should always be mindful of the fact that an approximation that works for one model or set of parameters, may not work so well for another model or set of parameters.

4 Discussion

Using the normative framework of the DDM and 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 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 which are 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 which 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 (e.g. Chang & Cooper, 1970; Smith, 2000; Voss & Voss, 2008; 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 stronger constraints when fitting models, than predictions made for large groups of trials at once (Park et al., 2016), which has until now been the standard approach (see Section 1). Computationally cheap predictions may also allow us to use techniques which require predictions to be evaluated many times, such as cross-validation and Markov chain Monte Carlo (MCMC; Bishop, 2006).

For readability, we have kept the dependence of the predictions for confidence on model parameters implicit, but it will be by adjusting these parameters that we can fit the model described to data (Table 2). Additionally, by constraining parameters to certain values we can construct model variants for model comparison. For example, we could ask whether metacognitive noise is an important source of variability by comparing a model in which we fit all parameters, to a model in which the standard deviation of metacognitive noise is set to zero. There is special flexibility with the decision threshold, because the modeller can choose its shape and how to parameterise it. Using these derivations we have recently compared a variety of models, including models in which the decision threshold is flat, and models with a decreasing decision threshold (Calder-Travis, Charles, Bogacz & Yeung, 2020).

At the outset we noted that 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 the scaled log-posterior ratio. 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 (39). It is informative to consider the case where the observer scales their readout so that, in the absence of metacognitive noise, it perfectly matches the log-posterior ratio, x_{lp}/K. Substituting in all abbreviations, and the scaling, we have that the most likely confidence report is

$$\frac{E}{K\theta(t_e)} = \frac{2\nu_0 E}{s^2 + t_e\sigma_\nu^2} = \frac{2t_f\Delta\mu E}{t_f^2\sigma_{acc}^2 + t_f\sigma_E^2 + t_e\Delta\mu^2\sigma_\varphi^2}\,. \tag{91}$$

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).

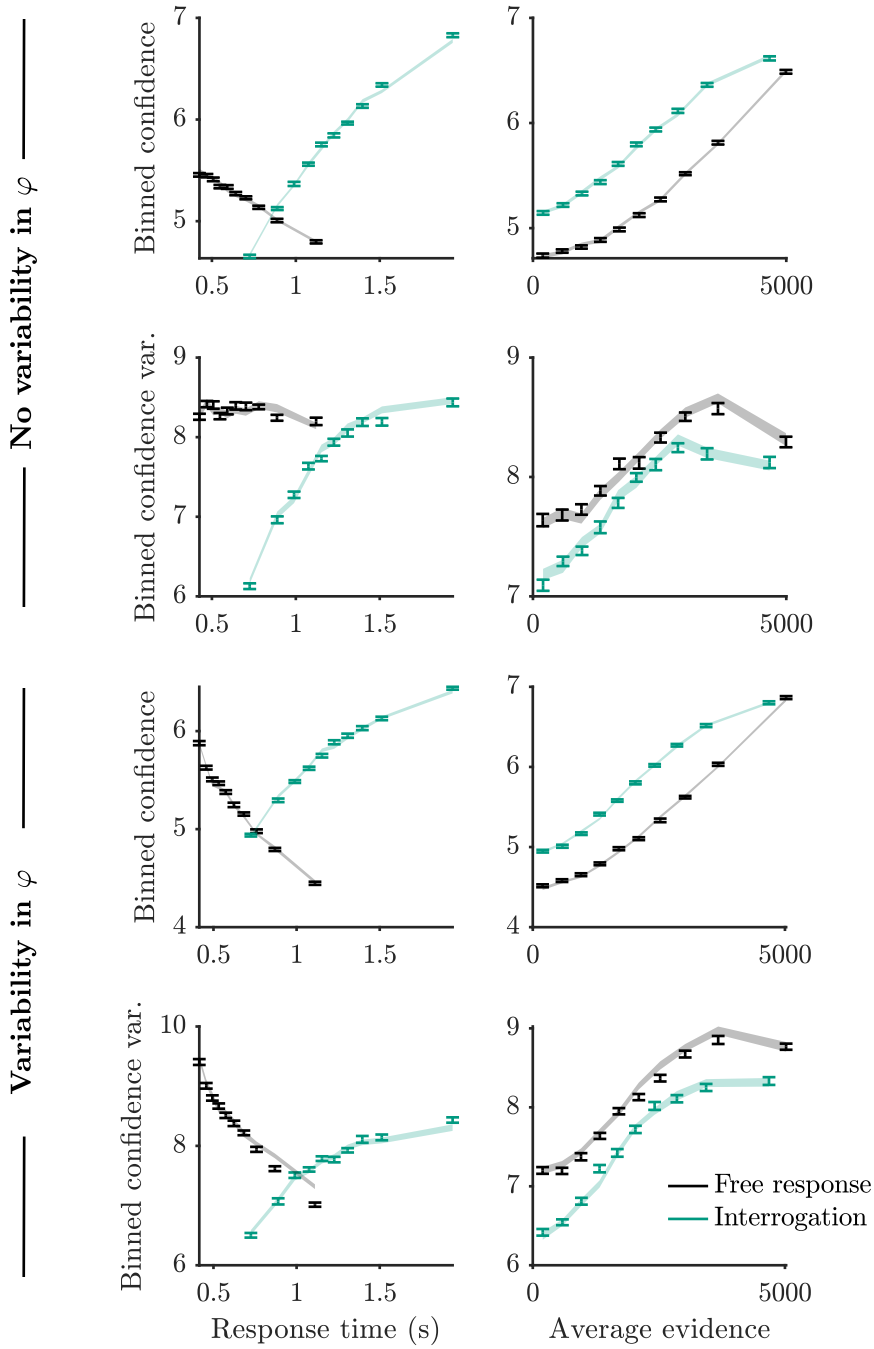

Figure 7: Mean and variance of binned confidence, produced via simulation of the model (error bars), and through the derived predictions (shading). Details of the simulation and plotting are provided in appendix D. Predictions matched the simulation closely. There were some signs that approximations used to derive the predictions lead to a small overestimation of variability in confidence, when there is variability in standardised drift rate (φ).

It would be simple to extend the derivations to account for the case where observers incorrectly estimate variance in the stimulus. We would 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 (scaled) log-posterior ratio, x_{lp}. 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, and the observer scales their report so it aligns with the log-posterior ratio. In the absence of metacognitive noise, the most likely confidence report is given by (86),

$$\frac{1}{K\theta(t_e)}\left(x_d + \Delta E\frac{\sigma_{acc}^2 + x_d\sigma_\varphi^2\overline{E}}{\sigma_{acc}^2 + t_d\sigma_\varphi^2\overline{E}^2}\right) = \frac{2t_f\Delta\mu}{t_f^2\sigma_{acc}^2 + t_f\sigma_E^2 + t_e\Delta\mu^2\sigma_\varphi^2}\left(x_d + \Delta E\frac{\sigma_{acc}^2 + x_d\sigma_\varphi^2\overline{E}}{\sigma_{acc}^2 + t_d\sigma_\varphi^2\overline{E}^2}\right). \tag{92}$$

Breaking this expression down, we see that the prefactor which multiplies the term in brackets is the same as the prefactor which 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 counterintuitive 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 (Fig. 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.

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 which captures the path of the accumulator following the decision (Pleskac & Busemeyer, 2010). Note that if standardised drift rate variability, σ_φ, is zero, this second term,

$$\Delta E\frac{\sigma_{acc}^2 + x_d\sigma_\varphi^2\overline{E}}{\sigma_{acc}^2 + t_d\sigma_\varphi^2\overline{E}^2}, \tag{93}$$

simply reduces to ΔE. In this case, the most probable level of confidence is determined by the state of the accumulator at decision time, plus the evidence presented following a decision. This consideration also shows that the fraction multiplying ΔE, which is 1 when $\sigma_\varphi = 0$, is a term which accounts for the effect of drift variability on confidence. This term provides a mathematical description of the fact 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 (92). 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). 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 which previously appeared difficult to explain. 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. 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. $\Delta E = 0$.

Kiani et al. (2014) suggested that simultaneous decisions and confidence may have been key to the pattern they observed. Recent work has added weight to the idea that the time at which confidence reports are collected is key, and has provided evidence that a combination of the DDM with a Bayesian readout for confidence can account for these apparently discrepant findings (Desender et al., 2020; Khalvati, Kiani & Rao, 2020). Our derivations further support these ideas.

To see why this is the case we look again at our expression for most likely confidence, (92), but now consider the situation in which $\Delta E = 0$. We have,

$$\frac{2t_f \Delta\mu}{t_f^2 \sigma_{acc}^2 + t_f \sigma_E^2 + t_e \Delta\mu^2 \sigma_\varphi^2} x_d \,. \tag{94}$$

If we have a flat threshold then x_d is constant, and the only variable which changes between trials is t_e, the duration of evidence presentation. Duration of evidence presentation is the same as response time in the free response task. Consider a set of trials with one shared value for signal strength (i.e., same average evidence strength from the stimulus, and same standardised drift rate). For these trials, a flat decision threshold, which is symmetric for the two options, generates response times for correct and error trials which are on average identical (Shadlen, Hanks, Churchland, Kiani & Yang, 2006). Additionally, a greater signal strength will lead to faster responses. Hence, we predict that higher signal strength leads to both faster correct responses and faster error responses, i.e. lower t_e for correct and error trials. As we can see from the expression, the most likely confidence report will also be higher, due to the lower t_e, when signal strength is greater.

We can also explain why confidence on error trials decreases with signal strength, when the observer has time to process pipeline evidence. The processing pipeline contains a considerable amount of information from the stimulus (approximately 400ms of the stimulus prior to response; Ratcliff and McKoon, 2008; Resulaj et al., 2009). On error trials, this evidence will tend to favour the alternative (correct) option, decreasing confidence, and this effect will be stronger when signal strength is high. The implication is that confidence will decrease as signal strength increases.

To test the intuitions gained from studying the equations, we simulated response times, decisions, and confidence reports, under two different free response 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 D for details of the simulations). 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. 8; statistics in Appendix G).

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 from the value of the fitted metacognitive noise parameter (Shekhar & Rahnev, 2020). A successful existing method for inferring the level of metacognitive noise is based on the signal detection framework (Maniscalco & Lau, 2012). Therefore, this approach implicitly assumes that signal detection theory is an adequate description of decision making mechanisms (Shekhar & Rahnev, 2020). 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 a model 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 which affect the scope of the derivations themselves, and those which 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 (Bogacz et al., 2006; Moreno-Bote, 2010). Each accumulator corresponds to one of the choices, and the accumulators may

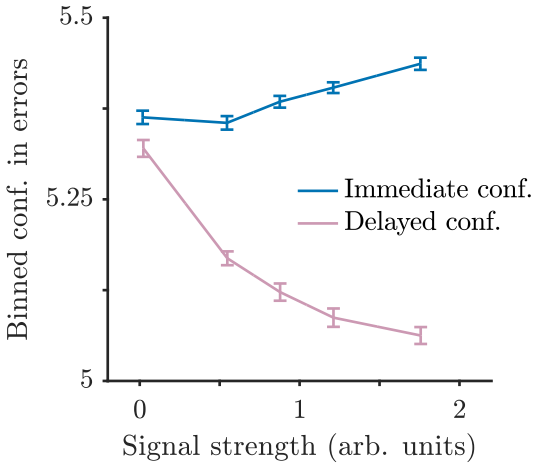

Figure 8: Confidence on free response trials that led to errors, as a function of signal strength. When confidence is reported immediately, and therefore does not reflect evidence measurements in the processing pipeline, confidence on error trials increases with signal strength. On the other hand, when 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. Simulation and plotting details in Appendix D.

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 (Shan, Moreno-Bote & Drugowitsch, 2019). Notwithstanding this consideration, we believe the DDM to be a particularly interesting case due to the normative properties of the diffusion mechanism (see Section 1). Given the evolutionary age of perceptual decision making (500 million years ago animals had complex eyes; Lee et al., 2011), it seems plausible that the algorithm used has normative properties.

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. Statistical properties of the stimuli are only used to derive confidence of the optimal observer, given an accumulator state and deliberation time. Statistical properties of the stimuli are not important for deriving predictions for confidence, once the observer's internal mapping from accumulator 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 which 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 which 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 the difficulties in deriving computationally cheap expressions for responses and response times (see Section 1). 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 will be able to make predictions for decisions and response times. This will allow us to examine whether a model that fits well to confidence nevertheless generates implausible response and response time data, and will provide an additional check of the model and its assumptions.

Regarding lapses, it would not be difficult to include a lapse rate parameter which describes some

probability that a random confidence report is given. It would be trickier to model lapses that also affect the response given (Adler & Ma, 2018; Ratcliff & Tuerlinckx, 2002). The probability that a given decision and response time resulted from a lapse would be difficult to calculate because, whilst we might know the probability distribution over decisions 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 for the two purposes described at the outset: Supporting deeper insight into the confidence of an important class of observers, and supporting trial-by-trial modelling. We have seen in this discussion that the derived expressions offer the potential to directly explore and understand the relationship between confidence and other variables. This level of explanatory power may be difficult to gain even after running simulations using a wide range of parameter values. The expressions found only require evaluating once per trial, making trial-by-trial modelling of dynamic stimuli more feasible. As a consequence, we hope these results will support efforts to develop models which make ever more precise and sophisticated predictions for behaviour.

Code availability

All code written for the study will be made publicly available upon publication, and will be accessible through doi:10.17605/OSF.IO/TK3VP.

References

Adler, W. T. & Ma, W. J. (2018). Comparing bayesian and non-bayesian accounts of human confidence reports. *PLOS Computational Biology*, *14*(11), e1006572. doi:10.1371/journal.pcbi.1006572

Ais, J., Zylberberg, A., Barttfeld, P. & Sigman, M. (2016). Individual consistency in the accuracy and distribution of confidence judgments. *Cognition*, *146*, 377–386. doi: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. doi:10.1371/journal.pcbi.1004519

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

Balsdon, T., Wyart, V. & Mamassian, P. (2020). Confidence controls perceptual evidence accumulation. *Nature Communications*, *11*(1). doi: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). doi:10.1037/xge0000511

Bishop, C. M. (2006). *Pattern recognition and machine learning*. Information Science and Statistics. New York: 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. doi: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. doi:10.1037/0033-295X.113.4.700

Bromiley, P. A. (2014). 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. doi:10.1016/j.jmp.2006.03.004

Calder-Travis, J., Charles, L., Bogacz, R. & Yeung, N. (2020). Bayesian confidence in optimal decisions. *PsyArXiv*. doi:10.31234/osf.io/j8sxz

Chang, J. & Cooper, G. (1970). A practical difference scheme for fokker-planck equations. *Journal of Computational Physics*, *6*(1), 1–16. doi: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. doi:10.1037/xhp0000583

Cox, D. R. & Miller, H. D. (1965). *The theory of stochastic processes*. London: Methuen.

De Martino, B., Fleming, S. M., Garrett, N. & Dolan, R. J. (2013). Confidence in value-based choice. *Nature Neuroscience*, *16*, 105–110. doi:10.1038/nn.3279

Desender, K., Boldt, A., Verguts, T. & Donner, T. H. (2019). Confidence predicts speed-accuracy tradeoff for subsequent decisions. *eLife*, *8*, e43499. doi: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. doi:10.1177/0956797617744771

Desender, K., Donner, T. H. & Verguts, T. (2020). Dynamic expressions of confidence within an evidence accumulation framework. *bioRxiv*. doi: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. doi:10.1016/S0022-2496(03)00003-8

Drugowitsch, J. (2016). Fast and accurate monte carlo sampling of first-passage times from wiener diffusion models. *Scientific Reports*, *6*(1), 20490. doi: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. doi: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. doi:10.1523/JNEUROSCI.4010-11.2012

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. doi: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. doi:10.1037/rev0000045

Ganupuru, P., Goldring, A. B., Harun, R. & Hanks, T. D. (2019). Flexibility of timescales of evidence evaluation for decision making. *Current Biology*, *29*(12), 2091–2097.e4. doi:10.1016/j.cub.2019.05.037

Gold, J. I. & Shadlen, M. N. (2007). The neural basis of decision making. *Annual Review of Neuroscience*, *30*(1), 535–574. doi:10.1146/annurev.neuro.29.051605.113038

Green, D. M. & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York ; London: 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). doi:10.1038/s41598-017-06116-z

Jang, Y., Wallsten, T. S. & Huber, D. E. (2012). A stochastic detection and retrieval model for the study of metacognition. *119*(1), 186–200. doi:10.1037/a0025960

Khalvati, K., Kiani, R. & Rao, R. P. N. (2020). *Bayesian inference with incomplete knowledge explains perceptual confidence and its deviations from accuracy*. Neuroscience. doi:10.1101/2020.09.18.304220

Kiani, R., Corthell, L. & Shadlen, M. N. (2014). Choice certainty is informed by both evidence and decision time. *Neuron*, *84*(6), 1329–1342. doi: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. doi: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. doi:10.1126/science.1169405

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. doi:10.1038/nature10097

Luce, R. D. (1986). *Response times: Their role in inferring elementary mental organization*. New York: Oxford University Press.

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. doi:10.1037/xge0000286

Maniscalco, B. & Lau, H. (2012). A signal detection theoretic approach for estimating metacognitive sensitivity from confidence ratings. *Consciousness and Cognition*, *21*(1), 422–430. doi: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*(1). doi:10.1093/nc/niw002

McMillen, T. & Holmes, P. (2006). The dynamics of choice among multiple alternatives. *Journal of Mathematical Psychology*, *50*(1), 30–57. doi:10.1016/j.jmp.2005.10.003

Moran, R. (2015). Optimal decision making in heterogeneous and biased environments. *Psychonomic Bulletin & Review*, *22*(1), 38–53. doi: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. doi: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. doi:10.1162/neco.2010.12-08-930

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. doi: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. doi:10.1016/j.jmp.2009.02.003

Owen, D. B. (1980). A table of normal integrals: A table. *Communications in Statistics - Simulation and Computation*, *9*(4), 389–419. doi:10.1080/03610918008812164

Park, H., Lueckmann, J.-M., von Kriegstein, K., Bitzer, S. & Kiebel, S. J. (2016). Spatiotemporal dynamics of random stimuli account for trial-to-trial variability in perceptual decision making. *Scientific Reports*, *6*(1). doi:10.1038/srep18832

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. doi: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. doi:10.1037/a0019737

Ratcliff, R. & McKoon, G. (2008). The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Computation*, *20*(4), 873–922. doi:10.1162/neco.2008.12-06-420

Ratcliff, R. & Smith, P. L. (2004). A comparison of sequential sampling models for two-choice reaction time. *Psychological Review*, *111*(2), 333–367. doi: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. doi: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. doi:10.3758/BF03196302

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

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. doi:10.1016/j.biopsych.2017.12.017

Sanders, J. I., Hangya, B. & Kepecs, A. (2016). Signatures of a statistical computation in the human sense of confidence. *Neuron*, *90*(3), 499–506. doi: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* (2017, pp. 208–237). doi:10.7551/mitpress/9780262042383.003.0010

Shan, H., Moreno-Bote, R. & Drugowitsch, J. (2019). Family of closed-form solutions for two-dimensional correlated diffusion processes. *Phys. Rev. E*, *100*(3). doi:10.1103/PhysRevE.100.032132

Shekhar, M. & Rahnev, D. (2020). The nature of metacognitive inefficiency in perceptual decision making. *Psychological Review*, No Pagination Specified–No Pagination Specified. Place: US Publisher: American Psychological Association. doi:10.1037/rev0000249

Shinn, M., Lam, N. H. & Murray, J. D. (2020). A flexible framework for simulating and fitting generalized drift-diffusion models. *eLife*, *9*, e56938. doi:10.7554/eLife.56938

Smith, P. L. (2000). Stochastic dynamic models of response time and accuracy: A foundational primer. *Journal of mathematical psychology*, *44*(3), 408–463. doi:10.1006/jmps.1999.1260

Tajima, S., Drugowitsch, J., Patel, N. & Pouget, A. (2019). Optimal policy for multi-alternative decisions. *Nature Neuroscience*, *22*, 1503–1511. doi:10.1038/s41593-019-0453-9

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. doi: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. doi:10.3758/BF03195402

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. doi:10.7554/eLife.12192

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. doi: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. doi:10.1016/j.cub.2016.10.021

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. doi: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. doi: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. doi:10.1214/aoms/1177730197

Wiecki, T. V., Sofer, I. & Frank, M. J. (2013). HDDM: Hierarchical bayesian estimation of the drift-diffusion model in python. *Frontiers in Neuroinformatics*, *7*. Publisher: Frontiers. doi:10.3389/fninf.2013.00014

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). doi:10.1007/978-3-642-45190-4_7

Zylberberg, A., Barttfeld, P. & Sigman, M. (2012). The construction of confidence in a perceptual decision. *Frontiers in Integrative Neuroscience*, *6*. doi: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*. doi:10.7554/eLife.17688

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. doi:10.1016/j.neuron.2018.07.035

A Product of normal distributions

We will repeatedly encounter a situation where we are interested in the probability distribution over z in the following equation,

$$p(z) \propto \int N(y; z, \sigma_A^2) N(y; \mu, \sigma_B^2) dy \,, \tag{95}$$

where N denotes the normal distribution. The product of two normal distributions is a scaled normal distribution (over y; Bromiley, 2014). The scaling is given by,

$$S = N(z; \mu, \sigma_A^2 + \sigma_B^2) \,. \tag{96}$$

Once we have integrated over y, only the scaling is left, as the normal distribution over y integrates to one. Hence,

$$p(z) = N(z; \mu, \sigma_A^2 + \sigma_B^2) \,. \tag{97}$$

B 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 (15), (16), and (17). Using Bayes rule the observer can infer the probability of the two possible values for S. Using a result from Drugowitsch et al. (2012), Moran (2015) 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,

$$p(S|\boldsymbol{\delta x}) \propto p(\boldsymbol{\delta x}|S)p(S) \tag{98}$$

$$\propto \int \prod_{i=1}^{N} p(\delta x_i|\nu)p(\nu|S)d\nu\,, \tag{99}$$

where we have used the flat prior over S in (1). Looking at the product in (99), using (17), and using standard formulae for the product of normal distributions (e.g. see Bromiley, 2014) we have,

$$\prod_{i=1}^{N} p(\delta x_i|\nu) = \prod_{i=1}^{N} N(\delta x_i; \nu\delta t, s^2\delta t) \tag{100}$$

$$\propto N(\nu; \frac{x}{t}, \frac{s^2}{t})\,, \tag{101}$$

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 (99) we have,

$$p(S|\boldsymbol{\delta x}) \propto \int N(\nu; \frac{x}{t}, \frac{s^2}{t})N(\nu; \pm\nu_0, \sigma_\nu^2)d\nu\,, \tag{102}$$

where the sign on ν_0 is determined by S. Using the result in appendix A,

$$p(S|\boldsymbol{\delta x}) \propto N(\pm\nu_0; \frac{x}{t}, \frac{s^2}{t} + \sigma_\nu^2)\,. \tag{103}$$

Rearranging, we find the (not scaled) log-posterior ratio is given by,

$$\log \frac{p(S = 2|\boldsymbol{\delta x})}{p(S = 1|\boldsymbol{\delta x})} = \frac{2x\nu_0}{s^2 + t\sigma_\nu^2}\,. \tag{104}$$

C Effect of posterior ratio scaling

In the main text we considered a model in which the observer uses a noisy readout of the scaled log-posterior ratio to determine their confidence. A noisy readout of the scaled log-posterior ratio (x_{lp}), is equivalent to a scaled version of a noisy readout of the unscaled log-posterior ratio (x_{lp}/K). First consider the readout in the main text in (22):

$$p(x_c|x_{lp}) = N(x_c; x_{lp}, \sigma_m^2)\,. \tag{105}$$

Then consider a noisy readout of the same form but of the unscaled log-posterior ratio. Denote this new readout $\widetilde{x}_c$. We have,

$$p(\widetilde{x}_c|\frac{x_{lp}}{K}) = N(\widetilde{x}_c; \frac{x_{lp}}{K}, \widetilde{\sigma}_m^2)\,, \tag{106}$$

where $\widetilde{\sigma}_m$ is the standard deviation of the metacognitive noise affecting this readout.

Our claim is that the distribution over $\widetilde{x}_c$ when scaled by K, is identical to the distribution over x_c. Consider a scaled version of $\widetilde{x}_c$, $\widehat{x}_c = K\widetilde{x}_c$. Using the rule for changing the variable of a probability density function, where the relationship between the variables being swapped is monotonic, then,

$$\begin{aligned}
p(\widehat{x}_c|\frac{x_{lp}}{K}) &= \left|\frac{\mathrm{d}\widetilde{x}_c}{\mathrm{d}\widehat{x}_c}\right| N(\widetilde{x}_c; \frac{x_{lp}}{K}, \widetilde{\sigma}_m^2) && (107)\\
&= \frac{1}{K}\frac{1}{\sqrt{2\pi\widetilde{\sigma}_m^2}}\exp\left\{-\frac{(\widetilde{x}_c - \frac{x_{lp}}{K})^2}{2\widetilde{\sigma}_m^2}\right\} && (108)\\
&= \frac{1}{\sqrt{2\pi K^2\widetilde{\sigma}_m^2}}\exp\left\{-\frac{(K\widetilde{x}_c - x_{lp})^2}{2K^2\widetilde{\sigma}_m^2}\right\} && (109)\\
&= N(\widehat{x}_c; x_{lp}, K^2\widetilde{\sigma}_m^2)\,. && (110)
\end{aligned}$$

Hence, the scaled readout of the unscaled log-posterior ratio, $\widehat{x}_c$, has the same distribution as the readout of the scaled log-posterior ratio, x_c. The metacognitive noise parameters in the two cases are related to each other through $\sigma_m = K\widetilde{\sigma}_m$.

D Simulation and plotting details

We simulated the task described in the main text with the parameters shown in Table 3. 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 midpoint between the means of the two evidence signals, referred to as "reference value" and denoted $\overline{\mu}/2$, 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	$p(\overline{\mu}/2)$	See text
Mean difference between evidence signals	$\Delta\mu$	90
Standard deviation of evidence in a frame	σ_E	311

Table 3: Parameters for the task used in the simulation.

The parameters used for the observer are shown in Table 4. Within a simulation, all participants were simulated using the same parameters. Except for Fig. 8, half of the trials were simulated for the free response condition, and half for the interrogation condition. For Fig. 8 all trials were from the free response condition, but two different versions of the free response condition were used, as described in the main text.

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 4 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.

Variable	Expression	Fig. 3	Fig. 7	Fig. 8
Participants per simulation		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 (separate simulations)	1 or 0 (separate simulations)	0.6
Standard deviation of metacognitive noise	σ_m	0	1500	1500
Threshold height		2000	2000	2000
Threshold slope		0	$-800\ s^{-1}$	0

Table 4: Parameters of the simulated observers.

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 from the distribution in (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. 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 (21), (22) and Fig. 2). 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 set the values of the parameters in these expressions (σ_{acc}, σ_{φ}, σ_m, I, threshold height and threshold slope) to the true values used to simulate the data. We set the bin edge parameters (d_i) to the values that were used for dividing the simulated confidence reports into 10 bins.

Prior to plotting, data for the x-axis was binned separately for each participant and data series. For each participant, data series, and bin, the average value of the x-variable was computed. Using these averages, the average value across participants was computed, and this determined the x-location of each bin. 10 bins were used apart from in Fig. 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.

E Rearranging interrogation condition result

For the probability distribution over C in the interrogation condition, we found in (55) that

$$p(C=i|R,\boldsymbol{E}) = L\int_{Ld_i}^{Ld_{i+1}} dx_c \; \frac{1}{\Phi(L\frac{\mu_{lp}}{\sigma_{lp}})} N(x_c;\mu_{lp},\sigma_m^2+\sigma_{lp}^2)\Phi\left(L\frac{x_c\sigma_{lp}^2+\mu_{lp}\sigma_m^2}{\sigma_m\sigma_{lp}\sqrt{\sigma_m^2+\sigma_{lp}^2}}\right). \tag{111}$$

It is possible to rearrange this expression into a form which is easier to evaluate numerically.

Consider a change of variables,

$$z = \frac{x_c-\mu_{lp}}{\sqrt{\sigma_m^2+\sigma_{lp}^2}}\,. \tag{112}$$

Then we have,

$$dz = \frac{dx_c}{\sqrt{\sigma_m^2+\sigma_{lp}^2}} \tag{113}$$

and,

$$N(x_c;\mu_{lp},\sigma_m^2+\sigma_{lp}^2) = \frac{1}{\sqrt{\sigma_m^2+\sigma_{lp}^2}}N(z;0,1) \tag{114}$$

$$= \frac{1}{\sqrt{\sigma_m^2+\sigma_{lp}^2}}\phi(z)\,, \tag{115}$$

where $\phi()$ indicates the standard normal distribution. Also denote,

$$g = \frac{L\mu_{lp}\sqrt{\sigma_m^2+\sigma_{lp}^2}}{\sigma_m\sigma_{lp}} = L\frac{\mu_{lp}\sigma_{lp}^2+\mu_{lp}\sigma_m^2}{\sigma_m\sigma_{lp}\sqrt{\sigma_m^2+\sigma_{lp}^2}} \tag{116}$$

$$h = \frac{L\sigma_{lp}}{\sigma_m} = \frac{L\sigma_{lp}^2}{\sigma_m\sigma_{lp}}\,. \tag{117}$$

Then,

$$g+hz = L\frac{x_c\sigma_{lp}^2+\mu_{lp}\sigma_m^2}{\sigma_m\sigma_{lp}\sqrt{\sigma_m^2+\sigma_{lp}^2}}\,. \tag{118}$$

Hence, in (111) we have,

$$p(C=i|R,\boldsymbol{E}) = L\frac{1}{\Phi(L\frac{\mu_{lp}}{\sigma_{lp}})}\int_{e_i}^{e_{i+1}} dz\;\phi(z)\Phi(g+hz)\,, \tag{119}$$

where,

$$e_i = \frac{Ld_i-\mu_{lp}}{\sqrt{\sigma_m^2+\sigma_{lp}^2}}\,. \tag{120}$$

Using result 10010.1 from Owen (1980), gives us,

$$p(C=i|R,\boldsymbol{E}) = L\frac{1}{\Phi(L\frac{\mu_{lp}}{\sigma_{lp}})}\Bigg[\quad BvN\Big(\frac{g}{\sqrt{1+h^2}}, e_{i+1}; \frac{-h}{\sqrt{1+h^2}}\Big) \tag{121}$$

$$- \quad BvN\Big(\frac{g}{\sqrt{1+h^2}}, e_i; \frac{-h}{\sqrt{1+h^2}}\Big)\Bigg]\,, \tag{122}$$

where,

$$BvN(l_1, l_2, \rho) = \int_{-\infty}^{l_1} \int_{-\infty}^{l_2} \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}(x^2+y^2-2xy\rho)} dxdy\,, \tag{123}$$

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

$$\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{124}$$

$$\Sigma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}. \tag{125}$$

This integral can then be numerically evaluated using standard functions.

F Testing the approximations in other models

In Fig. 7 we compared the mean and variance of confidence produced via simulations, and confidence produced through the predictions derived in the main text. The aim was to test whether the approximations made during the derivations were reasonable. We found little discrepancy between the simulated and predicted confidence. For completeness, we ran a similar analysis using a model not considered in the main text. Specifically, we considered an observer who did not use a Bayesian readout for confidence. Instead their confidence was directly determined by the final state of the accumulator: More accumulated evidence favouring the choice generated higher confidence. We adjusted the reported derivations to cover this case by setting $\theta(t) = 1$ (see equation 18).

For this model we found similar discrepancies between the simulations and the derived predictions. Figure 9 provides an example. Again, discrepancies were in general small. Nevertheless, we should bear in mind that just because an approximation works well for some model, or at some parameter values, does not mean it is guaranteed to work well in all models or at all parameter values.

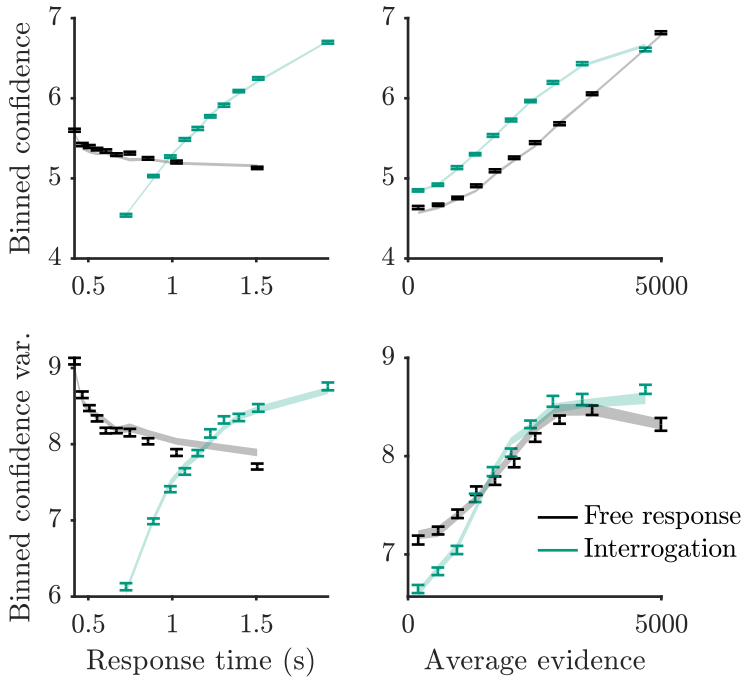

Figure 9: Mean and variance of binned confidence, produced via simulation of an alternative model (error bars), and produced through the derived predictions adjusted for an alternative model (shading). The alternative model featured confidence that was a direct readout of the evidence accumulated, rather than a Bayesian readout for confidence.

G Statistics for Fig. 8

We verified that the patterns observed in Fig. 8 were statistically significant. To do this, for each participant and condition, we calculated the correlation coefficient between the signal strength and binned confidence on error trials. We then compared the correlation coefficients to zero across participants. In the "Immediate confidence" condition binned confidence increased with signal strength ($t(39) = 7.1$, $p = 1.4 \times 10^{-8}$), while in the "Delayed confidence" condition binned confidence decreased with signal strength ($t(39) = -17$, $p = 7.7 \times 10^{-20}$).