Metacognitive Computations for Information Search: Confidence in Control

Lion Schulz¹, Stephen M. Fleming²,³,⁴, and Peter Dayan¹,⁵

¹ Department of Computational Neuroscience, Max Planck Institute for Biological Cybernetics, ² Department of Experimental Psychology, University College London ³ Wellcome Centre For Human Neuroimaging, UCL Queen Square Institute of Neurolog, University College London ⁴ Max Planck Centre For Computational Psychiatry and Ageing Research, University College London ⁵ Department of Computer Science, Eberhard Karls University of Tübingen

The metacognitive sense of confidence can play a critical role in regulating decision-making. In particular, a lack of confidence can justify the explicit, potentially costly, instrumental acquisition of extra information that might resolve uncertainty. Human confidence is highly complex and recent computational work has suggested a statistically sophisticated tapestry behind the information that governs both the making and monitoring of choices. However, the consequences of the form of such confidence computations for search have yet to be understood. Here, we reveal extra richness in the use of confidence for information seeking by formulating joint models of action, confidence and information search within a Bayesian and reinforcement learning framework. Through detailed theoretical analysis of these models, we show the intricate normative downstream consequences for search arising from more complex forms of metacognition. For example, our results highlight how the ability to monitor errors or general metacognitive sensitivity impact seeking decisions and can generate diverse relationships between action, confidence, and the optimal search for information. We also explore whether empirical search behavior enjoys any of the characteristics of normatively derived prescriptions. More broadly, our work demonstrates that it is crucial to treat metacognitive monitoring and control as closely linked processes.

Keywords: Computation, confidence, metacognition, information search, decision making

After carefully deliberating between Scotland and the Cote d’Azur, you have decided to spend your next summer holiday in the north of Britain. You are rather confident about this choice. Before it comes to booking your train tickets, an article pops up in your feed: "Skye or Saint-Tropez – the ultimate comparison". Do you spend money and time on reading this article? Or do you purchase the tickets right away? Conflicts like this are sadly commonplace: Do you read another news story before heading to the polls? Do you consult a doctor before heading to the pharmacy? Each time, we have to balance accuracy with (monetary or temporal) cost (Cohen, McClure, & Yu, 2007; Dayan & Daw, 2008; Wald, 1949). The decision to gather further information rather than making the choice based on our current knowledge thus depends critically on the initial choice’s expected rectitude, given current information – which is a form of subjective confidence (Pouget, Drugowitsch, & Kepecs, 2016). In this paper, we examine formal relationships between confidence and information search.

Humans enjoy a sophisticated and explicit sense of their expected accuracy, in forms of metacognition (Fleming & Daw, 2017; Shekhar & Rahnev, 2020; Yeung & Summerfield, 2012). Explicit metacognition refers to conscious representations of performance that are available for flexible usage in behavioural control or communication to others (Shea et al., 2014). Metacognitive evaluations duly accompany a wide range of decisions, from basic judgments of perception and memory to reflective evaluations of our knowledge or the "goodness" of subjective choices (De Martino, Fleming, Garrett, & Dolan, 2013; Fischer, Amelung, & Said, 2019; Nelson & Narens, 1990; Rahnev et al., 2020).

Recent research has begun to reveal complexities in how these metacognitive judgments arise – both in terms of within-subject decision processes and between-subject factors (Fleming & Daw, 2017;
Of particular interest are the processes that contribute to drops in confidence following errors. Such error monitoring can occur even in the absence of external feedback, and can rely on a purely internal evaluation mechanism (Atiya et al., 2020; Boldt & Yeung, 2015; Rabbitt, 1966; Yeung, Botvinick, & Cohen, 2004). Moreover, the quality of confidence judgements differs substantially between individuals even when their objective decision performance is equivalent (Fleming & Lau, 2014; Shekhar & Rahnev, 2020). This indicates personal-level influences on metacognition – a finding with implications for phenomena ranging from psychiatric disorders to political radicalisation (David, Bedford, Wiffen, & Gilleen, 2012; Hoven et al., 2019; Rollwage et al., 2018).

Models of confidence have attempted to address the diversity of human confidence, often treating confidence as an (approximately) Bayesian readout of decision correctness. However, while Bayesian models in which the same information underlies both decision and confidence can account for some confidence phenomena (Cartwright & Festinger, 1943; Kepecs, Uchida, Zariwala, & Mainen, 2008; Sanders, Hungya, & Kepecs, 2016), they struggle to capture error monitoring or differences in the quality of people’s metacognition. A particular focus of modelling has thus been placed on dissociating action and confidence: In essence, these accounts allow the specific information underlying choice and confidence to differ. For example, some propose that extra inputs are available for the confidence rating that accrue after, or in parallel, to the decision itself (Moran, Teodorescu, & Usher, 2015; Navajas, Bahrami, & Latham, 2016; Pleskac & Busemeyer, 2010). Others have suggested a range of covariance structures governing the underlying information sources (Fleming & Daw, 2017; Jang, Wallsten, & Huber, 2012).

In turn, the choice to collect more information has been shown to be causally controlled by metacognitive estimates. For instance, in a study of perceptual decision-making, Desender, Boldt, and Yeung (2018) used a perceptual manipulation to induce higher and lower levels of confidence in different conditions, while keeping subjects’ objective performance equal. In the condition with lower confidence, subjects were more likely to seek additional information, providing key causal evidence for the role of confidence in the collection of information. In the memory domain, artificially boosting people’s confidence when learning word pairs makes them less likely to choose to study those pairs again, even though performance remains unchanged (Metcalf & Finn, 2008). Other studies also support this close relationship between confidence and information search. For example, neural markers of confidence have been linked to variability in information search (Desender, Murphy, Boldt, Verguts, & Yeung, 2019) and different forms of confidence are proposed to influence the trade-off between exploring new options and exploiting old ones (Boldt, Blundell, & De Martino, 2019; Wilson, Geana, White, Ludvig, & Cohen, 2014; Wu, Schulz, Speekenbrink, Nelson, & Meder, 2018).

However, like metacognition, information-seeking behavior is highly complex and differs substantially between individuals. This has important implications, for example for psychiatric symptoms such as paranoia (Ermakova et al., 2018; Garety & Freeman, 2013; So, Siu, Wong, Chan, & Garety, 2016), or for patients suffering from obsessive compulsive disorder (Baranski & Petrusic, 2001; Hauser, Moutoussis, Dayan, & Dolan, 2017; Navajas et al., 2016; Tolin, Abramowitz, Brigidi, & Foa, 2003). Inter-individual differences in information search are also linked to real-world attitudes, as is evident in a relationship between lowered search and dogmatism (L. Schulz et al., 2020). The close coupling between metacognitive monitoring and control (Nelson & Narens, 1990) makes it compelling to study how the complexities of the former and the latter relate to each other. However, they have yet to be studied within a unified framework.

To do this, we probe the consequences for information search of a recent, rather general, account of metacognitive monitoring described by Fleming and Daw (2017). Fleming & Daw’s main proposal, a
second-order model of self-evaluation, posits that confidence formation depends on a "rater" equipped with an inferential mechanism that evaluates the (covarying) evidence supporting an "actor's" choice. Through this mechanism, the second-order model can account for diverse aspects of confidence, from error monitoring to variations in both metacognitive sensitivity and overconfidence. Following Fleming and Daw (2017), we present the second-order model prefaced with two architecturally simpler treatments, the first-order and postdecisional models, which help introduce and illuminate the extra richness of the second-order model.

In general, the purpose of this paper is not to make specific judgements about the merits of one particular model of metacognitive monitoring over another. Such arguments have been the focus of much debate in the metacognitive literature, and we point the interested reader to these works (Fleming & Daw, 2017; Khalvati, Kiani, & Rao, 2021; Rahnev et al., 2020; Shekhar & Rahnev, 2022; Webb, Miyoshi, So, Rajananda, & Lau, 2021; Yeung & Summerfield, 2012). Rather, the intention of this paper is to probe the consequences of what it means to possess more complex forms of confidence, however they might arise, for metacognitive information search and control more generally. In this endeavour, the three models we investigate are intended to be broadly representative of larger groups of models.

This paper adopts a theoretical and computational perspective that aims to elucidate what more complex forms of metacognition should normatively mean for search. To do so, we start by introducing the core components of confidence and information seeking at an abstract level, outlining the intuitions behind the relationships between action, confidence and informa-
tion search. We then zoom in on these computations in more detail, first discussing different theories of monitoring as delineated by Fleming and Daw (2017), and then considering the downstream consequences that arise in optimal control computations for information seeking. In our main results section, we investigate how normative metacognitive search should optimally proceed under a diversity of situations implied by these accounts. Finally, we seek to build a bridge from our theoretical accounts of optimal metacognitive information search to empirical data by analyzing suitable elements of a large existing dataset of human confidence and information search behaviour.

The Information-Seeking Problem

General overview

Action, confidence, and information seeking can be investigated in minimal settings such as the bare-bones perceptual task presented in Figure 1A. There, participants are presented with a noisy stimulus (for example, two boxes each with a different number of flickering dots), about which they have to first make an initial binary decision (more dots in the left or right box). They then express their confidence in this decision. Following this, they can decide whether to (1) see another helpful stimulus before making a final decision or whether they want to (2) make this final decision without any additional evidence. Seeing the second sample is associated with a cost, and the final (and possibly the initial) decision is rewarded.

Such a set-up is similar to the controlled environments previously used to study confidence and information seeking (Desender et al., 2018; Desender, Murphy, et al., 2019; L. Schulz et al., 2020). In these tasks, human subjects have been shown to modulate their information-seeking act might then be asking a friend for advice, or deciding to study further. The information-seeking act could equally address other sensory modalities. More abstractly, the available actions could also represent two objects of noisy value between which the actor decides (De Martino et al., 2013; Lee & Daunizeau, 2021), or a judgment of learning (Metcalfe & Finn, 2008). The information-seeking act might then be asking a friend for advice, or deciding to study further.

Formalising action and metacognitive monitoring

We can frame the decision-making problem as a partially observable Markov decision problem, or POMDP (Monahan, 1982; Sutton & Barto, 2018). In this, the actor’s first task is to use its cue $X_I$ to infer which of two states of the world $I_d$ (with $d \in \{-1, 1\}$) it inhabits. These two states can represent a multitude of stimuli and task configurations, including more dots in the left ($I_d = -1$) or right ($I_d = 1$) box in the task of Figure 1A, but equally any other binary judgement. The actor’s cue $X_I$ only affords partial information about $d$ and is conventionally thought to be drawn from a normal distribution with mean $d$ and standard deviation $\sigma_I$. 

In these terms: first, the rater perceives some evidence $X_I$, and then makes a decision, $a_I \in \{-1, +1\}$, where $a_I = -1$ represents choosing left and $a_I = 1$ right. The rater then publicly expresses its confidence, $c_I \in [0, 1]$ in this decision, based on the information to which it has access. This information may or may not include $X_I$ and/or some unique information of its own $Y_I$. Third, the seeker decides whether more information should be sought ($s_I \in \{0, 1\}$, with $s_I = 0$ representing no search and $s_I = 1$ representing search). The actor then makes a final decision, $a_F \in \{-1, +1\}$. In our simple formulation, $a_F$ can be based on $X_I$ along with $c_I$ (since the rater’s confidence judgment is veridical and public) and, if extra information was sought, a further sample, $X_F$. We refer the reader to Table 1 for an overview of the notation used throughout.

We now unpack the computations behind these steps further, first discussing models of the initial decision and confidence, as outlined by Fleming and Daw (2017), before elucidating their consequences for how the seeker should optimally decide to search for information. While we use a visual decision-making task for illustration, the underlying computational problem is more wide-reaching, and our computational-level analysis is itself agnostic to the setting in which actor, rater, and seeker are placed. For example, the stimuli could equally address other sensory modalities. More abstractly, the available actions could also represent two objects of noisy value between which the actor decides (De Martino et al., 2013; Lee & Daunizeau, 2021), or a judgment of learning (Metcalfe & Finn, 2008). The information-seeking act might then be asking a friend for advice, or deciding to study further.

Formalising action and metacognitive monitoring

We can frame the decision-making problem as a partially observable Markov decision problem, or POMDP (Monahan, 1982; Sutton & Barto, 2018). In this, the actor’s first task is to use its cue $X_I$ to infer which of two states of the world $I_d$ (with $d \in \{-1, 1\}$) it inhabits. These two states can represent a multitude of stimuli and task configurations, including more dots in the left ($I_d = -1$) or right ($I_d = 1$) box in the task of Figure 1A, but equally any other binary judgement. The actor’s cue $X_I$ only affords partial information about $d$ and is conventionally thought to be drawn from a normal distribution with mean $d$ and standard deviation $\sigma_I$.
Table 1

Notation. We distinguish between (A) aspects of the task, (B) random variables, and (C) quantities/actions that the agent computes

<table>
<thead>
<tr>
<th>Notation</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A. Experiment</strong></td>
<td></td>
</tr>
<tr>
<td><strong>States</strong></td>
<td></td>
</tr>
<tr>
<td>$I_d \in {I_{-1}, I_1}$</td>
<td>Initial actual state of the problem</td>
</tr>
<tr>
<td>$F_d \in {F_{-1}, F_1}$</td>
<td>Final state of the problem</td>
</tr>
<tr>
<td>$d \in {-1, 1}$</td>
<td>Underlying state of the stimulus</td>
</tr>
<tr>
<td><strong>Rewards and costs</strong></td>
<td></td>
</tr>
<tr>
<td>$r_I, r_F$</td>
<td>Reward for the initial and final decisions</td>
</tr>
<tr>
<td>$r_S$</td>
<td>Cost for obtaining the additional stimulus</td>
</tr>
<tr>
<td><strong>B. Random variables and their attributes</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Random variables</strong></td>
<td></td>
</tr>
<tr>
<td>$X_I, X_F$</td>
<td>Actor’s stimuli at $I_d$ and $F_d$</td>
</tr>
<tr>
<td>$Y_I$</td>
<td>Rater’s stimulus at $I_d$ (in postdecisional and second-order model)</td>
</tr>
<tr>
<td>$Z_I$</td>
<td>Combination of $X_I$ and $Y_I$</td>
</tr>
<tr>
<td>$Z_F$</td>
<td>Combination of $X_I$, $Y_I$ and $X_F$</td>
</tr>
<tr>
<td><strong>Noise terms associated with random variables</strong></td>
<td></td>
</tr>
<tr>
<td>$\sigma_I, \sigma_F$</td>
<td>Standard deviation of $X_I$ and $X_F$</td>
</tr>
<tr>
<td>$\tau_I$</td>
<td>Standard deviation of $Y_I$</td>
</tr>
<tr>
<td>$\zeta_I, \zeta_F$</td>
<td>Standard deviation of $Z_I$ and $Z_F$</td>
</tr>
<tr>
<td>$\rho_I$</td>
<td>Correlation between $X_I$ and $Y_I$</td>
</tr>
<tr>
<td>$\Sigma_I$</td>
<td>Covariance matrix for $X_I$ and $Y_I$ with $\sigma_I, \tau_I$ and $\rho_I$</td>
</tr>
<tr>
<td><strong>C. Agent</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Actions and expressions</strong></td>
<td></td>
</tr>
<tr>
<td>$a_I, a_F \in {-1, 1}$</td>
<td>The actor’s initial and final decisions</td>
</tr>
<tr>
<td>$c_I \in [0, 1]$</td>
<td>The rater’s confidence in the actor’s initial decision</td>
</tr>
<tr>
<td>$s_I \in {0, 1}$</td>
<td>The seeker’s decision whether (1) or not (0) to seek</td>
</tr>
<tr>
<td>$a_{F,S_I}$</td>
<td>The actor’s final decision conditioned on the seeking decision</td>
</tr>
<tr>
<td><strong>Values and action values</strong></td>
<td></td>
</tr>
<tr>
<td>$Q_S(s_I)$</td>
<td>Action values for seeking $Q_S(1)$ or not seeking $Q_S(0)$</td>
</tr>
<tr>
<td>$V_{F,Z_F}$</td>
<td>Value of having a specific cue $Z_F$ at final state $F$</td>
</tr>
<tr>
<td>$V_{F,Z_I,S_I}$</td>
<td>Value of having a specific cue $Z_I$ at final state $F$, conditioned on whether agent will seek or not</td>
</tr>
<tr>
<td>$V_{F,Y_I,S_I}$</td>
<td>Value of having a specific cue $Y_I$ at final state $F$, conditioned on whether agent will seek or not</td>
</tr>
</tbody>
</table>

$$X_I \sim \mathcal{N}(d, \sigma_I)$$ \hspace{1cm} (1)

In a task only capturing the first decision, we might incentivize the initial decision $a_I$ with a pay-off of $r_I$ points for correct choices and 0 points for incorrect choices. In this case, a reward-maximising actor should optimally compare its sensory sample against a threshold. Under our stimulus and pay-off regime with equal noises and pay-offs and equally prevalent underlying states, this optimal threshold is set to 0, implying that:

$$a_I = \text{sgn}(X_I)$$ \hspace{1cm} (2)

More complex schemes for pay-offs (e.g., more reward for correctly identifying $d = 1$) or asymmetric sources will impact this decision rule (Dayan & Daw, 2008), but we will focus on this simple set-up for clarity. Our reward criterion also disregards any further notions of timing. In general, the expected performance of the actor is determined by $\sigma_I$, with higher values associated with more mistakes, on average (see below for more details).

The rater’s task is now to compute a confidence, $c_I$, in $a_I$. We assume that the rater follows Bayesian precepts and reports its belief that $a_I$ was the correct choice, given its information (which we here denote by $C$) and the task parameters $\theta$:

$$c_I = P(d = a_I|C; \theta)$$ \hspace{1cm} (3)

We note that here we use the term "confidence" to refer only to this specific posterior probability that an
action was correct given the rater’s information. This delineates it from the broader notion of ‘certainty’ that refers to estimates of precision about noisy sensory or cognitive variables (Fleming & Daw, 2017; Pouget et al., 2016). As previewed, Fleming and Daw (2017) discuss three different models for the nature of C, the first-order, the postdecisional and the second-order model. We now recapitulate these models before adapting them to the information-seeking problem. Since the first-order model is a special case of the postdecisional model, we discuss them jointly.

### Postdecisional and first-order models

In the postdecisional model, the rater knows the actor’s information \( X_i \) and action \( a_i \). It also receives independent postdecisional information, \( Y_i \) (see also Figure 1B). This postdecisional cue \( Y_i \) is sampled from a distribution with the same mean \( d \) but with its own standard deviation \( \tau_i \),

\[
Y_i \sim N(d, \tau_i)
\]

The first-order model is an instance of the postdecisional model in which \( \tau_i = \infty \). In other words, in its case, the rater has no extra postdecisional information over and above the actor.

The rater first combines its sample with the actor’s sample in a precision weighted fashion, leading to a sufficient statistic \( Z_i \) which has a standard deviation of \( \zeta_i \):

\[
Z_i = \frac{X_i + Y_i}{\sigma_i^2 + \tau_i^2} \sim N(d, \zeta_i)
\]

The rater’s confidence in the actor’s choice then comes from the posterior distribution obtained through Bayes’ rule. Here, the distance between the threshold and \( Z_i \) becomes a proxy for the rater’s confidence:

\[
c_i = P(d = a_i | Z_i; \zeta_i) = \frac{p(Z_i | d = a_i; \zeta_i)}{\sum_d p(Z_i | d; \zeta_i)} = \frac{1}{1 + e^{-2dZ_i/\zeta_i}}
\]

An important facet of the postdecisional model is that \( Z_i \) and \( a_i \) can “contradict” each other. In other words, the rater might have information that favours one judgement (e.g. \( Z_i = 0.7 \)) while the actor might have had information that favoured the other (e.g. \( X_i = -0.2 \)). Such a disagreement will lead confidence to be lower than 0.5, triggering what is known as error monitoring (Boldt & Yeung, 2015; Fleming & Daw, 2017; Yeung & Summerfield, 2012) as we see in Figure 1F.

In the first-order model, with \( \tau_i = \infty \), the actor and rater have the same information \( (Z_i = X_i) \), but the confidence computations outlined in equation 6 still hold. As a consequence, the rater will always endorse the actor’s choice in the first-order model. This, in turn, prevents it from exhibiting error monitoring. Furthermore, and inconsistent with empirical observations of dissociations between performance and metacognition (Rahnev et al., 2020; Shekhar & Rahnev, 2020), it ensures the actor and the rater’s accuracy remain coupled, as we will discuss in more detail below. 1

### Second-order model

Fleming & Daw’s postdecisional rater is particularly well endowed with information: it knows exactly what the actor used to make its decision, plus some additional information (if \( \sigma_i < \infty \)). This assumption might not hold under several scenarios, for example different neural pathways for action and confidence formation. It also does not allow the rater to know less than the actor, a fact that will be important when capturing empirical metacognitive hyposensitivity, as we will see later. The second-order model solves this problem by denying the rater direct access to the actor’s variable \( X_i \), and rather only allowing correlational access to it. That is, rather like two humans interacting, the second-order rater only observes the actor’s binary decision \( a_i \), and has to use this in concert with its own personal information \( Y_i \) to form a confidence estimate in this decision. This is facilitated by the fact that, in contrast to the postdecisional model, the actor’s \( (X_i) \) and rater’s \( (Y_i) \) information can be correlated, as is visible in Figure 1E:

\[
\begin{bmatrix}
X_i \\
Y_i
\end{bmatrix} \sim N \left( \begin{bmatrix} d \\ 1 \end{bmatrix}, \Sigma_i \right)
\]

\[
\Sigma_i = \begin{bmatrix}
\sigma_i^2 & \rho_i \sigma_i \tau_i \\
\rho_i \sigma_i \tau_i & \tau_i^2
\end{bmatrix}
\]

Knowledge about this informational set-up and the actor’s decision rule allows the rater to make partial inferences about the value of \( X_i \) through its observation of the actor’s action. Specifically, because the actor makes its decision based on its cue’s sign (as per

---

1Our postdecisional model somewhat extends Fleming and Daw (2017) but uses a formally equivalent architecture. Specifically, whereas Fleming and Daw (2017) only discuss cases where \( \tau_i = \sigma_i \) (and thus \( \zeta_i = \frac{\sigma_i}{\tau_i} \)), we allow the actor and noise to vary independently and describe how the two can be optimally combined. This additional flexibility enables us to subsume the first-order model within the postdecisional model and will become key in our results when we describe different levels of metacognitive insight.
equation 2), the rater can immediately exclude one half of the stimulus space (either positive or negative) after observing the action alone. Depending on the correlational structure linking \(X_t\) and \(Y_t\), the rater can further pinpoint the location of \(X_t\) by leveraging knowledge about the pattern of correlations between the samples. In short, the rater’s confidence combines two actual, and one inferential, source of information about \(d\): the action \(a_t\), its own \(Y_t\), and the information provided by these variables about \(X_t\) via the covariance between \(X_t\) and \(Y_t\):

\[
c_t = P(d = a_t|Y_t, a_t; \Sigma_t)
\]

(9)

We recapitulate the details of this computation, including the inference of \(X_t\) in the appendix (section A). As with the postdecisional model, the second-order model also supports error monitoring, and can give rise to different levels of metacognitive insight. However, in contrast to the postdecisional model, the rater can have worse information than the actor, which, as we will see, is critical to produce more complex forms of confidence.  

A crucial aspect of the second-order model is that an agent (here, the rater) has to "infer the causes of its own action" (Fleming & Daw, 2017). This (partial) decoupling of action and confidence information gives it more flexibility than the postdecisional model and makes the action a crucial input to the computation, in turn boosting confidence for ambiguous \(Y_t\)'s. This is visible in Figure 1G, and discussed at length in Fleming and Daw (2017).  

Interim discussion of confidence computations

The previous section has recapitulated the accounts of metacognitive monitoring investigated by (Fleming & Daw, 2017). Their main idea is a (partial) dissociation between action and confidence. Many non-exclusive accounts exist about the source of these dissociations. For example in perceptual decision-making, evidence might further accumulate or degrade after a motor action is initiated. In memory- or value-based decision-making more information might arise through additional pondering. Furthermore, rating and acting might rely on partially different neural pathways (see Rahnev et al. (2020) for a recent overview of dissociations between confidence and action). We note that, in keeping with Fleming and Daw (2017), we here largely stick to Marr’s computational level (Marr, 1982). This more abstract perspective also means that we are a priori agnostic towards implementational and algorithmic level questions, and the specific source of information or error.

Modelling complex confidence phenomena through a dissociation between evidence sources underlying acting and rating is of course not unique to Fleming and Daw (2017). However, locating other accounts relative to the postdecisional and second-order models requires a careful look at their respective informational architectures. Perhaps the most significant family which is at least subtly different includes those models that assume that evidence accumulates continually (Moran et al., 2015; Pleskac & Busemeyer, 2010; Resulaj, Kiani, Wolpert, & Shadlen, 2009; van den Berg, Zylberberg, Kiani, Shadlen, & Wolpert, 2016). These relate to our models in two structurally different ways.

For an example of the first, consider what Pleskac and Busemeyer (2010) call a two-stage model. Superficially, this looks like our postdecisional model: the actor makes its decision based on one source of information \(X_t\), and the rater bases its confidence on a \(Z_t\), which is \(X_t\) plus some additional, independent, information \(Y_t\) (whose precision is usually governed by the time that passes between the action and confidence). However, in Pleskac and Busemeyer (2010), the actor uses an algorithm based on diffusion-to-bound, and so \(X_t\) is perfectly predicted by \(a_t\). Consequently, whereas our \(X_t\) can be accompanied by different degrees of (first-order) certainty, the accumulation bound fixes this certainty. As a result, the rater can use the actor’s decision as a sufficient statistic for the rater’s random variable, and will know (as a function of the bounds) how accurate this decision is on average. In turn, this informational set-up for the rater is an instance of what we would call a second-order model with \(\rho_{Z_t} = 0\). There, the rater also only knows the average accuracy of the actor, and receives uncorrelated evidence which it combines with

---

2We note that theoretically, nothing prevents correlations between the samples \(X_t\) and \(Y_t\) of the the postdecisional model. However, as we will see, the key distinction between the second-order and postdecisional model is at the level of the general access that the rater has to the actor’s information. In the second-order model, knowledge of such correlations between rater and actor becomes crucial to allow for more precise confidence computations. In the postdecisional model, in contrast, introducing such a correlation simply degrades the additional information carried by \(Y_t\). As a result, a hypothetical postdecisional \(\rho_{Z_t}\) parameter and the already existing \(\tau_t\) would only trade-off in the computations of the rater and seeker and not add any additional subtleties. In keeping with Fleming & Daw, we therefore do not further discuss the role of correlation in the postdecisional model.

3We note that the specific distributions used for this simple form of second-order model have some previously unexplored peculiarities at various limiting values. Because these are not essential for our investigations, we discuss them in the appendix, section C.
The second structural relationship is to note that the action threshold in dynamical diffusion-style models already implements an implicit case of the optional information seeking computation that we study explicitly—allowing more information to be collected (typically at the expense of time) given insufficient confidence.

While confidence models with dynamically accumulating evidence can provide some additional insights (for example into reaction times), we here focus on the second-order model, and its postdecisional sibling. This is because its flexible computational-level framework provides a broader view of metacognitive monitoring through the lens of partially shared information between rating and acting and therefore subsumes many other accounts.

### Formalizing the information-seeking problem

Regardless of the model of metacognitive monitoring, after a confidence estimate is formed, the seeker needs to decide whether the actor should see additional information before making its final decision \( a_F \) about \( d \). To conceptualize this more formally, we extend our POMDP (Dayan & Daw, 2008; Gottlieb, Oudeyer, Lopes, & Baranes, 2013) by adding a second pair of states \( F_d \) that deterministically follow \( I_d \) \((L_1 \rightarrow F_{-1}, I_1 \rightarrow F_1)\). If the seeker decides to seek, the actor receives a second stimulus \( X_F \) at \( F_d \) which it can use to make its final decision. We again assume this second cue to be sampled from a normal distribution with mean \( d \) and an associated standard deviation \( \sigma_F \):

\[
X_F \sim \mathcal{N}(d, \sigma_F) \tag{10}
\]

A final correct decision again comes with a remuneration \( r_F \) whereas an incorrect choice leads to 0 points. Finally, and crucially, seeking incurs a cost, \( r_s \) \(^4\).

Regardless of the metacognitive information configurations outlined above, the seeker’s choice involves the same basic question: Is seeking worth the cost? To decide, it computes two action-values, \( Q_5(s_I) \): one for seeking \( Q_F(1) \) and one for not seeking \( Q_F(0) \). In short, these involve predicting how accurate the actor will be on the final decision, with or without \( X_F \). We will next outline these computations, first explaining the details in the simpler case of the postdecisional model, and then highlighting differences in second-order computations. Across these models, we assume that the seeker has the same information as the rater, i.e., that there is a computational symmetry between metacognitive monitoring and control. This will allow us to capture key unique contributions of confidence to information search over and above the actor’s objective performance.

To illustrate the computations involved, we follow recent studies (Desender et al., 2018; L. Schulz et al., 2020) who provide no reward for the initial decision \( r_I = 0 \). We fix the reward for the final decision at \( r_F = 1 \) and will show different costs for the additional stimulus \( r_s \). Furthermore, we assume a noisier first \((\sigma_I = 1.5)\) than second \((\sigma_F = 1)\) stimulus.

#### Seeking in the postdecisional / first-order case

To compute the two \( Q \)-values, we first need to consider how the final decision \( a_F \) might be made at \( F_d \) with and without \( X_F \).

If the seeker decides to collect no further information, the actor’s final decision will be based on the same information as the rater’s initial confidence. As a result, the final decision will just repeat its initial decision \( a_{F,0} = a_I \) if \( Z_I \) and \( X_I \) agree. If they contradict each other, the actor will correct what it assumes to be an initial mistake and change its mind. In confidence space, this transition occurs at \( c_I = 0.5 \)

To compute the associated action value for not seeking, \( Q_F(0) \), the seeker first computes the optimal expected value \( V^*_F(z_{F,0}) \) of having a specific \( Z_F \) at \( F_d \) conditioned on its non-seeking behavior. This involves multiplying the reward obtained through a correct final decision with the probability of making a final correct decision based on \( Z_I \) (assuming that incorrect decisions incur no cost). In the postdecisional model, this probability is simply \( \max(c_I, 1 - c_I) \). Figure 2 shows the sub-components of the postdecisional seeker. For clarity, we there assume that \( r_I = \infty \), reducing the problem to the first-order model. Figure 2A depicts the posteriors and the associated values. Importantly, this is the equivalent of the curves for the first-order confidence. Since not seeking costs nothing, the optimal action-value for not seeking is just this value:

\[
Q_F^*(0) = V^*_F(z_{F,0}) = \max(P(d = -1|Z_I), P(d = 1|Z_I))r_F = \max(c_I, 1 - c_I)r_F \tag{11}
\]

In contrast, if the seeker decides to seek, the actor can use the additional stimulus \( X_F \) to disambiguate \( d \) further for its final decision \( a_{F,1} \). It does this by first forming a final combined variable \( a_{F,1} \) in a precision weighted fashion equivalent to equation 5:

\(^4\)We use this pay-off scheme for simplicity, but note the possibility of others (including temporal discounting). Furthermore, we assume that the agent has a linear utility function, which precludes forms of risk-aversion.
Subcomponents of the information-seeking computation in a first-order / postdecisional model. (A) To compute the Q-value of not seeking, $Q_S(0)$ (bold line), the agent computes the max of the two posteriors over $d$ from $Z_I$, $P(d = -1|Z_I)$ and $P(d = 1|Z_I)$. (B) If the seeker decides to search, it receives $X_F$ which it combines in a precision-weighted fashion with $Z_I$ to form a $Z_F$. Here, we plot the posterior $P(d = 1|Z_F)$. Because $Z_I$ is noisier than $X_F$ ($\zeta_I > \sigma_F$), the apparent slope of this posterior is not $-1$. Rather, $X_F$ is weighted more than $Z_I$. The converse posterior $P(d = -1|Z_I) = 1 - P(d = 1|Z_I)$ is the remaining probability. (C) The seeker computes the value associated with a given $Z_F$ from the maximum of these two possible posteriors. (D) Because it needs to decide whether to seek or not before receiving $X_F$, the agent needs to predict $X_F$. It does this by summing the two possible source distributions $N(d, \sigma_F)$ weighted by their individual confidence values. (E) To compute the value for seeking $Q_S(0)$, the agent averages over the two quantities in C and D, based on its $Z_I$. We here display the $Q$-value for not seeking and for seeking overlayed, with the latter shown as a function of the seeking cost $r_S$. Note how the maximum of the $Q$-value for seeking $Q_S(0)$ is defined by this cost. (F) The agent seeks when searching is more valuable than not seeking. We here display the difference between the two values, transformed into confidence space. As confidence increases, the benefit of seeking decreases. Partially adapted from Dayan and Daw (2008). Parameters set at $\zeta_I = \sigma_I = 1.5, \sigma_F = 1$.

\[ Z_F = \frac{Z_I + X_F}{\frac{1}{\sigma_I^2} + \frac{X_F}{\sigma_F^2}} \sim N(d, \zeta_F) \] where \[ \zeta_F = \sqrt{\frac{1}{\frac{1}{\zeta_I^2} + \frac{1}{\sigma_F^2}}} \] (12)

Similarly to the first decision, the agent can then compare $Z_F$ against a threshold (again, optimally $Z_F = 0$ given our pay-off regime) to make the final decision. We plot the posterior associated with this value for $d = -1$ in Figure 2B. There, the threshold for $a_F$ is where $P(d = 1|Z_F) = 0.5$.

In our example, recall that we set the initial stimulus $Z_I$ to be noisier than the final stimulus $X_F$ ($\sigma_I = 1.5$ and $\sigma_F = 1$). As a result, $X_F$ is given more weight than $Z_I$ in the posterior. For example, an $X_F = 1$ will increase the $P(d = 1|Z_F)$ posterior more than an equivalent $Z_I = 1$. Similarly, a less extreme $X_F$ will be necessary to overturn a $Z_I$ of a different sign. This is evident in the tilt of the posterior, which is not fully diagonal but rather slants towards $X_F$.

As for $Z_I$ in the no-seeking calculations, we compute the expected value of a given combination of $Z_I$ and $X_F$ from the maximum of the two possible posteriors (where $P(d = -1|Z_F) = 1 - P(d = 1|Z_F)$):\[ V_{F,|Z_F}^* = \max(P(d = -1|Z_F), P(d = 1|Z_F))r_F \] (13)
We plot this value $V^*_F(Z_t)$ in Figure 2C as a function of $Z_t$ and $X_F$. Again, the slope of the relationship is determined by the greater contribution to $Z_F$ of $X_F$ than that of $Z_t$.

Crucially, however, the seeker has to decide whether it wants to seek before the actor has seen $X_F$. It therefore needs to predict this second cue. The resulting distribution $p(X_F|Z_t)$ is a function of how likely the seeker believes that the actor is to receive a stimulus from one of the two means, or a sum of the two possible source distributions weighted by the rater's initial confidence $c_t$ (see appendix A). Figure 2D shows this distribution as function of $Z_t$, a mixture of two Gaussians.

To compute the expected value, $V^*_{F,Z_t}$, without having seen $X_F$, the seeker then integrates over this distribution and the previously defined value function for its value of $Z_t$ given the prospect of seeking. Based on this mean value the seeker can now work out the action-value for seeking by considering the cost of the search:

$$Q^*_S(1) = r_S + V^*_F(Z_t, 1) = r_S + \int_{X_t} p(X_F|Z_t) V^*_F(Z_t) dX_F \quad (14)$$

This value is shown in Figure 2E as a function of $Z_t$ and for different seeking costs $r_S$. It is highest when the seeker expects the final choice to be likely correct, that is when it is relatively sure about the identity $X_F$. With more ambiguous values of $Z_t$, this prediction can only be made with less certainty. The ceiling of the seeking value is defined by the cost $r_S$. We plot the value for not-seeking in Figure 2E. It approaches 0.5 as $Z_t$ becomes less distinctive, and $c_t$ therefore becomes lower. The larger of the two $Q$-values then determines the seeking choice:

$$s_t = \begin{cases} 
1 & \text{if } Q^*_S(1) > Q^*_S(0) \\
0 & \text{otherwise.} 
\end{cases} \quad (15)$$

For ambiguous values of $Z_t$, seeking is useful and will likely produce a better final outcome, even when taking into account the additional cost. When we transform the difference between the two values into confidence space (Figure 2F), we notice that seeking is more valuable than not seeking in lower confidence ranges, highlighting a crucial role of confidence in guiding the decision to seek.

**Seeking in the second-order case**

The second-order model entails some additional subtleties stemming from the different sources of information of actor, rater, and seeker. Recall that, in the second-order model, the rater only observes $a_I$ and $Y_I$ but does not have full access to the actor’s random variable $X_I$ (compare Figure 1C). Similarly, one might assume that the actor does not directly know $Y_I$ but only observes the rater’s utterance, $c_I$. However, because the actor knows its own first action, $a_I$, it can leverage the knowledge about the rater’s confidence algorithm to infer the initial confidence variable, $Y_I$, underlying $c_I$. It can then combine this random variable with $X_I$ to form $Z_I$, taking into account the cues’ relative precisions and their covariance (see appendix A). The reason the actor can extract $Y_I$ from $c_I$ but the rater cannot infer $X_I$ from $a_I$ is that confidence $c_I$ is continuous, whereas the action $a_I$ is discrete.

In the case of no seeking, the actor makes its decision based on $Z_t$ in a similar vein to the postdecisional model. In contrast to the postdecisional model, such a change of mind is not necessarily coupled to $c_I < 0.5$ given specific stimulus configurations, because of the additional information possessed by the actor at the second stage. Regardless, the value computations for holding a given $Z_t$ are equivalent to the postdecisional model (we detail this $V^*_F$ in the appendix section A). However, the seeker does not know $Z_t$ because it does not have access to $X_I$. It therefore has to marginalize out this quantity in a similar manner to the postdecisional model’s seeking computations:

$$Q^*_S(0) = V^*_{F,Y_I} = \int_{Z_t} p(Z_t|Y_I, a_I) V^*_F(Z_t) dZ_t \quad (16)$$

$$= \int_{X_I} p(X_I|Y_I, a_I) V^*_F(Z_t) dX_I \quad (17)$$

When the seeker decides to seek, the actor receives $X_F$ (again as per equation 10) which it combines with $Z_t$ to form a joint variable $Z_F$ (because there is no correlation between $X_F$ and $Z_F$ this is optimally done in a manner analogous to the postdecisional model, equation 12). This final variable $Z_F$ can then again be compared against a threshold $a_{F1}$ and is used to compute a value. Similarly to the first-order and postdecisional models, the seeker does not know all the parts of $Z_F$ and has to marginalize over the unknowns. As before these mean values are then used to compute the $Q$-values associated with seeking and not seeking:

$$Q^*_S(1) = r_S + V^*_{F,Y_I} = r_S + \int_{Z_F} p(Z_F|Y_I, a_I) V^*_F(Z_F) dZ_F \quad (18)$$

$$= r_S + \int_{X_I} \int_{X_F} p(X_I, X_F|Y_I, a_I) V^*_F(Z_F) dX_F dX_I \quad (20)$$
Commonalities and differences between the models’ seeking computations

While their details diverge, the different models still share some key commonalities. Crucially, they all employ their current confidence, $c_I$, to predict the future location of the second stimulus, $X_F$, which is then combined with the final value of a stimulus combination to form the $Q$-value for seeking. For the postdecisional (and first-order) model confidence is also a determinant of the value of not seeking. The second-order model additionally uses the confidence to compute the non-seeking value, albeit by harnessing it to predict the location of $X_t$, similarly to the postdecisional seeking value. All this highlights the crucial role metacognition and confidence play in optimal seeking decision.

Theoretical Results

In the following, we discuss how these models behave in our information-seeking task thereby showing intricate facets of optimal metacognitive information search. The task allows us to investigate several markers of action, confidence, and information search. With regard to the initial decision, we can observe (1) the average initial decision performance, (2) the initial confidence, and (3) an agent’s metacognitive accuracy (their ability to tell apart correct from incorrect choices through their confidence). With regards to the information-seeking decision, we can investigate an agent’s (1) average level of information search (2) its seeking criterion as well as (3) how calibrated search is to their initial decision accuracy. Finally, we can observe how accurate an agent is in its final decisions. Our models produce specific patterns of interactions between these behavioural markers.

We note that the optimal model behavior we discuss provides an upper bound as to how an agent could optimally harness its metacognition to seek information. These theoretical results thereby should not be taken as strong predictions for human choices, which need not be optimal. Rather, they reveal limits and possibilities of what metacognitive monitoring might mean for metacognitive control in the context of information search. We investigate more general patterns of metacognitive search and their link to human behavior in the second, more empirically focused, results section. We also discuss broader deviations from this normative behaviour in our closing discussion section.

Initial accuracy, average confidence and information seeking

If an agent has perfect insight into its average levels of correctness, it should use this insight to guide its search decisions: In essence, the more likely it is to make a mistake, the more additional information should benefit it.

To investigate this in the context of our models, we now first fix the quality of the second stimulus as well as the cost for seeking, and investigate an agent’s average confidence and information search. We show these markers as a function of the initial accuracy which, across all models, is a function of $\sigma_I$:

$$P(\text{correct}) = \phi(\sigma_I) = \int_0^{\infty} p(X_I|d=1;\sigma_I)dX_I \quad (21)$$

A plot showing this function is displayed in Figure 3A: The lower the actor’s noise $\sigma$ becomes, the more accurate the objective decision. Of note is that in the current model set-up, average confidence is correctly calibrated and so tracks the objective accuracy (Figure 3B). For example, when the actor correctly responds in 71% of cases, the rater’s average confidence will also be 71%. Thus, it is worth noting that the relationships between average initial accuracy and average search will be the same as between average confidence and average search. We discuss aberrations to this perfect calibration in a later section.

Postdecisional and first-order models

Figure 3C shows how the first-order ($\tau_I = \infty$) and more general postdecisional models prescribe a relationship between accuracy and information seeking: the lower the initial accuracy, the more likely it is to seek out information. In fact, average search approaches an asymptote of 1 below a certain accuracy for the first-order model given the final stimulus precision and information cost used here. In other words, when objective accuracy is low it will almost always be worthwhile for a first-order agent to seek despite the cost.

The extra information provided by the postdecisional stimulus $Y_I$ impacts this average seeking over and above the average objective accuracy. Specifically, we see a marked reduction in seeking with lower $\tau_I$ in comparison to the first-order model. This arises because the joint noise $\xi$ associated with $Z_I$ decreases as $Y_I$ becomes more precise, and since, in our model, $a_I$ can be informed by $Z_I$ even without extra seeking. In

---

"In terms of signal detection theory, the sensitivity $d'$ is proportional to the ratio of the difference between the mean signal for the two states and the standard deviation $\sigma$. We followed Fleming and Daw (2017) in adjusting $d'$ (and hence the accuracy) by adjusting $\sigma_I$; we could equivalently have adjusted the difference between the means (Fleming, Putten, & Daw, 2018)."
Figure 3
Initial accuracy, average confidence and information search. (A) Across models, the accuracy of the agent’s initial decision is governed by $\sigma_I$ through the function $\phi(\sigma_I)$. (B) In all models, the average initial confidence matches the average initial accuracy. (C,D,E) Average seeking decreases with increasing initial accuracy for all models. However, the precision of the rater’s stimulus, $\tau_I$, moderates this relationship differently depending on the model. (C) In the postdecisional model, lower rater noise leads to less seeking across the accuracy spectrum because the rater and seeker have more additional information. (D, E) In the second-order model, high rater noise is associated with reduced seeking at higher levels of initial accuracy. This is because the seeker lacks direct access to the actor’s actual cue $X_I$ and thus has to trust its decision. The correlation between the two ($\rho_I$) modulates this effect: (D: $\rho_I = 0.2$; E: $\rho_I = 0.5$). In the seeking plots, final stimulus noise and cost are fixed at $\sigma_F = 1$ ($\phi(\sigma_F) = 0.84$) and $r_S = -0.1$, respectively. The effect of $\rho_I$ is shown further in section C of the appendix.
tion seeking of a first-order agent whose actor noise $\sigma_I$ is equivalent to the postdecisional model’s rater noise $\tau_I$. In other words, when the actor knows almost nothing, the average seeking behaviour of a postdecisional agent will still resemble a first-order model whose decision accuracy would be governed by $\tau_I$.

**Second-order model**

The optimal seeking behaviour of the second-order model differs in key aspects from the postdecisional model (see Figure 3D;E). While normative second-order agents should, broadly speaking, reduce their search with increasing initial accuracy, their optimal behaviour exhibits a marked interaction between actor and rater noise. Specifically, the average seeking curves appear similar to those of the postdecisional model when the initial accuracy of the second-order actor is relatively low. There, rater/seekers with higher $\tau_I$ cues will seek more than those with lower $\tau_I$. Strikingly however, when the actor is more reliable (the initial accuracy is higher), second-order agents with higher $\tau_I$ will seek less than those with more precise rater information.

The peculiar interaction between objective accuracy and confidence-noise arises from the second-order model’s informational set-up: Whereas the postdecisional model makes use of both $X_I$ and $Y_I$, the second-order architecture only affords the rater access to a single cue, $Y_I$, and the actor’s initial decision $a_I$. This leaves it to make inevitably imperfect inferences about $X_I$.

When the actor is relatively accurate (e.g. $\sigma_I = 1$) and the rater’s information relatively inaccurate (e.g. $\tau_I = 3$), the rater has little information about the actor, but knows that the decision is likely correct (because $\phi(\sigma_I = 1) = .84$) – which will even be the case if the rater has an entirely ambiguous or even somewhat contradictory $Y_I$. As a result, its confidence will remain high, even when $Y_I$ and $a_I$ contradict each other (see appendix C and Figure 6F). In other words, the rater will essentially resort to “trusting” the actor’s action across a wide range of its own information $Y_I$. Because the seeker is equipped with the same information as the rater, it will likewise have too little information to justify the cost of seeking. Consequently, it will either fully trust or distrust the actor’s initial decision. In extreme cases, when $\tau_I$ approaches $\infty$, the relation between initial accuracy and average seeking will in fact resemble a step-function.

Relatedly, there is a marked lack of seeking for high accuracies in the second-order model when keeping $\tau_I$ constant. Notice how under the conditions of the cost of sampling and the accuracy of the second sample in Figure 3C, the first-order model will still search on up to a quarter of trials at 90% initial accuracy. In comparison, our normative second-order agent does not seek at all beyond that point with any but the most insightful values of $\tau_I$ (Figure 3D;E). This again comes down to the fact that the rater and seeker have no alternative but to trust the actor’s decision when $\tau_I \gg \sigma_I$. When the objective accuracy is very high (e.g. $\phi(\sigma_I = 8) \approx 90\%$) such an imbalance arises even when the rater noise $\tau_I$ is objectively low.

While these general trends hold across different values of the correlation $\rho_I$ (see panels D and E of Figure 3) we still note this parameter’s importance. In general, $\rho_I$ shapes both the additional information afforded by combining $X_I$ and $Y_I$ as well as the confidence rating process itself. Briefly, one way $\rho_I$ impacts normative information seeking is by increasing the step-like nature of the high $\tau_I$ curves which is visible in the difference between the $\rho_I = 0.2$ and $\rho_I = 0.5$ settings we depict. Also somewhat visible in our figures is the fact that with lower rater noise $\tau_I$ and with low accuracy, information seeking will in fact slightly decrease. Both these aspects arise from intricacies in the way signal and noise trade off in bivariate normal distributions. Because we focus on the cognitive rather than specifically mathematical implications of our models here, we save the discussion of these aspects for the appendix (section C).

**Cue reliability and information seeking**

The decision to seek out additional information should naturally not only be influenced by the quality of the stimuli we have encountered, but also by the quality of the stimuli that we will encounter in the future. With regard to the latter, there is room between two extremes: The second piece of information might always perfectly disambiguate the judgment (small $\sigma_F$) or it might carry almost no information whatsoever (large $\sigma_F$). While an optimal agent should want to almost always consult the former, it won’t profit much from the latter.

**Postdecisional and first-order models**

The first-order and postdecisional models capture this intuition in their normative behaviour, as evident in the first-order model depicted in Figure 4A. There, we show the average seeking for different levels of accuracy afforded by the actor’s initial and final stimulus $\phi(\sigma_I)$ and $\phi(\sigma_F)$ while again keeping cost constant at $r_S = -1$. As before the agent will seek more as the initial cue becomes noisier (right to left). In turn, decreasing final cue noise (higher values of $\phi(\sigma_F)$; top to
bottom) increases the usefulness of the additional cue and with it the average information seeking for a given level of initial actor noise $\sigma_I$.

Figure 4B shows different levels of postdecisional noise. This produces similar patterns to the first-order model, albeit with some added complexity. The imprecision $\tau_I$ of the postdecisional cue again has considerable influence on the maximum possible optimal average seeking behaviour of the postdecisional agent. When the rater’s cue contains little noise (low $\tau_I$), almost no search is necessary. This is regardless of initial and final stimulus reliability. In turn, the normative information-seeking profile begins to again resemble that of a first-order agent as $\tau_I$ becomes larger.

Second-order model

While optimal second-order search broadly traces the postdecisional pattern arising from the interplay of the three noise parameters $\sigma_I$, $\tau_I$, and $\sigma_F$, we can observe some further intricacies. Specifically, the second-order model’s seeking does not progress as smoothly from high to low information seeking with lower $\sigma_I$ and higher $\sigma_F$, especially with high levels of rater noise, $\tau_I$ (as we have previously observed when only varying accuracy). Rather, it begins to resemble more of a step-function as the rater knows less and less. For example, compare the highest levels of $\tau_I = 3$ in panels B and C of Figure 4. Whereas the postdecisional model smoothly transitions from high to low search, the second-order model remains with a high propensity to search relatively long before terminating search more abruptly. The reason for this can again be found in the limited information of the rater: When the rater knows little and the actor surpasses a specific relative uncertainty, the decision to sample becomes more binary across the objective accuracy range.

Intermediate summary: Accuracy and search

In the two preceding sections, we demonstrated how metacognitive search is normatively governed by the information available to the seeker and the information expected to be gained through search. Broadly, the less information the seeker has and the more it can expect to gain from the final cue, the more it should seek. We highlighted how this relationship is complicated in a second-order architecture. There, the seeker does not have full access to what the actor already knows. When the accuracies of the seeker/rater and the actor are particularly imbalanced, this can give rise to what looks close to step-functions in the average search profiles. In
CONFIDENCE IN CONTROL

Figure 5

Confidence-seeking thresholds (A,B) The confidence at which an agent stops sampling (confidence threshold) is largely independent of initial accuracy in the postdecisional and second-order models. Rather, one governing factor is the cost of the additional stimulus $r_S$. By decreasing the $Q$-value of seeking (compare Figure 2E and F), higher costs reduce the space of confidence where it is worth probing. In the second-order model, we can also see the effect of the transition from seeking into no-seeking where the two confidence thresholds begin moving together. We set $\phi = 0.84$, and for the second order model $\tau_I = \sigma_I$ as well as $\rho = 0.5$. (C,D) The two main factors governing the confidence threshold are $r_S$ and the noisiness of the final stimulus $\sigma_F$, as is visible when we plot the upper confidence threshold as a function of the two (C: postdecisional; D: second-order). Specifically, the more expensive and the less reliable the information becomes, the lower the threshold is set and the less information is sought given the same initial stimulus statistics. When the agent does not seek at all we mark the threshold as 50%. Panels C and D use $\sigma_I = 3$ and $\tau_I = \infty$ for the postdecisional model and $\sigma_I = \tau_I = 3$ and $\rho_I = 0.5$ for the second-order model respectively.

other words, the seeker either fully trusts or distrusts the actor, leading it to seek information almost always or almost never.

Search threshold in confidence space

Apart from the average seeking propensity, another important feature of an agent’s behaviour in our task is its internal confidence threshold for search. Put differently, how confident should an agent optimally be to decide it has seen enough information? Our models allow us to investigate this phenomenon by finding the value of the rater’s internal variable for which the $Q$-values for seeking and not-seeking intersect and computing the confidence at this point. For a better intuition, compare Figure 2F, where the difference between the two values is plotted: The threshold is the point where this difference is 0. Importantly, turning this threshold into a marginalized prediction about how often an agent seeks information is not completely straightforward, as will be apparent when we later consider the underlying confidence distribution in more detail.

Postdecisional and first-order models

In the postdecisional model, this threshold is normatively largely independent of the initial rater and ac-
order model’s optimal thresholds remain mostly unchanged across a range of objective accuracies for the postdecisional case, for a constant final stimulus noise $\tau_F$. This confidence variations appear between the first-order instance of the postdecisional case, for a constant final stimulus noise $\tau_F$. This confidence variation varies neither as a function of accuracy nor postdecisional noise (which we do not display here).

This counterintuitive result arises from the Markovian property of the first-order and postdecisional models where the confidence $c_I$ is equivalent to a belief state summarising all the previous information. In the $Q$-value computations, only this belief matters, and not how it came about. Put differently, it is unimportant whether $c_I$ was based on a large $Z_I$ and large $\zeta_I$ or smaller $Z_I$ and smaller $\zeta_I$.

Rather than the initial stimulus statistics, the determining factors for the optimal placement of the threshold are the cost of the additional information and its precision. For intuition, consider the $Q$-value functions in Figure 2E (and Figure 2F). There, the cost alters the intersection between the two $Q$-values, with higher cost reducing the space of $Z_I$ in which seeking is worthwhile and thus lowering the threshold. This influence is apparent in panel A of Figure 5. In turn, smaller $\sigma_F$ afford less noisy predictions of $X_F$ in Figure 2D for a given $Z_I$, especially when this $Z_I$ is relatively unambiguous. Consequently, the $Q(s_I = 1)$ curve becomes steeper which leads to the intersection appearing for lower confidences. The joint influence of the final cue’s noise and cost are plotted in Figure 5C. The less expensive and the more precise the final stimulus set-up, the higher the boundary.

A subtle difference regarding the lower threshold appears between the first-order instance of the postdecisional model and regular postdecisional models with $\tau_I < \infty$. In the first-order version, the minimum confidence is bounded at 50% because the rater has exactly the same information as the actor. The rater will thus always endorse its decision. As a result, we only observe one set of confidence-space thresholds for seeking in the first-order model, namely the upper ones. In contrast, when the postdecisional cue contains information, the lower bound is simply the opposite of the upper bound. This is because the net uncertainty of an initial decision made with 50% confidence is essentially the same as one made with 55%. In turn, if the rater has high confidence that the actor has made a mistake, then it can safely turn down the opportunity to acquire additional information: The actor can change its choice $a_F$ without any additional external information.

**Second-order model**

Similarly to the postdecisional model, the second-order model’s optimal thresholds remain mostly unaffected by the initial stimulus statistics, as is visible in Figure 5B. There, we show the seeking threshold for a model whose rater noise $\tau_I$ always equals its actor noise $\sigma_F$ across a range of initial accuracies. Rather, it is again the cost and noise associated with the additional stimulus that determine where the threshold optimally falls (see Figure 5B and D). In fact, given their differences in knowledge, it is striking that this threshold is largely equivalent between the postdecisional and second-order models, at least for low initial accuracy values. Additionally, because the second-order model also produces confidence levels below 50% just like the postdecisional model, it possesses a lower threshold that mirrors the upper one.

As discussed above, the second-order model differentiates itself from the postdecisional model by producing behaviour where it does not seek at all. This allows us to investigate what optimally happens in the transition to this state of uniform non-seeking. In these cases, as we can observe that with higher cost levels in Figure 5B, the two confidence cut-offs begin moving closer together until they end up meeting at 50%. At this point, seeking stops. While the baseline threshold for low initial accuracy is thus unaffected by the initial stimulus set-up, different $\sigma_F$’s can produce different initial accuracies at which seeking becomes too costly. This thus affects when the two thresholds begin moving toward each other.

The influence of the final cue noise $\sigma_F$ and the seeking cost $r_s$ on the confidence cut-off is equivalent between the postdecisional and the second-order model despite their somewhat different $Q$-value computations (compare Figure 5C and D). This is because the second-order’s additional task of predicting the $X_I$ value is required to evaluate the $Q$-values for and against seeking. The net effect is that this extra step does not impact the threshold.

**Metacognitive accuracy and information search**

Metacognitive accuracy broadly describes an agent’s ability to discriminate its mistakes from its successes. In our task, this manifests in distinct confidence distributions for correct and incorrect choices: Agents with high metacognitive accuracy tend to have high confidence ratings when they are correct and low confidence ratings when they made a mistake.

We can delineate two measures of metacognitive accuracy: metacognitive sensitivity and metacognitive efficiency (Fleming & Lau, 2014). Metacognitive sensitivity describes the aforementioned separation of confidence distributions, with less overlap between the two functions a hallmark of high metacognitive sensitivity. In our framework, this sensitivity is largely gov-
Metacognitive sensitivity, efficiency and search (A,D) Seeking as a function of the metacognitive sensitivity. (A) As the metacognitive efficiency in the postdecisional model ($\sigma_I/\zeta_I$) increases, the need to seek more information decreases. (Note that metacognitive hypo-sensitivity is not possible within the discussed postdecisional model, in which either the same or additional information is available for the confidence rating as for the decision). (B,C,D,E) Distributions of confidence ratings after correct and incorrect decisions with areas in which an agent will seek displayed in grey. (B) In the first-order model (or a postdecisional model with $\tau_I = \infty$), metacognitive sensitivity is tied to the objective accuracy ($\sigma_I$) and thus has no independent influence on search. Confidence also has a lower bound at 50%. (C,D,E) When keeping the objective accuracy ($\sigma_I$) fixed, noise $\tau_I$ associated with the rater’s additional stimulus can produce diverse confidence distributions in the postdecisional model (C) and second-order model (E,F). Notice in particular how these models can correct their own mistakes without the need for additional search when confidence is below 0.5. In contrast to the postdecisional model, the second-order model can produce metacognitive hypo-sensitivity when $\sigma_I/\tau_I < 1$. (E) When initial accuracy is high and metacognition particularly inefficient in the second-order model, the confidence distribution shifts almost entirely out of the seeking-zone. (F) As a result, metacognitive hypo-sensitivity can prescribe both increasing and decreasing information search in the second-order model. Metacognitive hypersensitivity is still related to reduced search. For seeking averages and thresholds all plots use $\sigma_F = 1$ ($\phi(\sigma_F) = 0.84$) and $r_S = -0.1$. In the second-order case, $\rho_I = 0.5$. 

Figure 6
cerned by the quality of the rater’s information ($\xi$ in the postdecisional case, $\tau$ in the second-order case), with higher values of $\xi$ and $\tau$ resulting in lower metacognitive sensitivity.

While metacognitive sensitivity provides a useful marker of the quality of an agent’s metacognition, it is often confounded with objective accuracy. Easier tasks allow more insight into the quality of our decisions – such that when objective (e.g.) perceptual sensitivity is high, metacognitive sensitivity also tends to be high (Fleming & Lau, 2014). Metacognitive efficiency controls for this link between objective and metacognitive sensitivity by normalizing the latter by the former. This statistic is expressed as a ratio, with values less than 1 indicating metacognitive hyposensitivity, where metacognitive sensitivity is worse than would be expected based on objective performance, and values greater than 1 indicating metacognitive hypersensitivity, in which case metacognitive sensitivity is higher than expected based on objective performance (Fleming & Daw, 2017; Fleming & Lau, 2014). We note that what we refer to as metacognitive hyposensitivity has also been discussed under the label of metacognitive inefficiency (Shekhar & Rahnev, 2020).

The fact that the rater has different, possibly additional, sources of information from the actor is what licenses varying metacognitive efficiencies in our framework. The different models operationalize this slightly differently. In the postdecisional model, metacognitive efficiency can be expressed through the ratio $\sigma_1/\xi_1$. The larger this ratio, the more additional information the postdecisional rater has, and the higher its metacognitive efficiency. Note that in the present postdecisional model, with its optimal calculations, this ratio can never be below 1, precluding any forms of metacognitive hyposensitivity. The metacognitive efficiency of the second-order model is determined by $\sigma_1/\tau_1$ (for a constant $\rho_1$), again because of the restricted informational access of the second-order model. 6

**First-order and postdecisional models**

To understand the relationship between seeking and metacognitive accuracy, we first need to recapitulate in detail how metacognitive accuracy arises in our models. To illustrate this better, we plot distributions of confidence ratings conditioned on accuracy in Figure 6B-E. These illustrate the overlap between the distributions of confidence ratings for correct and incorrect answers.

In the first-order model (panels B; $\tau = \infty$), objective accuracy and metacognitive sensitivity are welded together. That is, higher objective performance (lower $\sigma_1$) results in more clearly distinguishable confidence distributions and thus increasing metacognitive sensitivity. By design, the ratio $\sigma_1/\xi_1$ is also always 1 in the first-order model, pinning down metacognitive efficiency.

In Figure 6B, we demonstrate the relationship between metacognitive accuracy and search in the first-order model. We plot the optimal seeking-thresholds we introduced above in black and the zone of confidence values where the agent seeks in grey. Recall that these are not influenced by the statistics of the first decision. Because the confidence distributions shift together for decreasing accuracy, this normatively results in more search. In essence, this relationship simply recapitulates what we have seen in the first section on objective accuracy and average search. Notably, sensitivity will appear to be related to decreased information search in the first-order model, but this is fully explained by the coupling of metacognitive and objective accuracy. Finally, there is no relation between search and metacognitive efficiency, as the latter is invariant in the first-order model.

In contrast to the first-order model, the postdecisional model with $\tau_1 < \infty$ can produce different levels of metacognitive efficiency. Figure 6C demonstrates this by keeping the objective accuracy ($\sigma_1$) constant, but increasing the quality of the rater’s information through $\tau_1$. In essence, these plots take the first order model of a given objective accuracy ($\tau_1 = 2$; middle of right top row of panel B), but give the rater additional information. The impact of this additional information is clearly visible: A well-endowed rater with a low $\tau_1$ and thus highly accurate postdecisional information is almost perfectly able to distinguish its correct from incorrect decisions, as expressed through its confidence. The confidence for correct decisions will be very high on average whereas the confidence for incorrect decisions will almost always indicate an error, that is be below 0.5. As $\tau_1$ increases, the rater’s additional information decreases, resulting in a confidence distribution very similar to the first-order model when noise is very

---

6 In the experimental literature, a plethora of measures assay metacognitive sensitivity and/or efficiency (Fleming & Lau, 2014). Most prominently, the meta-$d'$ statistic (Maniscalco & Lau, 2012) allows metacognitive sensitivity to be estimated within a signal detection theoretic (SDT) framework. Briefly, this approach estimates the $d'$ from a first-order SDT model that best fits the observed confidence distributions. This metric, known as meta-$d'$, can then be compared to the $d'$ calculated from the participant’s choices to produce a ratio meta-$d'/d'$, a typical measure of metacognitive efficiency (Fleming & Lau, 2014). Both meta-$d'$ and meta-$d'/d'$ scale with our parameters $\tau_1$ and, depending on the model, the ratio $\sigma_1/\xi_1$ or $\sigma_1/\tau_1$ (Fleming & Daw, 2017). However, for clarity we will in this theoretical section use our parameters $\tau_1$, $\sigma_1$, and their relationship to characterize sensitivity and efficiency.
high (leftmost plot of panel C with \( \tau_1 = 10 \)), albeit one which still allows for some confidence values that are less than 0.5.

The distributions in Figure 6C also provide insight into the relationship between metacognitive accuracy and optimal seeking in the postdecisional model. The highly separated distributions that result from low \( \tau_1 \)'s mean confidence is pushed outside of the thresholds on both ends of the confidence range. Mistakes will likely be accompanied by a strong error signal (very low confidence) that enables a change of mind without the need for additional information seeking. In turn, correct decisions will likely trigger confidences so high that no information seeking is deemed necessary either.

We can further investigate the relationship between objective accuracy, metacognitive efficiency and search by quantifying metacognitive efficiency in the postdecisional model as the ratio \( \sigma_I / \zeta_I \). Importantly, the ratio is always equal to or greater than 1, because the postdecisional stimulus set-up only allows for additional knowledge (metacognitive hyper-sensitivity) but not for reduced knowledge (metacognitive hypo-sensitivity). Figure 6A illustrates the effect of this hypersensitivity on search. We again observe a main effect of decision accuracy as governed by \( \sigma_I \) (differently shaded lines), but can now see the additional effect of metacognitive efficiency: Higher levels of metacognitive efficiency give rise to reduced search on average.

**Second-order model**

The rather diverse confidence distributions produced by the second-order model are visible in panels E-F of Figure 6. In E, we again hold the objective accuracy at \( \sigma_I = 2 \), but vary \( \tau_1 \). The rater noise importantly plays a different role in the second-order model, because the rater does not have direct access to the actor's cue \( X_I \). We again observe how low values of \( \tau_1 \) give rise to clearly distinct confidence distributions and increase the chance of successful error monitoring. However, in addition the second-order model also allows for metacognitive hypo-sensitivity, when \( \tau_1 < \sigma_I \). In these cases, the rater has less information than the actor. The consequences of this are visible when comparing the first-order plot with \( \sigma_I = 2 \) plots in panel B to the second-order plot with \( \sigma_I = 2 \) and \( \tau_1 = 3 \) in panel D: The second order model’s two distributions are less distinguishable than the first-order model because of the hypo-sensitivity produced by \( \tau_1 > \sigma_I \).

In general, the relationship between metacognitive accuracy and optimal average search holds in the second-order model. The increased levels of metacognitive insight resulting from lower values of \( \tau_1 \) push the confidence distributions out of the zone defined by the aforementioned seeking threshold. This results in a lowered propensity for search. Consequently, the effects of metacognitive hypersensitivity on seeking in the second-order model are comparable to those in the postdecisional model (see Figure 6F) – when keeping objective accuracy constant, higher efficiency again results in less need for additional information. The trend continues into metacognitive hypo-sensitivity. However, striking non-linear effects appear. These are again triggered by the specific knowledge states of the second-order actor and rater: Recall that if the actor is very accurate, and the rater has less knowledge, the agent’s confidence will begin to be relatively constant. In essence, the rater will begin to always trust the actor. This is visible in the leftmost plot in panel F, where \( \sigma_I = 1 \) and \( \tau_1 = 2 \). Here, the rater will know significantly less than the actor. The confidence ratings will thus closely congregate around the actor’s average accuracy, \( \phi(\sigma_I) = .84 \), which represents the rater’s best guess given its limited knowledge. This in turn will lead to most of the confidence ratings to be above the confidence threshold and reduce the average information seeking in comparison to a rater with more information. This effect of higher metacognitive efficiency is particularly visible in the panel E of figure 6: With the higher metacognitive efficiency of \( \tau_1 = 1.3 \) compared to \( \tau_1 = 2 \), there is more confidence mass within the seeking interval.

In summary, this means that search can be reduced in the second order model through two distinct mechanisms. When a second-order agent becomes more metacognitively hypsensitive, it will seek less because it has more information. However, counterintuitively, a second-order agent might also seek less when it is metacognitively hypo-sensitive, but this time because it has less, or to be more precise too little, information.

**Metacognitive accuracy in search**

Since search reflects metacognition, we can use the quality of search as a measure of the quality of metacognition that is formally distinct from the sort of metacognitive sensitivity measured by meta-d’ or other measures that target an agent’s confidence. Here, the quality of search is assessed by how likely the agent is to search on a trial when it was initially incorrect in comparison to when it was initially correct (see Figure 7A for illustration). Searching when already correct is, of course, a costly waste: Intuitively, an agent would like only to seek information on trials when their initial decision was incorrect, not paying the cost on occasions where search would simply affirm an initially correct decision. While such perfectly targeted seeking is of
Figure 7

**Sensitivity of search to accuracy** (A) When an agent is initially incorrect, it should ideally seek more information than when it was initially correct. We can use the difference between these two conditional search probabilities normalized with an agent’s average search as a measure of an agent’s search sensitivity. (B, C) Both in the postdecisional and in the second-order model, this sensitivity in metacognitive control is determined by the sensitivity in metacognitive monitoring, as indexed by $\tau_I$, and the actor’s accuracy $\sigma_I$. Plots use $\sigma_F = 1$, $r_S = -0.1$, and $\rho_I = .3$.

Intermediate summary: Confidence and search

In the preceding two sections, we discussed the intricate relationships between confidence and search. In both models, the threshold of confidence at which search is triggered is largely independent of the initial stimulus characteristics due to its (quasi-)Markovian property. Rather, the zone in confidence space where seeking is adequate is governed by the cost and precision of the additional information that can be collected.

The confidence thresholds are however crucial when considering the confidence distributions that fall within or outside of them. In both the postdecisional and the second-order models, metacognitive hypersensitivity shifts confidence outside of the seeking zone, reducing search. In the second-order model, metacogni-
tive hyposensitivity can trigger both increased and decreased search by either shifting more of the confidence distribution into or above the seeking zone.

Increased sensitivity in the confidence rating in turn enables an agent to target its search better to those decisions that were initially incorrect. This shows the key advantage that increased metacognitive sensitivity has for an agent, essentially enabling it to not waste its resources on ‘useless’ search. It also leads to a new measure of metacognitive quality, namely search sensitivity.

Final accuracy

The accuracy of its ultimate, overall judgement, $a_f$, constitutes a last crucial aspect of of an agent’s behaviour in our task. This final accuracy of course depends on the decision of the seeker – but in a potentially complex manner, because the seeker’s decision-making in turn is partially a function of its estimates of the benefits for this final accuracy of further search. We show these relationships in Figure 8A.

We distinguish between two kinds of final accuracy: The final accuracy when the agent decided not to seek additional information, and the final accuracy when the agent decided to do so. For brevity, we will in the following refer to these as the without-search accuracy and the with-search accuracy.

Postdecisional and first-order model

To understand how the final accuracy comes about, we need to consider what kind of cues are available to the actor at the final time point (see Figure 8A). Recall that in the postdecisional model, seeking is a function of $Z_I$: Ambiguous values of $Z_I$ (e.g. $Z_I = 0$) give rise to seeking, whereas extreme values of $Z_I$ (e.g. $Z_I = 3$) already come with sufficient confidence to make information search unnecessary. Thus, as seen in Figure 8A, with-search final decisions will be made with ambiguous, more intermediate, values of $Z_I$. In contrast, without-search final decisions will only be based on more extreme values.

This division of the $Z_I$ clearly impacts the final accuracy: In the without-search case, accuracy will be higher than would be expected when making a decision on the entire $Z_I$ space ($\phi(\zeta_I)$). This increase is triggered because the error-prone, intermediate $Z_I$ values are excluded (by the fact of seeking). In contrast, in the with-search case, the actor will have relatively poor information before receiving $X_F$ due to the ambiguity associated with these intermediate values. The results of this division are visible in Figure 8B where we show the two final accuracies as a function of the initial accuracy and the rater noise $\tau_f$.

Let us first inspect the with-search accuracy, pictured in red. Strikingly, this is not influenced by either the initial accuracy or $\tau_f$. The reasons for this lie in the aforementioned stimulus set-up and relate to the confidence threshold: Because the agent only has relatively ambiguous values $Z_I$ before it receives $X_F$, it will not have a strong preference for either option prior to search. This in turn means that the accuracy of the final stimulus $\sigma_f$ is the crucial determinant of the with-search accuracy. For example, in Figure 8B, $\sigma_f$ would afford an accuracy of around 84% ($\phi(\sigma_f) \approx 0.84$), and the without-search accuracy is only marginally higher. This slight boost over the accuracy afforded by a solitary $X_F$ is in fact governed by the cost, which if lower, increases the range of $Z_I$’s passed onto to seeking and therefore decreases the quality of information prior to the receipt of $X_F$.

In contrast, the relationship between initial accuracy and final without-search accuracy (plotted in blue) is linear – at least in the first-order model. That is the without-search accuracy is as accurate as the initial accuracy plus an additional boost. This again results from the $Z_I$’s available which tend to be less ambiguous when the agent does not seek. Introducing additional information through $\tau_f < \infty$ strongly modulates the without-search accuracy. This is again the case because in these cases, the agent makes its decision not based on a stimulus with $\sigma_f$ but on a combined stimulus with $\zeta_I$ which will always be more precise than $\sigma_f$. Low noise in the postdecisional information will thus considerably boost the final accuracy through the aforementioned capability for error monitoring.

Curiously, the model sometimes produces a behaviour where it makes worse decisions with additional information than without it. At first, this can appear implausible: information should serve to increase performance. However, the agent of course has to balance the gained accuracy with its cost, triggering it to seek out information only when it is not very confident that it has made a correct choice.

Second-order model

The limited informational access of the second-order model becomes especially crucial when investigating its final accuracy. Recall that the postdecisional seeker has access to $X_I$, and so already fully knows the quality of the final decision if it were not to seek. In contrast, the second-order seeker is less well informed. It only has access to the rater variable $Y_I$ and can make noisy inferences about the actor variable $X_I$ based on the actor’s decision. The seeker thus lacks the perfect insight afforded by the postdecisional model. As a result, seeking is only a function of $Y_I$ in the second-order model.
Figure 8

Final accuracy conditioned on search (A) Values of $Z_I$ that get passed on to the final accuracy with (red) and without seeking (blue) in the postdecisional model. When seeking, the $Z_I$ values tend to be ambiguous whereas when the agent decides against search, the values tend to be more extreme and therefore offer better accuracy. (B) As a result, in the postdecisional model, the final accuracy with seeking (red) is independent of the initial stimulus. The final accuracy without seeking (blue to purple) is governed by both the accuracy afforded through the initial decision as well as the extra information contained in the postdecisional cue $Y_I$. (C, D) Because the seeking decision is made without knowledge of the action variable $X_I$, the final accuracy differs in the second-order model. The final accuracy after seeking receives a boost through unambiguous values of $X_I$ that slip through. This fact also lets the final accuracy without seeking remain relatively stable until the agent doesn’t seek any information at all, at which point it becomes a function of the initial accuracy. B - D fix final stimulus noise and cost at $(\phi(\sigma_F = 1) = 0.84)$ and $r_S = -0.1$. Note the different scaling in the x- and y-axis for visibility.

rather than the full $Z_I$ in the postdecisional model (compare Figure 1, panels D and E). As a result, the stimulus space is not, as in the postdecisional model, divided along the crucial variable for the without-search accuracy ($Z_I$), but only along part of it, $Y_I$.

The key problem for the second-order seeker resulting from its limited access to $X_I$ is that potentially unambiguous $Z_I$ values can slip under its radar. As an example, picture the extreme case when the rater obtains a relatively ambiguous cue ($Y_I = 0.3$). Under most parameter combinations, this will result in low confidence and trigger the agent to seek. We can broadly think about two possible cases based on this: In one case, the actor itself might have received an ambiguous cue (e.g. $X_I = 0.1$). In this case, in the counterfactual scenario where the agent would not have searched, its final decision would have been based on a rather ambiguous $Z_I$. Here, seeking would have been a good decision. In another case, the actor might in fact have observed a very distinct cue ($X_I = 3$). In this case, the actor would have already had a rather unambiguous cue $Z_I$ for a final decision in the counterfactual non-seeking scenario. Here, seeking wouldn’t be of much benefit. Whereas the postdecisional agent would know this and thus not seek, the second-order rater has no access to $X_I$ and will thus sometimes sample even though it might not have been necessary.

This divergent knowledge results in a different pattern for the with-search accuracy, with influences for both $\sigma_I$ and $\tau_I$. The less relative insight the seeker has
(higher $\tau_l$) the more the $X_l$ "leakage", which is visible in panel C of Figure 8. Recall there that a constant level of $\tau_l$ results in decreasing metacognitive efficiency when increasing the objective accuracy. As a result, more $Z_l$’s which would lead to no seeking on the part of a postdecisional agent with full insight are assigned to seeking by the second-order model. This "unnecessary" seeking increases the with-search accuracy until the highly metacognitively inefficient agent stops seeking entirely, as was visible in the average seeking figures (Figures 3 D & E).

The $X_l$ leakage inherent in the second-order model’s seeking computations also affects the without-search final accuracy, but to its disadvantage. Specifically, the unnecessarily good values of $X_l$ included through the myopic seeking are now no longer available to the actor when the final decision is made without search. The final accuracy thus does not increase with increasing initial accuracy while the agent still seeks (again, compare Figure 3). In fact, under certain stimulus configurations, the without-search accuracy can even slightly decrease as a result of the good $X_l$ being "stolen" by the seeking. It also worth noting that the amount of additional decision information afforded by $X_l$ also decreases with heightened correlation $\rho$ reducing the additional information available.\(^7\)

The without-search final accuracy begins to fall into a linear relationship with the initial accuracy once the agent entirely stops seeking. It is then simply a function of $\xi_l (\phi(\zeta_l))$; by contrast with the postdecisional model where it is greater than $\phi(\zeta_l)$. Before then, the baseline without-search accuracy is governed by $\tau_l$, again as a result of the increasing capability for error monitoring that comes along with increasing metacognitive sensitivity. The ignition point of this increase is governed by the baseline rater noise $\tau_l$ which impacts when the seeker will stop seeking entirely.

As before, the main patterns remain intact when altering $\rho_l$ in the second-order model (compare panels C and D). However, some additional subtleties arise which we will show in the appendix (section C). Briefly, it is worth noting that the $X_l$ leakage is higher for the increased $\rho_l$ because in this case more information is sought and the average seeking curve resembles more of a step-function.

Interim Discussion: Normative metacognitive search

In the preceding theory sections, we approached metacognitive information search through a normative lens. In doing so, we examined the consequences of diverse aspects of confidence for how metacognitive agents should optimally elect to collect more information in a partially observable decision problem through the lens of more complex models of human metacognition. This work shows how computations that have thus far been solely conceptualized in metacognitive monitoring can give rise to a number of non-trivial downstream effects when extended to serve control purposes - for example when the seeker must trust or not trust the actor. On the whole, our work raises questions about how information is generally represented and used within metacognitive systems. We will discuss these in more depth in our main discussion.

More broadly, these results also show advantages that metacognition can have for agents over and above simple benefits arising from non-metacognitive objective accuracy. Specifically, we showed that, if used appropriately, good metacognition can be be harnessed to allow agents to search less and target their search better to incorrect trials. This is a very practical reason to have good metacognition, a virtue that is more often extolled than exhibited in investigations of confidence.

Empirical results

Our theoretical considerations raise the question as to whether general patterns of optimal metacognitive search hold in human data. We stress that here it is not our intention to use search behaviour to arbitrate between specific models of confidence. Such arbitration will require more targeted experiments, for example ones that independently vary aspects of confidence and action, or in the case of more extreme parameter combinations, rely on specific sub-populations. These are issues to which we will return in the discussion.

What we can do, however, is to investigate whether general patterns of metacognitive search that straddle postdecisional and second-order models are evident in human choices. To this end, we next turn to a reanalysis of data from a study that employed a task similar to the one we have discussed so far. Specifically L. Schulz et al. (2020) investigated the confidence-based information seeking decisions of 734 participants (see Appendix B for a description of the methods).

We compare patterns of participant choices with theoretical proposals derived from the second-order model. Given the overlap between the postdecisional and second-order models’ predictions, we choose the latter as a convenient framework within which to generate both hypo- and hypermetacognitive sensitivity (Fleming & Daw, 2017) as well as the the sort of over- and under-confidence that is often evident in human data (Fleming & Lau, 2014; Johnson & Fowler, 2011;\(^7\)How $\zeta_l$ stands in relationship with $\rho_l$, $\sigma_l$ and $\tau_l$ is in fact more complex under certain more extreme parameter combinations, as we discuss in further detail in the appendix C.)
Kruger & Dunning, 1999; Rouault, Seow, Gillan, & Fleming, 2018)

To introduce the latter point briefly: over- and under-confidence comes about when the average confidence that individuals report having about their decisions deviates from their objective accuracy. These dissociations from objective reality, also often termed metacognitive bias, are conceptually distinct from changes in the metacognitive sensitivity we discussed above. The separation between actor and rater in the second-order model makes it straightforward to create such biases, and thus easily lets us probe the consequences of metacognitive bias for otherwise normative search Fleming and Daw (2017). Specifically, miscalibration arises when the rater’s belief (σᵣ) about the actor’s accuracy is incorrect, so σᵣ ≠ σ. For intuition, imagine how one person might misestimate the skill of another.⁸

To preview our findings, we find key overlap between model predictions of patterns of metacognitive search and empirical data on both a within-subject and trial-by-trial basis, but also divergences. We begin by investigating trial-by-trial correlates of metacognitive search before investigating between-subject variation in participants’ markers of average behavior.

Metacognitive search from the trial by trial perspective

Effects of confidence and cost

Our theoretical results showed that, when the accuracy of the additional information available from search is held constant across trials, as is the case in L. Schulz et al. (2020), two factors should drive an agent’s likelihood to seek out information on a trial-by-trial basis: The cost of the additional information, and an agent’s momentary initial confidence.

We already briefly discussed the within-trial effect of confidence and cost on search when we introduced the computations underlying the information seeking choice. These are visible in the action values for or against search in Figure 2F. In slightly broader terms, these values simply mean the following: When the Q-value for seeking is larger than the Q-value for not seeking, an agent which seeks will, over the long run, receive more overall reward than an agent who does not seek. Because an optimal agent follows the policy that maximizes its long-term reward, it will adhere to this difference.

In actual behaviour, an agent is unlikely to follow a noise-free greedy policy, where it purely decides as in equation 15. Rather, behavior is typically found to be more consistent with a softmax policy where the underlying difference in Q-values is passed through a sigmoid function to determine the probability of seeking or not seeking. This function then returns choice probabilities conditioned on an agent’s cost and confidence. We plot such probabilities for an example second-order model and the two costs in L. Schulz et al. (2020) experiments in Figure 9A. This shows that an agent is more likely to seek on trials when confidence is most uncertain (i.e. around 50 % in the binary setting), and when the additional information is cheaper. We note that these curves are insensitive to participant-level parameters governing over- and under-confidence, and instead fully rest on across-trial variation in subjective confidence, rather than any environmental parameters or distributions.

The participant’s choice probabilities in L. Schulz et al. (2020) reflect these choice probabilities, as is visible in Figure 9E. Participants sought less information as they became more confident in their choices and as information became more costly. This is highlighted by the results of a trial-by-trial mixed-effects model which shows negative effects of cost and confidence on search (βconfidence = −2.36, p < 10⁻¹⁵; βcost = −1.60, p < 10⁻¹⁵).⁹

Taken together, these results highlight qualitative features of how humans solve the underlying decision problem, taking into account both the cost and confidence. Similar trial-by-trial effects of confidence are also visible in other data, for example Desender et al. (2018) and Pescetelli, Hauperich, and Yeung (2021) although neither of these studies varied the cost of seeking.

Metacognitive search sensitivity

As we have seen, the trial-by-trial modulation of search by confidence enables an agent to adaptively search more when it was initially mistaken than when it initially made a correct choice. This is the case in our data. Participants generally sought less information when they were initially incorrect (Figure 9E, β =

⁸Disconnection between what a rater thinks about an actor is of course not limited to the second-order model, and might similarly be implemented in other models, like the postdecisional model. Nevertheless, in keeping with Fleming and Daw (2017), we will limit ourselves to the second-order model here. We also note that this form of miscalibration theoretically introduces additional noise in the rating and decision making process because cues are not optimally integrated any more. This can be expected to have a modest negative impact on metacognitive sensitivity.

⁹To better analyze how search was modulated, we only included participants who sought information on between 5 \% and 95 \% of trials (N = 568) in our analysis of trial-by-trial results. For the analysis of the task averages reported below, all participants are included.
Second-order model captures across-trial effects of metacognitive search (A) Trial-by-trial probabilities of search as a function of confidence in an example second-order agent ($\sigma_I = \tau_I = 1.25, \rho_I = .5, \sigma_F = 1, r_F = 100$). (B-D) Association of trial-by-trial accuracy and search. (B) Simulated agents seek more information when they are initially wrong, even when we introduce variability in over- and underconfidence (see appendix B for details on parameters). The difference of the search averages conditioned on accuracy, normalized by the overall average level of information seeking defines a measure of metacognitive search sensitivity. (C) In optimal agents, this measure is always positive, and (D) correlates with an agent’s metacognitive sensitivity, as indexed through meta-$d'$. (E-H) Participants in L. Schulz et al. (2020) show similar patterns. (E) On a trial-by-trial basis, they seek less information the more confident they become and the more costly the information is. (F) They also seek less information when they were initially incorrect, and (G) mostly have positive search sensitivities. (H) Participants’ search sensitivity correlates with their meta-$d'$.

Indeed this was the case for the great majority of participants as is visible in the distributions of search sensitivities displayed in Figure (9G) (89.3 % of participants have a positive value).

Crucially, our theoretical results showed how higher sensitivity in metacognitive monitoring should give rise to increased sensitivity in participant’s search. To test this in our participants, we correlated their search sensitivities with meta-$d'$ fit to their confidence ratings. Participants who had more sensitive confidence ratings as indexed by greater meta-$d'$ values also had greater search sensitivities ($R = 0.40, p < 10^{-15}$). This result shows that participants with higher metacognitive sensitivity also were more targeted in their search behaviour, spending fewer resources on unnecessary trials.

As we will discuss in more detail below, our participants showed considerable variation in their degrees of over- and under-confidence. Thus, we wanted to make sure that our theoretical models still also produced similar patterns when we introduced agents that were more or less confident than their accuracy should have licensed. To this end, we simulated second-order agents with randomly sampled combinations of actor parameters.
and overconfident agents before comparing these results to behaviour of the participants from L. Schulz et al. (2020).

**Theoretical effects of over- and under-confidence**

Figure 10A shows the distributions of objective decision accuracies and mean confidence ratings produced by the over- and under-confident second-order agents we simulated for our trial-by-trial analysis above. In this figure, we see both overconfident agents, meaning that the agent’s average confidence is higher than their objective accuracy (in purple, stemming from $\sigma_R^I < \sigma_I$), as well as under-confident agents (in orange, $\sigma_R^I > \sigma_I$).

We observe two effects of this miscalibration on average search. Whereas, thus far, objectively higher accuracy has gone hand in hand with lower search in our normative and well-calibrated models, this accuracy-search relationship becomes broken once accuracy and confidence are uncorrelated (Figure 10B). This is because overconfident agents (purple in the figure) sample less than their well-calibrated or under-confident peers, believing that their initial choice is already good enough. Under-confident agents (orange in the figure) do the opposite. When confidence and accuracy are dissociated, these two biases balance each other out, breaking the relationship between accuracy and search.

In contrast to this broken search-accuracy relationship, our simulations show that the agents’ subjective accuracies (their average confidence) still remain a key determinant for search (Figure 10C). Agents who are on average more confident display lower information search than those agents who are on average less confident. This is because in our model the rater’s subjective assessment of the actor’s initial decision quality (rather than objective signal quality) feeds into the seeker’s computations. Thus, if the rater/seeker believes an actor to be a good decision-maker, it will still search less, even when this faith in the actor’s accuracy is misplaced.

Finally, our results show that simulated agents who had higher meta-\(\text{d}'\) also sought less information, just as they would have if their confidence was not biased. In brief, this is because one rater by itself can still have better information than another (smaller $\tau_I$), even if those two might be mistaken about the true accuracy of the actor. We also note that the influence of over- and under-confidence is reduced when agents have high meta-\(\text{d}'\). This is because these agents generally seek little additional information because of the high quality information the rater already possesses.

Summarizing these results, we show the individual contributions of average confidence, accuracy and metacognitive sensitivity to average search in a mul-

---

**(Intermediate discussion: Trial-by-trial empirical search)**

In summary, we find substantial overlap between theoretical prescriptions of trial-by-trial metacognitive search and participant data in L. Schulz et al. (2020). Higher confidence and cost led to reduced search, and participants sought less information when they made a correct rather than incorrect choice. Finally, we find that sensitivity in metacognitive monitoring is related to search that is better calibrated to objective accuracy. This shows that humans are at least partially able to carry forward and adaptively employ their metacognitive abilities in the control of future-oriented behaviour.

**(Relationship of task averages)**

Following on from these trial-by-trial analyses, we next probed how key task averages were related to each other. To do so, we first investigated this in the context of our simulated optimally seeking, but under-
Figure 10

Over-/under-confidence and relationships between task averages in model and data. (A) We can simulate over- and under-confidence through dissociating subjective and objective actor noise $\sigma_I$. (B) This dissociation breaks the link between average accuracy and average information seeking. However, (C) average confidence is still inversely correlated with the degree to which an agent searches, as is (D) metacognitive sensitivity. (E) Regression showing the effects of the variables in (B-D) on search. (F) In contrast to model predictions, more accurate participants from L. Schulz et al. (2020) were also more likely to seek more information. However, following the model predictions, participants who were more confident sought less information. There was no significant effect of metacognitive sensitivity, as indexed by meta-$d'$, on average search.

Multiple regression whose results we plot in Figure 10E (standardized betas). Over and above our previous theoretical results, this highlights how confidence is a critical driver of information search, an effect which is particularly apparent when it begins to dissociate from accuracy. This is in contrast to the optimal models we have investigated before where average confidence and average accuracy perfectly correlate, and so average accuracy will appear to predict average search.

Relationship between empirical task averages and model simulations

As in our simulated population of agents, participants’ average confidence ratings did not strongly correlate with their average accuracy ($R = 0.07, p = 0.06$). Such a disconnect allowed us to ask how each of these features of average behaviour predicted search, and whether such relationships were aligned with our normative models. To test this, we first probed how participants’ average confidence, average accuracy, and metacognitive sensitivity predicted their average propensity to seek out information through a correlational analysis. In key agreement with the crucial role of confidence in information seeking, participants who were more confident also sought out less information ($R = -.27, p < 10^{-12}$). However, in a curious disagreement with the model’s behaviour, more accurate participants sought out more, rather than less, information ($R = .19, p < 10^{-6}$). We also found no significant correlation between participants’ level of metacognitive accuracy, as indexed through meta-$d'$, and their average search ($R = 0.005, p = 0.89$).

These results are also visible in a multiple regression analysis where we predicted average search with these three predictors after normalizing all variables (see Figure 10F). This revealed a positive effect of average accuracy on average search ($\beta = 0.19, p < 10^{-6}$), although the positive effect of accuracy was smaller than the negative effect of average confidence ($\beta = -.32, p < 10^{-14}$). Finally, the regressor for meta-$d'$ was not significant.
Intermediate discussion: Over- and under-confidence

Why might participants in L. Schulz et al. (2020) have searched as they did? First, in regard to the effect of average accuracy, one should note that L. Schulz et al. (2020) did not causally manipulate participant’s accuracy and rather relied on one staircased performance level. This in contrast to, for example, Desender et al. (2018) who, within each participant, varied the difficulty levels in a similar information seeking task. In this within-participant case, participant sought more information in conditions with lower accuracy showing that individuals can generally be sensitive to their decision accuracy in the direction predicted by the model. Indeed, this is related to our own within-participant finding that participants seek more on incorrect in comparison to correct trials.

However it remains unclear how we can account for the small positive effect of average accuracy on search. One possible explanation rests in the way the second stimulus strength was determined in L. Schulz et al. (2020). Specifically, the second stimulus strength was yoked to the initial stimulus strength, which was itself determined by a staircase. As a result, participants who saw a relatively easier initial stimulus were also more likely to see an easier second stimulus. Consequently, participants with higher initial accuracy were able to reach close to perfect average levels of final decision accuracy when they sought information (see Figure B2 for visualization). This possible certainty might have been a ‘bright line’ that gave participants an extra incentive to seek out additional information and might thus have biased the average effect of accuracy. This would have been especially relevant if participants, as we will consider in the main discussion section, had non-linear utility functions. This can be examined in more detail in further work.

Additionally, the second-order model, even when miscalibrated, cannot account for the lack of relationship between metacognitive sensitivity and search exhibited by the participants in L. Schulz et al. (2020). We consider various possibilities in the main discussion session, including the case that the seeker’s belief (\(r_s\)) about the key source of the rater’s confidence is miscalibrated in the same way that the rater’s belief about the actor’s accuracy can be. It would be interesting to extend the sort of methods that Desender et al. (2018) adopted in order to manipulate explicitly the factors that we suggest should drive search.

Further work will also have to investigate why meta-d’ does not correlate with average search but correlates with search sensitivity, as we showed in an earlier section. Recall that the difference between these measures is that average search represents a general propensity to seek whereas search sensitivity is a trial-by-trial measure of how well each trial’s search decision is in line with each trial’s objective accuracy. This difference in correlation might arise because the two measures capture different characteristics of metacognition, similar to the separation between metacognitive bias and sensitivity in confidence ratings (Fleming & Lau, 2014). Average search, like metacognitive bias, is a general measure of an agent’s tendency to seek information, similar to an agent’s tendency to have a specific confidence level. The similarity between these two forms of ‘bias’ is apparent in the inverse correlation between average confidence and average search. As in measures of average confidence, average search can be both closer to and farther away from optimality. However, we note while the average confidence should, across models, be simply equivalent to the average accuracy, optimal average search is additionally subject to forces outside of the average accuracy as we have explored throughout this paper.

In contrast to these average measures, search sensitivity is more a question of resolution – how well search distinguishes between correct and incorrect trials. This fact is apparent in the correlation between search sensitivity and meta-d’, a measure of the trial-by-trial accuracy of confidence ratings. Further work will have to provide a more comprehensive treatment of how these concepts can be optimally measured and how they interact.

Most importantly, however, we still find a prominent negative effect of average confidence, highlighting the key role that the subjective feeling of correctness play in our participants’ average information search, and in line with the predictions of a miscalibrated second-order model. More broadly these results underline the primary role of confidence in search over and above accuracy.

Apart from L. Schulz et al. (2020), a further interesting link between our theoretical analysis of over- and under-confidence in search and empirical data can be drawn with the results of Desender et al. (2018). As briefly outlined in the introduction, this study used the perceptual positive evidence effect to investigate causal links between confidence and information search in a similar perceptual setting. The positive evidence effect refers to experimental manipulations that can boost a participant’s average confidence while leaving their objective accuracy untouched (Boldt, de Gardelle, & Yeung, 2017; Peters et al., 2017; Zylberberg, Barttfeld, & Sigman, 2012). In essence, it creates two levels of metacognitive bias within a single participant. In our

\[ \beta = 0.08, p = 0.07. \]
model, this positive evidence effect could be conceptualized as biasing the rater’s/seeker’s appraisal of \( \sigma_t \), mistakenly increasing the rater’s belief in the actor’s accuracy. As we have demonstrated, this condition would lead an agent to decrease its information seeking. Indeed, Desender et al. (2018) show that this is the case: In the condition in which the positive evidence effect induced higher confidence, participants were less likely to seek information.

This intuition also applies beyond the perceptual domain. For example, Metcalfe and Finn (2008) manipulated participant’s confidence in their memory of words through an elegant order-effect manipulation which induced higher and lower levels of confidence. Similarly to Desender et al. (2018), participants sought less information in the condition with lower confidence. We could again conceptualize this as a bias on the rater’s/seeker’s part.

### Discussion

Computational models of metacognition have recently been highly successful in explaining many intricate facets of human confidence (Fleming & Daw, 2017; Rahnev et al., 2020; Yeung & Summerfield, 2012). However, it has long been noted that metacognitive monitoring exists to guide subsequent control of behaviour (Nelson & Narens, 1990), such as knowing when to invest time and effort in studying new material or seeking new information (Desender et al., 2018; Goupil, Romand-Monnier, & Kouider, 2016; Metcalfe & Finn, 2008; Pescetelli et al., 2021; L. Schulz et al., 2020). How these two processes of monitoring and control interface has attracted less attention from computational modelers. Here, we considered the rather diverse consequences that different assumptions about the informational structure underlying confidence have for optimal search and how metacognitive search manifests in human behavior. We did so by treating the process of remunerated inference and costly information acquisition in the face of uncertainty as a simple instance of a partially observable Markov decision problem (POMDP).

We extended model architectures suggested by Fleming and Daw (2017), exploiting the simplified version of drift diffusion-like decision making discussed by Dayan and Daw (2008). In the postdecisional models, the rating process that generates confidence judgements has access to at least the information underlying the original decision whose confidence it judges, as well as in most cases additional information (Moran et al., 2015; Navajas et al., 2016; Pleskac & Busemeyer, 2010). By contrast, in the second-order model, rater and actor share only part of each other’s information (Fleming & Daw, 2017; Jang et al., 2012). In our extension, this confidence is used to determine whether the agent should, depending on the expense of doing so, collect more information before gaining reward for a final choice.

Our theoretical results highlight how seemingly small changes in the informational architecture of acting, rating and seeking can lead to diverse profiles of what constitutes optimal search under these assumptions. The second-order model in particular contains a number of non-trivial and often non-linear relationships between action, confidence, and optimal information search. For example, the average normative willingness to search as a function of objective accuracy can resemble a step-function for some parameter values in this model. In addition, because of the specific distributions of confidence associated with the second-order models, metacognitive hypo-sensitivity can give rise to both increased or decreased information search under some more extreme parameter combinations, depending on the underlying objective accuracy.\(^1\)

Our empirical results reinforce the importance of carefully considering metacognition in information search. While, as we will discuss later, more targeted paradigms will be necessary to make statements about confidence models from search behavior, we show how the quality of metacognitive monitoring is intricately linked to the quality of search.

Here, we did not focus on the potential neural realization of the seeker, and its interaction with the likely regions involved in acting and rating (Fleming et al., 2018; Shimamura & Squire, 1986; Vaccaro & Fleming, 2018) as well as the neuromodulators involved in information search and the representation of uncertainty (Hauser, Moutoussis, Purg, Dayan, & Dolan, 2018; Vellani, de Vries, Gaule, & Sharot, 2020; Yu & Dayan, 2005). It would be interesting to probe the most obvious substrates, such as those regions involved in model-based and goal-directed control (Daw, Niv, & Dayan, 2005; Dickinson & Balleine, 2002) or state inference (Behrens

---

\(^1\)In fact, even the basic confidence judgements produced by the second-order model can have counter-intuitive characteristics in certain regimes – such as that the more the rater’s private information contradicts the actor’s choice, the more confident the rater can be that the actor’s decision was in fact correct. We mainly focused on regimes in which predictions are less unusual, in keeping with the likely psychological unreality of these extremes. However, we point the interested reader to the fuller picture in appendix (section C). It would furthermore be interesting to consider what kinds of consequences arise from the modest noise added to the system through the over- and underconfidence miscalibration in the second-order framework, and how this would manifest in search.
et al., 2018; Schuck, Cai, Wilson, & Niv, 2016), using signatures derived from behaviour as potential correlates of neural activity.

**Informational flow and access**

Our extensions to the postdecisional and second-order model make particular choices about how information flows after the confidence rating. That is, how is the new information $X_f$, if collected by the actor, integrated with the actor’s ($X$) and rater’s ($Y$) original information to make the final choice ($a_f$)?

In our formulation of the postdecisional model, the perfectly accumulating sequential sampling renders unreasonable anything short of the full integration of the three samples ($X_t, Y_t, X_f$). The optimal computations would naturally be altered if the accumulation were lossy or affected by noise, or the rater had less knowledge about the actor.

In contrast to the full access afforded by our postdecisional account, in the second-order model, the initial actor and rater are more separate. This in turn leaves various credible possibilities for their subsequent integration. We endowed the final actor with the substantial inferential ability of calculating the rater’s variable $Y_t$ from the reported confidence. However, especially if the rater is not required to report this information “publicly”, this may not be possible. If the information about $Y_t$ available to the final actor is less than we assume here, then the computations for search would differ, for instance limiting the benefit of low rater noise $\tau_f$.

A related question is whether and how information is further propagated in the second-order model. Here, we stopped at the final action, but the rater could, of course, also compute its confidence in this second decision. For brevity, we have not included this here, but note how an optimal rater would now need to infer both the actor’s first and second cue based on the dynamics of the first and second decision. Even in the case of no search, the rater could, in some cases, receive additional information about the actor’s first cue by observing how the actor reacts to the rater’s initial confidence (for example whether the actor changes its mind after an error signal by the rater). This raises broader issues about an internal recursive back- and forth inference between the actor and rater.

In contrast to humans, other animals’ metacognition cannot be directly assayed with confidence ratings. Experimentalists have attempted to remedy this through paradigms that indirectly probe representations of subjective correctness, such as post-decision wagering (Kepecs & Mainen, 2012), opt-out experiments (Hampton, 2001) or neural markers (Kepecs et al., 2008; Kiani & Shadlen, 2009; Nieder, Wagener, & Rinnert, 2020). Information-seeking tasks have also seen wide use (Call, 2010; Call & Carpenter, 2001). There, an animal is hypothesized to possess a form of metacognition if it seeks information in situations of uncertainty (which the experimenter controls), a behavior that already develops in human infancy (Goupil et al., 2016).

However, there is ambiguity about whether confidence-related behaviours and information search in animals reflect a capacity for explicit metacognition – the ability to form a distinct representation of confidence about one’s knowledge or performance (Birch, Schnell, & Clayton, 2020; Carruthers, 2008; Kornell, 2014). To the extent that second-order architectures map onto a richer capacity for creating and using explicit confidence representations, our computational models could allow inferences about the varieties of animal metacognition when applied to the kinds of tasks used in this domain.

**Ambiguity, computational noise, uncertainty and normativity**

The focus of our theoretical investigations was to highlight the role different assumed metacognitive architectures have on optimal information search. This shows how strongly even optimal metacognitive computations can affect seeking behaviour. However, various other factors can influence search and might be crucial in explaining some of the more idiosyncratic average results.

Following Fleming and Daw (2017), the agents in our POMDP have first-order uncertainty about the stimulus on a trial, but suffer no ambiguity (or second-order uncertainty) about the inaccuracy or correlation of their sources of information. The case in which subjects receive information whose accuracy they are uncertain about is common in dynamic decision-making problems where, for instance, the contrast of input stimuli may change in an unsignalled manner between trials (Fleming et al., 2018; Gold & Shadlen, 2001, 2007; Kiani & Shadlen, 2009). There has been work on this in the equivalent of the first-order case. For instance, the conventional reward-rate maximizing strategy for the drift diffusion decision-making model in which evidence accumulates up to a fixed threshold changes to one involving what is known as an urgency function (O’Connell, Shadlen, Wong-Lin, & Kelly, 2018; Ratcliff, Smith, Brown, & McKoon, 2016). In such a model, if the agent discovers from the length of time it is taking to reach the threshold that the information it is receiving is not very accurate, it can make a quick, potentially inaccurate, decision, and hope that the next problem will be easier (Drugowitsch, Moreno-Bote, Church-
It would be possible to extend our models in a similar manner, allowing separate informational accumulations over time for actor and rater; with the seeker judging when to stop and allow the actor to perform. The added complexity would be that the explicit communication between actor and rater that we allowed (with the actor’s first action \(a_t\) observed by the rater; and the rater’s confidence report \(c_j\) being observed by the second actor) would have to be adjusted.

Our models focused on noise coming from the signals themselves, and so we assumed entirely noise-free decision and confidence processes in our theoretical results and simulations. This allowed us to pinpoint the influences of different confidence models on search. However, it is of course also limiting. Decision noise is ubiquitous in behaviour (Mueller & Weidemann, 2008; Wilson et al., 2014), and noisy computations offer a different lens for understanding metacognitive inefficiencies (Shekhar & Rahnev, 2020) and exploration (Findling, Skvortsova, Dromnelle, Palminteri, & Wyart, 2019). It is particularly interesting whether agents might take into account such noise processes. For example, in contrast to our noise-free agent, consider an agent which would have the opportunity to collect very accurate information, but has difficulty translating this information to good actions. Here, the sources of such noise are not crucial and might include forgetting that degrades information over time, lossy accumulation of evidence or noisy computations and action selections. In turn, agents that know about their own inaccuracies should issue confidence judgements and seeking decisions that take them into account. For instance, Moutoussis, Bentall, El-Deredy, and Dayan (2011) hypothesized that people with paranoia appear to “jump to conclusions” by refusing to gather information, because decision noise renders such collection futile (although see Ermakova et al., 2018).

Throughout, we modelled the information-seeking choice as being made deterministically based on a threshold. As briefly discussed in our data section, other reinforcement learning scenarios would often assume a softmax- or add a lapse-parameter to this choice, making it stochastic. Both would naturally reduce the prevalence of extreme cases outlined in the theoretical results where the agent currently seeks on all or none of the trials.

A more complicated problem arises if agents are confused about, or even not fully aware of, their own metacognitive skill. Throughout, we assumed a form of well-calibrated meta-metacognitive knowledge, in that the rater knew the exact value of its variance \(\tau_j\). While humans are indeed able to use the uncertainty of internal or external stimuli and sensations in their decision-making (Körding & Wolpert, 2004; Whitney, Rinehart, & Hinson, 2008), these stimuli are what we would consider first-order. Whether such capabilities extend to our metacognitive sense is itself questionable – and issues about how agents tune this capacity, and its psychological and neural realizations have yet to be thoroughly examined. For example, agents might exist that are metacognitively highly accurate, but might be unaware of this skill, or have low confidence in it. Conversely, individuals might possess little metacognitive skill, but could consider themselves to be great raters, in essence a meta-Dunning-Kruger effect (Kruger & Dunning, 1999). If agents do not know whether they can trust their own confidence, this naturally also has implications for our information-seeking problem, and metacognitive control more broadly.

Apart from these inference issues, human decision-making based on inferred uncertain instrumental values is also widely known to be subject to distortions (Hertwig, Barron, Weber, & Erev, 2004; Tversky & Kahneman, 1992). Such distortions include temporal-discounting or risk-aversion, and have been shown to influence search (Gigerenzer & Garcia-Retamero, 2017; Sadeghiyeh et al., 2020). In general, our value-based approach allows for the integration of such effects, for example by applying a discount rate to the second decision, assuming a non-linear value function or even a more complex coherent risk measure (Gagne & Dayan, 2022). Interindividual differences in such parameters might exert a significant effect on average information seeking and might be key in explaining some of the unexpected characteristics of our average search results, especially if factors like risk-sensitivity or discounting are non-trivially related to factors governing metacognition and objective accuracy (Fleming & Dolan, 2010). We note however that our model is already rich in parameters and so dissociating individual parameter contributions might present a challenge.

Beyond our current minimal task, valence and motivational effects impact information search over and above the purely instrumental and accuracy-focused seeking we discuss. Prominently, humans are more likely to look for information that has positive valence (Gesiarz, Cahill, & Sharot, 2019; Hart et al., 2009; Jonas, Schulz-Hardt, Frey, & Thelen, 2001; Sharot & Sunstein, 2020). In turn, we tend to be reluctant to seek information that might have negative valence, but might in fact be instrumentally useful - like the results of a medical test (Gigerenzer & Garcia-Retamero, 2017; Thornton, 2008). Our models do not accommodate these aspects at the moment. However, one might combine our purely instrumental values with internal values for
certain beliefs – which may or may not be in line with the accuracy goals we specify (Bénabou & Tirole, 2016; Bromberg-Martin & Sharot, 2020).

Even when there is no valence attached to the beliefs, empirical work in paradigms close to the one we use here suggest that humans integrate cues that favour an initial judgement more than those that disconfirm it, especially when confidence is high (Rollwage et al., 2020). Such a confirmation bias can be straightforwardly modelled within our framework and might assist in explaining behaviour (Fleming et al., 2018). Recent simulation work (Rollwage & Fleming, 2021) has shown that this apparent confidence-induced confirmation bias can in fact be adaptive when an agent possesses second-order metacognitive hypersensitivity. Notably, Rollwage and Fleming (2021) used a different information flow for the final decision. However, this still raises interesting questions about what constitutes optimality in both the passive and active sampling of information.

Also weakening the tie to normativity are recent empirical findings that human confidence based on choices with more than two options does not necessarily resemble the full Bayesian posterior, but rather tracks the difference between the two most likely options (Li & Ma, 2020). This has interesting implications for more complex choices, and it will be important to consider how search manifests in these settings.

Future steps in dissociating action, confidence and search

While our reanalysis of L. Schulz et al. (2020) provided insights into the empirical use of metacognition when seeking, it left various open questions. As we previewed, we believe more targeted paradigms and experimental manipulations will be necessary to better disentangle the role of metacognition and its informational architecture in search. Among them, transcranial magnetic stimulation (TMS) (Fleming et al., 2015; Rounis, Maniscalco, Rothwell, Passingham, & Lau, 2010; Shekhar & Rahnev, 2018) or pharmacological manipulations (Clos, Bunzeck, & Sommer, 2019) are able to create dissociable effects on action and confidence. Metacognition can also be trained (Carpenter et al., 2019) and there are task conditions which selectively impact decision and confidence quality like the positive evidence bias we discussed above. (Bona & Silvanto, 2014; Desender et al., 2018; Graziano & Sigman, 2009; Spence, Dux, & Arnold, 2016; Vlassova, Donkin, & Pearson, 2014). Investigating how search would manifest following such manipulations might provide key insights into the interplay of metacognitive monitoring and control and their underlying computations, especially with more of a focus on within-subject effects.

Furthermore, some neurological (Del Cul, Dehaene, Reyes, Bravo, & Slachevsky, 2009; Fleming, Ryu, Golfinos, & Blackmon, 2014; Goldstein et al., 2009; Persaud, McLeod, & Cowey, 2007; Shimamura & Squire, 1986) and psychological disorders (David et al., 2012; Hoven et al., 2019; Rouault et al., 2018) as well as aging (Palmer, David, & Fleming, 2014; Weil et al., 2013) specifically affect an individual’s metacognition but leave their “object-level” abilities relatively untouched. These would have implications for information search. For instance, agents might over- or under-estimate the usefulness of the second cue or have higher thresholds for stopping to seek, or for returning to check that some action (such as turning off a gas stove) has been completed (Hauser et al., 2017; Tolin et al., 2003).

Experimental manipulations of or individual differences in metacognition might provide one way to get a better understanding of metacognitive search. Particularly, there are other specific aspects of second-order computations that warrant further investigation. Most prominently, computing second-order confidence relies on observing the actor’s decision (Fleming & Daw, 2017) and its insight is curtailed when it cannot do so, a corollary also supported by empirical evidence (Pereira et al., 2020; Siedlecka, Paulewicz, & Wierzchoń, 2016). Future experiments could follow up on this by varying whether participants make an initial decision. When participants only rate their confidence but do not perform an action, this should lead to reduced metacognitive insight, and the optimal seeking computations would be more akin to a first-order model. It would be especially interesting whether such conditions could give rise to the step-like average seeking curves when varying the underlying object-level accuracy.

The aim of this paper was not to make strong statements about the veracity of one model versus another. For simplicity, we presented a relatively limited post-decisional model that in its current form is unable to capture the sorts of hypo-sensitivity observed in the empirical data, but could do so with relatively simple changes (e.g., the addition of confidence noise). In contrast, we think that qualitative effects of action on confidence provide a more promising route for clean tests of the role of second-order inference – treating our own actions as “data” – on a confidence computation. Recent empirical studies have begun to identify the aforementioned action-specific contributions to perceptual confidence, which future modeling studies could harness (also in the context of information seeking) to offer a more precise test of different model architectures (Fleming et al., 2015; Pereira et al., 2020; Siedlecka et al.,...
For more definitive comparisons between individual confidence models, further research could also follow the lead of Shekhar and Rahnev (2022), who compared a host of metacognitive monitoring models in terms of their fit to a large existing database of confidence ratings (Rahnev et al., 2020). Richer tasks, such as those involving three or more possible choices, together with information search, may also aid in discriminating between different computational architectures (Li & Ma, 2020; Rahnev et al., 2022). Overall, we believe that novel and bespoke experimental designs tailored to test specific model predictions will be the most fruitful avenue for testing models of metacognition more broadly, such as those aimed at evaluating the tell-tale effect of self-action on confidence in the second-order model.

Links to other types of information search and metacognitive control

Here we addressed a very restricted information-seeking problem. In other laboratory tasks or in real world situations, information seeking is itself often embedded in more complex decision-making tasks (Mobbs, Trimmer, Blumstein, & Dayan, 2018; E. Schulz et al., 2019). For example, in reinforcement learning problems with several options of unknown value, agents face an exploration-exploitation dilemma (E. Schulz & Gershman, 2019; Sutton & Barto, 2018). Theoretical treatments of (optimal) exploration (Gittins, 1979; Schwartenbeck et al., 2019; Sutton & Barto, 2018) and empirical investigations (Boldt et al., 2019; Speekenbrink & Konstantinidis, 2015; Wilson et al., 2014; Wu et al., 2018) of human exploration also highlight the key role of uncertainty in this decision problem.

At first sight, more sequential tasks might seem far removed from the setting we discussed. However, along with tasks that not only ask whether to sample information but also where to sample information from, tasks with longer horizons in fact share the same computational problem as the simple task we focus on here. For example, belief states in the exploration-exploitation problem are over action-reward contingencies rather than ‘world’ states, and the penalty arises as an opportunity cost. Alternatively, an agent faced with a similar task as ours but with several possible information sources whose quality is unclear will have a belief state that quantifies both the uncertainty about the state as well as the quality of the sources (Pescetelli & Yeung, 2020a).

Optimal solutions to both our reduced task as well as those more elaborate tasks rest on planning (Callaway et al., 2021; Hunt et al., 2021). Essentially, our agent plans one step ahead, considering all possible stimuli associated with its one source to compute its action values (e.g. equation 14). In the richer tasks, planning extends over several steps and/or sources, but the key idea remains equivalent. Of note is that optimal solutions to these more complex planning problems quickly become intractable. Any heuristic solution, however, will try to approximate this optimal solution, or at least be measured against it.

Models of exploration behaviour or of more complex search almost always consider uncertainty in what we would characterise a first-order computation – at most wondering about the effect of different prior distributions over unknown quantities. It would be interesting to think about the equivalent of postdecisional and second-order models – where agents could gain some extra, partially independent, information about the quality of their actions, for instance by observing other agents (Zhang & Gläscher, 2020). It might then be possible to use the sort of methods we have discussed to draw out the implications for exploration.

On a shorter timescale, the basic computations we discussed line up with those performed in drift-diffusion models. There, participants can infer their chance of being correct from information that accumulates over time, and have to decide whether to stop or continue sampling evidence (Gold & Shadlen, 2001, 2007; Ratcliff & Rouder, 1998; Wald, 1949). These models have been highly successful in explaining the latent speed-accuracy trade-off present in many perceptual tasks where participants decide implicitly to sample information (Bogacz, Wagenmakers, Forstmann, & Nieuwenhuis, 2010; Ratcliff et al., 2016), and it would be interesting to see how they are sensitive to more metacognitive architectures.

Outside of areas related to information acquisition, confidence also plays a key role in controlling other processes. For example, cognitive offloading (Gilbert et al., 2020; Hu, Luo, & Fleming, 2019; Risko & Gilbert, 2016), such as setting reminders, is closely tied to our subjective feeling of future success. Humans also prioritize the completion of different tasks as a function of their confidence (Aguilar-Lleyda, Lemarchand, & de Gardelle, 2020) and use confidence to decide adaptively when to deploy attention (Desender, Boldt, Verguts, & Donner, 2019; van den Berg et al., 2016) or engage in reflection about the value of options (Lee & Daunizeau, 2021). All these decision problems essentially boil down to a planning problem akin to our information-seeking case where an agent needs to balance the benefits of an action like setting a reminder with its (opportunity) cost. It would be furthermore interesting to consider parallels to metacognitive search.
sensitivity in these areas, for example asking whether a reminder was set in vein or not.

On a longer time horizon, confidence also shapes learning (Bjork, Dunlosky, & Kornell, 2013; Metcalfe & Finn, 2008). Here, computational modelling has shown, that, on the one hand, we learn from our local confidence about our own broader skills (Rouault, Dayan, & Fleming, 2019). On the other, we use momentary estimates of uncertainty to steer how much we learn from errors (Behrens, Woolrich, Walton, & Rushworth, 2007; McGuire, Nassar, Gold, & Kable, 2014; Meyniel, Schlunegeger, & Dehaene, 2015; Purcell et al., 2010; Vaghi et al., 2017). Investigating the effects of confidence on learning and on controlling future courses of actions through a more detailed and integrated model of metacognitive monitoring and control might provide insights into both their function and dysfunction.

Whether in our paradigm or in exploration-exploitation, the collection of information serves to increase an agent’s reward and thus has a direct instrumental purpose. However, there is also a large literature dealing with what at first glance appears to be non-instrumental information seeking. Such “curiosity” for seemingly (at least currently) reward-irrelevant information has long been a puzzle to experimentalists and theoreticians (Gottlieb & Oudeyer, 2018; Igaya, Story, Kurth-Nelson, Dolan, & Dayan, 2016; Kidd & Hayden, 2015; Kobayashi, Ravaioli, Baranè, Woodford, & Gottlieb, 2019). As in instrumental information search, confidence often plays a key role in the treatment of such behaviour, although its role is contested. Whereas some propose a monotonic relationship between confidence and curiosity similar to our instrumental results (Berlyne, 1950; Lehman & Stanley, 2011), others argue that intermediate levels of confidence are most conducive to curiosity (Baranes, Oudeyer, & Gottlieb, 2014; Kang et al., 2009; Kidd, Piantadosi, & Aslin, 2012).11 Others have attempted to reconcile these two perspectives (Dubey & Griffiths, 2019). These various models might benefit from the sort of explicit treatment of the underlying confidence that we have discussed.

As briefly alluded to above, in the real world, information is often not solely provided by faceless sources, but by other agents with their own intentions. Over and above just being noisy (and indeed nosey), such social sources might have their own biases and interests of which successful agents need to be aware when evaluating whether they should invest in hearing their opinion and using them to inform themselves (Hütter & Ache, 2016; Pescetelli & Yeung, 2020b; van der Plas, David, & Fleming, 2019). This is a particular pressing issue when faced with mis- and dis-information (Lazer et al., 2018; Pennycook & Rand, 2021). Such scenarios will require adaptive metacognitive systems to make inferences not only about themselves but also about others. From a computational perspective, theories such as cognitive hierarchy (Camerer, Ho, & Chong, 2004), interactive POMDPs (Gmytrasiewicz & Doshi, 2004) or Rational Speech Acts (Goodman & Stuhlmüller, 2013) could be adapted to consider hierarchies of partially self-aware agents interacting with each other.

Finally, we note that hierarchies of ever more sophisticated sub-agents that model each other inside a single decision-maker constitute a form of theory of (an internal) mind that is somewhat reminiscent of these externally-directed cognitive hierarchies (Carruthers, 2009). If the internal sub-agents enjoy their own partially individual rewards – so, for instance, the rater might have an incentive to lie about its confidence if it faces an overwhelming loss for being wrong or because it believes that the lie might have adaptive benefits (Bénabou & Tirole, 2016; Johnson & Fowler, 2011; Kurvers et al., 2021) – we can expect very rich patterns of behaviour to emerge, with agents partially fooling themselves as well as others.

Conclusion

By offering a joint account of metacognitive monitoring and control, our work provides theoretical grounding for, and empirical evidence of, rich patterns of behavior that can emerge when considering both parts of a metacognitive process. Such an integrated treatment highlights the importance of considering the different building blocks of (meta-)cognition together rather than treating them as isolated processes. Our discussion has described the wide conceptual applicability of this integrated approach, and pointed to a range of remaining empirical and theoretical questions, from neural realizations to more detailed accounts of potential irrationalities in human choice. Our hope is that this will provide a better lens for understanding the richness of metacognitive control behavior, in search and beyond.

Acknowledgements

We would like to thank Chris Gagne, Nitay Alon, Kevin Lloyd, and Sherry Dongji Bao for comments on the paper and support with code.

11We observe inverse U-shapes under some extreme parameter settings, but stress that these are due to the signal and noise properties of the second-order model (see appendix C)
References


Cohen, J. D., McClure, S. M., & Yu, A. J. (2007). Should I stay or should I go? How the human brain manages the trade-off between exploitation and ex-


Feeling


Siedlecka, M., Paulewicz, B., & Wiercz˙o´n, M. (2016). But I was so sure! Metacognitive judgments are less accurate given prospectively than retrospectively. *Frontiers in psychology*, 7, 218.


So, S. H. w., Siu, N. Y. f., Wong, H. l., Chan, W., & Garety, P. A. (2016). ‘Jumping to conclusions’ data-gathering bias in psychosis and other psychiatric disorders — Two meta-analyses of comparisons between patients and healthy individu-


Appendix A
Model details

Postdecisional model

Predicting $X_F$

To predict the location of $X_F$ for the value of seeking, the seeker combines the two possible normal distribution weighted them by the associated confidence:

$$p(X_F|Z_t) = p(X_F|d = -1)P(d = -1|Z_t) + p(X_F|d = 1)P(d = 1|Z_t)$$  \hspace{1cm} (23)

Second-Order Model

Confidence

Fleming and Daw (2017) describe the computations underlying their second-order model. Here, we present them in our notation. Recall that the rater observes the actor’s decision $a_t$ and receives its own cue $Y_i$. The rater then has to use this information to compute the probability that the actor’s decision was the correct one:

$$c_I = P(a_I = d|Y_t, \Sigma_t) = \begin{cases} P(d = 1|Y_t, a_I; \Sigma_t) & \text{if } a_I = 1. \\ P(d = -1|Y_t, a_I; \Sigma_t) & \text{if } a_I = -1. \end{cases}$$  \hspace{1cm} (24)

As with the postdecisional model, we apply Bayes-rule to compute this. In the following, we suppress $\Sigma_t$ for clarity:

$$P(d|Y_I, a_I) = \frac{P(d|Y_I)P(a_I|Y_I, d)}{\sum_d P(d|Y_I)P(a_I|Y_I, d)}$$  \hspace{1cm} (25)

We begin teasing this apart, beginning with the second term:

$$P(a_I|Y_I, d) = \int_{X_I} P(a_I|X_I)P(X_I|Y_I, d)dX_I$$  \hspace{1cm} (26)

Because $P(a_I|X_I)$ is contingent on the threshold (so that $a_I = 1$ if $X_I > 0$), this can also be expressed as:

$$P(a_I|Y_I, d) = \begin{cases} \int_{\infty}^{\infty} P(X_I|Y_I, d)dX_I & \text{if } a_I = 1. \\ \int_{0}^{\infty} P(X_I|Y_I, d)dX_I & \text{if } a_I = -1. \end{cases}$$  \hspace{1cm} (27)

This is cumulative density function of the conditional density of a multivariate Gaussian. This conditional density of a multivariate Gaussian is itself simply a univariate Gaussian.

$$P(X_I|Y_I, d) \sim N(\mu_{X_I|Y_I}, \sigma_{X_I|Y_I})$$  \hspace{1cm} (28)

The conditional parameters of this distribution are defined as follows:

$$\mu_{X_I|Y_I} = d + \frac{\sigma_I}{\tau_I}p(Y_I - d)$$  \hspace{1cm} (29)

$$\sigma_{X_I|Y_I} = \sqrt{(1 - \rho^2)\sigma_I^2}$$  \hspace{1cm} (30)

The first term is the normalized likelihood of $Y_I$ conditioned on a $d$:

$$P(d|Y_I) \propto P(Y_I|d)$$  \hspace{1cm} (31)

$P(Y_I|d)$ in turn equals the density of a unidimensional Gaussian with mean $d$ and standard deviation $\tau_I$ at $Y_I$.

Optimal weighting of $X_I$ and $Y_I$ for $Y_I$ under covariance

In contrast to the postdecisonal model, we cannot simply weigh $X_I$ and $Y_I$ according to their variances when combining them to $Z_I$. Rather, we need to take into account their covariance (Oruç, Maloney, & Landy, 2003). As a result, $X_I$ and $Y_I$ are summed with their respective weights $w_{X_I}$ and $w_{Y_I}$:

$$Z_I = w_{X_I}X_I + w_{Y_I}Y_I$$  \hspace{1cm} (32)

These weights are functions of the reliabilities of the cues which in turn are corrected for the correlation.

$$w_{X_I} = \frac{r'_{X_I}}{r'_{Y_I} + r'_{X_I}} \quad \text{and} \quad w_{Y_I} = \frac{r'_{Y_I}}{r'_{Y_I} + r'_{X_I}}$$  \hspace{1cm} (33)

$$r'_{X_I} = r_{X_I} - \rho_I \sqrt{r_{X_I}r_{Y_I}} \quad \text{and} \quad r'_{Y_I} = r_{Y_I} - \rho_I \sqrt{r_{X_I}r_{Y_I}}$$  \hspace{1cm} (34)

$$r_{X_I} = \frac{1}{\sigma_I^2} \quad \text{and} \quad r_{Y_I} = \frac{1}{\tau_I^2}$$  \hspace{1cm} (35)

This way we can also define the standard deviation $\zeta_I$ of $Z_I$.

$$\zeta_I = \frac{1}{r_{Z_I}}$$  \hspace{1cm} (36)

$$r_{Z_I} = \frac{r_{X_I} + r_{Y_I} - 2\rho \sqrt{r_{X_I}r_{Y_I}}}{1 - \rho^2}$$  \hspace{1cm} (37)

This form of cue combination can give rise to several non-intuitive results which we discuss further below.
Value computations

In the following, we detail the value computations in the second-order model. First, if there is no seeking, the actor uses \( Z_t \) (see above) to make its decision. The value of this combined stimulus is defined as:

\[
V^*_F = \max(P(d = 1|Z_t), P(d = -1|Z_t))r_F
\]  
(38)

This is then used in the \( Q \)-value computations for the \( Q \)-value of not seeking (see equation 17).

However, the seeker does not know \( Z_t \) because it does not have access to \( X_t \). It therefore has to marginalize out this quantity:

\[
V^*_{F,Y,0} = \int_{Z_t} p(Z_t|Y_1, a_1)V^*_F dZ_t
\]  
(39)

\[
= \int_{X_t} p(X_t|Y_1, a_1)V^*_F dX_t \text{ where}
\]  
(40)

\[
p(X_t|Y_1, a_1) = p(X_t|Y_1, a_1, d = 1)P(d = -1|Y_1, a_1) + p(X_t|Y_1, a_1, d = 1)P(d = 1|Y_1, a_1)
\]  
(41)

Given seeking, the actor receives \( X_F \) (again as per equation 10) which it combines with \( Z_t \) to form a joint variable \( Z_F \) (see equation 12). This variable can then again be compared against a threshold for \( a_{F,1} \). Given this set-up, we can now consider the values that go into the individual \( Q \)-value computations.

\[
V^*_{F,Z_F} = \max(P(d = 1|Z_F), P(d = -1|Z_F))r_F
\]  
(42)

Similarly to the first-order and postdecisional models, the seeker does not know all the variables underlying \( Z_F \), when it decides whether to seek, and it also does not know \( Z_t \). Therefore, it has to marginalize over them both:

\[
V^*_{F,Y_1} = \int_{Z_F} p(Z_F|Y_1, a_1)V^*_{F,Z_F} dZ_F
\]  
(43)

\[
= \int_{X_t} \int_{X_F} p(X_t, X_F|Y_1, a_1)V^*_{F,Z_F} dX_t dX_F
\]  
(44)

\[
p(X_t, X_F|Y_1, a_1) = p(X_t|Y_1, a_1)p(X_F|d = -1)P(d = -1|Y_1, a_1) + p(X_F|d = 1)P(d = 1|Y_1, a_1)
\]  
(45)

Notice how both the with- and without-search value computation contain, \( P(d|Y_1, a_1) \), or the rater’s confidence.

Over- and underconfidence

As discussed in the main text, the second-order model can produce over- and underconfidence by way of dissociating the parameter underlying the sampling of the actual actor stimulus, and the parameter that the rater uses to invert the model and compute its confidence. In more detail, this means that we use an objective and subjective \( \sigma_t \), where the stimuli are still sampled with \( \sigma_t \) as in equations 7 and 8. In contrast to this, the rater and seeker proceed in their computations using the subjective \( \sigma_t^R \), for example when forming the confidence:

\[
c_t = P(a_t = d|Y_1; \Sigma_t^R)
\]  
(46)

\[
\Sigma_t^R = \begin{bmatrix}
(\sigma_t^R)^2 & \rho_t \sigma_t^R \tau_t \\
\rho_t \sigma_t^R \tau_t & \tau_t^2
\end{bmatrix}
\]  
(47)

This then also applies to the seeker’s computations outlined in equations 41-45.

Appendix B

Methods

We implemented our models, simulations, and data analysis in R. Our code and data is hosted openly on a dedicated GitHub repository (https://github.com/lionschulz/SchulzFlemingDayan). Our work only includes theoretical simulations and additional analysis of previously published data from Schulz et al. (2020). That study was not preregistered.

Simulation details

To simulate the effects of over- and underconfidence on optimal search, we sampled values of \( \sigma_t \) and \( \sigma_t^R \) to span the empirically observed accuracy range (60 - 85 %) from L. Schulz et al. (2020). Specifically, we sampled accuracies individually from a uniform distribution within this range and then transformed them to \( \sigma_t \) and \( \sigma_t^R \) using the inverse of function 21:

\[
\sigma_t = \frac{1}{\phi^{-1}(\text{Accuracy})} \quad \text{(48)}
\]

where \( \phi^{-1} \) denotes the inverse cumulative density function of the normal distribution. To simulate different degrees of metacognitive sensitivity, we furthermore sampled using the same procedure and range. The correlation \( \rho_t \) was fixed to 0.3, and the cost of the information \( r_S \) to 0.1 (\( r_F = 1 \)).

To probe agent’s average confidence in these simulations, we simulated 1000 trials for each stimulus combination. To compute the agent’s meta-\( d' \), we split these confidences into five equally spaced confidence bins (details on meta-\( d' \) fit below).
Data

Here, we reanalysed participant data from L. Schulz et al. (2020) who collected choices and confidence ratings from 734 participants in an information-seeking task that shares many commonalities with the theoretical task we discuss. A few subtle differences between the task we describe in the theory section and Schulz et al. (2020) exist: Our theoretical task uses a continuous confidence report made after the initial decision, offering an arbitrarily fine-grained picture of confidence. In contrast, L. Schulz et al. (2020) employed a discrete, three-step, scale that probed confidence conjointly with the judgement. This lowers the resolution of possible confidence values. Such discrete confidence ratings are usual in the field of metacognition (Rahnev et al., 2020). Because of the joint assessment of choice and confidence, confidence could not fall below 50%. Furthermore, L. Schulz et al. (2020) employed just one, stair-cased, strength of the initial (and yoked final) stimulus limiting our ability to probe changes in accuracy level to post-hoc analyses (see Figure B2 for the two accuracies). However, the task employed two levels of costs, letting us probe the effects of cost on participants search. A detailed description of this task can be found in L. Schulz et al. (2020).

As is usual in the literature (Fleming & Daw, 2017; Rahnev et al., 2020; Shekhar & Rahnev, 2020), participants displayed significant variation in their metacognitive sensitivity. We show the distribution of perceptual (d’) and metacognitive (meta-d’) sensitivities in Figure B1.

Data and simulation analyses

Trial-by-trial effects were computed using the mixed logistic models in the “afex” (Singmann et al., 2018). To analyze correlations in the data, we used Pearson correlation, and for linear regression models, we applied the lm() function. We employed HMeta-d’ (Fleming, 2017) using non-hierarchical fits to fit meta – d’.

Appendix C
Further Second-Order Results
Confidence and general stimulus conditions

Signal, noise correlation

In the second-order model, the correlation can give rise to counterintuitive confidence curves. This is visible in Fig C1 where we plot confidence values for a positive decision (aI = 1) varying the parameters individually. We observe a few aspects already reported by Fleming and Daw (2017):

- Panel A: Increasing the accuracy of the actor (lower σI) increases the boost that the confidence receives through the action. If the actor is very accurate, it takes a highly negative Yi to overturn the decision.
- Panel B: Higher rater noise (τI) means the confidence curves will be less well-tuned.
- Panel C: Higher correlations (ρI) also results in a
reduced sharpness in the confidence curves.

However, what has yet to be reported is the following: Under conditions of metacognitive hyposensitivity, that is when $\sigma_I$ is sufficiently smaller than $\tau_I$, and when $\rho_I$ is large enough, confidence will begin rising again with seemingly contradictory $Y_I$’s. This is particularly visible in the rightmost panel where $\rho_I$ is most pronounced, but is also visible in the most extreme cases in panels A and B. As an example, imagine the actor has received $X_I = 0.5$ and decides $a_I = 1$. If the rater receives $Y_I = -5$, this would usually be a strong error signal and the confidence in the initial decision very lower than when $Y_I$ would have had more intermediate values. However, under some parameter conditions, the exact opposite is the case: There, when $Y_I$ strongly contradicts the decision sign, confidence will in fact be higher for this very low $Y_I$ than for $Y_I = 0$.

While we note that the marginal probability of these cases is relatively low given the underlying correlation, such a pattern is striking. The reason for it lies in the way the two possible sources occupy the $X_I$ and $Y_I$ space and create signal and noise (compare Figure 1 E). A crucial aspect of this is the line on which the posterior based on $Z_I$ (i.e. the combination of $X_I$ and $Y_I$) is uniform, so that $P(d = 1|X_I, Y_I) = P(d = -1|X_I, Y_I) = 0.5$. It is this posterior that the rater only has partial information about. The equality line subdivides the space in two zones where the likelihood of $d = 1$ is larger than the likelihood of $d = -1$ (or vice versa). Given the equal prior, this line of equality in turn is defined by the points at which the two likelihoods equal each other.

$$p(X_I, Y_I|d = -1) = p(X_I, Y_I|d = 1)$$ (49)

The two likelihoods are defined by the bivariate normal distribution’s density:

$$p(X_I, Y_I|d) = \frac{1}{2\pi \sigma_I \tau_I \sqrt{1 - \rho_I^2}} e^{-\frac{(\bar{\mu}_{X_I} - X_I)^2}{2\sigma_I^2} - \frac{(\bar{\mu}_{Y_I} - Y_I)^2}{2\tau_I^2} + \frac{(\bar{\mu}_{X_I} - X_I)(\bar{\mu}_{Y_I} - Y_I)}{\sigma_I \tau_I}}$$ (50)

From this, we can can define the values of $Y_I$ for which the two posteriors equal each other as a function of $X_I$:

$$Y_I = -mX_I$$ (51)

where from equation 49, we get:

$$m = \frac{\frac{1}{\sigma_I^2} - \frac{\rho_I}{\sigma_I \tau_I}}{\frac{1}{\tau_I} - \frac{\rho_I}{\sigma_I \tau_I}}$$ (52)

We plot this in Figure C1 D-F for a range of parameter combinations. When there is no correlation $\rho_I = 0$, $m$ (panel F) this line is defined by $\frac{\tau_I}{\sigma_I^2}$ and the space is thus divided diagonally from a positive $Y_I$ to a negative $Y_I$ with the slope defined by the relationship between the two parameters. This general result holds, even when re-introducing the correlation. Importantly, what this division of space means is that for every possible actor cue $X_I$, more positive rater cues $Y_I$ will favour $d = 1$ and more negative $Y_I$ will favour $d = -1$. Crucially however, under metacognitive hyposensitivity ($\tau_I > \sigma_I$) this diagonal becomes steeper and steeper until it is fully vertical. This point is defined when:

$$\rho_I = \frac{\sigma_I}{\tau_I}$$ (53)

In other words, at this point, the decision rule based on $Y_I$ and $X_I$ is the same as based on $X_I$ alone – $Y_I$ thus affords no additional help with the decision. Beyond this vertical point, the space is again divided diagonally, but the dividing line now has a positive rather than negative slope. This only appears under relatively extreme parameter combinations, but will crucially flip the logic outlined above. Now, for every $X_I$, lower values of $Y_I$ will begin providing more evidence for $d = 1$ instead of $d = -1$. This then in turn gives rise to the confidence rising with seemingly contradictory values of $Y_I$. This phenomenon will appear once the equality lines have ‘flipped’, as is visible when comparing the confidence curves and slopes depicted in Fig. C1.

**Relationship between $\zeta_I$ and $\sigma_I, \tau_I, \rho_I$**

As alluded to in the main text, the joint standard deviation $\zeta_I$ produced from optimally combining $\sigma_I, \tau_I$ and $\rho_I$ stands in a non-trivial relationship with its subparts.

For context, recall how the $\sigma_I$ and $\tau_I$ are combined when there is no correlation (see equation 5). As we discussed in the main text, the maximum of $\zeta_I$ is then defined by the smaller of the two standard deviations $\sigma_I$ and $\tau_I$. Additionally, the smaller the larger of the two is, the smaller $\zeta_I$ becomes. In other words, the agent would benefit from a reduction of noise in both cases. For an illustration of this effect, see the yellowmost lines in Figure C2A that show a cue integration in accuracy space ($\phi(\zeta_I)$) as a function of $\phi(\sigma_I)$ for $\rho_I = 0$. Notice how lower $\tau_I$’s shift the baseline upwards and how the better accuracy of afforded by $\sigma_I$ increases the accuracy afforded by $\zeta_I$.

In most cases of optimal cue combination, two independent sources (low $\rho_I$) of information hold more information (lower $\zeta_I$) than two correlated sources
Figure C1

Second-order confidence across parameter regimes. (A-C) Second-order confidence as a function of $Y_I$ for $a_I = 1$. In general, note how confidence for a completely ambiguous rater cue ($Y_I = 0$) doesn’t necessarily mean that confidence $c_I$ will be 0.5. (A) This is mainly a function of the relationship between $\sigma_I$ and $\tau_I$. (B) High values of $\tau_I$ for a fixed $\sigma_I$ can lead to the confidence being less sensitive to $Y_I$. When $\tau_I$ is particularly large in relation to $\sigma_I$, the confidence will in fact again begin to rise for negative $Y_I$’s (which intuitively contradict $a_I$). Grey line in (A) and (B) highlights an equivalent parameter setting of $\tau_I = $ $\sigma_I = 2$, $\rho_I = 0.5$. (C) This rise of confidence with contradictory rater cues $Y_I$ is particularly pronounced for high correlations $\rho_I$. The rising confidence is tied to the way the correlation affects signal and noise and in extension the line on which the joint posterior $P(d|X_I, Y_I)$ is equivalent between the two $d$. This line is plotted in (D-F). When an agent is metacognitively hyposensitive ($\sigma_I < \tau_I$) and when the correlation $\rho_I$ between $Y_I$ and $X_I$ is high enough, confidence will not decrease for negative $Y_I$ but rather again rise.

Specifically, when the equality line ‘flips’, the posteriors get compressed differently between the two sources, allowing a better inference than in the classical separation of $X_I$, $Y_I$ space.

The effects of this “flip” are formally analogous to the way in population codes that correlations between the activities of units can either help or hurt discrimination and decoding depending on their alignment relative to the way that signals are coded (the mean difference) (Abbott & Dayan, 1999).

Seeking and final accuracy for the high $\rho_I$

The two aforementioned particularities of the second-order model also impact the agent’s search behaviour and final accuracy, which we depict in Figure
The fact that confidence rises again with contradictory values of $Y_I$ will result in U-shaped seeking curves for most $\tau_I$. This is because the rising confidence will favour not seeking, rather than seeking once the actor accuracy is below a specific value while keeping $\tau_I$ fixed.

With regards to the final accuracy, the maximum attainable accuracy from combining $X_I$ and $Y_I$ (and $X_F$) will be impacted by the combination of $\sigma_I$, $\tau_I$ and $\rho_I$ giving rise to $\zeta_I$ (discussed above). This will for example mean that more accurate (low $\tau_I$) raters can produce less accurate final judgements than noisier (high $\tau_I$) raters.
Figure C3

Effects of high correlation between actor and rater signal. (A) Average search by average initial accuracy and rater noise $\tau_I$. (B) Final accuracy by average initial accuracy and rater noise, and conditioned on whether the agent sought out information or not.