Human optional stopping in a heteroscedastic world

Hannah Tickle[1,2], Konstantinos Tsetsos[3], Maarten Speekenbrink[2] and Christopher Summerfield[1]*

[1] Dept. Experimental Psychology, University of Oxford, Oxford UK
[2] Dept. Experimental Psychology, University College London, London UK
[3] Medical School, Hamburg University, Hamburg, Germany

* to whom correspondence should be addressed. christopher.summerfield@psy.ox.ac.uk

Abstract

When making decisions, animals must trade off the benefits of information harvesting against the opportunity cost of prolonged deliberation. Deciding when to stop accumulating information and commit to a choice is challenging in natural environments, where the reliability of decision-relevant information may itself vary unpredictably over time (variable variance or "heteroscedasticity"). We asked humans to perform a categorisation task in which discrete, continuously-valued samples (oriented gratings) arrived in series until the observer made a choice. Human behaviour was best described by a model that adaptively weighted sensory signals by their inverse prediction error, and integrated the resulting quantities to a collapsing decision threshold. This model approximated the output of a Bayesian model that computed the full posterior probability of a correct response, and successfully predicted adaptive weighting of decision information in neural signals. Adaptive weighting of decision information may have evolved to promote optional stopping in hetereoscedastic natural environments.

Introduction

Time is of the essence for an agent who wishes to make good decisions. Longer deliberation promotes decision accuracy, by allowing precise estimates of noisy sensory variables to be formed over time. However, deliberation incurs an opportunity cost, postponing the receipt of the positive reinforcement signals that accompany good choices. Excessive deliberation can thus be pernicious, as exemplified by Buridan's classic fable in which a donkey who is both hungry and thirsty expires whilst choosing between proffered food and water. Successful decision policies must thus strike a delicate balance between the acquisition of sufficient choice-relevant information and the timely harvesting of rewards. Understanding how humans and other animals negotiate this problem has long been a central concern in psychology, neuroscience and behavioural ecology (Bogacz, 2007; Busemeyer and Rapoport, 1988; Drugowitsch et al., 2012; Edwards, 1965; Gluth et al., 2012; Hawkins et al., 2015; Kira et al., 2015; Malhotra et al., 2017; Moran, 2015; Murphy et al., 2016; Ratcliff and McKoon, 2008; Thura et al., 2012).

The Sequential Probability Ratio Test (SPRT)

Over the past half century, a normative framework has been developed to understand how an agent should decide to stop sampling information and commit to a binary choice (Bogacz et al., 2006; Drugowitsch et al., 2012; Frazier and Yu, 2008; Malhotra et al., 2017; Moran, 2015). The genealogy of many current theories can be traced back to the Sequential Probability Ratio Test (SPRT), first proposed more than half a century ago (Gold and Shadlen, 2002; Wald and Wolfowitz, 1949). The SPRT proposes that for binary choices, agents aggregate the (log) likelihood ratio of evidence up to a fixed threshold, at which point a response is initiated. To illustrate, consider a canonical problem in which the observer views k samples each conveying information X_k and is asked to decide from which of two Gaussian distributions $\mathcal{N}(\mu_A, \sigma_A^2)$ or $\mathcal{N}(\mu_B, \sigma_B^2)$ the samples are drawn. A key relevant decision variable is the (log) likelihood ratio $\mathcal{L} = log\left(\frac{p(X|A)}{p(X|B)}\right)$, which is turn given by the sum of log likelihood ratios for each individual sample drawn X_k. Setting symmetric bounds on $\mathcal{L}$ allows the observer to initiate a response at a desired level of certainty, with higher bounds allowing more correct responses and lower bounds terminating the decision process more promptly at the expense of accuracy. Under these simple assumptions, the SPRT allows the observer to map the stream of information X onto the probability that responding now (vs. later) will give rise to a correct response, $p(correct)$. Given further knowledge of the payoff matrix for correct and incorrect responses, and the likely delay between trials (or encounters) it is possible to compute the criterial level of certainty at which a response should be initiated in order to maximise overall rate of return (i.e. reward per unit time). In other words, the SPRT allows the observer to set their decision boundary (and thus speed-accuracy tradeoff) at a reward-maximsing level.

Variants of this model have been proposed to account for decision latencies in psychology, neuroscience and behavioural ecology. In the latter field, Optimal Foraging theory builds on the intuitions above to predict when animals should stop harvesting rewards from patch A (by analogy, a given trial) and switch to patch B (the next trial) given a fixed travel time (the inter-trial interval) (Stephens and Krebs, 1986). However, our focus here is the study of decisions and their latencies in psychology and neuroscience, where the study of mental chronometry has been a central theme since the time of Donders, and a mature theoretical framework has been built for understanding the timing of decisions.

The Drift Diffusion Model (DDM)

The key challenge for an agent wishing to produce decisions that maximise the rate of return is that in natural environments and laboratory experiments alike, the sufficient statistics of the generative distributions from which samples are drawn are often *a priori* unknown to the observer. The theoretical intuition that allows $p(correct)$ to be estimated without exact knowledge of the generative model – at least in the simplest case involving binary choices about samples drawn from one of two symmetric Gaussian distributions – is that the log-likelihood ratio for each sample is proportional to the momentary sensory information conveyed by each sample, i.e. $\mathcal{L}_k = g \cdot X_k$ where g is a scaling constant that depends inversely on the overlap between the generative distributions. A popular model, known as the "drift-diffusion" model (DDM), thus simply fits a free parameter d that encodes the average rate of accumulation (or drift) in a trial or condition, along with an additional noise term ε that is required to account for both sampling error and the intrinsically stochastic nature of choice in biological systems. The proposed momentary update δDV to the cumulative decision variable occurring on sample k is thus computed as $\delta DV_k = d + \mathcal{N}(0, \varepsilon)$, and this quantity is aggregated until a symmetric boundary at $\pm Z$ is breached. The fitting of the boundary as a further free parameter additionally relaxes the assumption that humans perfectly tailor the height of the bound to maximise rates of accuracy or reward. The DDM has become a ubiquitous tool for modelling human choices, and with some minor elaboration, has enjoyed striking success in jointly capturing the form of human reaction time (RT) distributions for both correct and error trials (Ratcliff and McKoon, 2008). For example, by fitting the (extended) DDM to response times, the parameters obtained are sufficient to predict the psychometric function that relates decision accuracy to the strength of sensory signals (Kang et al., 2017; Ratcliff and McKoon, 2008).

The intuition described above implies that the DDM is not only parsimonious but also optimal for categorisation experiments involving two choices and Gaussian noise. Unfortunately however, this intuition breaks down in the ubiquitous case where each category is composed of more than one distribution with unknown statistics. For example, in a canonical psychophysical paradigm in which observers discriminate whether the net motion direction of moving dots tends to the left or right of vertical (the Random Dot Motion (RDM) task), the coherence (or, alternatively, the incident angle of the dots with respect to the boundary) may be manipulated over multiple levels across trials. To illustrate, imagine that samples X are drawn from one of four distributions with fixed variance and means μ^{++}, μ^{+}, μ^{-} and μ^{--}, and the task is to report whether $\mu > 0$ or $\mu < 0$. In this case, $p(correct)$ cannot be inferred directly from $sum(X)$ alone, because $p(X|\mu^{+}) \neq p(X|\mu^{++})$. It follows that the "flat" bound on aggregrated evidence proposed in the classic formation of the SPRT (and DDM) is no longer optimal when the difficulty of the trial is unknown a priori to the observer. This prompts two major questions relating to optional stopping – one theoretical, and one empirical – that have given rise to considerable controversy in the recent literature.

The theoretical question concerns the optimal form of the bound when the difficulty is mixed among conditions and unknown to the observer. This question is challenging because a fully optimal solution to the problem of optimal stopping under mixed difficulties requires the observer to compare the expected value of responding now with that of responding later. This necessitates a prediction about how the trial will unfold over multiple theoretically possible future samples, a problem that can only be solved optimally with computationally costly recursive optimization methods. For example, Drugowitsch et al describe a dynamic programming solution for optimally computing the cost of responding A, B or "later", which fit empirical data well under the assumption that the agent's discount function (i.e. the subjective cost of time passing) was allowed to vary idiosyncratically (Drugowitsch et al., 2014; Drugowitsch et al., 2012). Alternatively, agents may use Bayesian inference to jointly compute the likelihood of the category (A vs. B) and the difficulty level (e.g. coherence; easy vs. hard), a recursive operation that depends on the unfolding sensory evidence in the trial (Deneve, 2012). An approximate solution is offered in the latter case that involves factorising the inference into two subproblems, one for category and one for signal strength (Sun and Landy, 2016). Nevertheless, the full forward inference implied by these models may be prohibitively computationally expensive for biological agents making rapid sensorimotor judgments on the fly (Malhotra et al., 2017).

The informativeness of the passage of time

Simpler (and potentially more biologically plausible) solutions to the optimal stopping problem under unknown sensory reliability assume that agents track the passage of time in order to "decide when to decide". Intuitively, the passage of time is itself informative about $p(correct)$, because in the case where $sum(X)$ does not diverge substantially from zero after lengthy deliberation, it is likely that samples are being drawn from a distribution paramerised by μ close to zero. This is important for the problem of optional stopping, because each additional sample harvested incurs an opportunity cost by delaying the onset of the subsequent trial, which is potentially richer in evidence. If the samples continue to yield information which is not decisive for choices, a good strategy may be simply to guess (with a 50% chance of positive feedback) and move on. For example, consider an experiment that in addition to μ^{+} and μ^{-} involves a condition μ^{0} in which the evidence favours neither response. On μ^{0} trials, the best policy is simply to guess as fast as possible (and move on to the next trial). As time within a trial unfolds, the probability that samples are drawn from μ^{0} increases, and thus the imperative to draw decisions to a close grows stronger. Indeed, it can be demonstrated analytically that evidence, e.g. $sum(X)$, and time elapsed are sufficient statistics for computing $p(correct)$ at any given point in the trial (Kiani et al., 2014) (but not, as emphasised above, for computing the full opportunity cost of responding now vs. later).

Empirically, the idea that animals decide when to decide under the joint influence of cumulative sensory evidence and time elapsed has considerable appeal. Firstly, time elapsed is computationally frugal to estimate, and human judgments of short intervals are approximately optimal (Jazayeri and Shadlen, 2010). Secondly, we know that when asking participants for their subjective confidence in a decision, their estimates are influenced by the time taken to respond, as if subjectively experienced $p(correct)$ is jointly computed from these two variables (Kiani et al., 2014). On the neural level, it is established that those neurons in the lateral intraparietal area (LIP) whose firing rates scale with the cumulative information during psychophysical stimulation are also sensitive to the passage of time (Jazayeri and Shadlen, 2015). Moreover, there is empirical support for the existence of neural signals which build up over time towards a decision independent of the strength of sensory information (Churchland et al., 2008; Hanks et al., 2014a; Heitz and Schall, 2012). Because these signals grow with the passage of time, over and above the accumulation of evidence, they suggest a mechanism by which both sensory signals and time could jointly drive decisions to a close. For example, in the RDM task, the existence of such signals may partly explain why trial averages of neural firing rates on 0% coherence trials diverge as a function of whether the monkey responds in or out of the receptive field of the LIP neuron, with a faster buildup for “in” trials. Similarly, when the training regime emphasises speed over accuracy, neural signals in LIP and primary motor cortex build up faster to a bound, as if spurred on by a time-dependent “urgency” signal (Hanks et al., 2014b). Related results have been reported in human electroencephalographic (EEG) recording, where a centro-parietal positivity thought to track the accumulation on information shows a faster and shaper excursion when decision speed is faster (Murphy et al., 2016; Spieser et al., 2018; Twomey et al., 2015).

Despite these suggestive findings, the question of whether monkeys and/or humans draw decisions to a close in a fashion that is determined by evidence alone (as proposed by the classic DDM) or jointly by evidence and the passage of time, has become a point of some controversy in psychology and the neurosciences. Hawkins et al (2015) conducted a systematic review of nine datasets (> 100,000 trials) that involved either the RDM paradigm, brightness discrimination, or judgments of dot separation. All tasks involved multiple levels of difficulty that were intermingled randomly over trials and thus unknown to the human or nonhuman primate observers. Perhaps surprisingly, given the neurophysiological evidence described above, the results provided strong support for a fixed-bound model in human data, whereas the monkey data did not clearly disambiguate the two accounts. The authors speculated that perhaps it is the extra training available to the experimental animals that allowed them to adopt a policy that accounted quasi-optimally for the unknown reliability in the data. An alternative explanation is that humans may be more motivated to make a correct response, rather than simply to maximise the rate of positive feedback per unit time (Maddox and Bohil, 1998).

Expanded judgement tasks as an alternative to psychophysical paradigms

A substantial challenge for understanding the factors that determine optional stopping during psychophysical judgment is that samples of information arrive either simultaneously or in very rapid succession, and decisions are typically on the order of under a second. This means that relative to intrinsic sources of variability in response times, the divergence in predictions between different classes of decision model are often slight, and so arbitrating among them consequently requires a large body of data. Voluminous data may be hard to obtain, in particular for humans, and where it is available it is possible that the policy itself is nonstationary over time due to the effect of extended practice. A related challenge arising from rapid presentation is that it is hard to capitalise on variability in model predictions for specific sequences of information. In other words, two streams of information X^{a} and $X^{a'}$ might be drawn from the same distribution, but lead to different model predictions about stopping on any given trial due to sampling variability. Approaches that simply model performance as a function of the mean information ignore the leverage that this sampling variability may have on choices.

One solution to these problems is to move beyond psychophysical tasks to examine optional stopping in a different class of paradigm, an expanded judgment task, in which samples arrive in sequence as discrete sensory pulses of information (Brunton et al., 2013; Cheadle et al., 2014; de Lange et al., 2010; Drugowitsch et al., 2016; Kira et al., 2015; Tickle et al., 2016). A number of recent studies have modelled stopping times in tasks that involve a stream of discrete binary outcomes (e.g. samples X^a and X^b), the object being to identify whether the samples are drawn from a distribution A containing majority X^a samples, or B containing majority X^b samples. This task is akin to the classic "urn and balls" paradigm that has been used to understand probabilistic inference in the literature on human judgment and decision-making. For example, Malhotra and colleagues (Malhotra et al., 2017) showed participants a series of rapid, discrete arrow cues (ISI = ~200ms) that signalled whether they should take a right or left fork in a virtual maze. A salient cue (an animal image) alerted participants to whether the cues were partially informative, uninformative, or of unknown reliability. An analysis of the predictive power of both time and evidence on the probability of commitment suggested that under unknown reliability, humans employed a bound that declined linearly with time (or a growing urgency signal), in a fashion that tended towards (but did not reach) optimality. Intriguingly, however, other studies have observed evidence for time-varying boundaries in expanded judgment tasks even where there is a single level of difficulty. Gluth et al. used a task that involved choosing to "buy" or "reject" stocks on the basis of discrete ratings, finding that the data were best fit by a variants of the SPRT that incorporated a bound that declined linearly with time (Gluth et al., 2012). Similarly, Kira et al (Kira et al., 2015) trained macaque monkeys to perform a free-response version of the weather prediction task that involved viewing a stream of shapes, each of which was associated with a relative probability of a leftwards vs. rightwards response being rewarded. Their data were best fit by a model that incorporated a time-varying urgency signal that drove decisions to a close independent of the evidence provided by the shapes, in a fashion that is mathematically equivalent to the collapsing bound described in the human studies. Confusingly, thus, there is evidence for both fixed and time-varying boundaries, under both conditions where they are warranted for optimal choices and those where they are not.

In psychophysical tasks, the reliability of the sensory information is proportional to the signal-to-noise ratio (SNR), for example the fraction of coherently vs. randomly moving dots on an trial of the RDM task. In binomial judgment tasks, such as that from Malhotra and colleagues described above, the reliability is proportional to the probability that a cue predicts the correct response. This raises a second difficulty of the use of psychophysical paradigms to assess decision latencies: by collapsing reliability into a single "difficulty" estimate, these framework(s) sidestep the fact that evidence may vary unpredictably in both its strength (i.e. mean) and reliability (i.e. variance). For example, X may be drawn from a distribution with a mean that varies over multiple levels (e.g. μ^{++}, μ^{+}, μ^{-} or μ^{--}) and standard deviation that similarly varies across trials (e.g. σ^{+}, σ^{++} or σ^{+++}). This joint variability in signal strength (e.g. mean) and signal reliability (e.g. variance) is ubiquitously observed in natural environments. For example, the evidence presented in a court of law may be strongly indicative of crime culpability or merely circumstantial (evidence strength), but the witness reporting the evidence might be trustworthy or untrustworthy (evidence reliability). Similarly, economic prospects, such as an investment opportunity, may vary unpredictably in terms of both the mean and variance of a distribution of possible outcomes, i.e. in both their expected value (strength) and risk (reliability). We call these settings "heteroscedastic", because the variance of the generative process is itself variable over trials, in addition to its central tendency.

The influence of heteroscedasticity

One question that naturally emerges, thus, is how evidence variability and signal strength influence choice certainty when they are manipulated orthogonally, as in the example above. In Fig.1, we plot analytic solutions for $p(correct)$ under three different settings. Firstly, we consider the case where samples X are drawn sequentially from one of two distributions with means μ^{+} and $\mu^{-}$and fixed variance σ (upper panels). As is well known, in this case $p(correct)$ after each new sample depends only on the aggregate evidence $sum(X)$. Secondly, we consider the case where the difficulty due to the mean varies idiosyncratically across trials. In other words, X may be drawn from μ^{++}, μ^{+}, μ^{-} or μ^{--} but σ remains fixed. As previously described (Kiani and Shadlen, 2009) in this case decision certainty depends both on the sum of evidence and the passage of time (middle panels). However, the novel intuition illustrated by these simulations is revealed in the lower panels. Here, we consider the full heteroscedastic case in which X is drawn from a distribution with unknown mean and unknown variance. As illustrated by the changing topography of $p(correct)$ over the five panels, $sum(X)$ and time are now no longer sufficient statistics for choice certainty. Rather, $p(correct)$ depends additionally on $sum(X^2)$, a quantity that scales with the heterogeneity of the information presented sequentially on each trial. Intuitively, this can be explained by the fact that, for a given value of $sum(X)$ obtained after k samples, the LLR will be closer to zero if the variance (which scales with $sum(X^2)$) of the sampling distribution is larger.

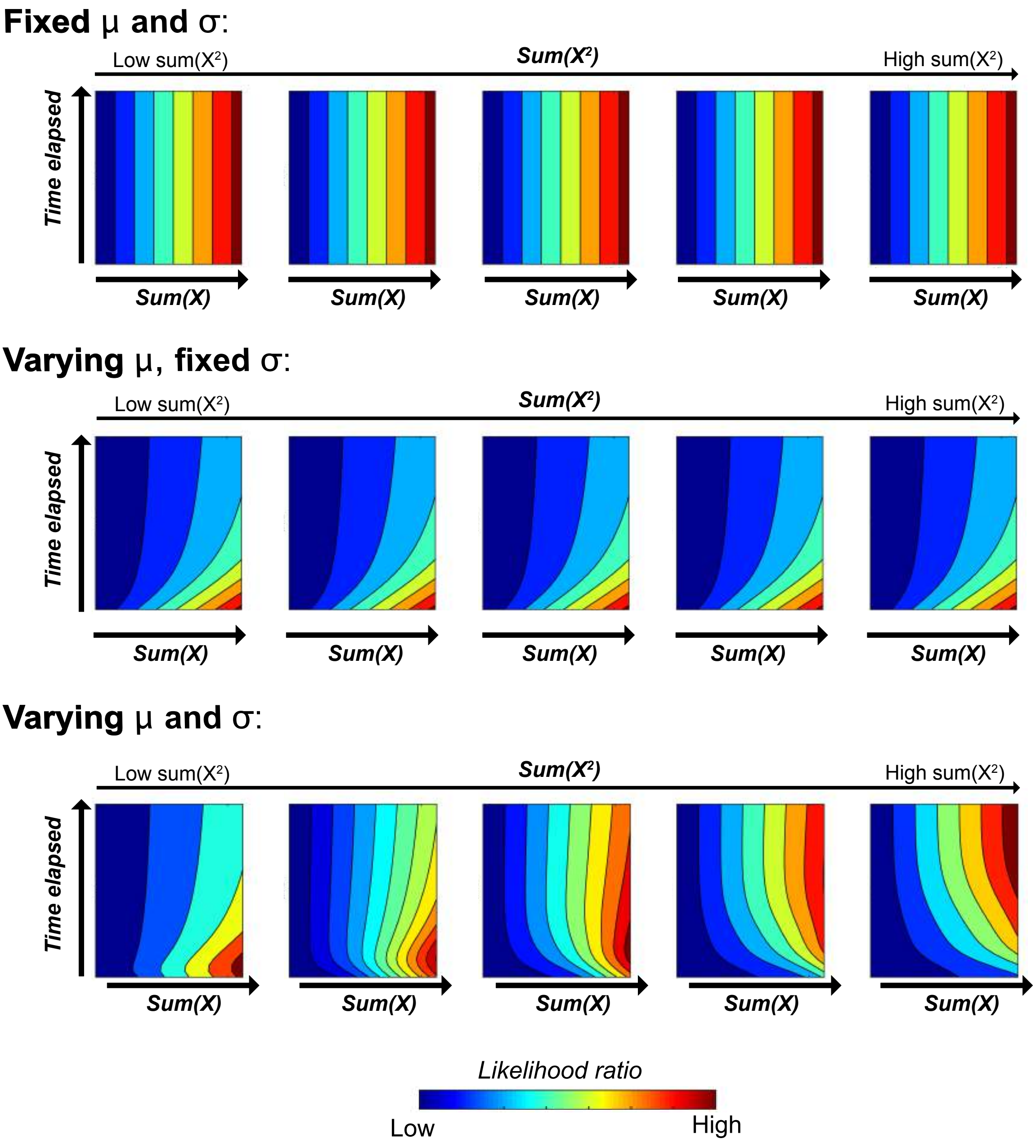


Figure 1: The influence of heteroscedasticity on choice certainty. Each subpanel plots the absolute likelihood ratio (which varies monotonically with $p(correct)$ as a function of the sum of evidence (x-axis) and the time elapsed (y-axis). Columns additionally show dependence on $sum(X^2)$. Each row shows how choice certainty varies under different knowledge of the generative statistics. Top row: fixed mean and variance. Middle row: unknown mean, fixed variance. Bottom row: unknown mean and variance.

This simulation shows that, under conditions of heteroscedasticity, an accurate measure of certainty cannot be gleaned without some estimate of evidence variance. Empirically however, it is not clear whether humans adopt this strategy, or instead use a more myopic estimate of certainty based only on the strength of evidence.

Understanding heteroscedasticity via an expanded judgement task

The primary goal of the current work was thus to understand, via model-based analysis, how humans "decide when to decide" in the heteroscedastic case – i.e. when evaluating pieces of evidence drawn from distributions with both unknown strength (mean) and reliability (variance). To this end, we employed an expanded judgement paradigm that allowed us to quantify the exact level of decision evidence available on a sample-by-sample basis. This permitted us in turn to evaluate and arbitrate between competing computational models of human decision latencies in a way that would have been more difficult using aggregate measures such as reaction time distributions.

The agent's goal on each trial of the experiment was to determine whether the *average* orientation of a stream of samples (tilted Gabor patches) fell clockwise (CW) or anticlockwise (ACW) with respect to a reference orientation, after having viewed as many samples as desired before committing to the choice. A schematic of this task is depicted in **Fig. 3a** (see below). Our approach differs from that those taken previously in at least three substantive ways. Firstly, we manipulated two distinct sources of uncertainty – the mean and the variance of the generative distribution. Specifically, the dispersion of the generative distribtions varied over 3 levels (σ^{+}, σ^{++} or σ^{+++}) and their means fell symmetrically around the reference with a separation that was either low, medium or high (μ^{+}/μ^{-}, μ^{++}/μ^{--}, , or μ^{+++}/μ^{---}), see **Fig. 3b.** Secondly, a novel feature of our task is that we allowed participants to draw samples (tilted gratings) at their own pace, and thirdly, we measured the resulting decision latencies in terms of number of samples. This allowed us to quantify precisely the available information at each point during decision formation, and to model decisions in a manner that is uncontaminated by variation in motor selection and execution times. This further allowed us to measure, using a single-trial analysis approach, how scalp electroencephalographic (EEG) signals accompanying each discrete sample covaried with both time and evidence during the formation of decisions. Specifically, we tested for the emergence of an "urgency" signal that grows with the occurrence of each discrete sample, up to the point at which a choice was made (Churchland et al., 2008; Drugowitsch et al., 2012; Hanks et al., 2014b; Heitz and Schall, 2012; Murphy et al., 2016; Thura et al., 2012)

Modelling decision latencies under heteroscedasticity

Our first goal was to understand the nature of the probabilistic computations occurring whilst humans performed the expanded judgment task. The probability of a correct response in the task described can be computed using a sequential Bayesian approach based on the sequential probability ratio test (SPRT). The model infers, after the occurrence of each sample, the posterior probability distribution over an exhaustive space of values for μ and σ, and then marginalises over this distribution to compute a (log) posterior ratio for CW vs ACW. For heteroscedastic inputs, the likelihood function that maps sensory features onto decision values evolves dynamically as beliefs about the most likely values of μ and σ change over the course of the trial. This model does not provide the strictly optimal policy, which would require estimation of the expected value of responding now vs. later using dynamic programming, but nonetheless provides a useful Bayesian benchmark for partially myopic (and potentially more realistic) behaviour.

In homoscedastic environments, the drift diffusion model (DDM) approximates the behaviour of the SPRT via a simpler mechanism that merely sums the sensory signals, accounting for human decision latencies without requiring the unrealistic assumption baked into the SPRT that agents have full knowledge of the generative structure of the task environment. Under the assumption that observers (erroneously) assume that information is drawn from a distribution with a single value for μ and σ that is fixed over all trials, thus, we considered the DDM here as it is the natural rival to the SPRT for explaining our data. If this model were to provide the best account of human behaviour, it would suggest that humans disregard the level of variance in the evidence, but that generates a choice via a computationally frugal mechanism that relies exclusively on evidence summation.

Finally, we proposed a related model that has successfully accounted for human decisions in several other settings: the Adaptive Gain model, according to whose mechanism samples that are inconsistent with the running mean (i.e. surprising samples) are down-weighted relative to those that concur with expectations (Cheadle et al., 2014; Li et al., 2018). In the context of this task, the process is implemented by summing the grating angles over time after modulation by a normalisation term that depends on prediction error. The Adaptive Gain model is influenced by differing levels of evidence variance as, when evidence variance is higher, these inconsistent samples are more likely to occur. We chose to include this model because a very similar account captures qualitative features of human behaviour in tasks involving a fixed number of samples. If this model were to provide the best account of human behaviour, it suggests that humans account for the level of variance in the evidence via an parsimonious mechanism.

The fact that the different quantities integrated by the SPRT and the Adaptive Gain models are affected in differing ways by the evidence strength and the evidence variance is illustrated in **Fig. 2** below**,** in which we plot the decision updates (DU) for these two models as a function of X (i.e. the sampled inputs, relative to the reference, in radians) for simulated data, separately for each level of variance of the generative distribution σ (for the DDM, the decision values are simply equal to X_k, and so the points would lie on the identity line). Note that for both the SPRT and adaptive gain models, the mapping from inputs to decision values varies with σ in a similar fashion, with a squashing (or compression) of more extreme values. These two models thus approximate each other and should be expected to perform similarly in explaining qualitative aspects of human behavioural data. As we shall see below, the key aspect that distinguishes the two is their neural predictions and (relatedly) the assumptions they make about the computational complexity of the quantity being integrated.

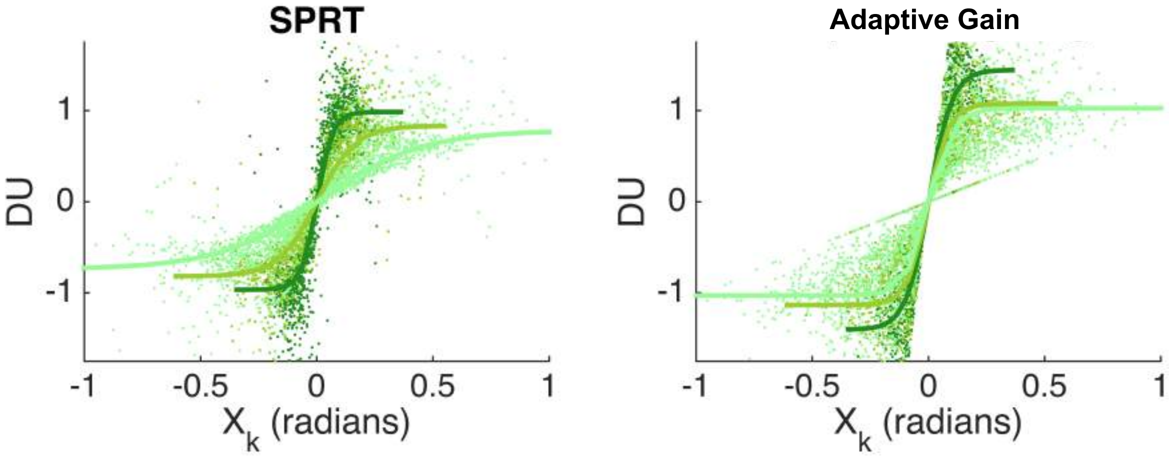


Figure 2: Model predictions for mapping function between evidence level and decision update. The mapping functions are depicted separately for low- (darker colour), medium- (middle colour) and high- (lighter colour) variance distributions, and are based on simulated data. The dots are individual trials and the lines are best cubic fits. The DDM is not displayed because all points for all levels of evidence lie on a straight line (i.e. there is a perfect linear mapping between the level of evidence in a sample (its angular disparity from the reference) and the decision update it generates.

The second key question addressed here pertained to the nature of the decision termination rule (i.e. at what point agents ceased to collect evidence and commit to a choice). Specifically, we wanted to understand whether decision latencies were best captured by integrating the three quantities outlined above to a fixed threshold, or to a threshold that collapsed over time (Hawkins et al., 2015). One biologically plausible means to implement the collapsing bound is to assume that decisions are pushed to termination by a time-varying “urgency” signal that does not depend on sampled evidence. We thus defined two variants of each of the three models that differed only according to whether whether or not they incorporated this ‘urgency’ signal or not. We also include, in a supplement, a normative consideration of the form of the bound in our task.

Advance summary of findings

For the reader’s convenience, we summarise our findings in advance. We found that human decisions and stopping times were captured by models in which evidence was accumulated to a bound that collapsed over time, but not by fixed-bound models. Over three experiments, collapsing-bound models fit the data qualitatively and quantitatively better than fixed-bound models on every metric tested. This behavioural finding was supported by the observation of an evidence-independent neural marker, observed over the parietal and central cortices, that built up over discrete samples towards a response. This signal is indicative of a time-varying “urgency” signal that drove decisions to a close, even in the absence of an explicit deadline (Churchland et al., 2008; Drugowitsch et al., 2012; Hanks et al., 2014b; Murphy et al., 2016; Thura et al., 2012).

We also found that models based on mere summation of sensory signals (i.e. based on the DDM) fared more poorly as explanations of human behaviour than those that computed posterior probabilities of the mean and variance of the generative distributions (i.e. based on the SPRT). We leveraged both behavioural and neural data to ask how such a computation might be implemented. Inspired by previous work involving categorisation of fixed-length sequences of samples, we show that the behaviour of the SPRT can be approximated by a model that accumulates sensory signals into a decision variable in a biased fashion, with reduced weight given to sensory samples that are conditionally surprising, implemented via an adaptive gain control process. The behaviour of this Adaptive Gain model approximated that of the SPRT in a computationally frugal fashion, and allowed it to mimic human behaviour on a range of qualitative and quantitative metrics. Moreover, the Adaptive Gain model successfully predicted that neural encoding of decision information in EEG signals is attenuated on samples which are inconsistent with current beliefs, as previously reported (Cheadle et al., 2014).

These findings suggest that human optional stopping in a heteroscedastic world is dictated by a decision policy that is based on summation of sensory signals, but adaptively down-weights decision information by the surprise it engenders (Summerfield and Tsetsos, 2015). This may be implemented via a normalisation mechanism that allows decision information to be efficiently coded neurally (Carandini and Heeger, 2012). Adaptive gain control in optional stopping tasks approximates the decision policy of a model, based on the SPRT, that optimally integrates posterior likelihood ratios to a collapsing bound.

Methods (Experiments 1-3)

Participants

Eighty-five participants, recruited from the University of Oxford (experiments 1 and 3, EEG version), and undergraduates from University College London (experiments 2 and 3, behaviour-only version), all in the 18-30 age range, took part in three separate iterations of the experiment (n=15, n=37 and n=33 respectively). All participants reported normal or corrected-to-normal vision, and reported no history of neurological or psychiatric disorders. Consent was given in accordance with the ethical guidelines of the Central University Research Ethics Committee at Oxford. Participants were either paid for their participation, or completed the task as part of an undergraduate research project (Table 1).

Exclusion criteria

We defined two criteria for participant exclusion: (i) noncompliance with instructions, and (ii) poor data quality during EEG recordings. Participants were explicitly instructed that to reach satisfactory performance levels, it would be insufficient to simply view the first few samples and hazard a guess. Nevertheless, some participants in experiment 2 (11/38) and experiment 3 (8/34) ignored this instruction as demonstrated by a modal response of 3 samples or fewer and/or a performance level that did not differ significantly from chance (range 49-57%). In experiment 3, we recorded EEG data from 17/34 participants, of whom two were excluded due to excessive movement and/or recording artefacts. Excluded participants were removed from all subsequent analyses, leaving n=15, n=27 and n=24 for experiments 1, 2 and 3 respectively (total n=65).

Task design and stimuli

Task overview

The three experiments reported here were very similar (subject to minor variations, see Table 1 below) and thus we report them together. The instructed goal of the task was to indicate whether the average orientation of a stream of oriented Gabor patches fell clockwise or anticlockwise with respect to a reference orientation. The reference orientation was drawn anew from a circular uniform distribution on each block. Participants viewed as many Gabor patches ('samples') as desired, before committing to a choice via a different button press. The task was challenging because the correct answer was determined by the mean of the underlying generative distribution (not the mean of the samples), such that it was possible for participants to see several samples in a row that appeared to favour the incorrect response. Thus, sufficient sampling of evidence was vital for task success, and a “one and done” strategy would not suffice (Vul et al., 2014). We instructed participants that simply viewing the first few samples would be unlikely to yield high levels of accuracy.

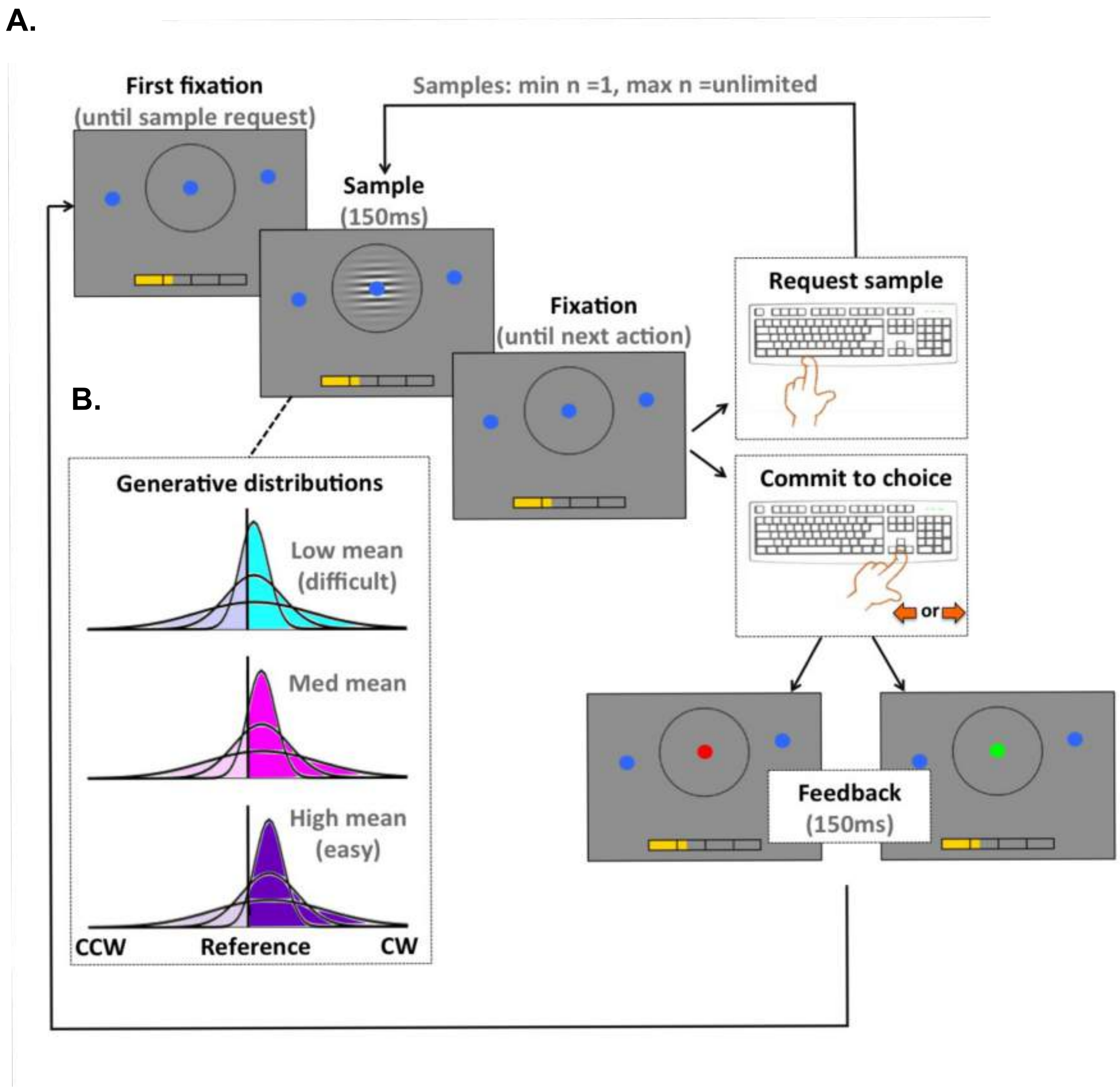


Figure 3: Method and generative structure overview. A. Schematic of an example trial. Participants first viewed a circular placeholder with a central blue fixation point, and two further blue dots indicating the reference orientation (frame 1). They pressed the space bar to request a sample, which appeared with the placeholder for 250ms (frame 2) followed by a further fixation screen (frame 3). After each sample, they could either press again to request another sample, in which case they returned to frame 2 (but seeing a different grating, drawn from the same distribution), or make a choice by pressing one of the two arrow keys (frames 4/5), upon which they received feedback in the form of the fixation dot tyrning

green (correct) or red (error). Cumulative winnings were shown in a yellow bar in the lower part of the screen; this bar increased by a given increment at the same time that the feeback dot changed colour if the answer was correct, and decreased by a larger increment at the equivalent time if the answer was incorrect. **B.** The generative distributions for low mean (cyan), medium mean (pink) and high mean (purple) for trials on which the true mean was clockwise of the reference. The darker colour on each plot reflects the probability of the stimulus having an orientation clockwise of the reference, the paler colour reflects the probability of the stimulus having an anticlockwise orientation. The distributions for trials on which the true mean was anticlockwise were identical, but with the reference line on the opposite side of the distribution means.

Task structure

The stimuli on each trial were drawn independently from a Gaussian orientation distribution whose mean was ±2, 4 or 6 degrees relative to a reference (±0, 2 or 4 degrees on experiment 1, with 66% of trials in the 0 mean condition leading to 'correct' feedback, on a pseudorandom basis), and whose variance was 4, 8 or 16 degrees. Mean and variance were generated independently and randomly on each trial. Sampling from Gaussian distributions was virtually identical to sampling from circular distributions because 99.8% of angles fell within ±45° of the reference. Participants completed 648 and 540 trials in experiments 1 and 2 respectively. In experiment 3, participants performed the task for exactly 1 hour, with no upper or lower limit on the number of trials they would complete.

Task procedure

All experiments took place in a quiet darkened room with participants' faces approximately 70cm from the presentation computer screen. The visual stimuli were presented using the Psychophysics Toolbox running in MATLAB. Before the experiment began, the instructions were explained clearly and participants had the opportunity to practise the task and ask questions. Each experiment was divided into blocks whose lengths were determined either by a time or trial limit (see Table 1). At all times, the display showed a grey screen with a central hollow black circular aperture (Fig 3a), flanked by two reference dots (diameter 5 pixels) that would bisect the centre of the circle if joined with a straight line. The location of the reference dots thus determined the reference angle against which participants compared all incoming stimuli. This location changed randomly block by block, and was drawn uniformly from the full range of possible angles (0-180 degrees), meaning that the reference could represent any possible angle within the circle. An additional dot was present in the centre of the circular aperture for clarity (experiment 3 only).

Each trial was initiated when the participants requested the first stimulus by pressing the spacebar (experiments 2 and 3); in experiment 1 the stream of stimuli onset automatically at a rate of 2Hz. Each stimulus consisted of an oriented Gabor patch that was displayed for 150ms within the circular aperture. The orientation of the Gabor patch was determined by the mean and the variance of the distribution relative to the reference from which it was drawn (see above). Participants continued to sample by pressing the spacebar for each new stimulus (or continuing to view the automatic stream in experiment 1) and committed to a choice (clockwise or anticlockwise) when they felt confident. Visual feedback was given in the form of a green or red central dot (5 pixel diameter) for correct and incorrect responses respectively.

At the bottom of the screen was a 'reward bar', which was partially full at the onset of the experiment, and which corresponded to the amount of bonus money a participant could receive. At the same time as the visual feedback, the bar increased by a given increment if participants were correct and decreased by a higher increment if they were incorrect (see table 1 below). This asymmetric payoff structure was necessary to ensure that the strategy to maximise reward rate was not simply to guess the answer after viewing just one stimulus. In experiments 2 and 3 there was a short delay after viewing the stimulus before participants were able to request another stimulus or commit to a choice. This ensured that stimuli couldn't be requested by accident, and gave time (in EEG) for motor signals to abate. Between trials there was a slightly longer delay during which participants could not initiate a new trial; in experiment 2 the aperture and reference dots disappeared during this time, in experiment 3 the feedback red or green remained on until participants were able to begin a new trial. In experiments 1 and 2, the maximum number of stimuli a participant could view per trial was 20; if they reached 20 without making a choice then "incorrect" feedback was shown and a new trial began. In experiment 3, the number of stimuli a participant could view on any trial was 100, this number was purposely chosen to be far higher than we (correctly) assumed any participant would go. However, had any participant requested 100 samples, pressing the spacebar would no longer have had an effect and they would have been obliged to commit to a choice in order to move on to the next trial.

Table 1: Summary of methodological differences between the three versions of the experiment

	Experiment 1	**Experiment 2**	**Experiment 3**
Aperture/gabor diameter	100px	100px	190px
Number of blocks	6 x 108 trials	5 x 108 trials	4 x 15 minutes

Experiment length	648 trials	540 trials	1 hour
Reward/penalty amount	+1 reward -1 penalty	+1 reward -3 penalty	+2 reward -4 penalty
Reward bar length at experiment onset	Empty	Half full (150 px), reset every block	One third full (100 px), did not reset
Min delay between stimuli	50ms	150ms	Jittered 0.2-0.3 seconds
Min delay between trials	1000ms	150ms	Jittered 1-2 seconds
Bonus structure	Could earn £1 per block, p(reward) corresponded to how full reward bar was	Could earn £2 per block, p(reward) corresponded to how full reward bar was	£2.50 per quarter of max length of reward bar at the end of experiment (e.g. if it was half full, bonus was £5)
Payment	£8 plus bonus	£5 plus bonus	£15 plus bonus (EEG), voluntary (part of course) (behaviour only)
Max stimuli viewable per trial	20	20	100 (limit never reached)
Response buttons for choice	R/L mouse button for CW/ACW	R/L mouse button for CW/ACW	R/L arrow key for CW/ACW

Models of decision latency/decision making

We focussed on three models of interest to predict human behaviour, as outlined above. Each of these models integrated a different quantity, as follows:

SPRT Model

The first model considered was the sequential probability ratio test (SPRT), which specifies the minimum decision time required in order to achieve a given level of accuracy under heteoscedastic conditions. To achieve this, an estimate of $p(correct)$ is required. This was computed using a sequential Bayesian approach that infers, after the occurrence of each stimulus, the posterior probability distribution over an exhaustive space of values for μ and σ and then marginalises over this distribution to compute a (log) posterior ratio for clockwise (CW) vs. anticlockwise (ACW) on each sample k:

$$DU_k^{SPRT} = log\left[\frac{p\big(X_k \sim N\ (\mu > 0, \sigma\ )\big)}{p\big(X_k \sim N\ (\mu < 0, \sigma)\big)}\right] + \mathcal{N}(0, \varepsilon)$$

Equation 1

Where ε is a noise parameter that varies freely. Given that the mean and variance distributions were unknown to participants on this task, we initiated the SPRT model on each trial with a flat prior across distributions with a range of values of mean and variance that spanned twice the maximum value of the generative distributions.

Drift Diffusion Model (DDM)

The second model considered, the drift diffusion model (DDM), simply summated a noisy estimate of the sensory information on each sample:

$$DU_k^{DDM} = X_k + \mathcal{N}(0, \varepsilon)$$

Equation 2

Adaptive gain model

The final model considered also bases its decision update on the information provided by each sample, but weights this information according to a prediction error term such that more deviant stimuli are incorporated with lower gain:

$$DU_k^{AG} = \frac{X_k}{c + \left|X_k - \hat{X}_k\right|} + \mathcal{N}(0, \varepsilon)$$

Equation 3

where $\hat{X}$ is the expected value of X on sample k, given by the mean tilt of all stimuli up to $k - 1$, and c is a small regularisation constant.

Stopping rules: fixed versus collapsing bounds

Two versions of each of the three models were defined: one that incorporated a fixed bound, and one that incorporated a collapsing bound, or equivalently, an urgency signal that grows with each sample independent of the sensory information it conveys. For both versions, decisions were triggered following any sample k when the model stopping variable (SV) exceeded a freely-varying threshold parameter τ. The stopping rule (SR) was computed for each model as follows:

$$SR_k^{model} = \left(\sum\nolimits_{k=1}^{j} DU_{model,k}\right) \cdot \gamma^{j-k} + \lambda k$$

Equation 4

Where k denotes the current sample, j is the total number of samples taken on that trial, and γ is a leak parameter (akin to the gradual decay of information in memory). The parameter λ denotes the gain of a stimulus-independent "urgency" signal that drives the stopping variable towards the bound even when the input is ambiguous. We set λ to zero (no effect) for fixed bound models, and γ to 1 (no effect) for collapsing bound models, thereby ensuring an equal number of free parameters across models.

Free parameter estimation

For all models, we allowed the boundary height ('threshold') and noise term ε to vary as free parameters. For fixed-bound variants, we also included a leak term that captured potential loss of information during integration. For collapsing bound models, we replaced the leak parameter with a term that allowed the threshold height to decline over time by a fixed amount on each step, leading to a linearly collapsing bound. We incorporated the leak because previous work has shown that a leak of the impact of evidence on behaviour over time is observed in human behaviour (Usher and McClelland, 2001; Wang, 2002) . The collapsing bound models posit a qualitatively different impact of the passage of time on behaviour, thus incorporating a leak into the fixed bound models provides a reasonable benchmark comparison to the collapsing bound urgency signal.

The best fitting parameter values – noise, threshold, and leak or urgency – were determined via an exhaustive search. We identified parameter values that gave rise to the lowest mean square error (MSE) between the model and human reaction time distributions (see below). These best fitting parameter values were then used for all further analyses. Model choices were generated on the basis of the sign of the model decision variable at the point that the decision was triggered. All formal model comparisons were conducted on log likelihoods using the VBA toolbox (Daunizeau et al., 2014). All models had an identical number of free parameters and thus no adjustments were required to the scores entered into these comparisons.

Normative considerations

None of the models proposed above is strictly optimal. An optimal model of this task requires estimation of likely future rewards under a policy that decides whether to stop now vs. later, a computationally costly estimation problem that relies on recursive estimation methods. It is nevertheless of interest what form the optimal bound might take in the heteoscedastic case. In the supplementary materials (**Fig. S1** and accompanying text) we solved this problem as a partially observable Markov decision process. For the statistics of the task described here, the optimal the bound collapses over the first few samples and then increases slightly over the remaining samples. We include this analysis for completeness; the major focus of the current work is to provide a descriptive, rather than a normative, account of decisions in the discrete-sample integration case.

Results

Behaviour data and model fitting

Using a first passage process – one that determines the first point at which the stopping rule was triggered for a given model on a given trial – we fit the SPRT, DDM and Adaptive Gain models to human stopping times on correct and error trials in the 3 behavioural experiments described. In what follows, we compared models via quantitative metrics (i.e. log likelihoods) and random effects analysis (i.e. Bayesian model selection), but we also simulated best-fitting variants of our models to generate predictions about qualitative features of the human data, allowing us to draw strong conclusions about the computations underlying optional stopping based on model falsification (Palminteri et al., 2017).

Human stopping time distributions are better predicted by collapsing rather than fixed bound models

We first compare the fits to stopping time distributions separately for correct and error trials (**Fig. 4** for experiment 3 (the EEG dataset), **Fig. S2** for experiments 1 and 2). Visual inspection suggests that the collapsing-bound models (upper panels) fit better than the fixed-bound models (lower panels) in all three experiments, and this was confirmed by frequentist comparison of the mean-squared error for every pairwise comparison among models (Exp.1: all t-values > 6.42, all p-values < 0.001; Exp. 2: all t-values > 5.18, all p-values < 0.001; Exp. 3: all t-values > 2.42 p-values < 0.05). For a more formal quantitative analysis we calculated model log likelihoods of the model having the same ST distribution as humans (calculated separately for correct and error trials and then combined). We then pooled over all three experiments to compare these model log likelihoods using Bayesian model selection. This analysis strongly favoured the collapsing bound models (expected frequencies for fixed-bound models were all less than 2%, whereas chance is 16.7% for the 6 models; and were 31.4%, 23.6%, and 41.3% respectively for the SPRT, DDM and Adaptive Gain models with collapsing bound).From the 6-way comparison among models, expected frequencies for the collapsing-bound SPRT and Adaptive Gain models were above chance (p < 0.002 and p < 0.001 respectively) but did not exceed chance for the DDM with collapsing bound (p = 0.08).

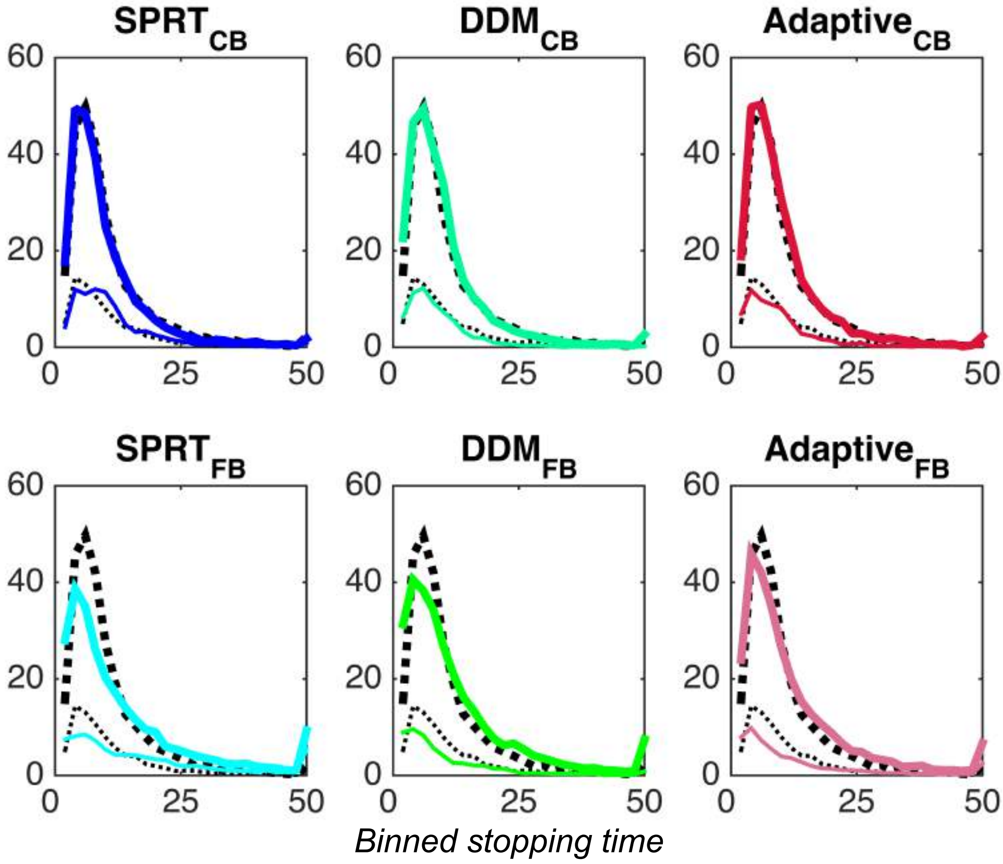


Figure 4: Stopping time distributions. The black dashed lines depict the human distributions of stopping times for correct (thicker dashed black line) and error (thinner dashed black line) trials, superimposed on the corresponding fits of collapsing-bound (upper panels) and fixed-bound (lower panels) variants of the SPRT (left panels), DDM (middle panels) and adaptive gain model (right panels)

The best-fitting variants of each collapsing-bound model were then used (without further assumptions, i.e. with the same parameters) to make predictions about four qualitative features of the human data: (i) condition-wise mean stopping times; (ii) predictors of next-sample commitment; (iii) adaptive weighting of information by the sample history; and (iv) weighting of sensory features in choice. For these comparisons, we focussed on collapsing-bound models because the fixed-bound models fit the stopping time distributions more poorly. However, for completeness, we present the full results in the **Fig. S3** (fixed-bound models provide an inferior fit to the human data on every measure considered here).

Evidence strength and reliability interact in predicting mean stopping times

We compared human data to the model predictions concerning mean stopping times for each level of $|\mu|$ and σ. Stopping times are log-transformed in these plots for compactness (**Fig. 5**). The SPRT and Adaptive Gain models predicted that stopping times should vary as an interaction of by $|\mu|$ and σ, with commitments delayed when the samples were drawn from more variable distributions, but this slowing exacerbated as $|\mu|$ deviates further from zero. This occurs because for both of these models, outlying feature values (i.e. tilts that are more deviant from the reference) are penalised more sharply during conversion to momentary decision signals. In the SPRT, this occurs because outlying tilts signal are more likely to be drawn from more variable generative distributions, decreasing certainty about the estimates of μ. In the Adaptive Gain model, this occurs because outlying tilts provoke larger prediction errors, leading to steeper normalisation of the update to the stopping variable. In both cases, this leads to shallower slopes of integration where samples are more variable, in particular when μ is further from zero. By contrast, the DDM predicts an inversion of this effect: that stopping times should be shortest when samples were more variable. This latter prediction follows from the dynamics of bounded integration under the DDM, whereby more variable evidence ensures that the diffusion process occupies a potentially wider range of states, and is thus more likely to contribute to a first passage process, in particular when evidence strength is low (Zylberberg et al., 2016). Human stopping time data displayed in the interaction between $|\mu|$ and σ predicted by the SPRT and Adaptive Gain models in all 3 experiments (Exp.1 $F_{2.6,36.9} = 11.39$, $p < 0.001$; Exp.2: $F_{3.3,85.4} = 5.37$, $p < 0.001$; Exp.3: $F_{3.6,82.1} = 4.41$ $p < 0.004$; ANOVA on log stopping times). Comparing least-squares fits of the models to these data aggregated over experiments, we found that the DDM performed overall worse than the other models (t-test on mean squared error: DDM vs. SPRT $t_{65} = 2.30$, $p < 0.02$; DDM vs. Adaptive Gain $t_{65} = 4.39$, $p < 0.004$). This confirms the pattern suggested by the quantitative fits, and argues against the DDM as a full description of human optional stopping in our experiment.

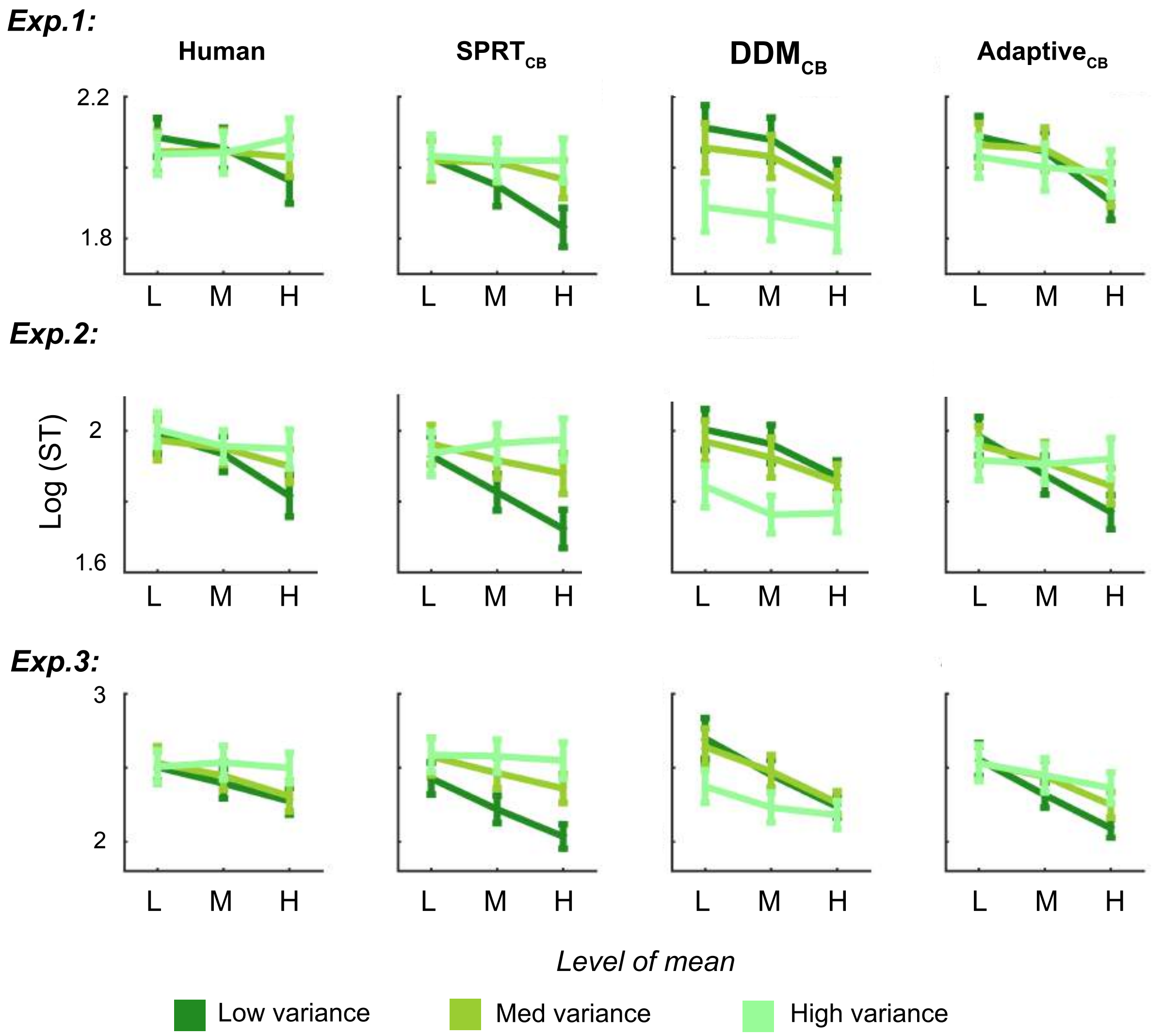


Figure 5: Stopping times as a function of evidence strength and variance. Average (log) stopping times (in samples) for trials with low (L), medium (M) and high (H) $|\mu|$, i.e. distance to the reference (x-axis) and low, medium and high σ (dark, medium and pale green lines respectively). Leftmost panel: humans; other panels, collapsing bound models. See **Fig. S3** for fixed bound models. Rows 1-3 are experiments 1-3 respectively. The predictions of the SPRT and Adaptive Gain qualitatively match the human data, where as the DDM does not.

Evidence variability delays next-sample commitment

Next, we capitalised on our use of a task that permitted precise quantification of the information (i.e. momentary tilt values) provided by each sample. We used probit regression to predict whether participants would decide to commit or defer on a given sample (Malhotra et al., 2017), using three time-varying quantities as predictors: (i) the absolute sum of feature values up to that sample $sum(X)$; (ii) the number of samples viewed thus far; and (iii) the sum of squared feature values $sum(X^2)$. This choice of predictors was motivated by the observation that in the heteroscedastic setting provided by our experiment, these three variables are sufficient statistics for the computation of $p(correct)$ (see **Fig. 1**). Once again, the DDM makes a prediction that diverges from the other models, in that it proposes that the sum of momentary evidence, rather than the passage of time through the trial (i.e. samples elapsed), should be the strongest predictor of next-sample commitment. The other models predict that samples elapsed is the strongest predictor of next-sample commitment, and that a larger sum of squared evidence has the effect of delaying commitment. As **Fig. 6** shows, this pattern was observed in human data (t tests against 0 for the beta weight corresponding to N and X^2 respectively, Exp 1: t_{14} = 10.34, p < 0.001 and t_{14} = -5.48, p < 0.001; Exp2: t_{26} = 7.35, p < 0.001 and t_{26} = -3.45, p = 0.001; Exp 3: t_{23} = 6.12, p<0.001 and t_{23} = -4.14, p = 0<0.001). This again provides evidence against the DDM, which does not predict this pattern, as an adequate model of human performance in our task. To assess the significance of the model fits to human behaviour, we also used coefficients from each model to predict human stopping time and compared the resulting model likelihoods using Bayesian model selection aggregated over the 3 experiments. Comparing among the 3 collapsing bound models, expected model frequencies were 54.6%, 7.7% and 37.8% for the SPRT, DDM and Adaptive Gain models respectively, indicating a clear disadvantage for the DDM.

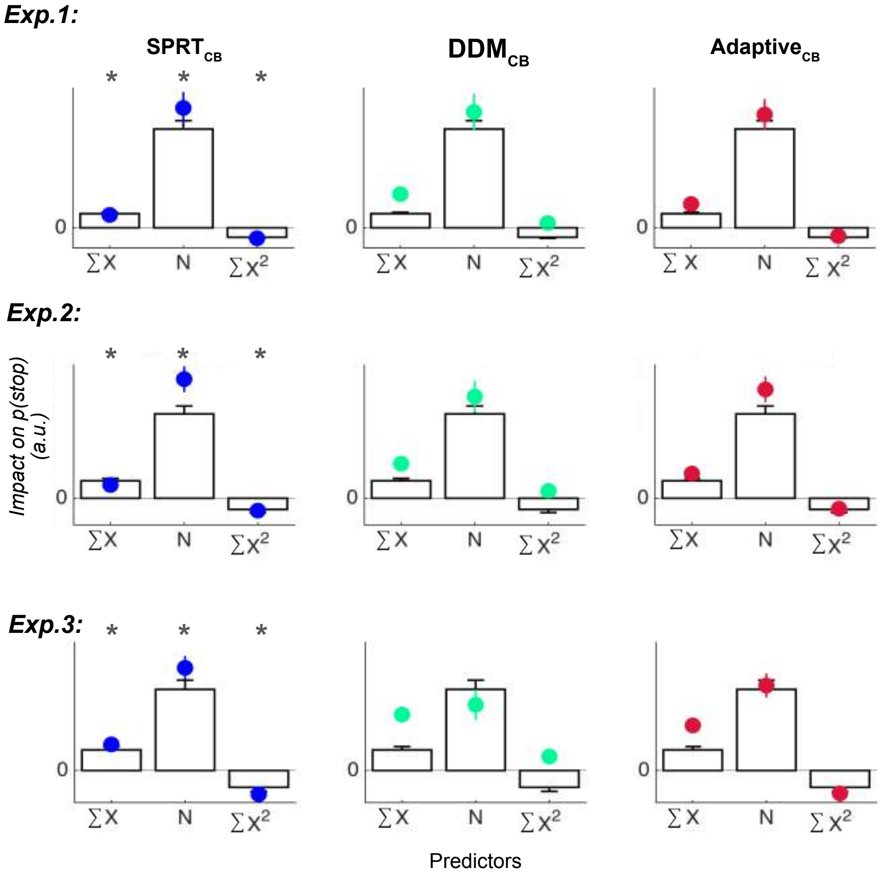


Figure 6: Predictors of next sample commitment. Coefficients from regressions predicting next-sample commitment based on sum(X), samples elapsed (N) and sum(X^2). White bars are human data and asterisks denote that the beta weight deviated significantly from 0 (i.e. the predictor in question had a meaningful impact on next sample commitment); note that these asterisks are only displayed on the leftmost panel but apply to all human data in a given row, as this is identical for all three panels. Each column shows the fit of a model (blue, green and red dots reflect SPRT, DDM and Adaptive Gain models respectively), and each row 1-3 depicts experiments 1-3 respectively. The SPRT_{CB} and Adaptive gain_{CB} models capture the negative influence of sum(X^2), but the DDM_{CB} does not.

Inconsistent samples have lower impact on choices

A third qualitative prediction that can be used to arbitrate among the models focuses on how the statistics of the samples observed on each trial determines their weight (or impact) on the categorical choice, i.e. CW vs. ACW response. It follows from the Adaptive Gain model that sensory samples that are incongruent with those occurring previously will carry less weight in eventual choices, because they provoke larger prediction errors. To quantify this effect in all 3 models, we used a previously described approach, in which probit regression is used to predict choices (CW vs. ACW) as a function of (i) the sum of sensory evidence (i.e. tilts), (ii) the absolute difference between each grating and its predecessor, and (iii) the interaction between (i) and (ii). In previous studies involving integration of a fixed number of samples, we have observed negative coefficients for (iii), indicative of a downweighting of surprising or inconsistent evidence (a "consistency" bias) (Cheadle et al., 2014). Here, the human consistency bias was significant in all 3 experiments (Exp.1: $t_{14} = 14.04$, $p < 0.001$; Exp. 2: $t_{26} = 6.50$, $p<0.001$; Exp. 3: $t_{23} = 6.08$, $p<0.001$, showing that the more inconsistent a sample was, the less weight it carried on the eventual choice (see **Fig. 7**). This pattern was replicated by the SPRT and normalisation models, but not by the DDM. Quantitative analysis using Bayesian model selection revealed lower expected frequencies for DDM than other models (2% vs 85% for SPRT and 12% for Adaptive Gain model; $p < 0.001$).

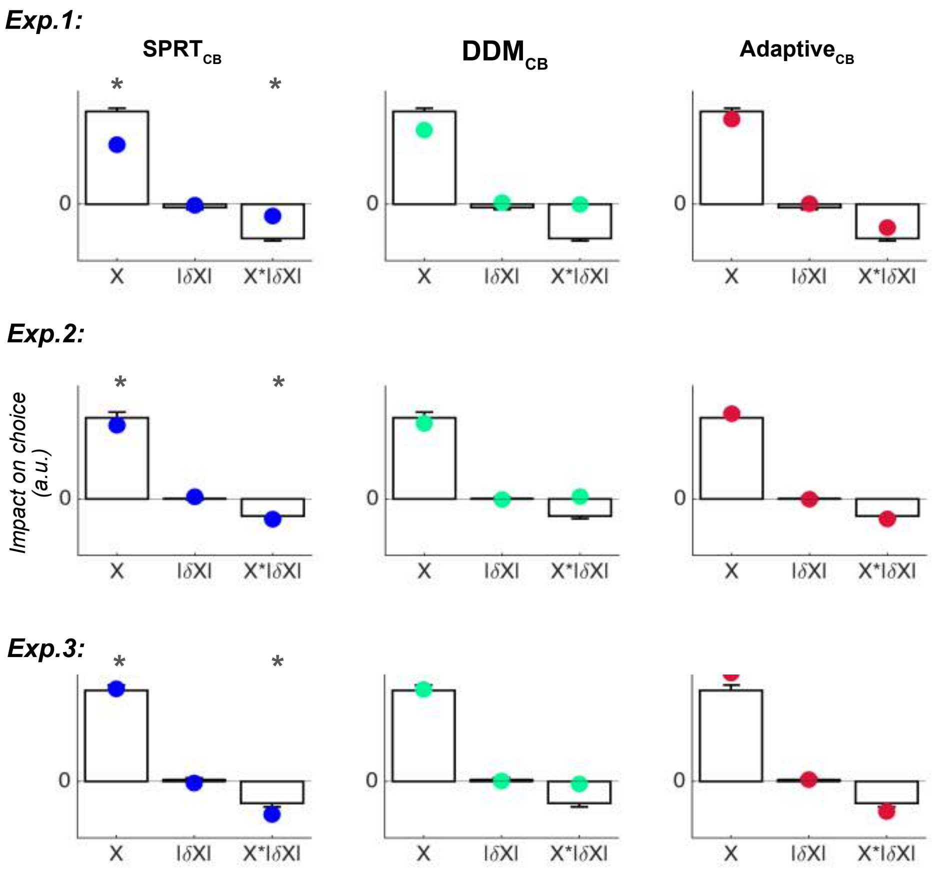


Figure 7: Predictors of choice. The impact on $p(choose\ clockwise)$ of sample information (X), the prediction error $|\delta X|$, and the interaction of the latter two. The Adaptive gainCB model predicts a negative interaction, meaning that the strength of impact of a stimulus on the decision decreases as the prediction error associated with that stimulus increases. White bars are human data and asterisks denote that the beta weight deviated significantly from 0 (i.e. the predictor in question had a meaningful impact on next sample commitment); note that these asterisks are only displayed on the leftmost panel but apply to all human data in a given row, as this is identical for all three panels. Each column shows the fit of a model (blue, green and red dots reflect SPRT, DDM and Adaptive Gain models respectively), and each row 1-3 depicts experiments 1-3 respectively.

Variance influences the weighting of sensory features in choice

A fourth prediction that differs qualitatively among the models concerns the form of the likelihood function that maps tilt values onto choices for different levels of σ. The expected form of this function for the SPRT and Adaptive Gain models is plotted in **Fig. 2**. In humans, this function was estimated by counting grating tilts that fell within 7 bins of angle with respect to the reference, and entering these tallies into a design matrix that was then used to predict choices via (robust) logistic regression. We conducted this analysis independently for trials defined by each level of σ**.** The resulting coefficients are plotted in **Fig. 8**, for humans as well as for the 3 collapsing bound models using an identical analysis of model choices. As noted above, the DDM predicts that this mapping function should be linear, and thus virtually identical for all 3 variance conditions. The SPRT and normalisation models, by contrast, predict a choice function that is sigmoidal (rather than linear) in form, with outlying tilt values "squashed" or downweighted. Moreover, they both predict that these sigmoidal choice functions will have a steeper slope for low variance than high variance conditions; this effect is more exaggerated for the SPRT model. The human data are shown for comparison in the leftmost column of the figure. We compared slopes of sigmoidal functions fit to human choice functions for each variance conditions separately, finding them to be reliably steeper for the low variance condition in all three experiments (Kruskall-Wallis test, Exp.1: $p < 0.02$, Exp.2: $p < 0.005$, Exp.3: $p < 0.001$). This finding argues against the DDM, which implements a decision policy that does not account for variability in the generative distribution on each trial.

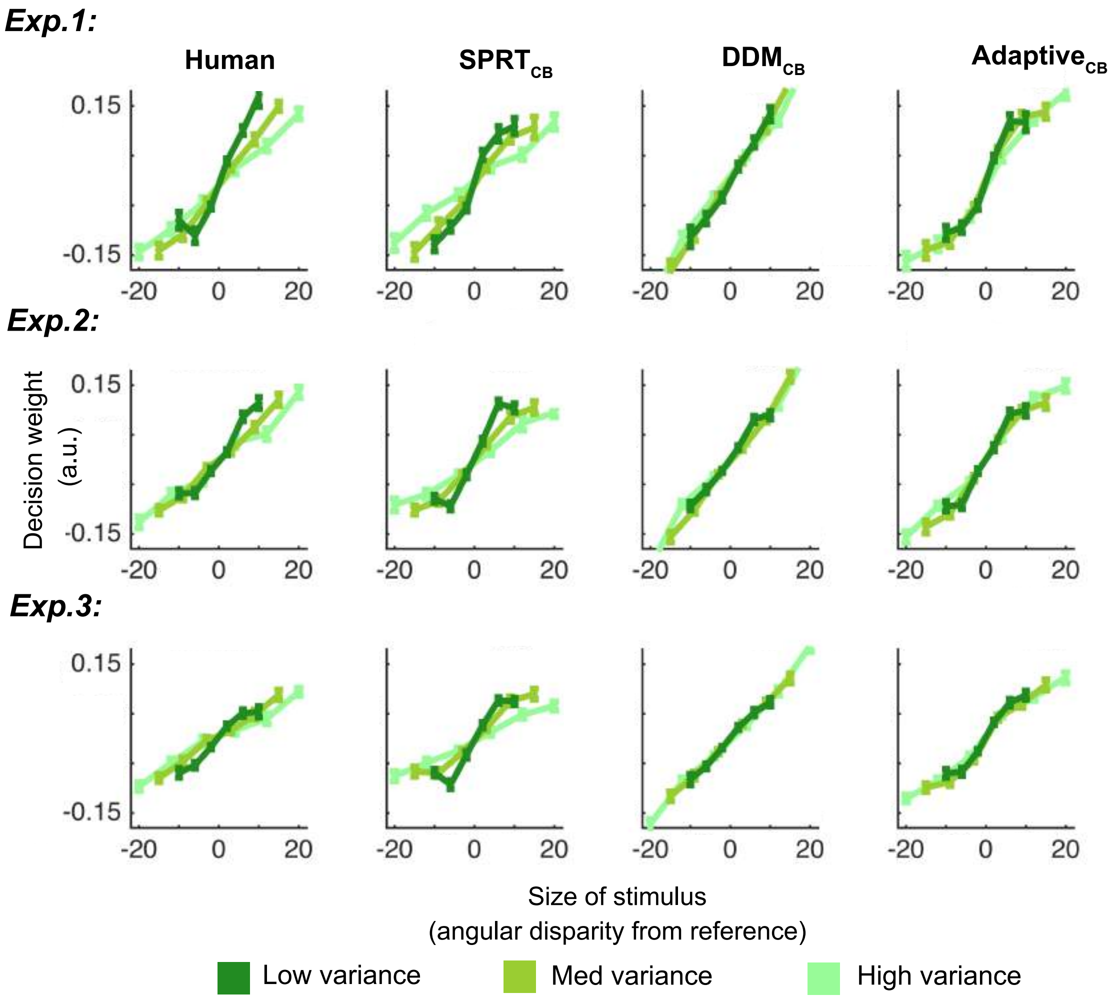


Figure 8: Subjective versus objective weighting of the decision evidence. The average subjective weighting the evidence carried on choice is depicted separately for the three conditions of variance

(low, medium and high in pale, medium and dark green respectively). The impact of more extreme angles, and angles from more variable streams, is lower in the human data. This is captured by the Adaptive Gain and SPRT models, but not the DDM. See also **Fig. 2**, which shows the expected pattern based on simulated data.

Summary of behaviour data

Together, these behavioural analyses provide a rich picture of the dynamics of human optional stopping in a heteroscedastic environment. We draw three conclusions. Firstly, models based on a fixed bound cannot jointly explain human stopping times and choices. Fixed-bound models fit more poorly than collapsing-bound models on every qualitative and quantitative metric used here, providing a notably clear demonstration that in a task involving integration of discrete samples of information, humans do not draw decisions to a close using a flat decision threshold (**Fig. 4**). Secondly, models that involve unbiased linear accumulation of evidence to a collapsing bound (e.g. based on the DDM) fail to explain the qualitative pattern of stopping times as a function of evidence strength and reliability (**Fig. 5**), fail to account for how the statistics of stimulation predict next-sample commitment (**Fig. 6**), and fail to capture how the disparity between current and previous sample value influences choices (**Fig. 7**), or to correctly predict the form of the choice function that humans use to map sensory features onto decision evidence (**Fig. 8**). These qualitative observations seem to rule out an unbiased, linear integration process as a good description of human choice computations in this task. Thirdly, we report that the slope of the human weighting function (transducer) varied as a function of the dispersion of sensory features in the stimulation sequence. This suggests that humans either inferred the generative statistics of the sequence online during stimulation (SPRT model), or used an approximation that allowed them to behave as if they were doing so (Adaptive Gain model). However, behavioural data alone is not able to distinguish between these two accounts.

Electroencephalographic (EEG) recordings

In this final section, we report EEG data collected from a subset of participants who took part in experiment 3. We first describe our EEG methods, and then go on to outline a series of analyses aimed at disambiguating the SPRT and Adaptive Gain models on the basis of how various model-predicted quantities are encoded in human brain signals.

EEG acquisition and preprocessing

EEG signals from 64 Ag-AgCl electrodes, plus four additional electrodes used in bipolar montage as horizontal and vertical EOGs, and two electrodes used as references at the mastoids, were recorded using a Neuroscan EEG system with NuAmps digital amplifiers. The EEG signals were recorded at a sampling rate of 1 kHz and were high pass filtered online at 0.1Hz. The impedances of all functioning electrodes were below 10 kΩ at recording onset.

All preprocessing was completed using MATLAB custom scripts and the EEGLAB toolbox (Delorme and Makeig, 2004). First, data were downsampled to 250Hz, band pass filtered between 1 and 40Hz, then split into epochs of -1 to 1 second from each stimulus (sample) onset. It was necessary to epoch sample-wise rather than trial-wise due to the uneven number of samples taken per trial, and therefore widely varying trial lengths. The data from all channels were then visually inspected to remove epochs containing high frequency noise typical of muscle artefacts or electrical surges. Rejected samples formed no further part of the EEG analyses. Data from consistently bad channels (observed either during recording, during visualisation of the data, or both) were interpolated spherically (range of channels interpolated per participant: 0-5, average 2.9 channels interpolated per subject). The data were then re-referenced to the average signal. An extended independent component analysis (ICA) was then conducted, and components with activity typical of blinks or electrical noise were rejected. Finally, data were baselined relative to the first 250ms in each epoch (-1000 to -750ms preceding sample onset).

Broadband EEG analyses: encoding

The logic of our analysis was that although the predictions of the Adaptive Gain and SPRT mdoels are similar for behaviour, they differ in the way that the output (decision update) is computed. The Adaptive Gain model calculates its decision update by summing the evidence provided on each sample after weighting it according to the prediction error (see **equation 3**). Thus it predicts the existence of neural signals encoding the momentary level of evidence (the angular disparity between the stimulus and the reference), the prediction error, and the interaction between the two. The SPRT model also predicts a representation of the level of evidence on the sample, as the quantity is central to its decision update process. Over and above this, the SPRT model predicts a neural representation of the log posterior ratio of evidence, i.e. the quantity the model actually integrates (see **equation 1**). To ascertain which of these predictions was best supported by our data, rather than calculating event related potentials (ERPs), we took parametric predictors within a general linear regression model to predict sample to sample fluctuations in the average EEG activity across frontal (F1, Fz, F2, FC1, FCz and FC2) and parietal (P5 P3

P1 Pz P2 P4 and P6) electrodes. We regressed 9 of these predictors, corresponding to decision information, time elapsed, and nuisance variables including pertinent information from the samples immediately preceding and following (see **equation 5**), against the single sample EEG activity, point by point over 500 evenly spaced datapoints that spanned from -1s to 1s from sample onset.

Our single-trial approach employed multiple linear regression (Wyart et al., 2012; Wyart et al., 2015). This allowed us to assess how the strength of the neural signal is parametrically influenced by decision-relevant quantities such as the DU predicted by each model. To do this, we evaluated the following regression model on single-trial EEG data:

$$EEG_{k,t} = \boldsymbol{\beta_1} \cdot \left|DU_k^{SPRT}\right| \dots$$
$$+\, \boldsymbol{\beta_2} \cdot |X_k| \; + \boldsymbol{\beta_3} \cdot |\bar{X}_k - X_{k-1}| + \boldsymbol{\beta_4} \cdot (|\bar{X} - X_{k-1}| \cdot |X_k|) \dots$$
$$+\, \boldsymbol{\beta_5} \cdot -\log(k) \dots$$
$$+\, \boldsymbol{\beta_6} \cdot |X_{k-1}| + \boldsymbol{\beta_7} \cdot |X_{k+1}| + \boldsymbol{\beta_8} \cdot |X_0| + \boldsymbol{\beta_9} \cdot \left|X_j\right| + \boldsymbol{\beta_0}$$

Equation 5

Equation 5 is a complex expression and so we have colour coded the terms for readers' convenience. The term in red refers to the predicted decision variable from the SPRT model: $|DU_k^{SPRT}|$ is the absolute of the log posterior ratio of evidence (from equation 1). Note that we use the absolute (unsigned) DU because we expect univariate EEG signals to vary with certainty, rather than differing according to whether evidence was clockwise or counterclockwise. The terms in blue encode the predictions of the Adaptive Gain model. $|X_k|$ represents the absolute (rectified) momentary decision information conveyed by the relevant sample k. The prediction error term $|\bar{X} - X_{k-1}|$ denotes the absolute angular difference between the current stimulus and the mean of all previous stimuli in the current trial up to that point, i.e. the prediction error term from the Adaptive Gain model. The Adaptive Gain model predicts an interaction between this prediction error term and the momentary evidence X_k, which is encoded by the coefficient β_4. We also included the (negative log) number of samples elapsed (green), which starts large on the first sample and gradually shrinks as the trial progresses, akin to an adaptation signal. The final four terms in black ($\beta_6 - \beta_9$) are nuisance predictors that signal decision information on the previous ($|X_{k-1}|$) and subsequent ($|X_{k+1}|$) samples, and variance associated with the first sample ($|X_0|$) and last sample, i.e. that which prompted commitment ($|X_j|$). Term β_0 is the intercept.

Each predictor was regressed against the EEG activity evoked by each sample to derive regression coefficients that encoded the slope of the relationship between the quantity of interest and the EEG signal at successive timepoints relative to sample onset. The resulting coefficients were then averaged across trials, from stimulus onset to 800ms following stimulus onset, for second-level (group) analysis; time periods at which these weights deviated significantly from 0 indicate that the relevant quantity is being encoded with above-chance strength, correcting for multiple comparisons by using a non parametric cluster correction technique with a familywise error of 0.05 (Maris and Oostenveld, 2007).

Given the entirely self-paced nature of experiment 3, epochs from adjacent samples sometimes had a degree of overlap. Although this should be accounted for by our regression approach, as an additional safeguard we included the evidence from the preceding and following samples as nuisance regressors (see **equation 5**). The methods we used for disambiguation are thus akin to those used in parametric event related functional neuroimaging designs (Josephs, Turner and Friston 1997). Entering all of the quantities in the same regression ensured that they competed for unshared variance, meaning that we could be confident that any effects were the result of the predictor's unique influence on the EEG signal.

Both the SPRT and the Adaptive Gain models predict an influence of the momentary decision evidence $|X|$, as this quantity is required to derive the decision update for both models. However, only the Adaptive Gain model predicts significant encoding of the prediction error term, as well as its interaction with the decision evidence $|X|$, as these quantities are used to compute the model's decision update. The DU of the SPRT, on the other hand, which is equivalent to the log posterior ratio of evidence, is predicted to have an influence on neural signals only under the SPRT.

Broadband EEG analyses: results

The results are show in **Fig. 9a** (parietal) and **9b** (frontal). Encoding of $|X|$ exhibited a characteristic negative-positive deflection at ~200/450ms over parietal electrodes (red trace; upper panels), and the sign reversed pattern in frontal electrodes (lower panels) that closely replicates the encoding of decision information described previously (Tickle et al., 2016; Wyart et al., 2012; Wyart et al., 2015). However, a number of other predictors explained variance in EEG data. Consistent with the predictions of the Adaptive Gain model, there is significant encoding of the prediction error signal over parietal and frontal electrodes (again with sign reversed pattern in the two regions) 400ms after stimulus onset. This signals a component of the EEG signal that was modulated when samples viewed were incompatible with the running stream of evidence. Importantly, in both regions there was also a significant interaction between the prediction error and the decision evidence, meaning that the strength of encoding of the decision information was modulated by the degree of prediction error, as predicted by the Adaptive Gain model. These findings replicate those previously reported for sequences of fixed length (Cheadle et al., 2014).

In contrast, parietal regions showed no significant encoding of the LPR during this time period, i.e. the quantity accumulated by the SPRT model. The lack of such a signal in parietal regions is particularly salient given previous links between this region of cortex and the accumulation of decision-relevant information (Hanks et al., 2006; Kiani et al., 2008; Kira et al., 2015; Roitman and Shadlen, 2002).

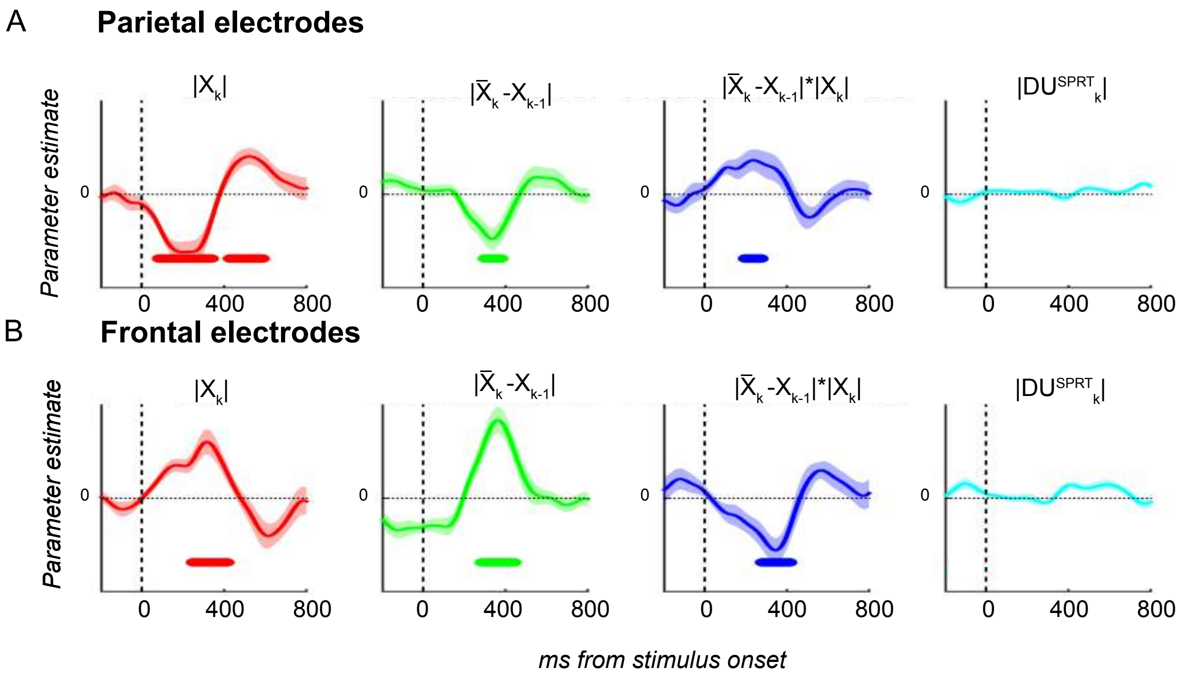


Figure 9: **Neural encoding of model-predicted information**. Curves depicting the encoding of model-predicted quantities in EEG signals (beta series) for (A) parietal and (B) frontal regions from -100 to 800ms following each sample. The shaded regions around the solid lines depict SEM. The coloured lines show periods of time during which the signal significantly deviated from zero, corrected for multiple comparisons. The Adaptive Gain model predicts the encoding of the prediction error signal that is seen here (2nd panel from left), as well as the inverse modulation of encoding of sensory information by prediction error (3rd panel from left). The SPRT model predicts encoding of the LPR, but no significant encoding is seen in either parietal or frontal regions.

Thus analysis of the signal over parietal and frontal regions supports the notion that the quantities computed by the Adaptive Gain model, rather than those from the SPRT model, are observed in human brain signals.

EEG analyses: Lateralised beta band activity

Previous studies have shown beta band desynchronisation in the build up to motor execution of a decision in the contralateral hemisphere to the hand making the choice. In our experiment, samples were requested with the left hand and a decision commitment was made by the right. Thus we would expect to see desynchronisation in the left hemisphere, i.e. contralateral to the response hand. Furthermore, this signal should be influenced by quantities that are predictive of choice execution. In particular, our paradigm allowed us to determine whether the cumulative decision information and/or time elapsed were more predictive of this lateralised desynchronisation motor preparation signal.

To this end, we computed the interhemispheric difference in neural activity across 10 logarithmically spaced frequency bands (~10-40Hz) at lateral central electrodes, by subtracting the spectral log power of C3+CP3 from C4+CP4. We then regressed the cumulate of the angular disparity, the log of the sample number (i.e. time elapsed) against the resulting signal, in order to ascertain whether there was an evidence-independent neural signature of urgency. As controls, we also included the inverse of the urgency signal, i.e. the quantity that started large and decreased over the course of a trial, to control for the large neural adaptation signal typically seen, and an indicator variable (1 or 0) that denoted whether that sample had immediately preceded the choice committment. (See **equation 6** below). The resulting parameter estimates were averaged across samples separately for each frequency band before being entered into group-level statistical analyses, in which we used a nonparametric cluster correction technique with familywise error of alpha = 0.05 to assess significance.

An independent signal corresponding to time elapsed ("urgency") is present in the lateralised frequency band activity

In behavioural comparisons, collapsing bound models captured reaction time data more faithfully than the fixed bound models, implying that a time-sensitive signal that is independent of the cumulative evidence is influential in driving choices to a close. Our next step thus sought to determine whether such an "urgency" signal was present in our data. We motivated the analysis on the basis of previous work indicating that desynchronisation in beta band activity in central regions of the hemisphere contralateral to the hand that executes the motor response is strongly present in the build-up to choice selection (Donner et al., 2009). Our paradigm was suited to addressing this question because by design, participants requested samples with the left hand and committed to choice with right hand. Thus, beta-band activity should gradually desynchronise over the left motor cortex (contralateral to the commitment) as proximity to the response increases, even after excluding the final sample where the response was executed. Thus, unlike previous psychophysical paradigms, our self-paced, discrete-sample integration task offered a unique opportunity to test for a neural urgency signal that was unconfounded with motor execution.

Using a wavelet filter, we computed relative (contralateral minus ipsilateral) activity in the alpha/beta bands (9-40 Hz) from central electrodes (C3/CP3 and C4/CP4). We then evaluated the following regression:

$$EEG\_beta_{k,t} = \beta_0 + \beta_1 \cdot \left|\sum_1^k X\right| + \beta_2 \cdot -\log(j-k) + \beta_3 \cdot \sum_1^k DU_{SPRT} + \beta_4 \cdot -\log{(k)}$$

Equation 6

In equation 6, we include coefficients for the cumulative sum of evidence (β_1), an “urgency” predictor that encodes proximity to commitment in a nonlinear fashion (β_2), the cumulative decision quantity predicted by the SPRT model (β_4), and a nuisance predictor in the form of a quantity which starts large and decreases as the trial progresses, to account for spurious adaptation-type neural effects (β_4). This regression was evaluated at each timepoint from -500 to 500ms following sample onset, for 10 logarithmically spaced frequency bands in the alpha/beta range. We found that there was a significant independent influence of proximity to the choice, indicative of an ‘urgency’ signal influencing motor response (**Fig. 10)**. However, there was no effect of cumulative decision information or of the quantity predicted by the LPR model.

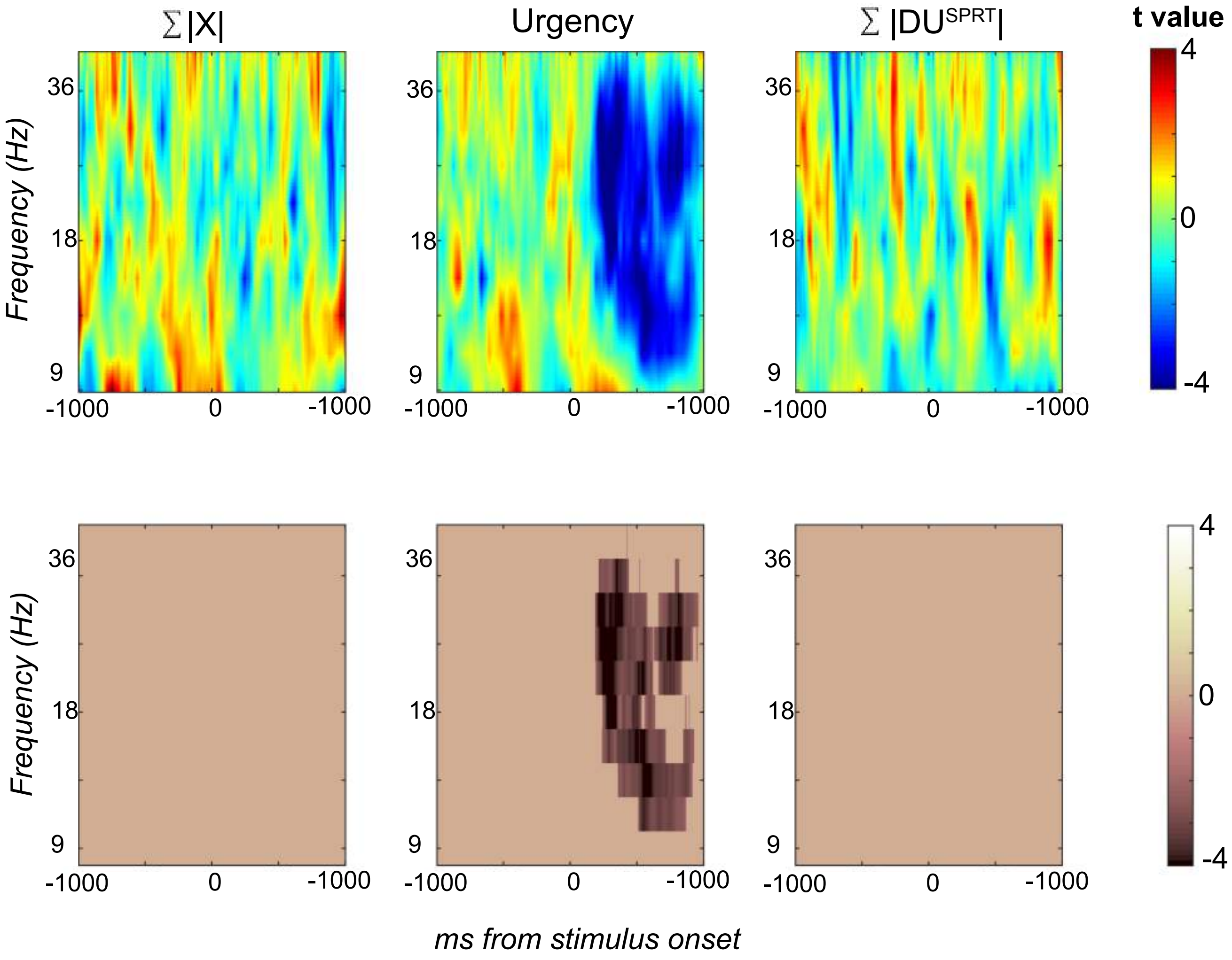


Figure 10: **Lateralised beta band activity.** T values derived from beta weights from a regression assessing the correlation between the cumulative decision information (left column) log time elapsed (middle column) and the cumulative information encoded by the LPR model (rightmost column) and the interhemispheric difference in the EEG signal from -500 to 1000ms following stimulus onset. Desynchronisation in beta band activity is associated with preparation to execute response; factors that influence this process should therefore cause greater desynchronization. Time elapsed has a highly significant impact on desynchronization; the upper panel reflects overall t values, the lower panel depicts only values which exceed a cluster corrected alpha of 0.05.

Discussion

Overly lengthy deliberation over a decision postpones the receipt of reward, and thus minimises the rate of return per unit time. On the other hand, overly hasty decisions often result in poorer choices.

How do agents find the right balance? It has been known for several decades that in order to achieve a given level of accuracy in a free response task, evidence should be accumulated to a fixed threshold (when trial difficulty is known). More recent work extended this notion for situations in which trial difficulty is not known, demonstrating that evidence should be accumulated towards a collapsing boundary in order to achieve this same accuracy level. In both of these situations, the sum of decision evidence and time elapsed are sufficient statistics to estimate the probability of being correct at any given moment (Kiani and Shadlen, 2009). However, decision difficultly is influenced by factors other than just strength of the evidence: for example choosing fastest route between 2 destinations is harder when the shortest route is sometimes laden with traffic, but clear at other times. When evidence variance influences decision difficulty over and above evidence strength, time elapsed and sum of evidence are no longer sufficient statistics to estimate the probability of making a correct choice (Fig. 1).

Here, we investigated theoretically and empirically what drives agents to commit to a decision when the evidence strength and variance are manipulated orthogonally from trial to trial. We found that an 'Adaptive Gain' model, which assumes agents sum the evidence provided by each stimulus in a weighted manner that causes inconsistent information to have lower impact on choices, captured human data closely. The qualitative predictions of this model approximated those of a Bayesian SPRT model, which integrates the posterior probability of each stimulus being drawn from the relevant distributions. Despite similarity in behavioural predictions, the two models imply different underlying computations. Analysis of the concurrent EEG activity allowed us to tease apart the model predictions in a manner that behaviour alone could not, and favoured the Adaptive Gain model as the best explanation for human evidence integration and decision termination.

The Adaptive Gain model implements a gain control mechanism in which the strength of evidence accumulation is modulated on the basis of the local statistics of the available information, in a manner reminiscent of the normalisation of low level sensory systems accoding to context (Bartlett, 1965; Carandini and Heeger, 2012; Fairhall et al., 2001). The model's ability to explain human data here replicates previous results from a related task that involved a fixed number of samples: as here, it was found that information was transformed in a nonlinear manner, with enhanced processing of that closest to the running tally of evidence. This down-weighting of statistically deviant information has been observed in a number of other studies: in averaging tasks in which all information is presented simultaneously, outlying information is 'robustly averaged' such that it carries lower eventual weight on choice (de Gardelle and Summerfield, 2011; Li et al., 2017). However, none of these paradigms were paced entirely by the participant – i.e. with each sample arriving at the participant's request until they decided they had seen sufficient information – therefore it was unclear whether, when the viewer dictates the pace, the down-weighting of outlying information is an element of the process that is sacrificed in favour of reaching the bound more quickly, and therefore making a faster decision. Here, we showed clearly that the down-weighting of discrepant information does indeed still feature strongly in self-paced decisions, despite this time cost.

We hope that our data are able to shed further light on the debate of the nature and existence of an 'urgency' signal. Behaviourally, we found that collapsing (rather than fixed) bound models fit our data better in every scenario investigated, supporting the idea that an evidence-independent time-dependent signal does indeed influence choices. However in many other situations, static bounds have been shown to capture evidence better (Hawkins et al., 2015). One possible explanation for these discrepant findings is the slower pacing of our task relative to many of those used in the previous literature. In rapid-paced psychophysical designs participants may have little time to adapt their policy on the basis of evolving decision variables. Our slower, self-paced design – reminiscent of many cognitive decisions in the real world – may allow more time for participants to adjust quasi-optimally to the arriving information. Of note, another task that employed a slower sequence of discrete samples similarly revealed evidence for a collapsing bound (Malhotra et al., 2017).

Neurally, we found evidence of an 'urgency' signal (in the beta-band components of the frequency-transformed EEG signal) that was independent of the evidence in the trial. We are not the first to show neural evidence for such a signal: for example, LIP activity apparently ramps to a bound even in a zero coherence condition, which may be attributed to an urgency effect (Churchland et al., 2008; Hanks et al., 2011). However, measuring an 'urgency' signal is non-trivial (Braunlich and Seger, 2016): over the course of a trial, the absolute values of many decision-related factors (such as the level of accumulated evidence) also increase. How can we be confident that we have uniquely isolated an 'urgency' signal from the mix? We used several techniques in an attempt to control for this. In the neural analysis, we included in our regressions a quantity that started high and logarithmically decreased – akin to a neural adaptation signal – as well as the quantity that started small and logarithmically increased, akin to an actual urgency signal. By including the former as an additional regressor, we could ensure that unique effects of general trial ramping were included in the adaptation signal. We also note the importance of interpreting this signal alongside behaviour; combining the clear behaviour evidence with the neural analysis adds extra strength to our interpretation of both.

Notably, the drift diffusion model (DDM), which postulates that evidence is accumulated linearly until it reaches a bound, provided a poor fit to the stopping time data in this environment. The DDM has been highly successful in explaining reaction times in humans and non human primates, and it has a wealth of neural support in particular from single unit recordings. Why then does it fare badly as an explanation here? We suggest three causes. The first pertains to the orthogonal influences of evidence strength and evidence variability on trial difficulty in our experiment. Previous studies have often used a measure of "coherence" to determine trial difficulty, whereby the variability itself is the determinant of evidence strength, and thus the two are not dissociable. The explanatory power of the DDM in its current form seems ill-suited to generalise to scenarios such as the one we employed, in which a number of influences can lead to an identical evidence state (decision variable). Furthermore, previous work has often focused on fitting the DDM to reaction time distributions (though see (Kang et al., 2017)). Our task also allowed us to interrogate the data with sharper precision: due to the entirely participant-paced nature of the experiment, the decision to view more information, as well as to commit to a decision, were both active choices. This allowed us to have high certainty at any given time the exact statistics of the decision information seen so far and at time of choice, and as such we were able to predict outcomes on a sample to sample basis, and to develop heretofore untested models that we put to the test using these rich data. Finally, it ought also to be noted that the DDM and the adaptive gain model are in nature very similar, with the key difference being the nonlinear weighting of information during integration.

A reward-maximising agent, when faced with the choice of whether to gather more information, or to commit to a choice now, should simulate the value of all possible future states and commit at the point that will yield the highest expected (Drugowitsch et al., 2012). However, whether or not such demanding computations are plausible for a biological agent remains questionable. We simulated the optimal form of the bound (**Fig. S1**), and found that it collapsed and then rose again after a few samples for the generative model that we had created. We did not attempt to fit such a bound to human data, but we think it is *a priori* unlikely to provide a good descriptive account of human choices.

Across the field, much is known about what decision people will make, and factors – optimal or otherwise – that influence choices. However, decisions are only good when they are made at the appropriate time: assessing every menu option of every café in town to find the perfect lunchtime sandwich will result in me making the best selection, but not until well after the cafes have shut for the day. In order to be worthwhile, choices must be executed in a timely manner, and relatively less is known about the factors that drive decisions to a close. Here we described human behaviour via a mechanism that executes choices when evidence, scaled by its consistency to that seen so far, reaches a bound that collapses over time to avoid prolonged deliberation under conditions of low certainty. By elucidating the mechanism, this opens an avenue to understanding what makes decision timing go "wrong", such as overly-impulsive choices associated with some mental illnesses, and opens the possibility of determining targeted interventions for such problems.

References

Bartlett, N.R. (1965). Dark and light adaptation. In Vision and visual perception New York: John Wiley and Sons, Inc; 1965, C.H. Graham, ed. (New York: John Wiley and Sons, Inc.).

Bogacz, R. (2007). Optimal decision-making theories: linking neurobiology with behaviour. Trends Cogn Sci *11*, 118-125.

Bogacz, R., Brown, E., Moehlis, J., Holmes, P., and Cohen, J.D. (2006). The physics of optimal decision making: a formal analysis of models of performance in two-alternative forced-choice tasks. Psychol Rev *113*, 700-765.

Braunlich, K., and Seger, C.A. (2016). Categorical evidence, confidence, and urgency during probabilistic categorization. Neuroimage *125*, 941-952.

Brunton, B.W., Botvinick, M.M., and Brody, C.D. (2013). Rats and humans can optimally accumulate evidence for decision-making. Science *340*, 95-98.

Busemeyer, J.R., and Rapoport, A. (1988). Psychological Models of Deferred Decision Making. Journal of Mathematical Psychology *32*, 91-134.

Carandini, M., and Heeger, D.J. (2012). Normalization as a canonical neural computation. Nat Rev Neurosci *13*, 51-62.

Cheadle, S., Wyart, V., Tsetsos, K., Myers, N., de Gardelle, V., Herce-Castañón, S., and Summerfield, C. (2014). Adaptive gain control during human perceptual choice. Neuron *81*, 1429–1441.

Churchland, A.K., Kiani, R., and Shadlen, M.N. (2008). Decision-making with multiple alternatives. Nat Neurosci *11*, 693-702.

Daunizeau, J., Adam, V., and Rigoux, L. (2014). VBA: a probabilistic treatment of nonlinear models for neurobiological and behavioural data. PLoS Comput Biol *10*, e1003441.

de Gardelle, V., and Summerfield, C. (2011). Robust averaging during perceptual judgment. Proc Natl Acad Sci U S A *108*, 13341-13346.

de Lange, F.P., Jensen, O., and Dehaene, S. (2010). Accumulation of evidence during sequential decision making: the importance of top-down factors. J Neurosci *30*, 731-738.

Delorme, A., and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. Journal of neuroscience methods *134*, 9-21.

Deneve, S. (2012). Making decisions with unknown sensory reliability. Front Neurosci *6*, 75.

Donner, T.H., Siegel, M., Fries, P., and Engel, A.K. (2009). Buildup of choice-predictive activity in human motor cortex during perceptual decision making. Curr Biol *19*, 1581-1585.

Drugowitsch, J., DeAngelis, G.C., Klier, E.M., Angelaki, D.E., and Pouget, A. (2014). Optimal multisensory decision-making in a reaction-time task. Elife *3*.

Drugowitsch, J., Moreno-Bote, R., Churchland, A.K., Shadlen, M.N., and Pouget, A. (2012). The cost of accumulating evidence in perceptual decision making. J Neurosci *32*, 3612-3628.

Drugowitsch, J., Wyart, V., Devauchelle, A.D., and Koechlin, E. (2016). Computational Precision of Mental Inference as Critical Source of Human Choice Suboptimality. Neuron.

Edwards, W. (1965). Optimal strategies for seeking information: Models for statistics, choice reaction times, and human information processing. J Math Psychol *2*, 312–329.

Fairhall, A.L., Lewen, G.D., Bialek, W., and de Ruyter Van Steveninck, R.R. (2001). Efficiency and ambiguity in an adaptive neural code. Nature *412*, 787-792.

Frazier, P., and Yu, A.J. (2008). Sequential hypothesis testing under stochastic deadlines. . Advances in neural information processing systems *20*, 465-472.

Gluth, S., Rieskamp, J., and Buchel, C. (2012). Deciding when to decide: time-variant sequential sampling models explain the emergence of value-based decisions in the human brain. J Neurosci *32*, 10686-10698.

Gold, J.I., and Shadlen, M.N. (2002). Banburismus and the brain: decoding the relationship between sensory stimuli, decisions, and reward. Neuron *36*, 299-308.

Hanks, T., Kiani, R., and Shadlen, M.N. (2014a). A neural mechanism of speed-accuracy tradeoff in macaque area LIP. Elife *3*.

Hanks, T.D., Ditterich, J., and Shadlen, M.N. (2006). Microstimulation of macaque area LIP affects decision-making in a motion discrimination task. Nat Neurosci *9*, 682-689.

Hanks, T.D., Kiani, R., and Shadlen, M.N. (2014b). A neural mechanism of speed-accuracy tradeoff in macaque area LIP. Elife *3*.

Hanks, T.D., Mazurek, M.E., Kiani, R., Hopp, E., and Shadlen, M.N. (2011). Elapsed decision time affects the weighting of prior probability in a perceptual decision task. J Neurosci *31*, 6339-6352.

Hawkins, G.E., Forstmann, B.U., Wagenmakers, E.J., Ratcliff, R., and Brown, S.D. (2015). Revisiting the evidence for collapsing boundaries and urgency signals in perceptual decision-making. J Neurosci *35*, 2476-2484.

Heitz, R.P., and Schall, J.D. (2012). Neural mechanisms of speed-accuracy tradeoff. Neuron *76*, 616-628.

Jazayeri, M., and Shadlen, M.N. (2010). Temporal context calibrates interval timing. Nat Neurosci *13*, 1020-1026.

Jazayeri, M., and Shadlen, M.N. (2015). A Neural Mechanism for Sensing and Reproducing a Time Interval. Curr Biol *25*, 2599-2609.

Kang, Y.H.R., Petzschner, F.H., Wolpert, D.M., and Shadlen, M.N. (2017). Piercing of Consciousness as a Threshold-Crossing Operation. Curr Biol *27*, 2285-2295 e2286.

Kiani, R., Corthell, L., and Shadlen, M.N. (2014). Choice certainty is informed by both evidence and decision time. Neuron *84*, 1329-1342.

Kiani, R., Hanks, T.D., and Shadlen, M.N. (2008). Bounded integration in parietal cortex underlies decisions even when viewing duration is dictated by the environment. J Neurosci *28*, 3017-3029.

Kiani, R., and Shadlen, M.N. (2009). Representation of confidence associated with a decision by neurons in the parietal cortex. Science *324*, 759-764.

Kira, S., Yang, T., and Shadlen, M.N. (2015). A neural implementation of Wald's sequential probability ratio test. Neuron *85*, 861-873.

Li, V., Herce Castanon, S., Solomon, J.A., Vandormael, H., and Summerfield, C. (2017). Robust averaging protects decisions from noise in neural computations. PLoS Comput Biol *13*, e1005723.

Li, V., Michael, E., Balaguer, J., Herce Castanon, S., and Summerfield, C. (2018). Gain control explains the effect of distraction in human perceptual, cognitive, and economic decision making. Proc Natl Acad Sci U S A *115*, E8825-E8834.

Maddox, W.T., and Bohil, C.J. (1998). Base-rate and payoff effects in multidimensional perceptual categorization. J Exp Psychol Learn Mem Cogn *24*, 1459-1482.

Malhotra, G., Leslie, D.S., Ludwig, C.J.H., and Bogacz, R. (2017). Overcoming indecision by changing the decision boundary. J Exp Psychol Gen *146*, 776-805.

Maris, E., and Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. J Neurosci Methods *164*, 177-190.

Moran, R. (2015). Optimal decision making in heterogeneous and biased environments. Psychon Bull Rev *22*, 38-53.

Murphy, P.R., Boonstra, E., and Nieuwenhuis, S. (2016). Global gain modulation generates time-dependent urgency during perceptual choice in humans. Nat Commun *7*, 13526.

Palminteri, S., Wyart, V., and Koechlin, E. (2017). The Importance of Falsification in Computational Cognitive Modeling. Trends Cogn Sci *21*, 425-433.

Ratcliff, R., and McKoon, G. (2008). The diffusion decision model: theory and data for two-choice decision tasks. Neural Comput *20*, 873-922.

Roitman, J.D., and Shadlen, M.N. (2002). Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. J Neurosci *22*, 9475-9489.

Spieser, L., Kohl, C., Forster, B., Bestmann, S., and Yarrow, K. (2018). Neurodynamic Evidence Supports a Forced-Excursion Model of Decision-Making under Speed/Accuracy Instructions. eNeuro *5*.

Stephens, D.W., and Krebs, J.R. (1986). Foraging theory (Princeton, N.J.: Princeton University Press).

Summerfield, C., and Tsetsos, K. (2015). Do humans make good decisions? Trends Cogn Sci *19*, 27-34.

Sun, P., and Landy, M.S. (2016). A Two-Stage Process Model of Sensory Discrimination: An Alternative to Drift-Diffusion. The Journal of neuroscience : the official journal of the Society for Neuroscience *36*, 11259-11274.

Thura, D., Beauregard-Racine, J., Fradet, C.W., and Cisek, P. (2012). Decision making by urgency gating: theory and experimental support. J Neurophysiol *108*, 2912-2930.

Tickle, H., Speekenbrink, M., Tsetsos, K., Michael, E., and Summerfield, C. (2016). Near-optimal Integration of Magnitude in the Human Parietal Cortex. J Cogn Neurosci *28*, 589-603.

Twomey, D.M., Murphy, P.R., Kelly, S.P., and O'Connell, R.G. (2015). The classic P300 encodes a build-to-threshold decision variable. Eur J Neurosci *42*, 1636-1643.

Usher, M., and McClelland, J.L. (2001). The time course of perceptual choice: the leaky, competing accumulator model. Psychol Rev *108*, 550-592.

Vul, E., Goodman, N., Griffiths, T.L., and Tenenbaum, J.B. (2014). One and done? Optimal decisions from very few samples. Cogn Sci *38*, 599-637.

Wald, A., and Wolfowitz, J. (1949). Bayes Solutions of Sequential Decision Problems. Proc Natl Acad Sci U S A *35*, 99-102.

Wang, X.J. (2002). Probabilistic decision making by slow reverberation in cortical circuits. Neuron *36*, 955-968.

Wyart, V., de Gardelle, V., Scholl, J., and Summerfield, C. (2012). Rhythmic fluctuations in evidence accumulation during decision making in the human brain. Neuron *76*, 847-858.

Wyart, V., Myers, N.E., and Summerfield, C. (2015). Neural mechanisms of human perceptual choice under focused and divided attention. J Neurosci *35*, 3485-3498.

Zylberberg, A., Fetsch, C.R., and Shadlen, M.N. (2016). The influence of evidence volatility on choice, reaction time and confidence in a perceptual decision. Elife *5*.

Supplementary materials

Optimal model based on partially observable Markov decision process (POMDP)

We assume an agent performs the task for an (unknown) number of rounds $n = 1, \dots, N$. In each round, the agent observes a sequence of samples $X_{n,t} \sim N(\mu_n, \sigma_n)$, from a Gaussian distribution with an unknown mean μ_n and standard deviation σ_n. The task of the agent is to decide whether the mean of the distribution is positive or negative. Each round starts with a sample $X_{n,1}$, after which the agent can make a decision, or request another sample. After each sample $t = 1, \dots, T_n$, the agent thus has three actions: decide $\mu_n > 0$, decide $\mu_n < 0$, or collect another sample. If the agent decides, she gets a reward c if the decision is correct or a penalty e if the decision is incorrect. In the experiment, the means were $\mu_n \in \{-6, -4, -2,2,4,6\}$ and the standard deviations were $\sigma \in \{4,8,16\}$.

In statistics, such a task is known as a sequential hypothesis testing problem. For the case of two simple hypotheses, the optimal sequential testing procedure is the Sequential Probability Ratio Test (SPRT; Wald & Wolfowitz, 1948). The SPRT yields the minimum average decision time for a given probability of error. In the case of compound hypotheses (e.g., $H_0: \mu > 0$, rather than $H_0: \mu = \mu_0$) and unknown standard deviations, such a general optimal procedure is not available. However, by focusing a set of simple hypotheses $S = (H_1, \dots, H_S)$, $H_j: \mu = \mu_j, \sigma = \sigma_j$, and formulating the problem as a partially observable Markov decision process (POMDP), an approximately optimal strategy can be found.

POMDP solution

A POMDP (e.g., Littman, 2009) is a tuple $< S, A, X, T, O, b_0, R, \gamma >$. Here, S is the set of states, A the set of actions the agent can take, O the set of possible observations, T the transition probability distribution $T(s, a, s') = P(S_{t+1} = s' \mid S_t = s, A_t = a)$, O the observation probability distribution $O(x, s) = P(X_t = x \mid S_t = s)$, b_0 the initial belief state $b_0(s) = p(S_1 = s)$, R the reward function $R(a, s) \rightarrow R$, and $\gamma \in [0,1]$ the discount factor. In a POMDP, the agent cannot directly observe the current state S_t of the environment, but it can infer the state from the (noisy) observations X_t. This inference is reflected in the belief state $b_t(s) = P(S_t = s \mid X_1, \dots, X_t)$.

The objective is to find a policy $\pi = p(A_t \mid b_t)$ that maps belief states to actions and maximises the expected discounted future reward

$$E[R(A_t, S_t) + \gamma R(A_{t+1}, S_{t+1}) + \gamma^2 R(A_{t+2}, S_{t+2}) + \cdots] = E\left[\sum_{k=0}^{\infty} \gamma^k \, R(A_{t+k}, S_{t+k})\right]$$

In the following analysis, we set:

- $S = \{\mu \in (-10, -8, -6, -4, -2, -.001, .001, 2, 4, 6, 8, 10)\} \times \{\sigma \in (4, 8, 16, 32)\}$. The state-space is set of hypotheses considered, formed as the Cartesian product of a set of considered means and a set of considered standard deviations. The sets include the true values used in the experiment, but also some more and less extreme values.
- $A = \{decide\ \mu < 0, decide\ \mu > 0, sample\}$
- $X = (-50, -48, -46, \ldots, 0, \ldots, 50)$. A discretized set of possible observations covering the range of likely values.
- $P(s' \mid a = sample, s = s') = 1$, $P(s' \vee a = sample, s \neq s') = 0$, $P(s' \mid a \neq sample, s) = \frac{1}{|S|}$. If the agent samples, the environment stays in the same state. After an agent makes a decision about the mean, a new round is started, with state a draw from the uniform distribution over all possible states.
- $P(o \mid s) = N(o \mid \mu_s, \sigma_s)$. The observation distribution is a normal distribution.
- $P(S_1 = s) = \frac{1}{|S|}$. The initial belief state b_0 is uniform distribution over all states.
- $R(sample, s) = 0$, $R(decide\ \mu < 0, s \in \{s: \mu_s < 0\}) = 1$, $R(decide\ \mu < 0, s \in \{s: \mu_s > 0\}) = -3$, $R(decide\ \mu > 0, s \in \{s: \mu_s > 0\}) = 1$, $R(decide\ \mu > 0, s \in \{s: \mu_s < 0\}) = -3$. A correct decision provides a reward of 1, while an incorrect decision provides a penalty of -3. Sampling does not incur a cost.
- $\gamma = .95$. A reward after the next sample is worth 95% of the same reward now.

We used the pomdp-solve software to find an approximation to the optimal strategy for this POMDP, accessed via the R package pomdp (Kamalzadeh & Hahsler, 2019). We used a grid-based solution method, placing 5000 grid points on the belief space $B_t = P(S_t \mid x_1, \ldots, x_t)$. Code is available at http:

The approximately optimal policy prescribes an action for each belief state b_t. Below, we will depict the policy in terms of the probability of asking for another sample as a function of the summed evidence $\sum_{k=1}^{t} X_t$, as is common for evidence accumulation models (e.g. Drugowitz *et al.*, 2012). We should note that while there is a one-to-one mapping from belief states to optimal action, there is a many-to-one mapping from the history of samples $X_1, \ldots, X_t$ to belief states b_t, and a many-to-many mapping from summed evidence to optimal actions. As such, unlike for more tradional evidence accumulation models, there is no clearly defined deterministic decision bound in the space of summed evidence.

Fig. S1 shows, at each sample point t, the proportion of "sample" actions over all belief states that occur together with a particular value for the summed evidence, conditional upon not having made a decision about the mean previously. The probabilities were computed by averaging over all versions of the task (i.e. all true values of the mean and standard deviation). It can be seen that the probability of asking for another sample decreases after the first sample, and then increases again after the third sample.

Fig. S1

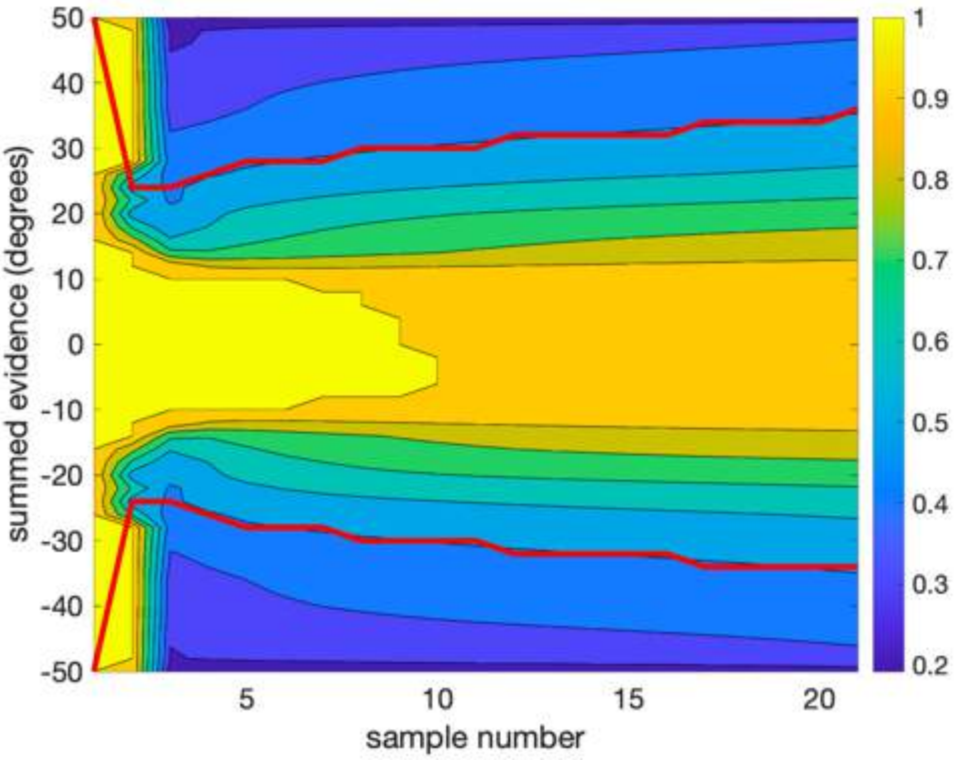
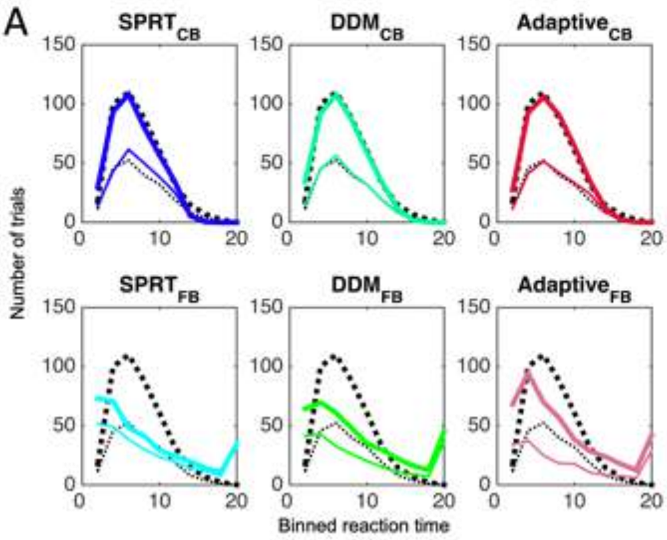


Fig. S1. Results of the computation of the optimal bound. The image plots for an optimal observer the probability of taking an additional sample, conditional on the number of samples elapsed and the sum of evidence (in degrees of visual angle) up to that sample. The red line traces the bound, i.e. the lowest (positive or negative) evidence value where participants are indifferent to taking another sample or not.

Fig. S2



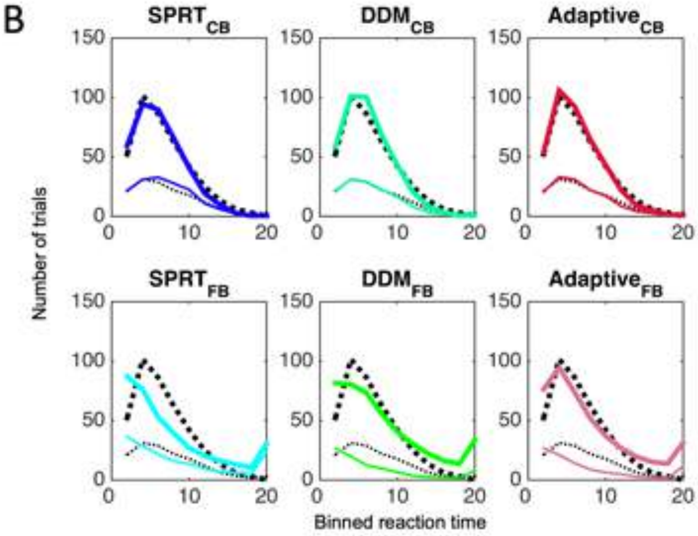


Figure S2: Stopping time distributions for experiments 1 and 2. The black dashed lines depict the human distributions of stopping times for correct (thicker dashed black line) and error (thinner dashed black line) trials, superimposed on the corresponding fits of collapsing-bound (upper panels) and fixed-bound (lower panels) variants of the SPRT (left panels), DDM (middle panels) and Adaptive Gain model (right panels)

Fig. S3

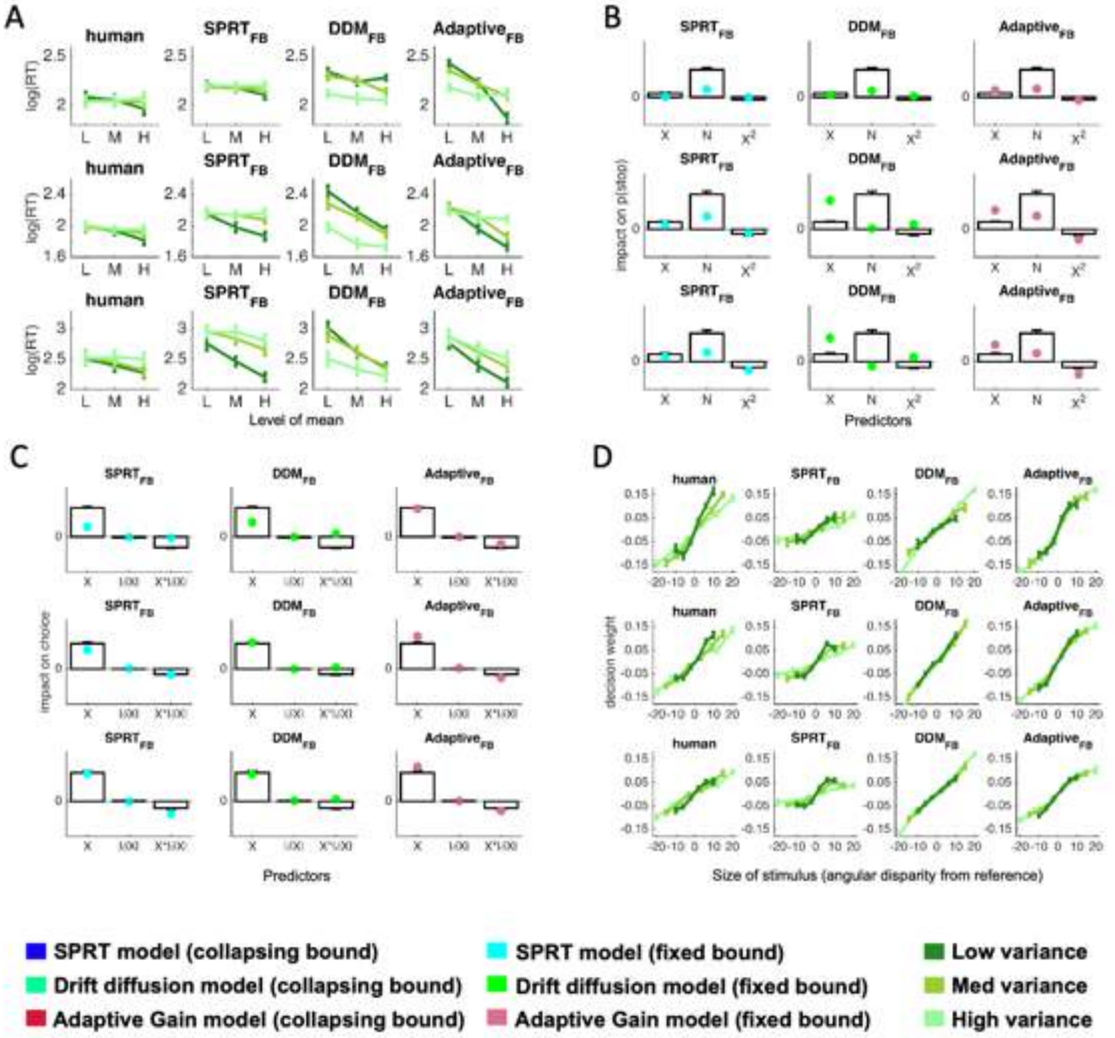


Figure S3. Fixed Bound models: qualitative data. A. Average (log) stopping times (in samples) for trials with low (L), medium (M) and high (H) $|\mu|$, i.e. distance to the reference (x-axis) and low, medium and high σ (dark, medium and pale green lines respectively). Leftmost panel: humans; other panels, fixed bound models. Rows 1-3 are experiments 1-3 respectively. The predictions of the SPRT and Adaptive Gain qualitatively match the human data, where as the DDM does not. B. Coefficients for statistics predicting next-sample commitment based on sum(X), samples elapsed (N) and sum(X^2) for the fixed bound models. White bars are humans; each panel shows the fit of a model (blue, green and red dots reflect SPRT, DDM and Adaptive Gain models respectively). Rows 1-3 are experiments 1-3 respectively. C. The impact on $p(choose\ clockwise)$ of sample information (X), the prediction error ($|\delta X|$), and the interaction of the latter two for the fixed bound models. D. Subjective versus objective weighting of the decision evidence, separately for the three conditions of variance (low, medium and high in pale, medium and dark green respectively) for the fixed bound models.

Supplementary References

Kamalzadeh, H. & Hahsler, M. (2019). pomdp: Solver for Partially Observable Markov Decision Processes (POMDP). R package version 0.9.2. https://CRAN.R-project.org/package=pomdp

Littman, M. L. (2009) A tutorial on partially observable Markov decision processes. Journal of Mathematical Psychology, 53, 119–125.