

1 Overcoming indecision by changing the decision boundary

2 Gaurav Malhotra

3 School of Experimental Psychology, University of Bristol

4 David S. Leslie

5 Department of Mathematics and Statistics, Lancaster University

6 Casimir J. H. Ludwig

7 School of Experimental Psychology, University of Bristol

8 Rafal Bogacz

9 MRC Brain Network Dynamics Unit, University of Oxford

10 Author Note

11 Word count: 16673. Correspondence concerning this article should be addressed to
12 Gaurav Malhotra, gaurav.malhotra@bristol.ac.uk, School of Experimental Psychology,
13 University of Bristol, 12a Priory Road, Bristol BS8 1TU, United Kingdom. Some of the
14 experiments reported in this study were previously presented at the 46th Annual Meeting
15 of the Society of Mathematical Psychology (SMP) in Potsdam, Germany and at
16 Decision-making Bristol, 2014.

Abstract

The dominant theoretical framework for decision-making asserts that people make decisions by integrating noisy evidence to a threshold. It has recently been shown that in many ecologically realistic situations, decreasing the decision boundary maximises the reward available from decisions. However, empirical support for decreasing boundaries in humans is scant. To investigate this problem, we used an ideal observer model to identify the conditions under which participants should change their decision boundaries with time in order to maximize reward rate. We conducted six expanded-judgement experiments that precisely matched the assumptions of this theoretical model. In this paradigm participants could sample noisy, binary evidence presented sequentially. Blocks of trials were fixed in duration and each trial was an independent reward opportunity. Participants therefore had to trade off speed (getting as many rewards as possible) against accuracy (sampling more evidence). Having access to the actual evidence samples experienced by participants enabled us to infer the slope of the decision boundary. We found that participants indeed modulated the slope of the decision boundary in the direction predicted by the ideal observer model, although we also observed systematic deviations from optimality. Participants employing sub-optimal boundaries do so in a robust manner, so that any error in their boundary setting is relatively inexpensive. The use of a normative model provides insight into what variable(s) human decision-makers are trying to optimise. Furthermore, this normative model allowed us to choose diagnostic experiments and in doing so we present clear evidence for time-varying boundaries.

Keywords: decision-making, decision threshold, decreasing bounds, optimal decisions, reward-rate

Overcoming indecision by changing the decision boundary

Disclosures and Acknowledgments

This research was carried out as part of the project ‘Decision making in an unstable world’, supported by the Engineering and Physical Sciences Research Council (EPSRC), Grant Reference EP/1032622/1. Additionally, RB was supported by Medical Research Council grant MC UU 12024/5. The funding sources had no other role other than financial support.

All authors contributed to the design of the experiment, the analysis of the results and the writing of the manuscript. All authors have read and approved the final manuscript.

All authors state that there are no conflicts of interest that may inappropriately impact or influence the research and interpretation of the findings.

We would like to thank E.J. Wagenmakers, Daniel Wolpert, Andreas Jarvstad and Iain Gilchrist for their insightful comments during the development of this research. We would also like to thank Caitlin Molloy for help with data collection in some of the experiments presented in this study.

Introduction

In an early theory of decision-making, Cartwright and Festinger (1943) modelled decision-making as a struggle between fluctuating forces. At each instant, the decision-maker drew a sample from the (Gaussian) distribution for each force and computed the difference between these samples. This difference was the *resultant force* and no decision was made while the opposing forces were balanced and the resultant force was zero. Cartwright and Festinger realised that if a decision was made as soon as there was the slightest imbalance in forces, there would be no advantage to making decisions more slowly. This was inconsistent with the observation that the speed of making decisions traded-off with their accuracy, a property of decision-making that had already been recorded (Garrett, 1922; Johnson, 1939; Festinger, 1943) and has been repeatedly observed since (e.g., Howell & Kreidler, 1963; Pachella, 1974; Wickelgren, 1977; Luce, 1986; Bogacz, Wagenmakers, Forstmann, & Nieuwenhuis, 2010). Cartwright and Festinger addressed the speed-accuracy trade-off by introducing an internal *restraining force* – also normally distributed and in the opposite direction to the resultant force – which would prevent the decision-maker from going off “half-cocked” (Cartwright & Festinger, 1943, p. 598). The decision-maker drew samples from this restraining force and did not make a decision until the resultant force was larger than these samples. The restraining force was adaptable and could be adjusted based on whether the decision-maker wanted to emphasize speed or accuracy in the task.

In the ensuing decades, Cartwright and Festinger’s theory fell out of favour due to several shortcomings (see Irwin, Smith, & Mayfield, 1956; Vickers, Nettelbeck, & Willson, 1972) and was superseded by the signal detection theory (Tanner & Swets, 1954) and sequential sampling models (Stone, 1960; LaBerge, 1962; Laming, 1968; Link & Heath, 1975; Vickers, 1970; Ratcliff, 1978). These models do not mention a restraining force explicitly, but this concept is implicit in a threshold, which must be crossed before the decision-maker indicates their choice. Just as the restraining force could be adjusted based on the emphasis on speed or accuracy, these models proposed that the threshold could be

lowered or raised to emphasise speed or accuracy. This adaptability of thresholds has been a key strength of these models, a feature that has been used to explain how distribution of response latencies changes when subjects are instructed to emphasise speed or accuracy in a decision (for a review, see Ratcliff & Smith, 2004; Bogacz, Brown, Moehlis, Holmes, & Cohen, 2006; Ratcliff, Smith, Brown, & McKoon, 2016).

Introducing a restraining force or a threshold to explain the speed-accuracy trade-off answers one question but raises another: how should a decision-maker select the restraining force (threshold) for a decision-making problem? Should this restraining force remain constant during a decision? This problem was examined by Wald (1947) who proposed that, for an isolated decision, an optimal decision-maker can distinguish between two hypotheses by choosing the desired ratio of Type 1 and Type 2 errors and then using a statistical procedure called the sequential-probability-ratio-test (SPRT). In the SPRT, the decision-maker sequentially computes the ratio of the likelihoods of all observations given the hypotheses and the decision process terminates only once the ratio exceeds a threshold (corresponding to accepting the first hypothesis) or decreases below another threshold (corresponding to accepting the second hypothesis). The values of these thresholds do *not* change as more samples are accumulated and they determine the accuracy of decisions. Wald and Wolfowitz (1948) showed that the SPRT requires a smaller or equal number of observations, on average, than any other statistical procedure, for a given accuracy of decisions.

The SPRT gives a statistically optimal procedure to set the threshold for an isolated decision. However, in many real-world decision problems – a bird foraging for food, a market trader deciding whether to keep or sell stocks, a professor going through a pile of job applications or, indeed, a psychology undergraduate doing an experiment for course credits – decisions are not made in isolation; rather, individuals have to make a sequence of decisions. How should one set the threshold in this situation? Is the optimal threshold still given by SPRT? If decision-makers accrue a reward from each decision, an ecologically

sensible goal for the decision-maker may be to maximise the expected reward from these decisions, rather than to minimise the number of samples required to make a decision with a given accuracy (as SPRT does). And for sequences which involve a large number of decisions or sequence of decisions that do not have a clearly defined end point, it would make sense for the decision-maker to maximise the *reward-rate*, the expected amount of rewards per unit time. In fact, under certain assumptions, including the assumption that every decision in a sequence has the same difficulty, it can be shown that the two optimisation criteria – SPRT and reward-rate – result in the same threshold (Bogacz et al., 2006). That is, the decision-maker can maximise reward-rate by employing the SPRT and maintaining an appropriately chosen threshold that remains constant within and across trials. Experimental data suggest that people do indeed adapt their speed and accuracy to improve their reward-rate (Bogacz, Hu, Holmes, & Cohen, 2010; Simen et al., 2009). This adaptability seems to be larger for younger than older adults (Starns & Ratcliff, 2010, 2012) and seems to become stronger with practice (Balci et al., 2011) and guidance (Evans & Brown, 2016).

However, maintaining a fixed and time-invariant threshold across a sequence of trials cannot be the optimal solution in many ecologically realistic situations where the difficulty of decisions fluctuates from trial-to-trial. Consider, for example, a situation in which there is very little information or evidence in favour of the different decision alternatives. Accumulating little evidence to a fixed threshold might take a very long time. The decision-maker risks being stuck in such an impoverished trial because they are unable to choose between two equally uninformative options, like Buridan’s donkey (see Lamport, 2012), who risks being starved because it is unable to choose between two equally palatable options. Cartwright and Festinger foresaw this problem and noted that “there is good reason to suppose that the longer the individual stays in the decision region, the weaker are the restraining forces against leaving it” (Cartwright & Festinger, 1943, p. 600). So they speculated that the mean restraining force should be expressed as a *decreasing* function of

time but they were not prepared to make specific assumptions as to the exact nature of this function.

In cases where the restraining force may change with time, the concept of a fixed threshold may be replaced by a time-dependent *decision boundary* between making more observations and choosing an alternative. A number of recent studies have mathematically computed the shape of decision boundaries that maximise reward-rate when decisions in a sequence vary in difficulty and shown that the decision-maker can maximise the reward-rate by decreasing this decision boundary with time (Drugowitsch, Moreno-Bote, Churchland, Shadlen, & Pouget, 2012; Huang & Rao, 2013; Moran, 2015). It can also be shown that the shape of the boundary that maximises reward-rate depends on the mixture of decision difficulties. Indeed, based on the difficulties of decisions in a sequence, optimal boundaries may decrease, remain constant or increase (Malhotra, Leslie, Ludwig, & Bogacz, under review¹; also see below).

The goal of this study was to test whether and under what circumstances, humans vary their decision boundaries with time during a decision. More generally, we assessed the relationship between the bounds employed by people and the optimal bounds – i.e., the boundary that maximises reward-rate. Importantly, we adopt an experimental approach that is firmly rooted in a mathematical optimality analysis (Malhotra et al., under review) and that allows us to infer the decision boundary relatively directly based on the sequences of evidence samples actually experienced by decision makers.

Previous evidence on whether people change decision boundaries at all during a trial, much less adapt it to be optimal, is inconclusive. Some evidence of time-dependent boundaries was found early on in studies that compared participant behaviour with Wald’s optimal procedure. These studies employed an *expanded-judgement* paradigm in which the participant makes their decision based on a sequence of discrete samples or observations presented at discrete times – e.g., deciding between two deck of cards with different means

¹Manuscript available at <https://osf.io/9t76q/>

1 based on cards sampled sequentially from the two decks (see e.g., Irwin et al., 1956; Becker,
2 1958; Manz, 1970; Busemeyer, 1985; Smith & Vickers, 1989; Vickers, 1995; Pleskac &
3 Busemeyer, 2010). The advantage of this paradigm is that the experimenter can record not
4 only the response time and accuracy of the participant, but also the exact sequence of
5 samples on which they base their decisions. In an expanded-judgement paradigm Pitz,
6 Reinhold, and Geller (1969) found that participants made decisions at lower posterior odds
7 when the number of samples increased. Similar results were reported by Sanders and
8 Linden (1967) and Wallsten (1968). Curiously, participants seemed to be disregarding the
9 optimal strategy in these studies, which was to keep decision boundaries constant. We will
10 discuss below why this behaviour may be ecologically rational when the participant has
11 uncertainty about task parameters.

12 The shape of decision boundaries has been also analysed in a number of experiments
13 using a paradigm where the samples drawn by the participant are implicit, i.e., hidden
14 from the experimenter. In these paradigms, the data recorded is limited to the response
15 time and accuracy, so one can distinguish between constant or variable decision boundaries
16 only indirectly, by fitting the two models to the data and comparing them. These tasks
17 generally involve detecting a signal in the presence of noise. Therefore, to distinguish these
18 experiments from the expanded-judgement tasks, we will call them *signal detection* tasks.
19 Examples of this paradigm include lexical decisions (Wagenmakers, Ratcliff, Gomez, &
20 McKoon, 2008), basic perceptual discrimination (e.g. brightness; Ratcliff & Rouder, 1998;
21 Ludwig, Gilchrist, McSorley, & Baddeley, 2005) and numerosity judgements (Starns &
22 Ratcliff, 2012). Pike (1968) analysed data from a number of psychophysical discrimination
23 studies and found that this data is best explained by the accumulator model (Audley &
24 Pike, 1965) if subjects either vary decision bounds between trials or decrease bounds during
25 a trial. Additional support for decreasing boundaries was found by Drugowitsch et al.
26 (2012) who analysed data collected by Palmer, Huk, and Shadlen (2005). Finally, data
27 from non-human primates performing a random dot motion discrimination task (Roitman

1 & Shadlen, 2002) were best fit by a diffusion model with decreasing boundaries (Ditterich,
2 2006).

3 In contrast to these studies which found evidence favouring decreasing boundaries,
4 Hawkins, Forstmann, Wagenmakers, Ratcliff, and Brown (2015) analysed data from
5 experiments on human and non-human primates spanning a range of experiments using
6 signal detection paradigms and found equivocal support for constant and decreasing
7 boundaries. They found that overall evidence, especially in humans, favoured constant
8 boundaries and that, crucially, experimental procedures such as the extent of task practice
9 seemed to play a role in which option was favoured. Therefore, what seems to be missing is
10 a more systematic analysis of the conditions under which people decrease the decision
11 boundary within a trial and understanding why they would do so.

12 In this study, we took a different approach: rather than infer the decision boundaries
13 indirectly by fitting different boundaries to explain reaction times and error rates, we used
14 the expanded-judgement paradigm, where the experimenter can observe the exact sequence
15 of samples used by the participant and record the exact evidence and time used to make a
16 decision. This evidence and time should lie on the boundary. This allowed us to make a
17 more direct estimate of the decision boundary used by the participants and compare this
18 boundary with the optimal boundary. We found that, in general, participants modulated
19 their decision boundaries during a trial in a manner predicted by the maximisation of
20 reward-rate. This effect was robust across paradigms and for decisions that play out over
21 time scales ranging from several hundreds of milliseconds to several seconds. However,
22 there were also systematic deviations from optimal behaviour. Much like the
23 expanded-judgement tasks discussed above, in a number of our experiments participants
24 seemed to decrease their decision boundary even when it was optimal to keep them
25 constant. We mapped these strategies on to the “reward landscape” predicted by the
26 theoretical model – i.e. the variation in reward-rate with different settings of the decision
27 boundary. These analyses suggest that participants’ choice of decision boundary may be

1 guided not only by maximising the reward-rate, but also by robustness considerations.
2 That is, they appear to allow for some “error” in their boundary setting due to uncertainty
3 in task parameters and deviate from optimality in a manner that reduces the impact of
4 such error.

5 While it has been argued that the results from an expanded-judgement task can be
6 generalised to signal detection paradigms, where sampling is implicit (Irwin & Smith, 1956;
7 Edwards, 1965; Sanders & Linden, 1967; Vickers, Burt, Smith, & Brown, 1985; Pleskac &
8 Busmeyer, 2010), these tasks usually use a slow presentation rate and elicit longer
9 response latencies than those expected for perceptual decisions. It is possible that
10 attention and memory play a different role in decision-making at this speed than at faster
11 speeds at which perceptual decisions occur. To address this possibility, we adapted the
12 expanded-judgement task to allow fast presentation rates and consequently elicit rapid
13 decisions.

14 The rest of the paper is split into five sections. First, we summarise the theoretical
15 basis for the relationship between a boundary and reward-rate. In the next three sections
16 we describe a series of expanded-judgement tasks, each of which compares the boundaries
17 employed by participants with the theoretically optimal boundaries. In the final section we
18 consider the implications of our findings as well as the potential mechanisms by which
19 time-varying boundaries may be instantiated. Data from all experiments reported in this
20 article is available online at: <https://osf.io/f3vhr/>.

21 Optimal shape of decision boundaries

22 We now outline how an expanded-judgement task can be mathematically modelled
23 and how this model can be used to establish the relationship between the task’s parameters
24 and decision boundaries that maximise reward-rate. We summarize the key results from a
25 theoretical study of Malhotra et al. (under review), provide intuition for them, and state
26 predictions that we tested experimentally.

Consider an expanded-judgement task that consists of making a sequence of decisions, each of which yields a unit reward if the decision is correct. Each decision (or trial) consists of estimating the true state of the world based on a sequence of noisy observations. We consider the simplest possible case in which the world can be in one of two different states, call these *Up* or *Down*, and each observation of the world can be one of two different outcomes. Each outcome provides a fixed amount of evidence, δX , to the decision-maker about the true state of the world:

$$\delta X = \begin{cases} +1 & \text{with probability } u \\ -1 & \text{with probability } 1 - u \end{cases} \quad (1)$$

where u is the *up-probability* that governs how quickly evidence is accumulated and depends on the true state of the world. We assume throughout that $u \geq 0.5$ when the true state is *Up* and $u \leq 0.5$ when the true state is *Down*. Note that the parameter u will determine the difficulty of a decision – when u is close to 0.5 the decision will be hard while when u is close to 0 or 1 the decision will be easy.

Figure 1 illustrates the process of making a decision in this expanded-judgement task. As assumed in sequential sampling models, decision making involves the accumulation of the probabilistic evidence, so let x be the cumulative evidence, i.e. the sum of all δX outcomes. The accumulation continues till x crosses one of two boundaries θ corresponding to the options, so that the decision maker responds *Up* when $x > \theta$, and *Down* when $x < -\theta$. During the expanded-judgement task described above, the state of the decision-maker, at any point of time, is defined by the pair (t, x) , where t is the number of observations made. In any given state, the decision-maker can take one of two *actions*: (i) make another observation – we call this action *wait*, or (ii) signal their estimate of the true state of the world – we call this the action *go*. As shown in Figure 1, taking an action *wait* can lead to one of two *transitions*: (i) the next observed outcome is +1; in this case, make a transition to state $(t + 1, x + 1)$, or (ii) the next observed outcome is -1; in this case, make a transition to state $(t + 1, x - 1)$. Similarly, taking the action *go* can also lead to one

1 of two transitions: (i) the estimated state is the true state of the world; in this case collect
 2 a reward and make a transition to the *Correct* state, or (ii) the estimated state is not the
 3 true state of the world; in this case make a transition to the *Incorrect* state. After making
 4 a transition to a *Correct* or *Incorrect* state, the decision-maker starts a new decision, i.e.
 5 returns to the state $(t, x) = (0, 0)$ after an inter-trial delay D_C following a correct choice
 6 and D_I following the incorrect choice.

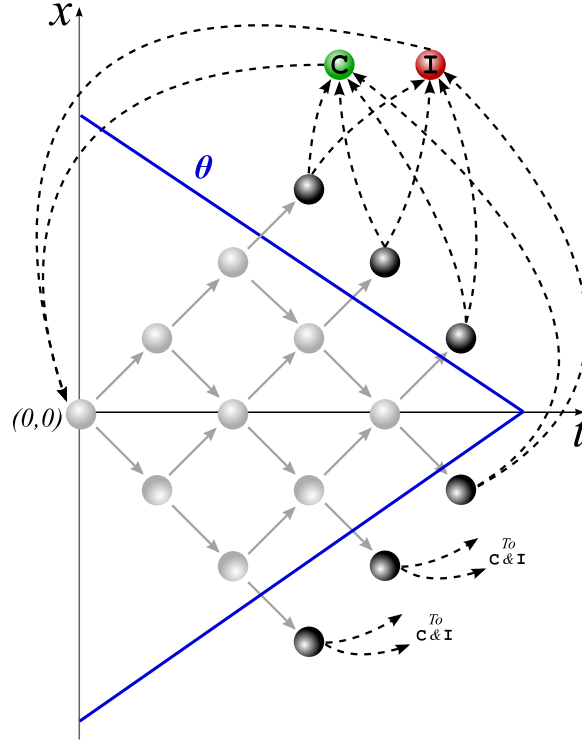


Figure 1. Evidence accumulation and decision-making as a Markov Decision Process: states are shown by circles, transitions are shown by arrows and actions are shown by colour of the circles. The solid (blue) line labeled θ indicates a hypothetical decision boundary. The policy that corresponds to the boundary is indicated by the colour of the states. Black circles indicate the action *go* while gray circles indicate *wait*. Dashed lines with arrows indicate transitions on *go* while solid lines with arrows indicate transitions on *wait*. The rewarded and unrewarded states are shown as **C** and **I**, respectively (for *Correct* and *Incorrect*.)

These set of state-action pairs and transitions between these states defines a Markov Decision Process (MDP) shown schematically in Figure 1. In this framework, any decision boundary which is a function of time, $\theta = f(t)$, can be mapped to a set of actions, such that action *wait* is selected for any state within the boundaries and action *go* for any state on or beyond the boundaries. The mapping which assigns actions to all possible states is called a *policy* for the MDP.

We assume that a decision maker wishes to maximise the reward-rate (which we defined as the expected number of rewards per unit time). The reward-rate depends on the decision boundary: If the boundary is too low, the decision maker will make errors and miss possible rewards, but if it is too high, each decision will take a long period, and the number of reward per unit of time will also be low.

The policy that maximises average reward can be obtained by using a dynamic programming procedure known as *policy iteration* (Howard, 1960; Ross, 1983; Puterman, 2005). Several recent studies describe how dynamic programming can be applied to decision-making tasks to get a policy that maximises reward-rate (see e.g., Drugowitsch et al., 2012; Huang & Rao, 2013; Malhotra et al., under review). We now summarize how the optimal shape of decision boundary depends on task’s parameters based on the analysis given in Malhotra et al. (under review). Let us first consider a class of tasks in which the difficulty of the decisions is fixed. That is, evidence can point either towards *Up* or *Down*, but the quality of the evidence remains fixed across decisions: $u \in \{\frac{1}{2} - \epsilon, \frac{1}{2} + \epsilon\}$, with ϵ corresponding to the *drift*. The drift can take values in range $\epsilon \in [0, \frac{1}{2}]$ and it determines the difficulty of each trial with higher drift corresponding to easier trials. For *Single-difficulty* tasks, ϵ remains fixed across trials.

In the *Single-difficulty* tasks, reward-rate can be optimised by choosing a policy such that the decision boundary remains constant during each decision. Intuitively, this is because the decision maker’s estimate of the probability that the world is in a particular state depends only on integrated evidence x , but not on time elapsed within the trial t .

Therefore the optimal action to take in each state only depends on x but not t , so *go* actions are only taken if x exceeds a particular value, leading to constant boundaries.

The optimal height of the decision boundary in the Single-difficulty tasks depends on task difficulty in a non-monotonic way. For very easy tasks (ϵ close to $\frac{1}{2}$) each outcome is a very reliable predictor of the state of the world, so very few outcomes need to be seen to obtain an accurate estimate of the state of the world (Figure 2a). As the difficulty increases, more outcomes are required, and the optimal boundary increases (compare Figures 2a and 2b). However, when the task becomes very difficult (ϵ close 0), there is little benefit in observing the stimulus at all, and for $\epsilon = 0$ the optimal strategy is not to integrate evidence at all, but guess immediately, i.e. $\theta = 0$ (compare Figures 2d and 2e).

Let us now consider a *Mixed-difficulty* task, in which half of the trials are *Easy* with drift ϵ_e and the other half of the trials are *Difficult* with drift ϵ_d , where $\epsilon_e > \epsilon_d$. We assume that during Mixed-difficulty tasks, the decision-maker knows that there are two levels of difficulty (either through experience or instruction), but does not know if a particular trial is easy or difficult. Indeed, a key assumption of the underlying theory is that the difficulty level is something the decision maker has to infer during the accumulation of evidence.

In Mixed-difficulty tasks, reward-rate is optimised by employing boundaries that may decrease, increase or remain constant based on the mixture of difficulties. Intuitively, this is because the decision maker's estimate of the probability that the world is in a particular state, given the existing evidence, depends on their inference about the difficulty of the trial. Time becomes informative in Mixed-difficulty tasks because it helps the decision-maker infer whether a given trial is *Easy* or *Difficult* and hence the estimate of the true state of the world depends not only on the evidence, x , but also on the time, t . The optimal decision-maker should begin each decision trial assuming the decision could be *Easy* or *Difficult*. Therefore, θ at the beginning of the trial should be in between the optimal boundaries for the two difficulties. As they make observations, they will update their estimate of the task difficulty. In particular, as time within a trial progresses, and the

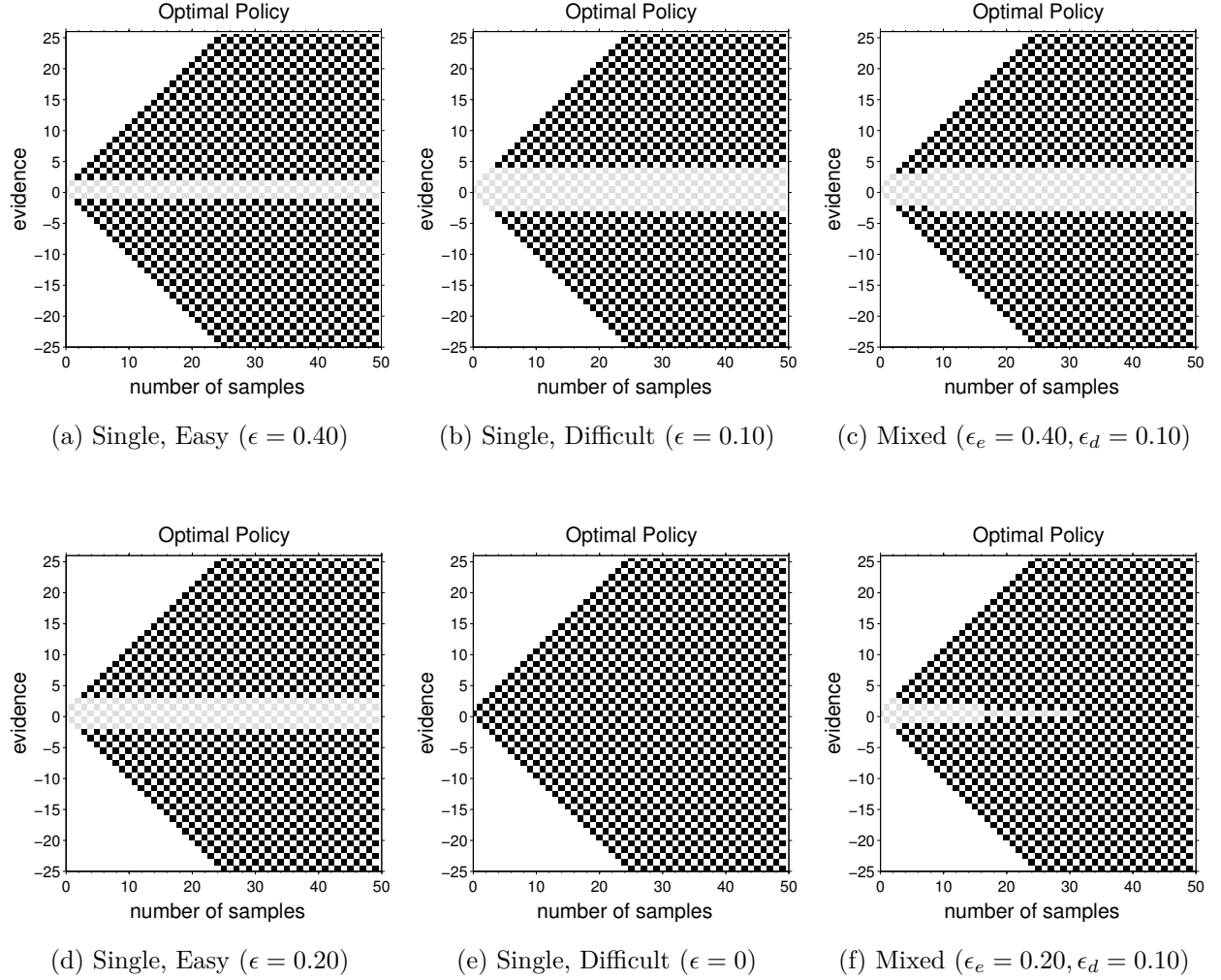


Figure 2. Optimal policies for Single and Mixed-difficulty tasks. Black squares indicate states of the MDP where the optimal action is to *go* – i.e., choose an alternative – while gray squares indicate states where the optimal action is to *wait* – i.e., collect more evidence. In each row, the two panels on the left show optimal policies for Single-difficulty tasks with two different levels of difficulty and the right-most panel shows optimal policy for Mixed-difficulty task obtained by mixing the difficulties is the two left-hand panels. The inter-trial intervals were $D_C = D_I = 70$ for the top row (panels (a)–(c)) and $D_C = D_I = 50$ for the bottom row (panels (d)–(f)).

1 decision boundary has not been reached, the estimated probability of the trial being
 2 *Difficult* increases and the decision boundary moves towards the optimal boundary for the
 3 *Difficult* trials.

4 The above principle is illustrated in Figures 2c and 2f showing optimal boundaries for
 5 two sample Mixed-difficulty tasks. Figure 2f shows the optimal boundary for a task in
 6 which half of the trials have moderate difficulty and half are very difficult (the optimal
 7 bounds for Single-difficulty tasks with corresponding values of drift are shown in Figures 2d
 8 and 2e). As the time progresses the optimal decision maker infers that a trial is likely to be
 9 very difficult, so an optimal strategy involves moving on to the next trial (which may be
 10 easier), i.e. decreasing the decision boundary with time in the trial.

11 In contrast, when the boundary for the *Difficult* task is higher than the *Easy* task
 12 (the *Difficult* task is *not* extremely hard; Figures 2a and 2b), the optimal boundary in the
 13 Mixed-difficulty task will again start at a value in-between the boundaries for the *Easy* and
 14 *Difficult* tasks and approach the boundary for the *Difficult* task (Figure 2c). In this case,
 15 the boundary for the *Difficult* task will be higher than the *Easy* task meaning that the
 16 optimal boundary will *increase* with time.

17 In summary, the mathematical model makes three key predictions about the
 18 normative behaviour: (i) optimal decision boundaries should stay constant if all decisions
 19 in a sequence are of the same difficulty, (ii) it is optimal to decrease decision boundaries if
 20 decisions are of mixed difficulty and some decisions are extremely difficult (or impossible),
 21 and (iii) it may be optimal to keep decision boundaries constant or even increase them in
 22 Mixed-difficulty tasks where the *Difficult* decision is not too difficult. In the next three
 23 sections, we compare human behaviour with these normative results.

24 Experiment 1

25 In order to compare human behaviour with the normative behaviour described above,
 26 we designed an experiment that involved an evidence-foraging game which parallels the

expanded-judgement task described in the previous section. We modelled this evidence-foraging game on previous expanded-judgement tasks, such as Irwin et al. (1956) and Vickers, Burt, et al. (1985), where participants are shown a sequence of discrete observations and required to judge the distribution from which these observations were drawn. We modified these expanded-judgement paradigms so that (i) the observations could have only one of two values (i.e. drawn from the Bernoulli distribution), (ii) the reward-structure of the task was based on performance, and (iii) the task had an intrinsic speed-accuracy trade-off (SAT). We introduced a speed-accuracy trade-off by using a fixed-time blocks paradigm: the experiment was divided into a number of games, the total duration for each game was fixed, and participants could attempt as many decisions as they like during this period. Therefore, if a participant takes a very long time for each decision they are likely to be accurate, but will not be able to complete many decisions during a game. If a participant decides very quickly, they are likely to perform worse in terms of accuracy, but will have more reward “opportunities” during the game. The goal of the participants was to collect as much reward as possible during each game, so they need to find a balance between these two strategies.

In this expanded judgement task, we are able to record the exact sequence of stimuli presented to the participants and the position in state-space (t, x) at which participants made their decisions. Based on the location of these decisions, we inferred how the decision boundary for a participant depended on time. According to the above theory, the optimal decision boundary should be independent of time in Single-difficulty tasks, but could vary with time during Mixed-difficulty tasks. By comparing the inferred decision boundary with optimal boundaries in each type of task, we assessed whether participants adjusted their decision boundaries to maximise reward-rate.

1 Methods

2 **Description of task.** Twenty four participants from the university community
 3 were asked to play a set of games on a computer. The number of participants was chosen
 4 to give a sample size that is comparable to previous human decision-making studies² and
 5 kept constant during all of our experiments. Each game lasted a fixed duration and
 6 participants made a series of decisions during this time. Correct decisions led to a reward
 7 and participants were asked to maximise the cumulative reward. The game was
 8 programmed using Matlab® and Psychtoolbox (Brainard, 1997; Pelli, 1997; Kleiner et al.,
 9 2007) and was played using a computer keyboard. The study lasted approximately fifty
 10 minutes, including the instruction phase and training.

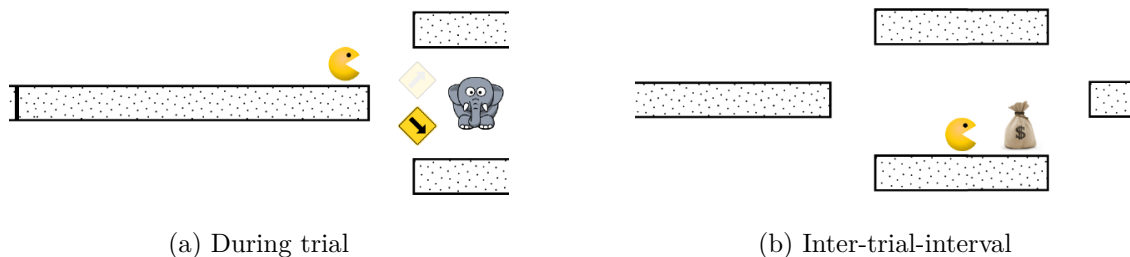


Figure 3. Two screenshots of the display during the experiment. The left panel shows the display during the evidence-accumulation phase of a trial. Participants chose whether pacman goes up or down after seeing a sequence of cues (arrows) pointing up or down. The elephant next to the arrow indicates that this is an Easy game, so the arrow points to the reward-holding path with probability 0.70. The right panel shows a screenshot during the inter-trial interval. The participant has chosen the lower path and can now see that this was the correct (rewarded) decision.

²See, for example, Palmer et al. (2005) who test six participants with ~ 560 trials for each participant and Ratcliff and McKoon (2008) who test fifteen participants with ~ 960 trials per participant. We used a performance-based reward paradigm (outlined below), which meant that the number of trials varied between participants and experiments but were between ~ 150 and ~ 650 trials per participant for each of the experiments reported below.

During each game, participants were shown an animated creature (pacman) moving along a path (Figure 3). A trial started with pacman stationary at a fork in the path. At this point pacman could jump either up or down and the participant made this choice using the ‘up’ or ‘down’ arrow keys on the keyboard. One of these paths contained a reward, but the participant could not see this before making the decision. Participants were shown a sequence of cues and they could wait and watch as many cues as they wanted before making their choice. The display also showed the total reward they accumulated in the experiment and a progress bar showing how much time was left in the current game (not shown in Figure 3).

Once the participant indicated their choice, an animation showed pacman moving along the chosen path. If this path was the rewarded one, a bag with a \$ sign appeared along the path (right panel in Figure 3). When pacman reached this bag, the reward was added to the total and pacman navigated to the next next fork and this started the next trial. If the participant chose the unrewarded path, the money bag appeared along the other path.

The inter-trial interval (ITI) started as soon as the participant indicated their choice. We manipulated the inter-trial interval for correct and incorrect decisions by varying pacman’s speed. Participants were told that pacman received a “speed-boost” when it ate the money bag so that inter-trial interval for correct decisions was smaller than that for incorrect decisions. Values for all parameters used during the game are shown in Table 1.

Cue stimuli. When pacman reached a fork, cues were displayed at a fixed rate, with a new cue every 200ms. We call this delay the inter-stimulus interval (ISI). During these 200ms, the cue was displayed for 66ms, followed by 134ms of no cue. Each cue was the outcome of a Bernoulli trial and consisted of either an upwards or a downwards pointing arrow. This arrow indicated the rewarded path with a particular probability.

Next to the cues, participants were shown a picture of either an elephant or a penguin. This animal indicated the type of game they were playing. One of the two

Parameter name	Value
Inter-stimulus interval (ISI)	200 msec
Inter-trial interval, correct ($ISI * D_C$)	3 sec
Inter-trial interval, incorrect ($ISI * D_I$)	10 sec
Reward	2p
Drift for Easy condition (ϵ_e)	0.20
Drift for Difficult condition (ϵ_d)	0
Block duration, training	150 secs
Block duration, testing, Easy	240 secs
Block duration, testing, Difficult	300 secs
Block duration, testing, Mixed	300 secs

Table 1

Values of parameters used during the game.

animals provided cues with a probability 0.70 of being correct, while the other animal provided cues with a probability 0.50 of being correct. Thus, the two animals mapped to the two Single-difficulty conditions – Easy (with $\epsilon = 0.20$) or Difficult (with $\epsilon = 0$) – shown in Figure 2d and 2e. The mapping between difficulties and animals was counterbalanced across participants.

We chose the values of *up-probability* so that the optimal decision boundaries in the Mixed-difficulty case have the steepest slope, making it easier to detect if participants decrease decision boundaries. The theory in the previous section shows that decision boundaries decrease only when the Difficult decisions are extremely difficult. In the experiment we set the *up-probability* for Difficult condition to the extreme value of 0.5, i.e. $\epsilon_d = 0$ – therefore, the cues do not give any information on the true state of the world. Using this value has two advantages: (i) it leads to optimal decision boundaries in Mixed-difficulty games with the steepest decrease in slope, and (ii) it makes it easier for

participants to realise that the optimal boundary in the Difficult condition is very low (in fact, the optimal strategy for Difficult games is to guess immediately). Optimal boundaries should also decrease (although with a smaller slope) when decisions are marginally easier (e.g., $\epsilon_d = 0.03$). But we found that participants frequently overweight evidence given by these low probability cues (perhaps analogous to the overweighting of small probabilities in other risky choice situations, e.g. Tversky & Kahneman, 1992; Gonzalez & Wu, 1999) and need a large amount of training to establish the optimal behaviour in such extremely difficult (but not impossible) games. In contrast, when $\epsilon_d = 0$, participants could learn the optimal strategy Difficult games with a small amount of training.

The experiment consisted of three types of games: ‘Easy’ games where only the animal giving 70% correct cues appeared at each fork; ‘Difficult’ games where only the animal giving the 50% cues appeared at the fork and ‘Mixed’ games where the animal could change from one fork to the next. Participants were given these probabilities at the start of each game and also received training on each type of game (see ‘Structure of experiment’ below). Importantly, during Mixed games participants were shown a picture of a wall instead of either animal and told that the animal was hidden behind this wall. In other words, they received no information indicating whether a particular trial during a Mixed game was ‘easy’ or ‘difficult’ other than the cues themselves, so that they had to infer the type of trial based on these cues. This corresponds to the Mixed-difficulty task shown in Figure 2f.

Reward structure. Participant reimbursement was broken down into three components. The first component was fixed and every participant received £6 for taking part in the study. The second component was the money bags accumulated during the experiment. Each money bag was worth 2p and participants were told that they could accumulate up to £5 during the experiment. The third component was a bonus prize of £20 available to the participant who accumulated the highest reward during the study. Participants were not told how much other participants had won until after they took part

1 in the study.

2 **Structure of experiment.** The experiment was divided into a training phase and
 3 a testing phase. Participants were given training on each type of game. The duration of
 4 each training game was 150 seconds. This phase allowed participants to familiarise
 5 themselves with the games and probability of cues as well as understand the speed
 6 accuracy trade-off for each type of games. The reward accumulated during the training
 7 phase did not count towards their reimbursement.

8 The testing phase consisted of six games, two of each type. Participants were again
 9 reminded of the probabilities of cues at the start of each game. The order of these games
 10 was counterbalanced across participants so that each type of game was equally likely to
 11 occur in each position in the sequence of games. The duration of the Easy games was 240
 12 seconds, while the Difficult and Mixed games lasted 300 seconds each. The reason for
 13 different durations for different types of games was that we wanted to collect around the
 14 same amount of data for each condition. Pilot studies showed that participants generally
 15 have faster reaction time during the Easy games (see Results below). Therefore, we
 16 increased the length of the Difficult and Mixed blocks. By using these durations,
 17 participants made approximately seventy to ninety choices during both Easy and Mixed
 18 conditions. In the middle of each game, participants received a 35 seconds break.

19 **Eliminating non-decision time.** We preprocessed the recorded data to eliminate
 20 *non-decision time* – the delay between making a decision and executing a response. Due to
 21 this non-decision time, the data contained irrelevant stimuli that were presented after the
 22 participant had made their decision. To eliminate these irrelevant stimuli, we estimated the
 23 non-decision time for each participant based on their responses during the Easy games.
 24 Appendix A illustrates the method in detail; the key points are summarised briefly below.

25 For each participant, we reversed the sequence of stimuli and aligned them on the
 26 response time. Let us call these ordered sequence $(s_1^i, s_2^i, \dots, s_T^i)$, where i is the trial
 27 number and $(1, 2, \dots, T)$ are the stimulus indices before the response. Each stimulus can

be either ‘up’ or ‘down’, i.e. $s_t^i \in \{\text{up}, \text{down}\}$. At each time step, we estimated the correlation (across trials) between the observing a stimulus in a particular direction and making a decision to go in that direction. That is, we computed p_t at each stimulus index, $t \in \{1, 2, \dots, T\}$, as the fraction of trials where the response $r^i \in \{\text{up}, \text{down}\}$ is the same as s_t^i . So, for each participant, the values (p_1, p_2, \dots, p_T) serve as an estimate of the correlation between the stimulus at that index and the response.

If stimuli at a particular index, t , occurred after the decision, i.e. during the non-decision time, we expected them to have a low correlation with response and consequently p_t to be below the drift rate, 0.70. We determined the first index in the sequence with p_t larger than 0.75; that is, the first index with more than 75% of stimuli in the same direction as the response. This gave us an estimate of the number of stimuli, ND , that fall in the non-decision period. We used this estimate to eliminate the stimuli, s_1, \dots, s_{ND} from each recorded sequence for the participant. See Figure A1 in Appendix A.

For twenty-one out of twenty-four participants, we estimated $ND = 1$, that is, a non-decision delay of approximately 200ms. For two subjects the non-decision delay was two stimuli and for one participant no stimuli were excluded.

Exclusion of participants. To ensure that each participant understood the task, we conducted a binomial test on responses in the Easy and Mixed-difficulty games. This test checked whether the number of correct responses during a game were significantly different from chance. Two participants failed this test during Mixed-difficulty games and were excluded from further analysis.

Analysis method. We now describe how we estimated the decision boundary underlying each participant’s decisions. In signal-detection paradigms, the experimenter cannot observe the exact sequence of samples based on which the participant made their decision. Therefore, parameters like boundary are obtained by fitting a sequential sampling or accumulator model to the reaction time and error distributions. In contrast, the expanded-judgement paradigm allows us to observe the entire sequence of samples used to

make each decision. Therefore, our analysis method takes into account not only the evidence and time at which the decision (‘up’/‘down’) was made, but also the exact sequence of actions (*wait/go*) in response to the sequence of cues seen by the participant. It also takes into account the trial-to-trial variability in the behaviour of participants: even when participants saw the exact same sequence of cues, they could vary their actions from one trial to next.

If a participant makes a decision as soon as evidence crosses the boundary, the value of time and evidence, (t, x) during each decision should lie along this boundary. Therefore, one way to recover this boundary is by simply fitting a curve through the values of (t, x) for all decisions in a block. However, note that participants show a trial-to-trial variability in their decision-making. Sequential sampling models account for this trial-to-trial variability by assuming noisy integration of sensory signals as well as variability in either drift, starting point or in threshold (see Ratcliff, 1978; Ratcliff & Smith, 2004). We chose to model this variability by assuming there is stochasticity in each *wait/go* decision. That is, instead of waiting when evidence was below the boundary and going as soon as evidence crossed the boundary, we assumed that a participant’s decision depended on the outcome of a random variable, with the probability of the outcome depending on the accumulated evidence and time.

Specifically, we define two predictor variables – the evidence accumulated, $X = x$, and the time spent in the trial, $T = t$ – and a binary response variable, $A \in \text{wait}, \text{go}$. The probability of an action can be related to the predictor variables using the following logistic regression model:

$$\log \frac{\mathbb{P}(A = \text{go})}{\mathbb{P}(A = \text{wait})} = \beta_0 + \beta_T * T + \beta_X * X \quad (2)$$

where β_T and β_X are the regression coefficients for time and evidence, respectively, and β_0 is the intercept. Given the triplet (X, T, A) for each stimulus in each trial, we estimated for each type of game and each participant the $\hat{\beta}_0, \hat{\beta}_T$ and $\hat{\beta}_X$ that maximised the likelihood of the observed triplets.

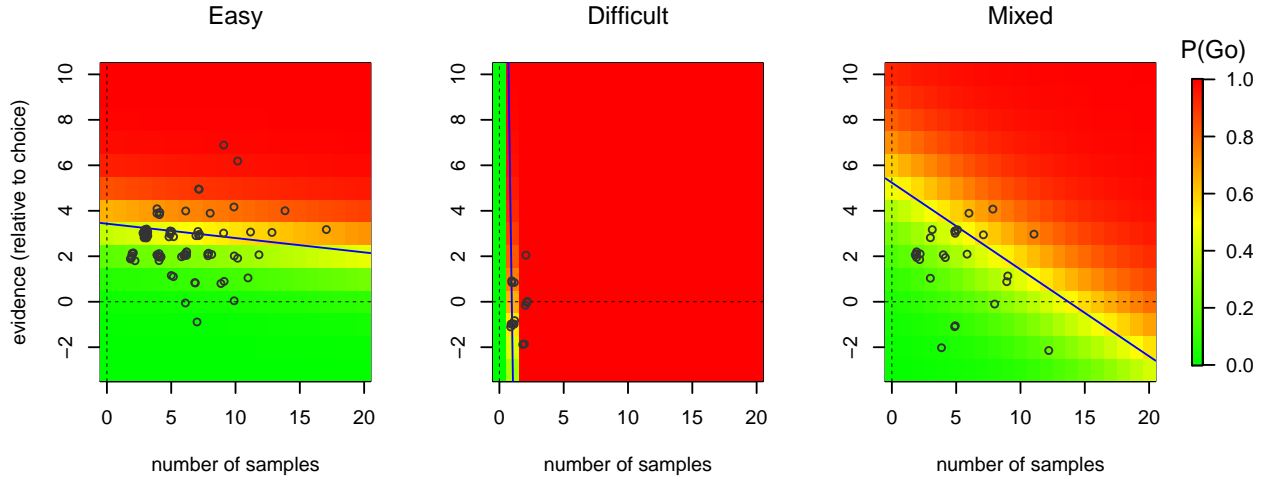


Figure 4. The decisions made by a subject during Experiment 1 and the inferred boundaries based on these decisions. Each scatter-plot shows the values of evidence and time where the subject made decisions during a particular game (only easy trials considered during Mixed game). These values have been slightly jittered for visualisation. The heat-map shows the $P(\text{Go}|X_t = x)$ for each x and t inferred using logistic regression (see text). The solid line shows a “line of indifference” where $P(\text{Go}|X_t = x) = P(\text{Wait}|X_t = x)$ and serves as a proxy for the subject’s boundary (see Appendix B).

Figure 4 shows the results of applying the above analysis to one participant. The data are split according to condition - Easy, Difficult or Mixed. Each circle shows the end of a random walk (sequence of stimuli) in the time-evidence plane. These random walks were used to determine the (maximum likelihood) regression coefficients, $\hat{\beta}_0$, $\hat{\beta}_X$ and $\hat{\beta}_T$, as outlined above. These estimated coefficients are then used (Equation 2) to determine the probability of going at each x and t , which is shown as the heat-map in Figure 4.

This heat-map shows that, under the Easy condition, this participant’s probability of going strongly depended on the evidence and weakly on the number of samples. In contrast, under the Difficult condition, the participant’s probability of going depends almost exclusively on the number of samples – most of their decisions are made within a

couple of samples and irrespective of the evidence. Under the Mixed condition, the probability of *going* is a function of both evidence and number of samples.

Since we were interested in comparing the slopes of boundaries during Easy and Mixed conditions, we determined a *line of indifference* under each condition, where $P(A = go) = P(A = wait)$, that is, the participant was equally likely to choose actions *wait* and *go*. Substituting in Equation 2 gives the line:

$$X = -\frac{\hat{\beta}_T}{\hat{\beta}_X} * T - \frac{\hat{\beta}_0}{\hat{\beta}_X} \quad (3)$$

with slope $-\frac{\hat{\beta}_T}{\hat{\beta}_X}$ and intercept as $-\frac{\hat{\beta}_0}{\hat{\beta}_X}$. We used the slope of this line as an estimate for the slope of the boundary. Appendix B reports a set of simulations that tested the validity of this assumption and found that there is a systematic relationship between this inferred slope and the true slope generating decisions. Importantly, these simulations also demonstrate that even if the variability in data is due to noisy integration of sensory signals (rather than trial-to-trial variability in decision boundary), this inferential method still allows us to make valid comparisons of slopes of boundaries in Easy and Mixed games.

Each panel in Figure 4 also shows the line of indifference for the condition. The slope of the line of indifference is steepest under the Difficult condition followed by the Mixed condition and most flat for the Easy condition. Note that for the Mixed condition, we only considered the “easy” trials – that is, trials showing cues with correct probability = 0.70. This ensured that we made a like-for-like comparison between Easy and Mixed conditions.

A quantitative comparison of slopes between conditions can be made by taking the difference between slopes. However, a linear difference is inappropriate as large increasing slopes are qualitatively quite similar to large decreasing slopes – both indicate a temporal, rather than evidence-based boundary (e.g. the Difficult condition in Figure 4). Therefore, we compared slopes in the Mixed and Easy conditions by converting these slopes from gradients to degrees and finding the circular difference between slopes:

$$\Delta m = ((m_e - m_m + 90) \bmod 180) - 90 \quad (4)$$

where m_e and m_m are the slopes in Easy and Mixed conditions, respectively; Δm is the difference in slopes and mod is the modulo operation. Equation 4 ensures that the difference between slopes is confined to the interval $[-90, +90]$ degrees and large increasing slopes have a small difference to large decreasing slopes.

The above analysis assumes that evidence accumulated by a participant mirrors the evidence presented by the experimenter – so there is no loss of evidence during accumulation and the internal rate of evidence accumulation remains the same from one trial to next. In Appendix C we performed simulations to verify that inferences using the above analysis remain valid even when there is loss in information accumulated and when the drift rate varied from one trial to next.

Results

The mean reaction times during Easy, Difficult and Mixed games were 1444ms ($sem = 23ms$), 1024ms ($SEM = 47ms$) and 1412ms ($SEM = 22ms$), respectively, where SEM is the within-subject standard error of the means. Note that ‘reaction time’ here refers to ‘decision time’, i.e. the raw response time minus the estimated non-decision time. As noted above, the non-decision time for most participants was approximately 200ms. Figure 5 compares the slopes for the lines of indifference in the Easy and Mixed games (black circles). Error bars indicate the 0.95 percentile confidence interval³. Like the participant shown in Figure 4 the estimated slope for most participants was more negative during the Mixed games than during Easy games, falling below the identity line. A paired t-test on the difference in slopes in the two conditions (using Equation 4) confirmed that there was a significant difference in the slopes ($t(21) = 5.24, p < 0.001, m = 15.94, d = 1.20$), indicating that the type of game modulated how participants set their decision boundary.

³Each confidence interval is based on percentiles of the bootstrap distribution of the lines of indifference. Each bootstrap distribution is obtained by generating 1000 independent bootstrapped data sets (per condition) and computing the slope for the line of indifference on the data set. Each data set consists of N sampled trials, where N is the number of trials (for that condition) seen by the participant

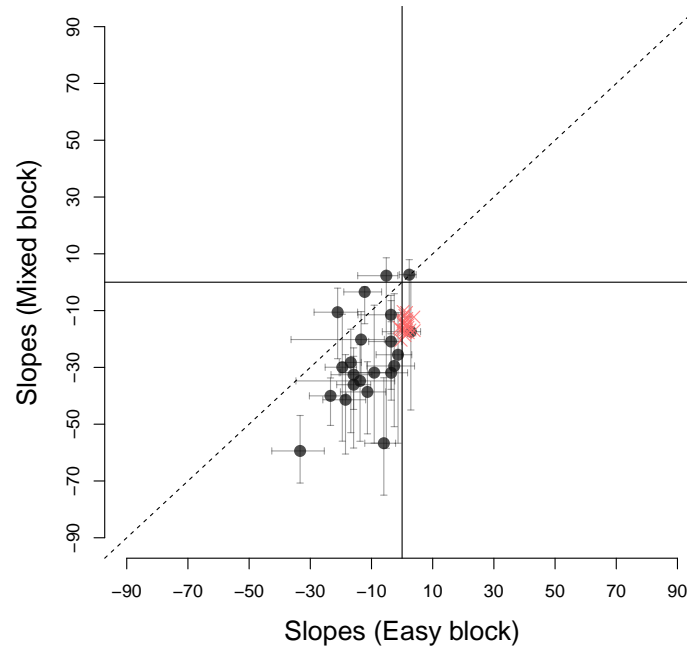


Figure 5. Each circle (black) compares the estimated slope in Easy and Mixed games for one participant. Circles below the dashed line were participants who had a larger gradient of the inferred boundary during the Mixed games as compared to the Easy games. Error bars indicate the 0.95 percentile bootstrapped confidence intervals for the estimated slopes. Crosses (red) show twenty-four simulated participants – decisions were simulated using a rise-to-threshold model with optimal boundaries shown in Figures 2d and 2f.

Figure 5 also shows the relationship between the slopes of Easy and Mixed games for twenty-four simulated participants (red crosses) who optimise the reward-rate. Each of these participants had slopes of boundary calculated using dynamic programming (Malhotra et al., under review) and made decisions based on a noisy integration of evidence to this optimal boundary. The slopes in each condition were then inferred using the same procedure as for our real participants. These optimal participants, like the majority of participants in our study, had a larger (negative) slope in the Mixed condition than the Easy condition. However, in contrast to the optimal participants, the majority of

participants also exhibited a negative slope during the *Easy* games, indicating that they lowered their decision boundary with time during this condition. A t-test confirmed that the slope during easy condition was less than zero ($t(21) = -5.51, p < 0.001, m = -11.47$). Participants also showed substantial variability in the decision boundary in the easy condition, with slopes varying between 0 and 45 degrees.

An alternative possibility is that participants change their decision boundary during the experiment, adopting a higher (but constant) boundary towards the beginning and lowering it to different (constant) boundary during the experiment. In order to check for this possibility, we split the data from each condition into two halves and checked whether the mean number of samples required to make a decision changed from the first half to second half of the experiment. During Easy games, we found that participants observed 7.5 and 6.8 samples, on average, during the first and second half of the experiment, respectively. During Mixed games, these mean observations changed to 6.8 and 6.2 samples, on average, during the first and second half of the experiment. A two-sided paired t-test which examined whether the mean number of samples were different in the two halves of the experiment found no significant difference in either the Easy games ($t(21) = 1.76, p = 0.09, m = 0.73$) or in the Mixed games ($t(21) = 1.40, p = 0.18, m = 0.64$).⁴

We checked the robustness of these results by performing a model comparison exercise, pitting a time-varying decision boundary against a fixed boundary model. The latter simply involves a logistic regression in which the decisions to *wait* or *go* were based on evidence only. The full details of this model comparison procedure and results are described in Appendix D. Based on a comparison of Bayesian Information Criteria (Schwarz, 1978; Wagenmakers, 2007), the time-varying model provided a better account of the behaviour of fifteen out of twenty-two participants in Mixed-difficulty games. For three participants, the evidence was ambiguous and for the remaining four participants the

⁴Analogous analysis was also done for Experiments 2a-2d as well and no significant differences were found.

1 simpler, fixed boundary model won. In Easy games, the model using time as a predictor
 2 was better at accounting for data from thirteen participants while the simpler model
 3 performed better to data for eight participants.

4 In order to understand why participants decrease the decision boundary in Easy
 5 games and why different participants show a large variation in their choice of boundary, we
 6 computed the reward-rate accrued by each participant’s choice of boundary and compared
 7 it to the reward-rate for the optimal policy. This gave us the cost of setting any
 8 non-optimal decision boundary. Figure 6 shows the landscape (heat-map) of the
 9 reward-rate for each type of game for a host of different boundaries, defined by different
 10 combinations of intercepts and slopes. The circles indicate the intercepts and slopes of the
 11 inferred line of indifference of each participant.

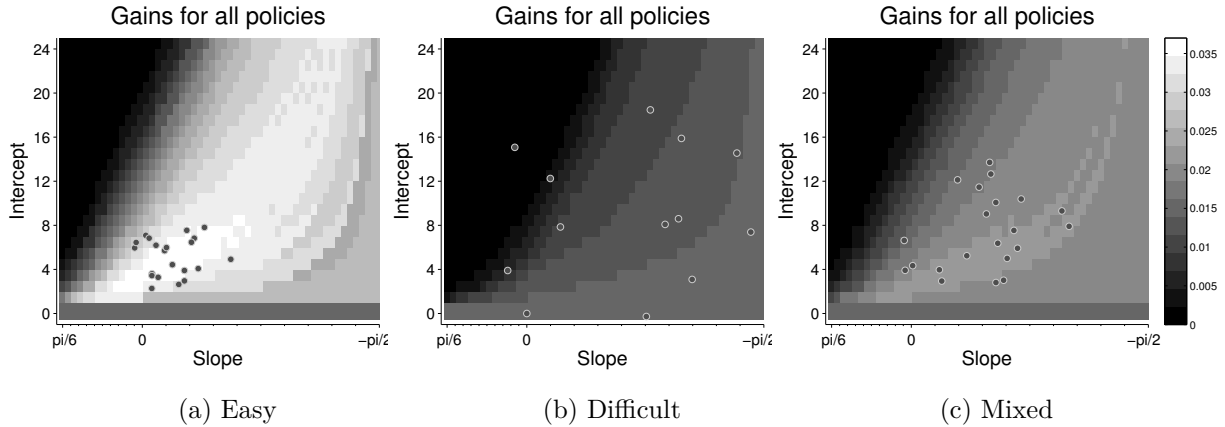


Figure 6. The reward-rates for different decision boundaries. In each panel, the slope and intercept determine a linear boundary. The actions of all states below the boundary are set to *wait* and all states above are set to *go*. The heat-map in each panel shows the reward-rate for each threshold. The circles show the inferred boundaries used by the participants in Experiment 1.

12 Notice, in particular, the landscape for the Easy games. Even though the peak of this
 13 landscape lies at the policy with zero slope (flat bounds), there is a “ridge” of policies on
 14 the landscape where the reward-rate is close to optimal. The policies chosen by most

participants in Experiment 1 seem to lie along this ridge – even though participants do not necessarily choose the optimal policy, they seem to be choosing policies that are close to optimal. A similar pattern holds in the Mixed games. In contrast, during Difficult games, the average reward is low, irrespective of the policy. Correspondingly, there is a large variability in the policies chosen by participants. We examine the effect of reward landscape on the policies chosen by participants in more detail at the end of Experiment 2.

Experiment 2

Experiment 1 established that people modulate their decision boundary based on task difficulty and variations in the reward landscape. However, our experimental paradigm – effectively an expanded-judgement task – is clearly very different from the dominant, typically signal-detection paradigms used to test rise-to-threshold models and time-varying boundaries (e.g., Britten, Shadlen, Newsome, & Movshon, 1992; Palmer et al., 2005; Ludwig, 2009; Starns & Ratcliff, 2012). In our paradigm, response times in Mixed games were generally between 1-2s, whereas in the perceptual decision-making literature, reaction times are typically between 0.5-1s (Palmer et al., 2005). It is possible that at this speed participants do not, or cannot, modulate their decision boundaries and instead adopt sub-optimal fixed thresholds.

Our aim in Experiment 2 then was to replicate and extend our findings to a more rapid task, where reaction times were similar to a signal-detection paradigm. More generally, we tested the robustness and generality of the results from the expanded-judgement task of Experiment 1 by introducing different (i) stimulus materials, (ii) inter-stimulus intervals and (iii) inter-trial intervals. The variation in inter-stimulus interval was designed to induce more rapid decision-making (with reaction times typically $< 1s$). Since the optimal policies are computed on a relative time scale (based on a unit inter-stimulus interval), we can scale both the inter-stimulus and inter-trial interval without affecting the optimal policy, but reducing the reaction time. The variation in inter-trial

interval (specifically: for correct decisions, D_C) was introduced to manipulate the reward landscape, without affecting the *optimal policy*. Bogacz et al. (2006) have previously shown that the optimal policy is invariant to change in D_C for Single-difficulty games. Malhotra et al. (under review) showed that this result generalises to the Mixed-difficulty scenario: optimal policy for Mixed-difficulty games depends only on the inter-trial interval for incorrect decisions, D_I , but is independent of the inter-trial interval for correct decisions, D_C , as long as $D_C < D_I$. If participants were optimising the reward-rate, they should not change their decision boundary with a change in D_C . However, as we will see below, changing D_C does affect the wider reward landscape around the optimal policy and we explored to what extent participants were sensitive to this change.

Parameter name		Value
Drift for Easy condition (ϵ_e)		0.22
Drift for Difficult condition (ϵ_d)		0
Reward		2p

ISI \ ITI	$D_I = ISI * 50$	
	$D_C = \frac{1}{3}D_I$	$D_C = D_I$
200msec	Experiment 2a	Experiment 2c
50msec	Experiment 2b	Experiment 2d

Table 2

Values of parameters for Experiment 2. The parameters that are common to all four sub-experiments are listed at the top. Each of the four sub-experiments has a different combination of inter-stimuli and inter-trial intervals, the values of which are listed at the bottom.

Permutations of varying these two parameters leads to four experiments, which we have labeled Experiments 2a –2d. The values of parameters for each experiment are shown in Table 2. Experiment 2a was a replication of Experiment 1 with exactly the same

parameters, but using the new paradigm (described below). In Experiment 2b, we scaled the ISI and ITI to elicit rapid decisions but kept all other parameters the same as Experiment 2a. In Experiment 2c, we increased the inter-trial-interval for correct responses to match that for incorrect responses. All other parameters were kept same as Experiment 2a. Finally, in Experiment 2d, we scaled ISI and ITI to elicit rapid decisions and also matched inter-trial intervals for correct and incorrect decisions.

Like Experiment 1, twenty-four healthy adults between the age of eighteen and thirty-five from the university community participated in each of these experiments, with no overlapping participants between experiments.

Methods

Decreasing the inter-stimulus interval increases two sources of noise in the experiment: (i) noise due to variation in attention to cues (i.e. there is a greater likelihood of participants “missing” samples when they are coming in faster) (ii) noise due to visual interference between consecutive cues in the same location. The second source of noise is particularly challenging for our purposes. That is, the analysis presented here assumes that each evidence sample is processed independently. However, if we were to present a sequence of cues in rapid succession, it is clear that, due to the temporal response properties of the human visual system, successive cues could “blend in” with each other (Georgeson, 1987). As a result, we could not simply speed-up the presentation of the arrow cues in the pacman task. We adapted the original task from Experiment 1 to another evidence-foraging game that retained the structure of the paradigm and that allowed for systematic variation of the various parameters of interest (i.e. inter-stimulus and inter-trial intervals).

Participants were again asked to maximise their cumulative reward by making correct decisions in a game. But now, during each trial participants focused on a fixation cross in the middle of the screen with gray background⁵ and were told that a reward was either on

⁵The luminance of the monitor was gamma-corrected so that luminance was a linear function of grayscale RGB value. The background luminance was fixed to 0.5 on a scale of $[0, 1]$.

the left or the right of the fixation cross. In order to make their choice, participants were shown cues that could appear either to the left or right of the fixation cross. In order to minimise interference (see below) cues could appear in two alternative locations on each side – ‘left-up’ or ‘left-down’ on the left and ‘right-up’ or ‘right-down’ on the right. A cue appeared on the same side as the reward with a given probability. Participants were given this probability at the beginning of the game. For Single-difficulty games, they were told that this probability was the same ($\frac{1}{2} \pm 0.22$) for all trials within this game. For the Mixed-difficulty games, they were told that a particular trial during the game could give cues with one of two different probabilities ($\frac{1}{2} \pm 0.22$ or $\frac{1}{2} \pm 0$) and they were given these possible probabilities at the start of each game (i.e., block). Participants were again told that they could see as many cues as they wanted during a trial before making a decision, but the total duration of the game was fixed. Figure 7 shows an example trial in which the participant makes the decision to go left after observing a series of cues.

Each cue was a Gabor pattern (sinusoidal luminance grating modulated by a 2D Gaussian window). We designed these cue patterns to minimise interference between consecutive patterns. The integration period of early visual mechanisms depends strongly on the spatiotemporal parameters of the visual patterns. But for coarse (i.e. low spatial frequency) and transient patterns it should be less than 100 ms (Georgeson, 1987; Watson, Ahumada, & Farrell, 1986). To ensure the low spatial frequency we fixed the nominal spatial frequency of the Gabor to 0.4 cycles/deg (we did not precisely control the viewing distance, so the actual spatial frequency varied somewhat between participants) and the size of the Gaussian window to 1.2 deg (2D standard deviation). The patterns had a vertical orientation. In the “fast” experiments (Experiments 2b and 2d), each cue was displayed for 10ms and the delay between onset of two consecutive cues (the inter-stimulus interval) was 50ms. To ensure that consecutive cues were processed independently by the visual system we (i) alternated the location on one side of the screen (e.g. ‘left-up’ and ‘left-down’) so that the smallest inter-stimulus interval at any one retinal location was

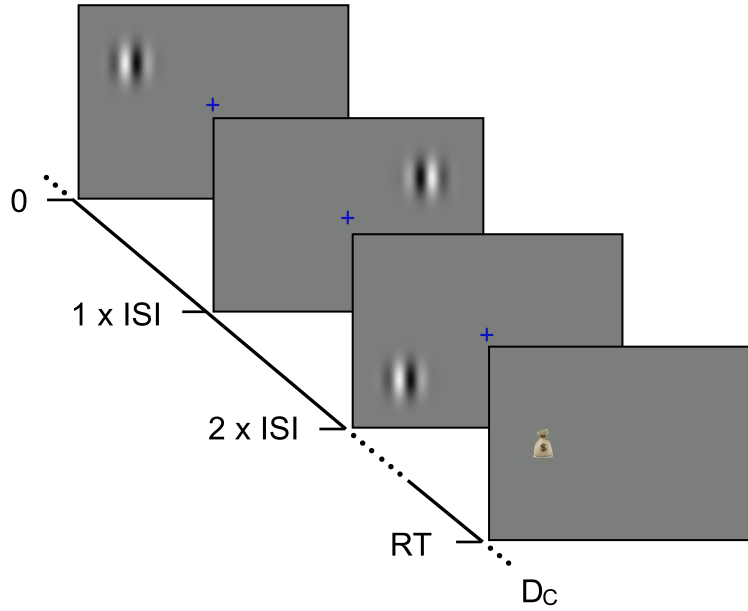


Figure 7. An illustration of the paradigm for Experiments 2a–2d. During each trial, participants chose left or right based on a sequence of cues. Each cue was a Gabor pattern displayed (for a fifth of ISI) in one of four possible locations, equidistant from the fixation cross. If the decision was correct (as in this example), a money bag was displayed on the chosen side of the fixation cross and the participant waited for the duration D_C before starting the next trial. If the decision was incorrect, no money bag was displayed and the participant waited for the duration D_I before starting the next trial.

1 90ms and (ii) alternated the phase of the patterns (90° and 270°).

2 Participants indicated their choice by pressing the left or right arrow keys on a
 3 keyboard. When the decision was correct, a money bag appeared on the chosen side.
 4 During the inter-trial interval, an animation displayed this money bag moving towards the
 5 bottom of the screen. When the decision was incorrect, no money bag appeared. All
 6 money bags collected by the participant remained at the bottom of the screen, so
 7 participants could track the amount of reward they had gathered during the current game.

8 The structure of the experiment was the same as Experiment 1, with the experiment
 9 consisting of a set of games of fixed durations and given difficulties. Each game consisted of

a sequence of trials where participants could win a small reward if they made the correct decision or no reward if they made an incorrect decision. Games were again of three different types: (i) Type 1, corresponding to Easy games from Experiment 1, (ii) Type 2 corresponding to Difficult games and (iii) Type 3 corresponding to Mixed games. The type of the game was indicated by the colour of the fixation cross – Type 1: Green, Type 2: Red and Type 3: Blue. The order of games was counterbalanced across participants.

In all four experiments the *up-probability* for Easy and Difficult games was 0.50 ± 0.22 and 0.50 ± 0 , respectively. During Mixed games, Easy and Difficult trials were equally likely. The reward-rate-optimal policy for Easy games was again to maintain constant threshold (Figure 2d) while for Difficult it was to guess immediately (Figure 2e). Similarly, the optimal policy for the Mixed condition was to start with a high boundary (similar to the boundary at the start of Easy games) and steadily decrease it, eventually making a decision at $x = 0$ (Figure 2f). Just like in Experiment 1, participants were given training on each type of game and the reward structure was divided into three components: £7 for participating, 2p for each correct response and £20 for the participant accumulating the largest number of money bags.

Results

We analysed data using the same method as Experiment 1 after removing the non-decision time. *Wait* and *go* actions were used to determine the probability of *going* at all combinations of evidence and time, which were then used to determine a line of indifference, where the probability of wait matched the probability of go. We compared the slopes of this line of indifference for Easy and Mixed games for each of the four experiments.

Experiment 2a. This experiment used the same parameters as Experiment 1, but replaced the Pacman game, with the evidence-foraging game described in Figure 7. Two participants failed the binomial test in the Mixed games and were excluded from analysis. The mean reaction times during Easy, Difficult and Mixed games were 1155ms (5.8

1 samples, $SEM = 17\text{ms}$), 618ms (3.1 samples, $SEM = 27\text{ms}$) and 1142ms (5.7 samples,
 2 $SEM = 19\text{ms}$), respectively. We estimated the average number of stimuli that fell in the
 3 non-decision period (ND) to be 0.96, i.e. a average non-decision delay of approximately
 4 192ms. For twenty-one out of the twenty-two participants, we estimated $ND = 1$ and for
 5 one participant no stimuli were excluded. Figure 8 (top-left panel) shows a comparison of
 6 the estimated slopes of lines of indifference in Easy and Mixed games. We observed that
 7 slopes were negative in both Easy and Mixed games for almost all participants and more
 8 negative during Mixed games than Easy games
 9 ($t(21) = 3.92, p < 0.001, m = 15.76, d = 0.84$). Again, circles show the slopes for estimated
 10 lines of indifference for each subjects. The 0.95 percentile confidence intervals on these
 11 slopes are obtained using the same bootstrap procedure described in Experiment 1.⁶ Note
 12 that the mean difference in slopes is virtually identical to Experiment 1, although the effect
 13 size (Cohen's d) was larger during Experiment 1. Thus, Experiment 2a replicated the
 14 results of Experiment 1 showing that the findings were robust to different formulations of
 15 the evidence-foraging game. A model comparison exercise concurred with these results,
 16 showing that the majority of participants in Mixed ($N = 18$) as well as Easy games ($N =$
 17 17) were better accounted by a logistic regression model using both evidence and time as a
 18 predictor than by a simpler model that used only evidence as the predictor (see
 19 Appendix D for details).

20 **Experiment 2b.** In the next experiment, we decreased the inter-stimulus interval
 21 to 50ms and scaled the inter-trial intervals accordingly. All other parameters were the same
 22 as Experiment 2a. All participants passed the binomial test in the Easy and Mixed games.
 23 The mean reaction time during Easy, Difficult and Mixed games were 337ms (6.7 samples,
 24 $SEM = 4.2\text{ms}$) 418ms (8.4 samples, $SEM = 9.8\text{ms}$) and 419ms (8.4 samples,

⁶When the estimated slope for a participant is really steep, a large negative slope is qualitatively similar to a large positive slope. For this reason the confidence intervals for some participants with large slopes seem extremely wide. We compare the slopes using a circular difference (Equation 4) which corrects for this problem.

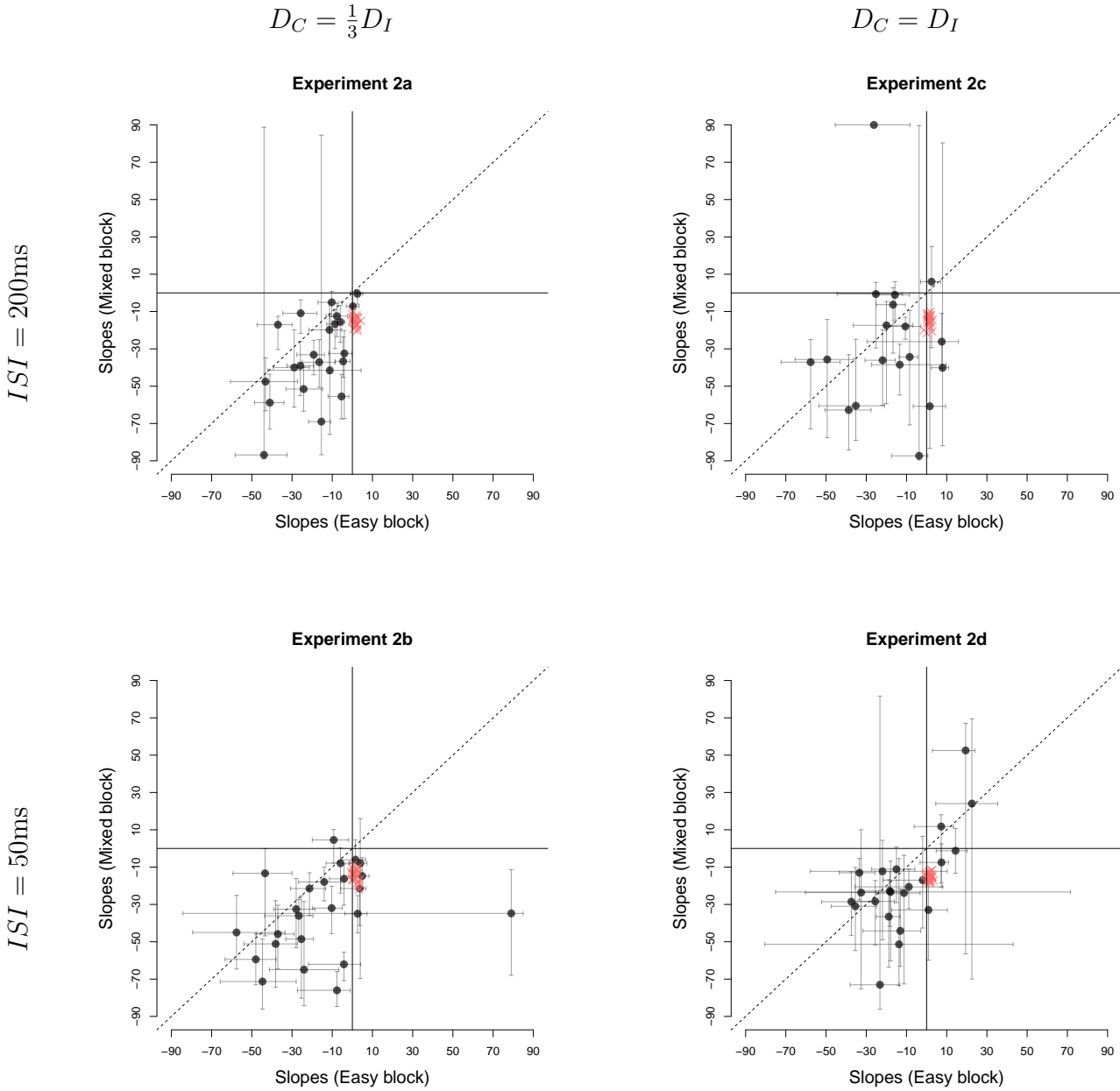


Figure 8. Slopes (in degrees) of estimated lines of indifference in Easy versus Mixed games in four experiments. The dashed line shows the curve for equal slope in Easy and Mixed games. Each circle (black) shows the estimated slopes for one participant. Error bars show 0.95 percentile bootstrapped confidence intervals. Crosses (red) show the estimated slopes for twenty-four simulated participants – decisions were simulated using a rise-to-threshold model with boundaries given by the optimal policy computed as described by Malhotra et al., under review.

$SEM = 5.3\text{ms}$), respectively, showing that this paradigm successfully elicited sub-second reaction times typically found in signal detection paradigms. We estimated the average number of stimuli that fell in the non-decision period to be 3.7, i.e., an average non-decision delay of approximately 183ms. We estimated $ND = 4$ for fifteen participants, $ND = 3$ for six participants, $ND = 5$ for two participants and $ND = 1$ for one participant. The bottom-left panel of Figure 8 shows the estimated slopes in Easy versus Mixed games. Like Experiment 2a, the slopes were negative for most participants in both Easy and Mixed games. Similarly, we also observed that the slopes were more negative in the Mixed games than in the Easy games, although the result was a little weaker than in Experiment 2a ($t(23) = 2.12$, $p = 0.044$, $m = 11.80$, $d = 0.47$). There are three possible reasons for this weaker result. First, the distribution for difference in slopes is more diffuse due the outlier at the right of the plot. Excluding this participant gave a clearer difference in slopes ($t(22) = 3.31$, $p = 0.003$, $m = 15.19$, $d = 0.72$) that was numerically highly similar to the slope difference observed in Experiments 1 and 2a. Second, for reasons discussed below, our estimates of non-decision time are likely to be less accurate in the “faster” paradigm. In turn, this error introduces variability in the accuracy of the actual evidence paths on a trial-by-trial basis that were used to derive our slope estimates. Lastly, it is possible that the process decreasing the boundary needs time to estimate the drift and adjust the boundary accordingly. With shorter ISI, this process may have less time to affect the decision process before the response is made, resulting in smaller difference in slopes between conditions.

Experiment 2c. Next, we changed the inter-stimulus interval back to 200ms (same as Experiment 2a) but increased D_C , the inter-trial interval for correct decisions, to the same value as D_I , the inter-trial interval for incorrect decisions (10s for both). Increasing the inter-trial interval decreased the reward per unit time and meant that participants had to wait longer between trials. Participants found this task difficult, we suspect because the inter-trial interval is so much longer than the typical reaction time. That is, participants

1 spend most of their time waiting for a new trial, but then those trials are over rather
 2 quickly. Perhaps as a result, the games lacked in engagement and six out of twenty-four
 3 participants failed the binomial test in Mixed games. For the eighteen remaining
 4 participants, the mean reaction times in Easy, Difficult and Mixed games were 807ms (4.0
 5 samples, $SEM = 27ms$), 858ms (4.3 samples, $SEM = 49ms$) and 954ms (4.8 samples,
 6 $SEM = 32ms$), respectively. We estimated the average number of stimuli that fell in the
 7 non-decision period to be 0.9, i.e., an average non-decision delay of approximately 176ms
 8 ($ND = 2$ for twenty-one participants and $ND = 0$ for the remaining three participants).
 9 The estimated slopes are shown in the top-right panel of Figure 8. We observed much
 10 greater variability in the estimated decision boundaries, though slopes were generally
 11 negative in Mixed as well as Easy games.⁷ The mean estimated slopes decreased more
 12 rapidly in Mixed games as compared to Easy games. However, given the large variability of
 13 responses and the number of participants that had to be excluded, this effect was
 14 comparatively weaker ($t(17) = 2.41$, $p = 0.028$, $m = 17.97$, $d = 0.60$). Nevertheless, the
 15 mean slope difference is very similar to that observed in all previous three experiments.

16 **Experiment 2d.** In this experiment we tested the final permutation of
 17 inter-stimulus and inter-trial intervals – we decreased the inter-stimulus interval to 50ms
 18 and matched the inter-trial intervals for correct and incorrect decisions (both 2.5s). Two
 19 participants failed the binomial test in Mixed games and were excluded from further
 20 analysis. The mean reaction times during Easy, Difficult and Mixed games were 240ms (4.8
 21 samples, $SEM = 4ms$), 308ms (6.2 samples, $SEM = 9ms$) and 294ms (5.9 samples,
 22 $SEM = 5ms$), respectively. We estimated the average number of stimuli that fell in the
 23 non-decision period to be 4.3, i.e., an average non-decision delay of approximately 216ms
 24 ($ND = 4$ for eighteen participants, $ND = 5$ for five participants and $ND = 7$ for one

⁷There is an outlier who seems to have a large *positive* slope in Mixed games. This may seem unintuitive,
 but, as noted above, there is little difference between a large positive and a large negative slope – in both
 cases the probability of *going* depends strongly on time (a temporal deadline) and weakly on evidence. The
 circular difference (Equation 4) accounts for cases like this.

participant). The bottom-right panel in Figure 8 compares the estimated slopes in Easy and Mixed games. The mean slope in either kind of game was negative ($t(21) = -3.10$, $p = 0.005$, $m = -11.66$ for Easy games and $t(21) = -3.42$, $p = 0.003$, $m = -18.82$ for Mixed games). However, in contrast to Experiment 2b, there was no significant difference in mean estimated slopes during Easy and Mixed games ($t(21) = 1.71$, $p = 0.10$, $m = 7.15$, $d = 0.32$).

Discussion

Experiments 2a-d revealed three key behavioural patterns: (i) participants generally decreased their decision boundaries with time, not only in the Mixed games, but also in the Easy games, (ii) this pattern held for the rapid task (Experiment 2b) but the variability of parameter estimates increased at faster reaction times, (iii) decreasing the difference between D_C and D_I decreased the difference in slopes between Easy and Mixed games.

Clearly, it is not optimal to decrease the decision boundary during fixed difficulty (Easy) games, but most participants seemed to do this. As noted in Experiment 1, a possible reason is that the reward-rate for sub-optimal policies is asymmetrical around the optimal boundary. Figure 9(a) shows the reward-rate landscape for all possible decision boundaries during Easy games in Experiment 2a, and maps the estimated boundaries for each participant onto this landscape.

Reward-rate is maximum at $(0, 3)$. When slope increases above zero the reward-rate drops rapidly. In contrast, when slope decreases below zero, reward-rate decreases gradually. This asymmetry means that participants pay a large penalty for a sub-optimal boundary with a positive slope, but a small penalty for a sub-optimal boundary with a negative slope. If participants are uncertain about the evidence gathered during a trial, or about the optimal policy, it is rational for them to decrease their decision boundary, as an error in estimation will lead to a relatively small penalty. Figure 9 suggests that most participants err on the side of caution and adopt policies with high (though not maximum)

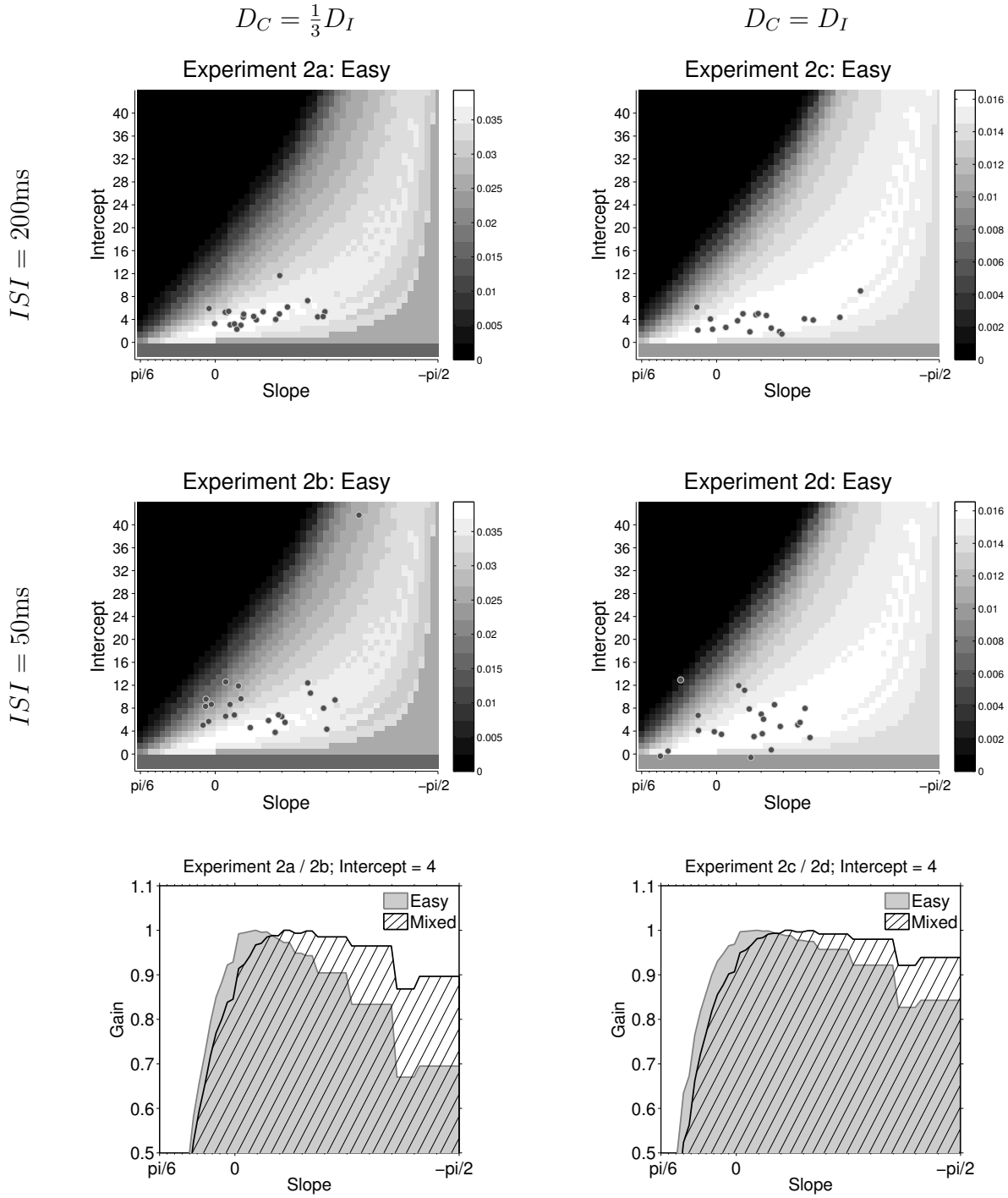


Figure 9. Reward-rate for Experiments 2a–2d. Each heat-map in the first two rows shows the “landscape” of reward-rate in policy space and dots show estimated policies adopted by participants in Easy games. The bottom row show profiles sliced through the (normalised) reward-rate landscape at a particular intercept. Shaded regions show profiles for Easy games, while hatched regions show profiles for Mixed games.

rewards and decreasing boundaries.

The shape of the reward landscape also sheds light on why participants behave differently when the inter-trial interval D_C is changed, even though changing this parameter does not affect the optimal policy. The first column in Figure 9 shows the reward-rate in experiments where $D_C = \frac{1}{3}D_I$, while the second column shows the reward-rate in experiments where $D_C = D_I$. The top two rows show the reward-rate landscapes in Easy games at all combinations of slopes and intercepts, while the bottom row compares the reward-rate in Easy and Mixed games at a particular intercept of decision boundary but different values of slope (i.e. a horizontal slice through the heat-maps above). Even though the optimal policy in all four experiments is the same, there are several ways in which the reward-rate landscape in the left-hand column (Experiments 2a and 2b) differ from the landscape in the right-hand column (Experiments 2c and 2d).

Firstly, the reward-rate landscape in Easy games is more sharply peaked when $D_C = \frac{1}{3}D_I$ (Experiments 2a and 2b). This is most clearly discernible in panels in the bottom row which shows the profile of the (normalised) reward-rate landscape at a particular intercept. If the participant adopts a boundary with large negative slope, the difference between the reward-rate for such a policy and the optimal reward-rate is larger when $D_C = \frac{1}{3}D_I$ (left panel) than when $D_C = D_I$ (right panel). So in Experiments 2a and 2b adopting a sub-optimal policy carries a larger ‘regret’ than in Experiments 2c and 2d. This means that the reward landscape constrains the choice of boundaries more in Experiments 2a and 2b than it does in Experiments 2c and 2d, even though the optimal policy for all experiments is the same.

The panels in the bottom row also compare the reward-rate profiles during Easy (shaded) and Mixed (hatched) games at a particular intercept. It can be seen that for both types of experiments the normalised reward-rate is larger in Mixed games than Easy games when slopes are more negative. Thus it is better (more rewarding) to have decreasing boundaries in Mixed games than in Easy games. However, the difference in Easy and

Mixed games is larger when $D_C = \frac{1}{3}D_I$ (left panel) than when $D_C = D_I$ (right panel). Correspondingly, we found a more robust difference in slopes during Experiments 2a and 2b than we did in Experiments 2c and 2d.

The third behavioural pattern was an increase in variability of slopes when the decisions were made more rapidly. There are two possible sources of this variability: internal noise and error in estimation of the non-decision time. Recall that we excluded stimuli that arrive during the non-decision time based on a single estimate of this time for each participant. It is likely that the non-decision time varies from trial-to-trial; indeed, this is a common assumption in models of decision-making (see Ratcliff & Smith, 2004). Any such variability means that on some trials we are including irrelevant samples (estimating a non-decision time too short) or excluding relevant samples (estimating a non-decision time too long). As a result, there is a discrepancy between the evidence paths that actually led to the participant’s decision and the one entered into the logistic regression model used to estimate the decision boundary. Importantly, this discrepancy will be much smaller in the experiments with a long inter-stimulus interval, because even an error in non-decision time of, say, 100 ms will at most introduce only one additional or excluded evidence sample. However, in the experiments with a much shorter inter-stimulus intervals, the same numerical error will result in several additional or missed evidence samples. Therefore, trial-to-trial variability in the non-decision times introduces more noise in the slope estimates for the faster experiments.

Experiment 3

In the above experiments, the optimal policy was to decrease decision boundaries in Mixed games but keep them constant in Single-difficulty games. Correspondingly, data suggested that participants adopted more strongly decreasing boundaries in Mixed-difficulty games than in Single-difficulty games, particularly when errors are costly (in terms of reward-rate). In Experiment 3 we changed the parameters so that the optimal

policy during Mixed-difficulty games was, in fact, to increase the decision boundary. Recall from the theory on optimal shapes of decision boundaries that the optimal policy in Mixed games is to decrease decision boundaries only when one of the decision types is extremely difficult. In contrast, when both types of decisions are easy or moderately difficult, the policy that optimises reward-rate is to increase decision boundaries or leave them constant (Figure 2c). Therefore, if participants were optimising their average reward, we expected estimated slopes in Mixed-difficulty games of this type to be either the same or larger than slopes in Single-difficulty games.

Experiment 3 used the same experimental paradigm as Experiment 2. The parameters for Experiment 3 are shown in Table 3. During this experiment, Easy games showed cues with *up-probability* $\frac{1}{2} \pm 0.40$ – so participants could make really rapid decisions in these games. And unlike Experiments 1 and 2, Difficult games showed cues with *up-probability* $\frac{1}{2} \pm 0.10$. The optimal boundaries in this case are higher for Difficult games than Easy games (Figures 2a and 2b) and the optimal boundary for Mixed games show a slight increase in evidence with time (Figure 2c).

Parameter name	Value
Drift for Easy condition (ϵ_e)	0.40
Drift for Difficult condition (ϵ_d)	0.10
Reward	2p
Inter-trial interval, correct	3.5 sec
Inter-trial interval, incorrect	3.5 sec
Inter-stimuli interval	50 msec

Table 3

Values of parameters used during Experiment 3.

Twenty four participants played blocks of Easy, Difficult and Mixed games with the objective of maximising their reward. Each correct decision was worth 2p and there was no

1 reward or penalty for incorrect decisions. The participant who collected the largest number
 2 of money bags received a bonus reward of £20. The inter-stimulus interval was 50 msec
 3 and inter-trial interval was 3.5 sec.

4 We used the same procedure as the above experiments to analyse the data. All
 5 participants passed the binomial test in Mixed games so no data was rejected. Figure 10
 6 shows the estimated slopes for lines of indifference in Easy, Difficult and Mixed games.
 7 Unlike the previous experiments, we compared the slopes in Mixed games not only to Easy
 8 games, but also to Difficult games, since the Difficult games in this case required
 9 participants to accumulate evidence before making a decision.

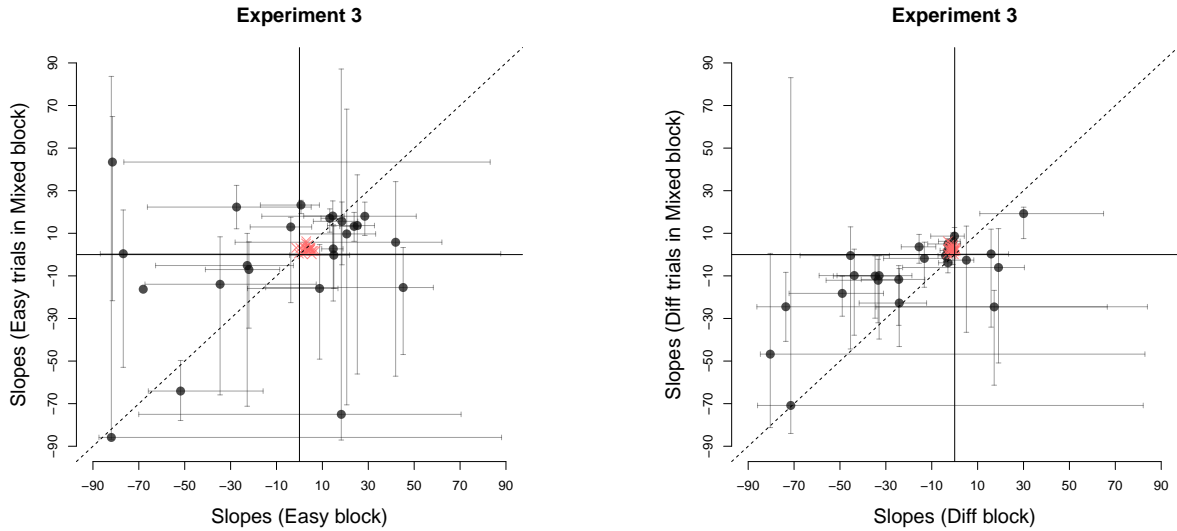


Figure 10. Slopes of estimated lines of indifference in Experiment 3. The panel on the left compares slopes for the Easy trial during Mixed games with slopes for all trials during Easy games while the panel on the right compares slopes for the Difficult trials in Mixed games with slopes for all trials during Difficult games. In each panel, the solid vertical and horizontal lines show lines of zero slope (flat threshold) and the dashed line shows the curve for equal slopes in the two types of games. Each circle shows the estimated slopes for one participant and crosses show estimated slopes from simulated optimal participants.

During Easy games, participants made really rapid (and accurate) decisions, with mean reaction times 107ms ($SEM = 4ms$), that is, based on two to three sample cues (after excluding non-decision time). Such fast responses are of course consistent with the model prediction of narrow decision boundaries in this condition. As discussed above, on this rapid time scale noise in the responses due to non-decision time or due to variability of the perceptual system has a large impact on the variability of estimated slopes. Indeed, we can see from Figure 10 (left panel), that the confidence intervals around estimated slopes are large and there was substantial (between-participant) variability in the mean estimated slopes.

A more accurate comparison between Single and Mixed-difficulty games is obtained by comparing the slopes in Difficult games with the slopes on Difficult trials in Mixed games. The panel on the right in Figure 10 shows this comparison. The mean reaction times in Difficult games was 397ms (8 samples, $SEM = 9ms$), while that in the Mixed games was 234ms (4.7 samples, $SEM = 6ms$). We estimated the non-decision time to be approximately 4.5 samples, i.e., 227ms ($ND = 5$ for thirteen participants and $ND = 4$ for the remaining eleven participants). Like previous experiments, the mean slope in Single-difficulty (here, Difficult) games was less than zero ($t(23) = -3.15, p < 0.001, m = -19.5$).

Crucially, in contrast to Experiment 1 and 2, but in agreement with the reward-rate optimising policy, we found that the estimated slopes in Mixed games were slightly *larger* (less negative) than in the Difficult games ($t(23) = -2.25, p = 0.034, m = -9.86, d = -0.39$). Indeed, model comparison suggested that, again in contrast to Experiment 1 and 2, the simpler logistic regression model using only evidence as the predictor provided a better account of data in Mixed games than the model using both evidence and time as predictors (see Appendix D). However, for the Difficult games, the evidence was more mixed in that for just over half the participants, a model that included time as a predictor performed better. These results are consistent with the slope

1 comparisons in that boundaries varied with time (slightly) in Difficult games, but were
 2 approximately constant in the Mixed games. Although we do not actually observe
 3 *increasing* boundaries, the shift from decreasing to approximately constant boundaries is a
 4 shift in the right direction.

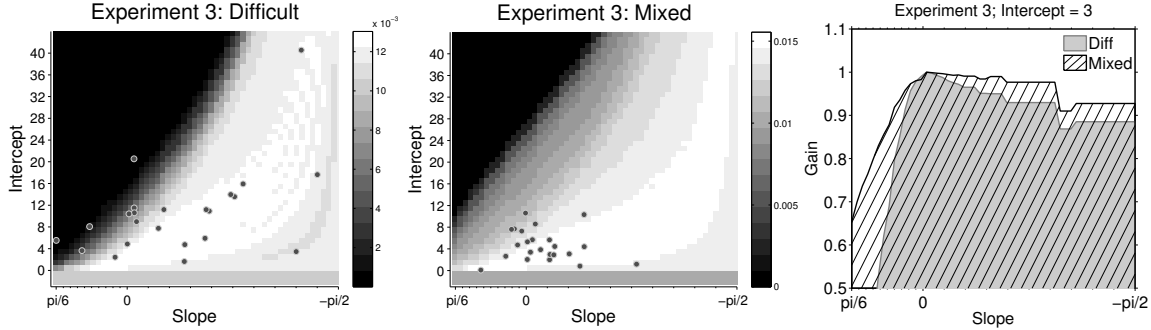


Figure 11. Reward-rate during Difficult and Mixed games during Experiment 3. Each heat-map shows the “landscape” of reward per unit time in policy space. Lighter colours show higher reward. Each circle shows the slope and intercept of the estimated line of indifference for a participant. The right-most panel compares the reward landscape in Difficult and Mixed games at a particular intercept.

5 In Figure 11, we have again plotted the estimated policies of all participants on the
 6 reward landscape. The key difference in behaviour between Mixed and Difficult games was
 7 that most participants were concentrated around the zero slope during Mixed games while
 8 the slope of boundaries chosen by participants in the Difficult games were spread over a
 9 large range with a number of participants choosing policies with large negative slopes. In
 10 the right-most panel, we have compared the profile of the landscape, slicing it at intercept
 11 $= 3$ (optimal policy in Mixed games had slope slightly above 0 and intercept between 3 and
 12 4). This profile shows that, like Experiment 2, reward-rate is an asymmetric function of
 13 slope in both the Difficult and Mixed games. However, in contrast to Experiment 2, the
 14 amount of asymmetry is now lesser during Mixed games than Difficult games. So during
 15 Mixed games, participants can choose policies in the neighbourhood of constant boundary
 16 with a lower regret, even in the presence of uncertainty about the evidence or the optimal

boundary. This could explain why the majority of participants in the Mixed games are concentrated around policies with zero slope. In contrast, the larger asymmetry during Difficult games seems to push a number of participants into adopting boundaries with large negative slopes – a lower risk strategy that nevertheless leads to a small loss in average reward.

General Discussion

Constant or Decreasing boundaries

Sequential sampling models have had a very successful history of fitting data in a variety of decision-making experiments (Ratcliff, 1978; Smith & Vickers, 1989; Ratcliff & Smith, 2004; Bogacz et al., 2006; Ratcliff et al., 2016). These models typically assume that decision boundaries remain constant during a trial, so introducing the possibility of changing boundaries adds further complexity to these models. The question is whether this complexity is warranted given existing data.

Recently, Hawkins et al. (2015) and Voskuilen, Ratcliff, and Smith (2016) conducted a model comparison based on data from a number of decision-making studies and found that introducing decreasing bounds did not generally improve the model fit. In this study, we took a different approach – instead of working out whether decreasing boundaries improves model fit, we used a mathematical model (Malhotra et al., under review) to establish the circumstances for changing decision boundary if the decision-maker wanted to maximise reward-rate. The key insight from this approach is that optimal decision boundaries decrease only in very specific scenarios – when one of the difficulty in a Mixed-difficulty task is extremely difficult, or even impossible. In other conditions, optimal boundaries for Mixed-difficulty tasks may increase or stay constant based on the difficulty of constituent decisions. An advantage of the model presented in this study is that it can be used for inferring the reward-rate of any given boundary, which can then be used to compare with the optimal boundary. Using this approach, we found that sub-optimal

1 policies were “asymmetrically distributed” near the optimal boundary in policy space. A
2 judicious decision-maker should consider this asymmetry in reward landscape to make
3 decisions that are robust to uncertainty in task parameters and to their own estimate of
4 the optimal policy. Six expanded-judgement experiments indicate that people may not only
5 be modulating how decision boundaries change with time, but may also be using such
6 robustness considerations to choose the value and shape of these boundaries.

7 So why do Hawkins et al. (2015) and Voskuilen et al. (2016) find no strong evidence
8 for changing decision boundaries and, indeed, why are models with constant decision
9 thresholds so successful at fitting data from a variety of experiments? There could be three
10 possible reasons. Firstly, the datasets analysed by Hawkins et al. (2015) and Voskuilen et
11 al. (2016) consist of Mixed-difficulty experiments with a variety of different difficulty levels.
12 For example, Experiment 1 conducted by Hawkins et al. (2015) was a
13 motion-discrimination task with six different difficulty levels (0%, 2.5%, 5%, 10%, 20% and
14 40%) while Experiment 1 from Ratcliff, Hasegawa, Hasegawa, Smith, and Segraves (2007)
15 was a brightness-discrimination task with three levels of difficulty (55%, 65% and 98%). It
16 is not clear in any of these experiments what the shape of boundaries that optimise
17 reward-rate should be. As we have discussed above, optimal boundaries do not necessarily
18 decrease in Mixed-difficulty trials and when they do decrease, the rate of decrease varies
19 over a broad range based on the levels of difficulty. So even if participants were optimising
20 reward-rate in the experiments considered by Hawkins et al. (2015) and Voskuilen et al.
21 (2016), this may not necessarily entail observing decreasing boundaries.

22 Secondly, each of our experiments carefully controls the cost / reward of each decision
23 and links performance to reward. This allows us to compute the optimal behaviour in the
24 task (in terms of reward-rate) and compare participant performance with this optimal
25 behaviour. In contrast, most studies considered by Hawkins et al. do not have a
26 performance-based reward structure. Participants are asked to emphasise speed, accuracy
27 or both and there is no explicit scale on which a participant can measure the expected

1 return of a policy. Exceptions to these are studies involving non-human primates, such as
2 Roitman and Shadlen (2002), Ratcliff, Cherian, and Segraves (2003) and Ditterich (2006),
3 where performance was explicitly linked to reward and interestingly, Hawkins et al. find
4 evidence for decreasing boundaries in these studies.

5 Decisions in ecologically realistic situations are typically accompanied by costs and
6 rewards and the structure of incentives can profoundly affect performance, as shown by a
7 series of studies in experimental economics (Camerer & Hogarth, 1999; Cubitt, Starmer, &
8 Sugden, 1998). Therefore, if we want to establish whether participants decrease decision
9 boundaries within a trial, we must determine what it is they stand to gain by changing
10 their decision boundaries during the experiment.

11 Lastly, note that the expanded-judgement paradigm used by us is different from the
12 signal detection paradigms used in studies analysed by Hawkins et al. (2015) and Voskuilen
13 et al. (2016). This is a key strength of our study as we are able to observe the exact
14 sequence of stimuli observed by the decision-maker and infer their decision boundaries
15 based on these observations. It has been demonstrated recently that constraining sequential
16 sampling models by the exact sequence of stimuli provides a closer description of reaction
17 times than that obtained from models in which the drift parameter is assumed constant
18 within a trial (Park, Lueckmann, Kriegstein, Bitzer, & Kiebel, 2016). However, using this
19 paradigm leaves open the possibility that the decision boundary is set differently when the
20 decision processes draw samples from an internal representation (e.g. in colour / brightness
21 / numerosity judgement tasks) and when samples drawn cannot be recorded by the
22 experimenter. Previous evidence suggests that results from expanded-judgement tasks can
23 be generalised to situations where sampling is internal (Vickers, Burt, et al., 1985; Vickers,
24 Smith, Burt, & Brown, 1985). However, these studies did not examine a signal-detection
25 task where reaction times are typically $< 500ms$. Thus, an important outstanding question
26 is whether people use different decision processes for internally and externally sampled
27 observations and whether this affects how they set their decision boundaries.

1 Data from several expanded-judgement tasks involving choice between multiple
2 alternatives have been successfully analysed using sequential sampling models with fixed
3 boundaries, which have been shown to capture the key interesting aspects of these data
4 (Brown, Steyvers, & Wagenmakers, 2009; Hawkins, Brown, Steyvers, & Wagenmakers,
5 2012, 2012). It would be interesting to extend the model fitting methodology presented in
6 this paper to the case of choice between multiple alternatives, and investigate if these data
7 are better described by a model with flat or decreasing boundaries.

8 **Individual differences**

9 In all of the above experiments, we observed variability in behaviour both between
10 individuals and between trials within a participant. We have already discussed two reasons
11 for the variability between trials: (i) non-decision time, which is estimated per individual
12 but may vary from trial-to-trial and (ii) internal noise, which could lead to a trial-to-trial
13 variability in drift rate. As mentioned above, a trial-to-trial variability in drift rate, starting
14 point or threshold has been shown to be essential for fitting reaction time distributions – in
15 particular, different patterns of error reaction times – using sequential sampling models (see
16 Ratcliff, 1978; Ratcliff & Smith, 2004). In addition to these, our study highlights another
17 source of variability between individuals – the shape of the reward landscape with its broad
18 region in which “acceptably” high reward-rates could be achieved. Reward rate was
19 asymmetrically distributed around the optimal policy in all the above experiments, with a
20 bias towards sub-optimal policies that yielded a reward-rate that was close to maximum.

21 A number of previous studies have compared individuals to optimal behaviour in
22 decision-making tasks and found that participants often use boundaries that are
23 sub-optimal (Pitz et al., 1969; Sanders & Linden, 1967; Wallsten, 1968; Simen et al., 2009;
24 Starns & Ratcliff, 2010; Bogacz, Hu, et al., 2010; Zacksenhouse, Bogacz, & Holmes, 2010).
25 It has also been observed that participants have a tendency to overvalue accuracy, setting
26 boundaries that are wider than those suggested by maximisation of reward-rate (Maddox

1 & Bohil, 1998; Bohil & Maddox, 2003; Myung & Busemeyer, 1989; Balci et al., 2011;
2 Starns & Ratcliff, 2012). To explain this behaviour, a set of studies have investigated
3 alternative objective functions (Bohil & Maddox, 2003; Bogacz et al., 2006; Zacksenhouse
4 et al., 2010). For example, Zacksenhouse et al. (2010) found that only about 30% of
5 participants achieve (reward-rate) optimality and the behaviour of the other 70% is better
6 explained by a robust strategy that maximises performance under presumed level of
7 uncertainty (the maximin strategy).

8 The behaviour of participants in our experiments is in line with such a robust
9 strategy: a small proportion of participants adopt policies that are close to optimal
10 (Figures 5, 8, and 10) but most participants adopt strategies that yield high, but not
11 maximum, reward-rate (Figures 6, 9 and 11). Since the gradient of reward-rate was larger
12 above constant boundary than below, this meant choosing a policy with a decreasing
13 boundary.

14 In the above experiments there can be several sources of uncertainty, leading to
15 adoption of a robust strategy: uncertainty in estimation of task parameters such as
16 ISI/ITI, uncertainty in the signal due to noise in the sensory system and uncertainty in the
17 estimate of reward-rate for the task. If participants use a hill-climbing learning mechanism
18 (Myung & Busemeyer, 1989; Simen, Cohen, & Holmes, 2006), these uncertainties will
19 introduce noise in the learning process and make it harder for participants to search for the
20 optimal policy, especially when the reward landscape has a low gradient, leading to the
21 observed differences in the choice of boundaries. With training, participants should be able
22 to reduce these uncertainties and approach optimal boundaries, as shown by previous
23 research (Myung & Busemeyer, 1989; Balci et al., 2011).

24 Conversely, when internal noise in the sensory system increases or when the estimate
25 of the task parameters becomes more uncertain, participants should find it more difficult to
26 locate the optimal policy in policy space. For example, it has been shown that the duration
27 estimates of older adults are more variable than younger adults (Block, Zakay, & Hancock,

1998) and visual perception declines with aging (Weale, 1963; Spear, 1993; Habak & Faubert, 2000; Owsley, 2011). These processes will increase the level of uncertainty in the (temporal) task parameters as well as the visual stimuli and could explain why older adults adopt boundaries that are farther from optimal (Starns & Ratcliff, 2010, 2012). Of course, it is also possible that the deviation from optimality is a consequence of not only an increase in visual and temporal noise but also a decline in the ability to flexibly set the boundary and more empirical studies would be required to tease apart the relative contribution of these two factors.

9 Mechanistic considerations

The behaviour of participants in the experiments above suggests that they adapt their decision-making mechanism to achieve near-maximal reward-rates. We do not claim participants are optimal – they are clearly not. Nor do we claim that the mathematical model we have used to derive the optimal policy is a *psychological* theory. The focus of our study was not on establishing the mechanism by which people achieve this behaviour but on comparing the normative behaviour with the empirical behaviour. In a manner similar to “ideal observer models” in the study of sensory systems (Geisler, 2003), specifying the optimal policy has helped us (i) identify experimental conditions that are best suited to empirically test time-varying decision boundaries, and (ii) identify sources for sub-optimal behaviour (or inefficiencies) through analysis of the reward landscape. Nevertheless, we finish with some considerations of the underlying mechanisms that may be responsible for the time varying boundaries observed in our study.

First of all, the reader may wonder whether the decreasing bounds we identify in our experiments may be accounted for by existing mechanisms in models that assume constant boundaries. In Appendix C we explore two such mechanisms – between-trial noise in the drift rate and imperfect integration of information. We simulated decisions using a rise-to-threshold model both with and without between-trial noise in drift rate and with

1 and without loss in integration of evidence. We then estimated the slopes of boundaries
2 using the method discussed above and found that the estimated difference in slopes
3 between Single and Mixed-difficulty conditions reflected the true difference, irrespective of
4 the noise in drift rate or loss in integration of evidence. Thus, our inferences about
5 difference in slopes remain valid even when these mechanisms are considered.

6 Next, the pattern of decision-making we observed in the expanded-judgement tasks is
7 compatible with a number of different mechanistic accounts. For example, it is possible
8 that participants did not weigh each cue equally and cues later in the decision carried a
9 larger weight. This mechanism has been recently suggested by the urgency-gating model
10 (Cisek, Puskas, & El-Murr, 2009; Thura, Beauregard-Racine, Fradet, & Cisek, 2012).
11 Similarly, it is also possible that participants maintained a constant threshold but also
12 employed a stochastic deadline. That is, they maintain an internal clock and make a
13 decision if evidence crosses a constant threshold before a deadline or choose the most-likely
14 alternative if the threshold is not crossed but a deadline is reached. This mechanism is
15 similar to the response signal paradigm (e.g., Ratcliff, 2006), with an internal instead of an
16 external deadline. Both these mechanisms will lead to decision boundaries that appear to
17 decrease with time. However, the urgency gating model does not assume integration of
18 sensory input over whole duration of trial, but rather rapid forgetting of previously
19 integrated input. It would be interesting to formally compare in a future study whether the
20 urgency gating model or an integration to boundary model better describes data from the
21 current study, which is freely available, as mentioned above.

22 However, note that the normative model does not always predict that decision
23 boundaries should decrease with time. In agreement with this, we found that many
24 participants in Experiment 3 did not appear to decrease their decision boundaries in
25 Mixed-difficulty condition (also see Figure D1). These findings are not straightforward to
26 reconcile in mechanistic accounts such as urgency-gating and stochastic deadline and
27 provide a good test for teasing apart these models.

1 The logistic regression model used to infer the boundary from data (Equations 2
2 and 3) assumes that people integrate evidence to a constant boundary but that the slope of
3 the boundary is allowed to vary. Under this assumption, participants appear to decrease
4 their decision boundaries when decreasing boundaries increases reward-rate. So the thrust
5 of our argument is that people seem sensitive to the normative behaviour and when the
6 normative behaviour changes (Single vs Mixed-difficulty conditions) participants seem to
7 adapt their decision mechanism in line with the normative standard. A separate and
8 important question is *how* people make this adaptation. Decreasing the decision boundary,
9 increasing the gain of observations or maintaining a stochastic deadline are all possible
10 mechanisms to achieve this goal and future research should examine what mechanisms are
11 employed by people.

References

- Audley, R., & Pike, A. (1965). Some alternative stochastic models of choice. *British Journal of Mathematical and Statistical Psychology*, 18(2), 207–225.
- Balci, F., Simen, P., Niyogi, R., Saxe, A., Hughes, J. A., Holmes, P., et al. (2011). Acquisition of decision making criteria: reward rate ultimately beats accuracy. *Attention, Perception, & Psychophysics*, 73(2), 640–657.
- Becker, G. M. (1958). Sequential decision making: Wald’s model and estimates of parameters. *Journal of Experimental Psychology*, 55(6), 628–636.
- Block, R. A., Zakay, D., & Hancock, P. A. (1998). Human aging and duration judgments: A meta-analytic review. *Psychology and Aging*, 13(4), 584.
- Bogacz, R., Brown, E., Moehlis, J., Holmes, P., & Cohen, J. D. (2006). The physics of optimal decision making: A formal analysis of models of performance in two-alternative forced choice tasks. *Psychological Review*, 113(4), 700–765.
- Bogacz, R., Hu, P. T., Holmes, P. J., & Cohen, J. D. (2010). Do humans produce the speed–accuracy trade-off that maximizes reward rate? *The Quarterly Journal of Experimental Psychology*, 63(5), 863–891.
- Bogacz, R., Wagenmakers, E.-J., Forstmann, B. U., & Nieuwenhuis, S. (2010). The neural basis of the speed–accuracy tradeoff. *Trends in neurosciences*, 33(1), 10–16.
- Bohil, C. J., & Maddox, W. T. (2003). On the generality of optimal versus objective classifier feedback effects on decision criterion learning in perceptual categorization. *Memory & Cognition*, 31(2), 181–198.
- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial vision*, 10, 433–436.
- Britten, K. H., Shadlen, M. N., Newsome, W. T., & Movshon, J. A. (1992). The analysis of visual motion: a comparison of neuronal and psychophysical performance. *The Journal of Neuroscience*, 12(12), 4745–4765.
- Brown, S., Steyvers, M., & Wagenmakers, E.-J. (2009). Observing evidence accumulation during multi-alternative decisions. *Journal of Mathematical Psychology*, 53(6),

1 453–462.

2 Busemeyer, J. R. (1985). Decision making under uncertainty: a comparison of simple
3 scalability, fixed-sample, and sequential-sampling models. *Journal of Experimental*
4 *Psychology: Learning, Memory, and Cognition*, 11(3), 538–564.

5 Camerer, C. F., & Hogarth, R. M. (1999). The effects of financial incentives in experiments:
6 A review and capital-labor-production framework. *Journal of risk and uncertainty*,
7 19(1), 7–42.

8 Cartwright, D., & Festinger, L. (1943). A quantitative theory of decision. *Psychological*
9 *Review*, 50(6), 595–621.

10 Cisek, P., Puskas, G. A., & El-Murr, S. (2009). Decisions in changing conditions: the
11 urgency-gating model. *The Journal of Neuroscience*, 29(37), 11560–11571.

12 Cubitt, R. P., Starmer, C., & Sugden, R. (1998). On the validity of the random lottery
13 incentive system. *Experimental Economics*, 1(2), 115–131.

14 Ditterich, J. (2006). Stochastic models of decisions about motion direction: behavior and
15 physiology. *Neural Networks*, 19(8), 981–1012.

16 Drugowitsch, J., Moreno-Bote, R., Churchland, A. K., Shadlen, M. N., & Pouget, A.
17 (2012). The cost of accumulating evidence in perceptual decision making. *Journal of*
18 *Neuroscience*, 32, 3612–3628.

19 Edwards, W. (1965). Optimal strategies for seeking information: Models for statistics,
20 choice reaction times, and human information processing. *Journal of Mathematical*
21 *Psychology*, 2(2), 312–329.

22 Evans, N. J., & Brown, S. D. (2016). People adopt optimal policies in simple
23 decision-making, after practice and guidance. *Psychonomic Bulletin & Review*, 1–10.

24 Festinger, L. (1943). Studies in decision. ii. an empirical test of a quantitative theory of
25 decision. *Journal of Experimental Psychology*, 32(5), 411.

26 Garrett, H. E. (1922). A study of the relation of accuracy to speed. *Archives of*
27 *Psychology*(56).

- 1 Geisler, W. S. (2003). Ideal observer analysis. *The visual neurosciences*, 10(7), 12–12.
- 2 Georgeson, M. A. (1987). Temporal properties of spatial contrast vision. *Vision research*,
3 27(5), 765–780.
- 4 Gonzalez, R., & Wu, G. (1999). On the shape of the probability weighting function.
5 *Cognitive psychology*, 38(1), 129–166.
- 6 Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York:
7 John Wiley and Sons.
- 8 Habak, C., & Faubert, J. (2000). Larger effect of aging on the perception of higher-order
9 stimuli. *Vision research*, 40(8), 943–950.
- 10 Hawkins, G., Brown, S. D., Steyvers, M., & Wagenmakers, E.-J. (2012). Decision speed
11 induces context effects in choice. *Experimental psychology*.
- 12 Hawkins, G. E., Brown, S. D., Steyvers, M., & Wagenmakers, E.-J. (2012). An optimal
13 adjustment procedure to minimize experiment time in decisions with multiple
14 alternatives. *Psychonomic bulletin & review*, 19(2), 339–348.
- 15 Hawkins, G. E., Forstmann, B. U., Wagenmakers, E.-J., Ratcliff, R., & Brown, S. D.
16 (2015). Revisiting the evidence for collapsing boundaries and urgency signals in
17 perceptual decision-making. *The Journal of Neuroscience*, 35(6), 2476–2484.
- 18 Howard, R. A. (1960). *Dynamic programming and markov processes*. New York, NY: John
19 Wiley & Sons.
- 20 Howell, W. C., & Kreidler, D. L. (1963). Information processing under contradictory
21 instructional sets. *Journal of Experimental Psychology*, 65(1), 39–46.
- 22 Huang, Y., & Rao, R. P. (2013). Reward optimization in the primate brain: a probabilistic
23 model of decision making under uncertainty. *PloS one*, 8(1), e53344.
- 24 Irwin, F. W., & Smith, W. A. S. (1956). Further tests of theories of decision in an
25 “expanded judgment” situation. *Journal of Experimental Psychology*, 52(6), 345.
- 26 Irwin, F. W., Smith, W. A. S., & Mayfield, J. F. (1956). Tests of theories of decision in an
27 “expanded judgement” situation. *Journal of Experimental Psychology*, 51(4),

1 261–268.

2 Johnson, D. M. (1939). Confidence and speed in the two-category judgement. *Archives of*
3 *Psychology*(241), 1–52.

4 Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007).
5 What's new in psychtoolbox-3. *Perception*, 36(14), 1–1.

6 LaBerge, D. (1962). A recruitment theory of simple behavior. *Psychometrika*, 27(4),
7 375–396.

8 Laming, D. R. J. (1968). *Information theory of choice-reaction times*. London: Academic
9 Press.

10 Lamport, L. (2012). Buridan's principle. *Foundations of Physics*, 42(8), 1056–1066.

11 Link, S., & Heath, R. (1975). A sequential theory of psychological discrimination.
12 *Psychometrika*, 40(1), 77–105.

13 Luce, R. D. (1986). *Response times* (No. 8). Oxford University Press.

14 Ludwig, C. J. (2009). Temporal integration of sensory evidence for saccade target selection.
15 *Vision research*, 49(23), 2764–2773.

16 Ludwig, C. J., Gilchrist, I. D., McSorley, E., & Baddeley, R. J. (2005). The temporal
17 impulse response underlying saccadic decisions. *The Journal of neuroscience*, 25(43),
18 9907–9912.

19 Maddox, W. T., & Bohil, C. J. (1998). Base-rate and payoff effects in multidimensional
20 perceptual categorization. *Journal of Experimental Psychology: Learning, Memory,*
21 *and Cognition*, 24(6), 1459.

22 Malhotra, G., Leslie, D., Ludwig, C., & Bogacz, R. (under review). A simple optimality
23 model for time-varying decision boundaries. *Psychonomic Bulletin and Review*.

24 Manz, W. (1970). Experiments on probabilistic information processing. *Acta Psychologica*,
25 34, 184–200.

26 Moran, R. (2015). Optimal decision making in heterogeneous and biased environments.
27 *Psychonomic bulletin & review*, 22(1), 38–53.

- 1 Myung, I. J., & Busemeyer, J. R. (1989). Criterion learning in a deferred decision-making
2 task. *The American journal of psychology*, 1–16.
- 3 Owsley, C. (2011). Aging and vision. *Vision research*, 51(13), 1610–1622.
- 4 Pachella, R. G. (1974). The interpretation of reaction time in information processing
5 research. In B. Kantowitz (Ed.), *Human information processing: Tutorials in*
6 *performance and cognition* (pp. 41–82). Erlbaum.
- 7 Palmer, J., Huk, A. C., & Shadlen, M. N. (2005). The effect of stimulus strength on the
8 speed and accuracy of a perceptual decision. *Journal of vision*, 5(5).
- 9 Park, H., Lueckmann, J.-M., Kriegstein, K. von, Bitzer, S., & Kiebel, S. J. (2016).
10 Spatiotemporal dynamics of random stimuli account for trial-to-trial variability in
11 perceptual decision making. *Scientific reports*, 6.
- 12 Pelli, D. G. (1997). The videotoolbox software for visual psychophysics: Transforming
13 numbers into movies. *Spatial vision*, 10(4), 437–442.
- 14 Pike, A. (1968). Latency and relative frequency of response in psychophysical
15 discrimination. *British Journal of Mathematical and Statistical Psychology*, 21(2),
16 161–182.
- 17 Pitz, G. F., Reinhold, H., & Geller, E. S. (1969). Strategies of information seeking in
18 deferred decision making. *Organizational Behavior and Human Performance*, 4(1),
19 1–19.
- 20 Pleskac, T. J., & Busemeyer, J. R. (2010). Two-stage dynamic signal detection: a theory of
21 choice, decision time, and confidence. *Psychological review*, 117(3), 864–901.
- 22 Puterman, M. L. (2005). *Markov decision processes: Discrete stochastic dynamic*
23 *programming*. New Jersey: John Wiley & Sons.
- 24 Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, 83, 59–108.
- 25 Ratcliff, R. (2006). Modeling response signal and response time data. *Cognitive psychology*,
26 53(3), 195–237.
- 27 Ratcliff, R., Cherian, A., & Segraves, M. (2003). A comparison of macaque behavior and

- superior colliculus neuronal activity to predictions from models of two-choice decisions. *Journal of neurophysiology*, 90(3), 1392–1407.
- Ratcliff, R., Hasegawa, Y. T., Hasegawa, R. P., Smith, P. L., & Segraves, M. A. (2007). Dual diffusion model for single-cell recording data from the superior colliculus in a brightness-discrimination task. *Journal of neurophysiology*, 97(2), 1756–1774.
- Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: theory and data for two-choice decision tasks. *Neural computation*, 20(4), 873–922.
- Ratcliff, R., & Rouder, J. N. (1998). Modeling response times for two-choice decisions. *Psychological Science*, 9(5), 347–356.
- Ratcliff, R., & Smith, P. L. (2004). A comparison of sequential sampling models for two-choice reaction time. *Psychological review*, 111(2), 333.
- Ratcliff, R., Smith, P. L., Brown, S. D., & McKoon, G. (2016). Diffusion decision model: Current issues and history. *Trends in cognitive sciences*, 20(4), 260–281.
- Roitman, J. D., & Shadlen, M. N. (2002). Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. *Journal of neuroscience*, 22(21), 9475–9489.
- Ross, S. (1983). *Introduction to stochastic dynamic programming*. Academic Press, New York.
- Sanders, A., & Linden, W. T. (1967). Decision making during paced arrival of probabilistic information. *Acta Psychologica*, 27, 170–177.
- Schwarz, G. (1978). Estimating the dimension of a model. *The annals of statistics*, 6(2), 461–464.
- Simen, P., Cohen, J. D., & Holmes, P. (2006). Rapid decision threshold modulation by reward rate in a neural network. *Neural networks*, 19(8), 1013–1026.
- Simen, P., Contreras, D., Buck, C., Hu, P., Holmes, P., & Cohen, J. D. (2009). Reward rate optimization in two-alternative decision making: empirical tests of theoretical predictions. *Journal of Experimental Psychology: Human Perception and*

1 *Performance*, 35(6), 1865.

2 Smith, P. L., & Vickers, D.(1989). Modeling evidence accumulation with partial loss in
3 expanded judgment. *Journal of Experimental Psychology: Human Perception and*
4 *Performance*, 15(4), 797–815.

5 Spear, P. D.(1993). Neural bases of visual deficits during aging. *Vision research*, 33(18),
6 2589–2609.

7 Starns, J. J., & Ratcliff, R.(2010). The effects of aging on the speed–accuracy compromise:
8 Boundary optimality in the diffusion model. *Psychology and aging*, 25(2), 377.

9 Starns, J. J., & Ratcliff, R.(2012). Age-related differences in diffusion model boundary
10 optimality with both trial-limited and time-limited tasks. *Psychonomic bulletin &*
11 *review*, 19(1), 139–145.

12 Stone, M.(1960). Models for choice-reaction time. *Psychometrika*, 25(3), 251–260.

13 Tanner, W. P., & Swets, J. A.(1954). A decision-making theory of visual detection.
14 *Psychological review*, 61(6), 401.

15 Thura, D., Beauregard-Racine, J., Fradet, C.-W., & Cisek, P.(2012). Decision making by
16 urgency gating: theory and experimental support. *Journal of Neurophysiology*,
17 108(11), 2912–2930.

18 Thurstone, L. L.(1927a). A law of comparative judgment. *Psychological review*, 34(4), 273.

19 Thurstone, L. L.(1927b). Psychophysical analysis. *The American journal of psychology*,
20 368–389.

21 Tversky, A., & Kahneman, D.(1992). Advances in prospect theory: Cumulative
22 representation of uncertainty. *Journal of Risk and uncertainty*, 5(4), 297–323.

23 Vickers, D.(1970). Evidence for an accumulator model of psychophysical discrimination.
24 *Ergonomics*, 13(1), 37–58.

25 Vickers, D.(1995). The frequency accrual speed test (fast): A new measure of ‘mental
26 speed’? *Personality and individual differences*, 19(6), 863–879.

27 Vickers, D., Burt, J., Smith, P., & Brown, M.(1985). Experimental paradigms emphasising

state or process limitations: I effects on speed-accuracy tradeoffs. *Acta Psychologica*, 59(2), 129–161.

Vickers, D., Nettelbeck, T., & Willson, R. (1972). Perceptual indices of performance: the measurement of ‘inspection time’ and ‘noise’ in the visual system. *Perception*, 1(3), 263–295.

Vickers, D., Smith, P., Burt, J., & Brown, M. (1985). Experimental paradigms emphasising state or process limitations: II effects on confidence. *Acta Psychologica*, 59(2), 163–193.

Voskuilen, C., Ratcliff, R., & Smith, P. L. (2016). Comparing fixed and collapsing boundary versions of the diffusion model. *Journal of Mathematical Psychology*, 73, 59–79.

Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic bulletin & review*, 14(5), 779–804.

Wagenmakers, E.-J., Ratcliff, R., Gomez, P., & McKoon, G. (2008). A diffusion model account of criterion shifts in the lexical decision task. *Journal of Memory and Language*, 58(1), 140–159.

Wald, A. (1947). *Sequential analysis*. New York, NY: John Wiley & Sons.

Wald, A., & Wolfowitz, J. (1948). Optimum character of the sequential probability ratio test. *The Annals of Mathematical Statistics*, 19(3), 326–339.

Wallsten, T. S. (1968). Failure of predictions from subjectively expected utility theory in a bayesian decision task. *Organizational Behavior and Human Performance*, 3(3), 239–252.

Wasserman, L. (2000). Bayesian model selection and model averaging. *Journal of mathematical psychology*, 44(1), 92–107.

Watson, A. B., Ahumada, A. J., & Farrell, J. E. (1986). Window of visibility: a psychophysical theory of fidelity in time-sampled visual motion displays. *JOSA A*, 3(3), 300–307.

Weale, R. A. (1963). *The aging eye*. Hoeber Medical Division, Harper & Row.

- 1 Wickelgren, W. A. (1977). Speed-accuracy tradeoff and information processing dynamics.
2 *Acta psychologica*, 41(1), 67–85.
- 3 Zacksenhouse, M., Bogacz, R., & Holmes, P. (2010). Robust versus optimal strategies for
4 two-alternative forced choice tasks. *Journal of mathematical psychology*, 54(2),
5 230–246.

Appendix A

Non-decision time

Figure A1 illustrates the reverse correlation procedure used to compute non-decision time across the experiments. For this example participant, the proportion of Easy-difficulty trials where the samples at time steps $(1, 2, 3, 4, 5, 6, \dots)$ before the response were in the same direction as the responses were $(0.62, 0.65, 0.63, 0.68, 0.85, 0.91, \dots)$. The first time-step where this proportion was larger than 0.75 was 5 and we estimated the non-decision time to be 4. Note that the correlation during the non-decision time is not 0.50 due to drift in the stimuli – the responses made by participants are generally correlated to the stimulus (provided their response rates were better than chance, which we check for in our exclusion criterion). Hence it is important to pick a threshold that is larger than the drift. We experimented with a number of different threshold values. Results remained similar for a range of threshold values that were above the drift rate but below the highest correlated stimuli for the participant. The threshold value chosen in this manner was kept constant across participants and experiments.

Note that this analysis method assumes a fixed non-decision time across trials. Since non-decision time would, presumably, vary from trial to trial, we carried out simulations to test how trial-by-trial variability in non-decision time affects our estimates of the slopes of the decision boundaries. We found that the analysis method outlined in the main text was robust under trial-to-trial variability in non-decision time: adding a trial-by-trial variability added a small amount of noise in our estimates but estimated slopes were still highly correlated with true slopes and inferences about difference in slopes between conditions remained the same irrespective of the variability of non-decision time.

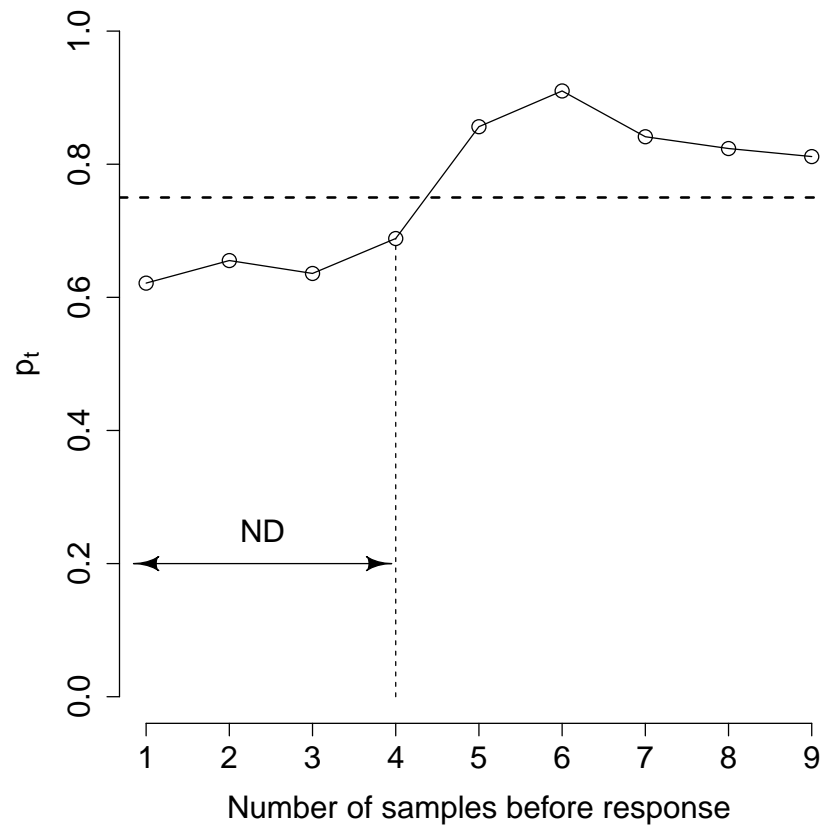


Figure A1. The plot shows the proportion of trials at each time step before the response (for a particular participant and condition) that are in the same direction as the response. The dashed horizontal line represents a threshold on this proportion used to compute the non-decision time. The dashed vertical line shows the last time-step where this proportion was below the threshold.

Appendix B

Parameter recovery

In this study, we estimated decision boundaries based on a *line of indifference* computed using a logistic regression model. We now show that this estimation method allows us to make valid comparisons about slopes of Single and Mixed-difficulty games and that our inferences about difference in slopes are valid irrespective of whether the noise in decisions originates from stochasticity in *wait/go* actions (as we assume) or from a noisy integration of sensory signals.

We evaluated estimated slopes by simulating decisions from known boundaries and comparing estimated values with known values. Decisions were simulated using two alternative models: (i) a rise-to-threshold model, with noisy accumulation of evidence but no noise in the boundary and (ii) a rise-to-threshold model with no noise in the accumulated signal but noisy wait-go decisions. Once decisions had been generated, slopes were estimated using two alternative fitting methods: the logistic regression method described in Equations 2 and 3 and a maximum likelihood fit to rise-to-threshold model described below. Below (Figure B1) we show a 2x2 comparison between the two estimation methods and the two models used to generate the decisions. Before discussing these results, we describe how simulated data was generated and how the slopes of boundary were estimated.

Simulating decisions

In the noisy accumulation model, each sample (cue) was generated using a Bernoulli process (Equation 1), with drift parameters, $u_e = 0.72$, $u_d = 0.50$. These samples were then integrated in a noisy decision-variable, V :

$$V_{t+1} = V_t + \delta X + \xi \quad (5)$$

where $\xi \sim \mathcal{N}(\cdot|0, \sigma_V^2)$ is a zero-mean Gaussian noise with variance σ_V^2 . This integration process generates a random walk, (X_1, \dots, X_t) and terminates when the decision-variable,

1 V_t , crosses a known boundary. We simulated random walks using 200 different values of
 2 boundary (standing for 200 simulated participants), with varying slopes and intercepts,
 3 generating 1000 random walks for each boundary (in both Easy and Mixed conditions).

4 For the stochastic boundary model, decisions were generated stochastically based on
 5 the distance from boundary. For each given boundary, θ , we determined the probability of
 6 going at any point (t, x) of the random walk, based on the distance, $f_\theta(t, x)$, of the point
 7 from the boundary:

$$\mathbb{P}(A = go) = \frac{e^{f_\theta(t, x)}}{1 + e^{f_\theta(t, x)}} \quad (6)$$

8 The key difference between data simulated using this model and the rise-to-threshold
 9 model is that this model does not assume any accumulation of noise – each *wait* / *go*
 10 action is independent and solely based on $\mathbb{P}(A = go)$ at that point.

11 Estimating slopes

12 The logistic regression method of estimating slopes has been described in the main
 13 text (Equations 2 and 3). We now describe how we generated maximum-likelihood
 14 estimates of slopes for the rise-to-threshold model with noisy accumulation of evidence.

15 According to this model, the value of the decision variable after accumulating t
 16 samples and evidence x is obtained by integrating Equation 5 and is given by the Gaussian
 17 distribution, $\mathcal{N}(\cdot|x, \sigma_d^2 t)$, with mean x and variance $\sigma_d^2 t$, where σ_d is a free parameter that
 18 needs to be estimated. The probability of observing a *go* at (t, x) will be given by the
 19 probability that the decision variable is greater than or equal to the boundary, θ , i.e.

$$\begin{aligned} \mathbb{P}(t^{go}, x^{go}) &= \int_{\theta(t, x)}^{\infty} \mathcal{N}(\kappa|x, \sigma_d^2 t) d\kappa \\ &= 1 - \Phi(f_\theta(t, x)|x, \sigma_d^2 t) \end{aligned} \quad (7)$$

20 where $f_\theta(t, x)$ is the distance to the current evidence, x , from the boundary and $\Phi(\cdot|x, \sigma_d^2 t)$
 21 is the cumulative Gaussian with mean x and variance $\sigma_d^2 t$. The boundary θ is
 22 parameterised by it's slope, m , and intercept, c . Both of these are free parameters of the

1 model. Similarly, the probability of *waiting* at (t, x) is given by $\Phi(f_{\theta(t,x)}|x, \sigma_d^2 t)$. If \mathcal{D}^{wait} is
 2 the set of all *wait* observations, $\{(t_1^{wait}, x_1^{wait}), \dots, (t_n^{wait}, x_n^{wait})\}$, and \mathcal{D}^{go} is the set of all *go*
 3 observations, $\{(t_1^{go}, x_1^{go}), \dots, (t_m^{go}, x_m^{go})\}$, then the likelihood of all observations is given by:

$$L(\mathcal{D}|m, c, \sigma_d) = \prod_{(t,x) \in \mathcal{D}^{go}} (1 - \Phi(f_{\theta(t,x)}|x, \sigma_d^2 t)) \prod_{(t,x) \in \mathcal{D}^{wait}} \Phi(f_{\theta(t,x)}|x, \sigma_d^2 t) \quad (8)$$

4 where \mathcal{D} includes both \mathcal{D}^{wait} and \mathcal{D}^{go} decisions. We obtained estimates, \hat{m} , \hat{c} and $\hat{\sigma}_d$ that
 5 maximised the likelihood function given in Equation 8.⁸

6 Evaluation of estimation methods

7 The panel on the top-left of Figure B1 compares the true and estimated difference in
 8 slopes when decisions were generated using the noisy accumulation model (Equation 5) and
 9 estimated using the logistic regression model (Equations 2 and 3). The estimated difference
 10 in slopes is not equal to the true difference and the deviation from truth does depend on
 11 the noise in the accumulation process, σ_V ; simulations showed that the difference
 12 approaches zero as σ_V approaches zero. Moreover, the estimated difference is
 13 approximately proportional to the true difference, so that if statistical test, such as the
 14 t-test, is valid on the true difference it will also be valid on the estimated difference.

15 The top-right panel in Figure B1 compares the true and estimated differences in
 16 slopes, when decision were again generated using the noisy accumulation model
 17 (Equation 5) and also estimated assuming a noisy accumulation to boundary (Equation 8).
 18 It can be seen that the difference in slopes estimated using this method is fairly similar to
 19 that estimated using the logistic regression model. The estimated difference is
 20 approximately proportional to the true difference and, like the estimates in top-left panel,
 21 these estimates also contains a deviation that depends on the diffusion parameter used to

⁸Note that we used both \mathcal{D}^{wait} and \mathcal{D}^{go} to obtain maximum-likelihood estimates. If we use only \mathcal{D}^{go} , which is what is available if the paradigm does not involve an expanded-judgement task, we get a more noisy estimate of the slope. In that case we also obtain a poor correlation between true and estimated differences when data was simulated using the stochastic boundary model (bottom-right) graph.

1 generate the data, σ_V . Simulations also showed (not shown in Figure B1) that, while the
 2 difference in slopes using this method was similar to that using the logistic regression
 3 method, this method overestimated the slopes when true slope was zero – that is, when
 4 bounds are in fact flat, it estimates them to be increasing, which is not the case with the
 5 logistic regression method used in this study.

6 The bottom-left panel shows that the difference in slopes estimated using the logistic
 7 regression model (Equations 2 and 3) are highly correlated with the true differences when
 8 data is simulated using the stochastic boundary model (Equation 6). This is not surprising
 9 since this method of generating the data inverts the logistic regression model. In contrast,
 10 the bottom-right panel shows that when slopes are estimated using the noisy accumulation
 11 process (Equation 8), the correlation between the true and estimated difference in slopes
 12 decreases and there is a bias in the estimated difference albeit, again, multiplicative.

13 To sum up, difference in slopes estimated using the logistic regression model, were at
 14 least as good as estimates using the maximum likelihood method and the logistic regression
 15 model was, in fact, robust to mis-specification of the model that generates the data.
 16 Furthermore, estimated differences were linearly related to true differences, which meant
 17 that it was valid to use the t-test on the estimated differences for inferring whether there
 18 was a difference in the true slopes of Single and Mixed-difficulty games.

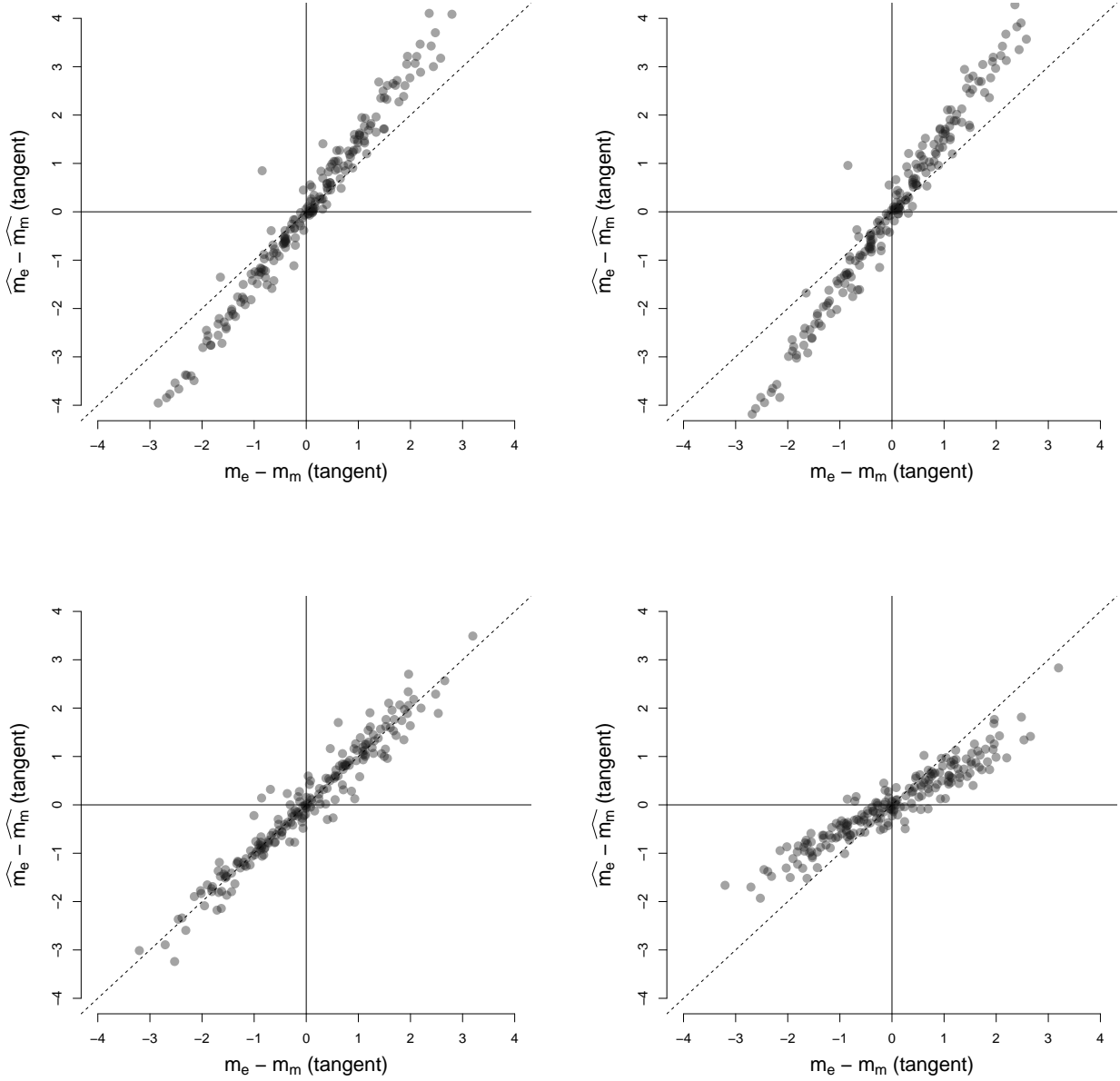


Figure B1. True versus estimated difference in slopes for 200 simulations. Each circle represents one participant simulated using either a rise-to-threshold model (first row) or a probabilistic boundary model (second row). The figures in the left column compare true difference with the difference estimated using logistic regression (as described in Section ‘Experiment 1’) while the panels in the right column make the same comparison but slopes are estimated by fitting a rise-to-threshold model.

Appendix C

Variable drift rate and Information loss

We have analysed the data based on the evidence and time of each *wait* and *go* decision, assuming that participants accumulated every cue provided to them and the internal drift rate was the same in every trial. In this Appendix we verified whether our inferences about difference in slopes would be valid if some evidence was lost during accumulation and the drift rate varied from trial-to-trial.

Consider a variable drift rate first. Even though our experiments use the expanded-judgement paradigm where the drift of the stimulus is controlled by the experimenter, it is possible that the internal drift rate varies from trial-to-trial due to fluctuations in attention and cognitive resources. Fluctuations in the effect of stimulation have been modelled as a random variable since Thurstone’s comparative and categorical judgement models (Thurstone, 1927a, 1927b) and form an integral part of signal detection theory (Tanner & Swets, 1954; Green & Swets, 1966) and sequential sampling models (Ratcliff, 1978; Ratcliff & Smith, 2004).

In order to check how variability in drift rate affected our results, we simulated decisions using a rise-to-threshold model both with and without variability in drift and estimated the slopes in each case using the logistic regression method (Equation 3). A comparison between the estimated slopes showed us how a variability in drift affects our estimate.

Decisions were simulated using the following method. Stimuli were generated using a Bernoulli process (Equation 1), with drift $\epsilon = \epsilon_0 + \nu$, where ϵ_0 was a constant drift parameter based on the type of game (e.g., $\epsilon_0 = 0.22$ for Easy games) and $\nu \sim \mathcal{N}(0, \sigma_{drift}^2)$ was a random variable drawn independently for every trial. The overall drift, ϵ , was bound between 0 and $\frac{1}{2}$. This stimuli was then integrated in a noisy decision-variable, V :

$$V_{t+1} = V_t + \delta X + \xi \quad (9)$$

where $\xi \sim \mathcal{N}(0, \sigma_V^2)$ is a zero-mean Gaussian noise with variance σ_V^2 . This integration

process generated a random walk, (X_1, \dots, X_t) and terminated when the decision-variable, V_t , crossed a known boundary. We simulated random walks using 200 different values of boundary (standing for 200 simulated participants), with varying slopes and intercepts, generating 1000 random walks for each boundary (in both Easy and Mixed conditions).

Figure C1(a) shows a comparison between true and estimated slopes in Easy games when $\sigma_{drift} = 0$, i.e. there was no trial-to-trial variability in drift, as well as when $\sigma_{drift} = 0.35$, i.e. there was a large trial-to-trial drift variability. When there was no trial-to-trial variability in drift, the estimated slopes were close to true slopes. In the presence of drift variability, the magnitude of slopes was systematically overestimated. However, constant slopes were still estimated as constant and increasing or decreasing slopes were also estimated as increasing or decreasing, respectively. Thus, irrespective of trial-to-trial variability in drift rates, a negative estimate of slope indicated that the true slope was also negative.

Figure C1(b) shows a comparison between true and estimated difference in slopes for $\sigma_{drift} = 0$ and $\sigma_{drift} = 0.35$. In the presence of drift variability, the bias in estimation of slopes results in a bias in estimation of difference in slopes, with the magnitude of estimated difference being larger than true difference. The estimated difference in slopes is approximately proportional to the true difference for both $\sigma_{drift} = 0$ and $\sigma_{drift} = 0.35$. Therefore, when there was no difference in true slopes, there was also no difference in estimated slopes. Similarly, when true difference in slopes was positive (negative), the estimated difference in slopes was also positive (negative). Therefore, even when there was trial-to-trial variability in drift, the estimated difference in slopes indicated a difference in true slopes.

We also checked whether our inferences were robust to loss of information in the integrative decision process. To do this, we simulated data from the binomial loss model (Smith & Vickers, 1989), where observations δX available to the decision-maker are only

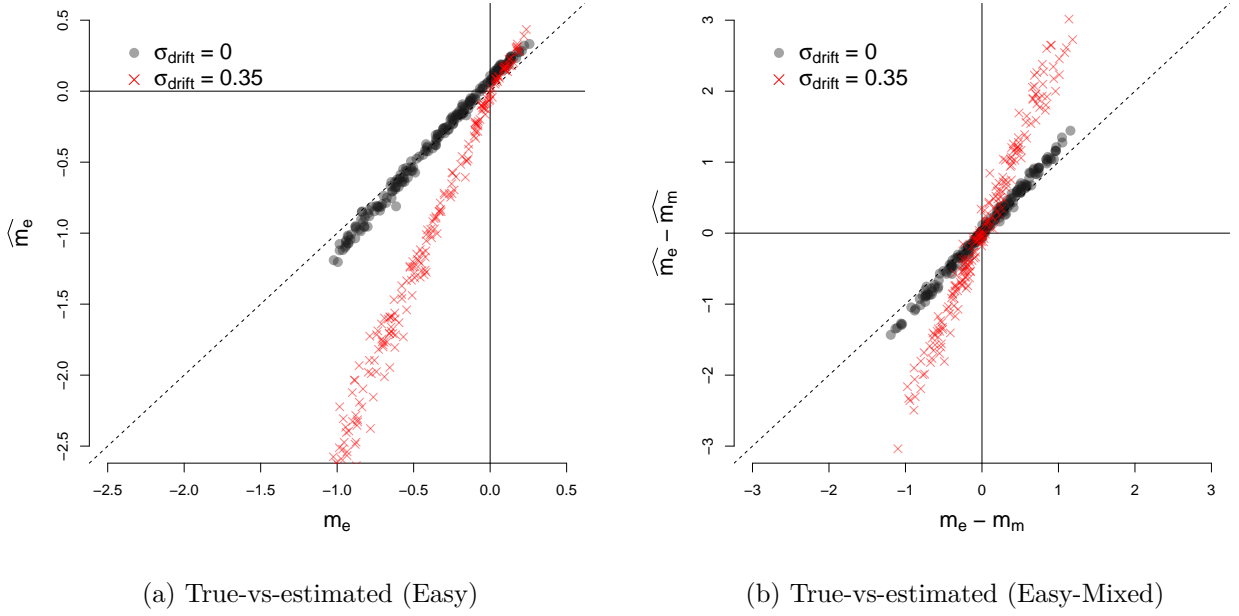


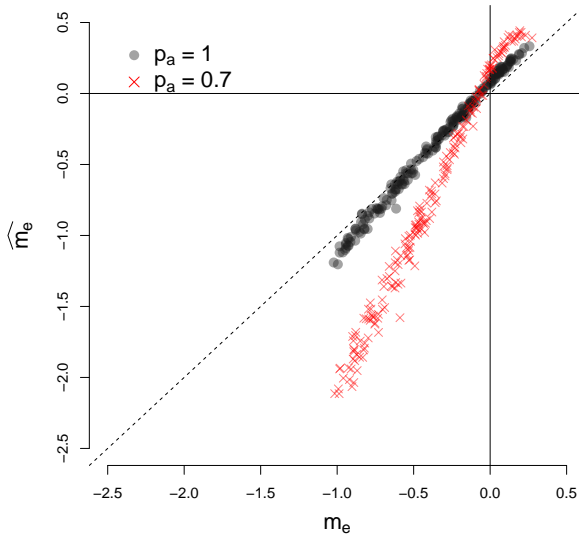
Figure C1. The effect of trial-to-trial variability in drift on estimation of slopes of decision boundary. Panel (a) shows a comparison of true and estimated slopes in a Single-difficulty task ($\epsilon = 0.22$). Panel (b) shows a comparison of true and estimated difference in slope during an Easy ($\epsilon = 0.22$) and Mixed ($\epsilon = 0.22/0$) task. In both panels, each dot represents a participant simulated using a rise-to-threshold model with no trial-to-trial drift and crosses represent participants simulated using a rise-to-threshold model with a large trial-to-trial variability in drift ($\sigma_{drift} = 0.35$).

1 accumulated with some fixed probability p_a , otherwise they are lost:

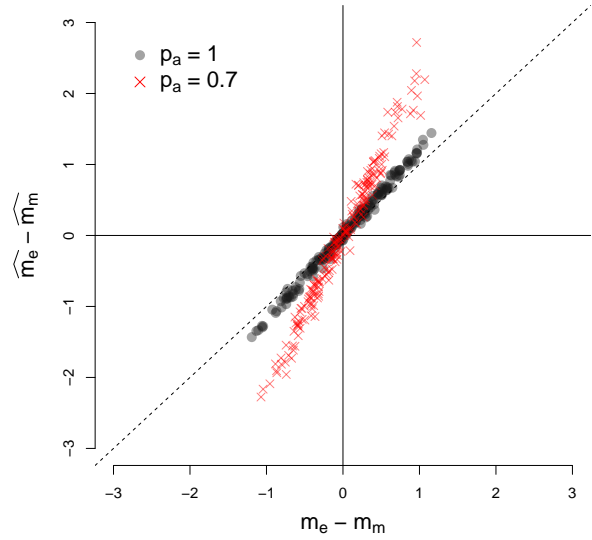
$$V_{t+1} = \begin{cases} V_t + \delta X + \xi & \text{with probability } p_a \\ V_t + \xi & \text{with probability } 1 - p_a \end{cases} \quad (10)$$

2 where V_t and ξ are defined in the same way as Equation 9. Figure C2(a) compares the
 3 estimates of slopes during Easy games for 200 simulated participants from the information
 4 loss model (Equation 10, $p_a = 0.7$) with 200 simulated participants from the
 5 rise-to-threshold model without any information loss (Equation 9). Introducing

1 information loss resulted in some systematic biases: estimated slopes were, in general,
 2 larger and shifted in the positive direction so that constant slopes were estimated to have a
 3 small positive value and slightly negative slopes were estimated to be constant while large
 4 negative and positive slopes were estimated to have a larger value than the true slopes.
 5 Note the direction of this bias – when slopes were estimated to be negative (as in the
 6 majority of experiments in this study), the true slopes were also negative while when they
 7 were estimated to be positive, they could in fact be constant. This makes sense: the total
 8 information loss will increase over time and since the drift is positive, decisions made at
 9 later points of time will seem to be at higher levels of evidence than internally integrated.



(a) True-vs-estimated (Easy)



(b) True-vs-estimated (Easy-Mixed)

Figure C2. The effect of information loss on estimation of slopes of decision boundary.

Panel (a) shows a comparison of true and estimated slopes in a Single-difficulty task ($\epsilon = 0.22$). Panel (b) shows a comparison of true and estimated difference in slope during an Easy ($\epsilon = 0.22$) and Mixed ($\epsilon = 0.22/0$) task. In both panels, each dot represents a participant simulated using a rise-to-threshold model without any information loss and crosses represent a participant simulated using a binomial loss model ($p_a = 0.7$).

Figure C2(b) compares the true and estimated difference in slopes between Easy and Mixed games. When information was lost, the estimate of slopes in each type of game was biased, therefore, the estimate for difference in slopes was also biased. However, the estimated difference in slopes was still approximately proportional to true difference in slopes, passing through the origin so that the estimated difference in slopes was proportional to the true difference in slopes even when data were simulated from the binomial loss model.

Appendix D

Model selection and recovery

A key finding of Experiments 1 and 2 is that participants seem to decrease their decision boundaries with time, especially in the Mixed-difficulty condition. In order to check the robustness of this result, we compared the logistic regression model presented in the main text (Equation 2), which uses both evidence and time to predict the probability of *go*, with a simpler model that uses only evidence to predict this probability:

$$\log \frac{\mathbb{P}(A = go)}{\mathbb{P}(A = wait)} = \beta_0 + \beta_X * X \quad (11)$$

where β_X is the regression coefficient for evidence and β_0 is the intercept. Note that this simpler model is equivalent to assuming that decision boundaries do *not* change with time as only evidence predicts whether a participant chooses the action *wait* or *go* during a trial. We inferred the preferred model for each participant and condition by computing the Bayesian Information Criterion (BIC) for each model (Schwarz, 1978; Wagenmakers, 2007). Following Wasserman (2000) and Hawkins et al. (2015), we approximated the posterior probability for a participant using each model under the assumption that both models are a priori equally likely:

$$P_{BIC}(M_i|Data) = \frac{\exp(-\frac{1}{2}BIC(M_i))}{\sum_{j=1}^m \exp(-\frac{1}{2}BIC(M_j))} \quad (12)$$

Figure D1 plots these posterior probabilities for data from both single-difficulty (left column) and mixed-difficulty conditions (right column) for each participant. Shaded (red) bars show the posterior model probability for the more complex model (using both evidence and time), while the white hatched bars show the complementary posterior probability for the simpler model (only on evidence).

In agreement with the slopes estimated using the line of indifference (Figures 5 and 8), we found that the model using both evidence and time provided the best account of the data in Experiments 1, 2a and 2b. This was especially true for the Mixed-difficulty

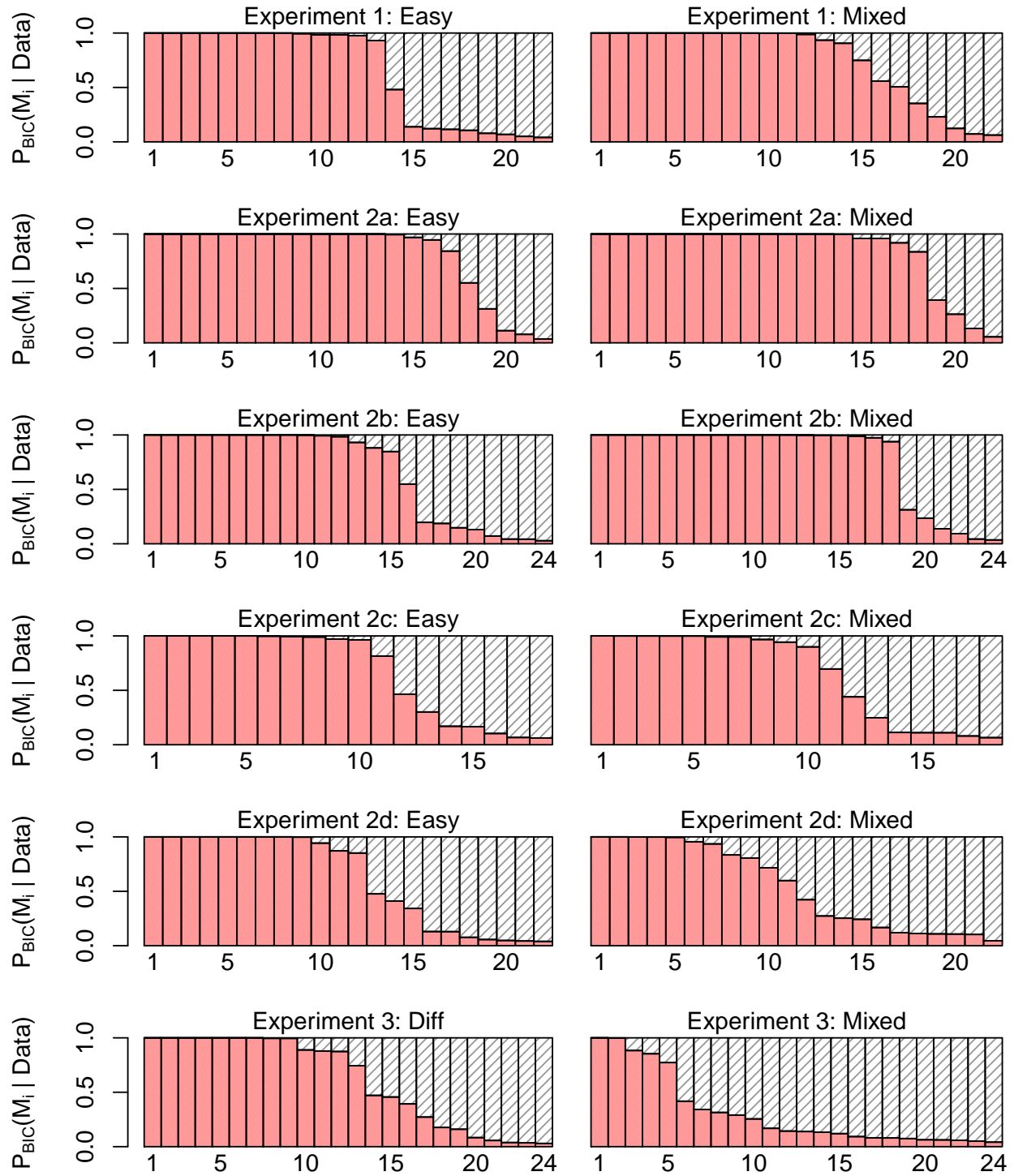


Figure D1. Each row shows the posterior model probability for the logistic regression models using only evidence as the predictor (hatched) and using both time and evidence as predictors (shaded, red) for all participants in an experiment. The left and right-hand columns show these posterior probabilities during the Single and Mixed-difficulty conditions, respectively.

games. Note that, in spite of the fact that the BIC rewards lower model complexity, the simpler model of Equation 11 provided the best account for only 5 participants during the Mixed-difficulty trials during these experiments. Data from most other participants was better accounted for by the model using time as an additional predictor variable.

In contrast, during Experiments 2c and 2d, where the inter-trial intervals were same for correct and incorrect decisions and reward landscapes were flatter (Figure 9), support for the two models was much more mixed. Finally, during Experiment 3, where the reward landscape favoured constant (or slightly increasing) decision boundaries with time (Figures 10 and 11), the simpler model, using only evidence to predict probability of *going*, provided the best account of the data, especially in the Mixed-difficulty condition.

We checked the validity of this model selection procedure using a model recovery analysis. We simulated decisions using two noisy accumulation to threshold models (see Equation 5 in Appendix B above). The first model used a constant threshold, while the second one used a threshold that decreased linearly with time. For each simulated participant, we generated 100 random walks with drifts drawn, with uniform probability, from the set $\epsilon \in \{0.22, 0\}$. Thus, these simulated random walks approximately matched the data collected and conditions for Mixed-difficulty trials in Experiment 2.

We then fit the two logistic regression models discussed above, one using only evidence as a predictor (Equation 11) and the other using evidence as well as time (Equation 2), to each of these simulated participants and computed the BIC values for each fit. Each plot in Figure D2 shows the distribution of difference in BIC values for 400 simulated participants, 200 of which are simulated using the fixed boundary model (hatched distribution), while the remaining 200 are simulated using the decreasing boundary model (shaded distribution). It can be seen from these plots that the decisions generated using a fixed boundary model are better fit (lower BIC_{fix}) by the logistic regression model using only evidence as the predictor, while the decisions generated using a decreasing boundary model are better fit (lower BIC_{var}) by the logistic regression model

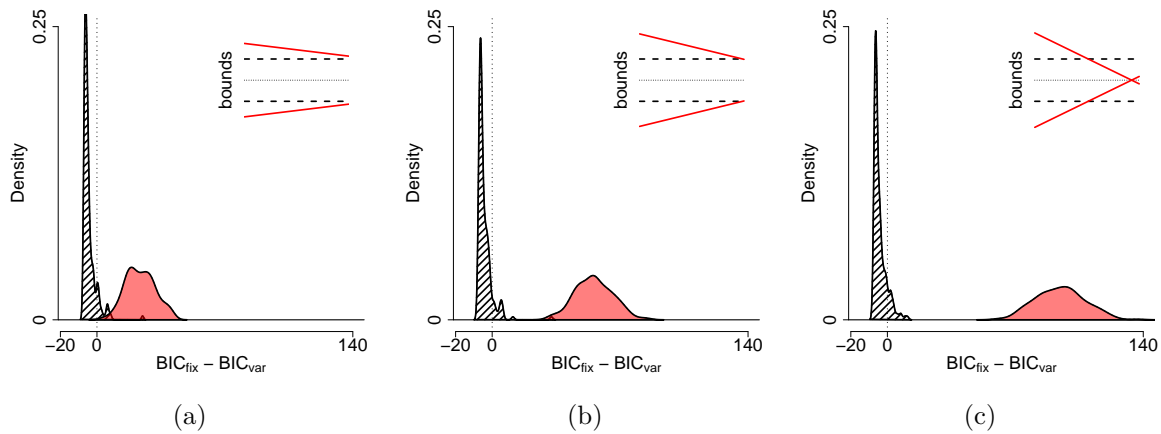


Figure D2. Each plot shows the distribution of difference in BIC values for fixed boundary model (only evidence used as predictor) and variable boundary model (both evidence and time used as predictors). Hatched distributions show this difference in BIC values when decisions are simulated using integration of evidence to a fixed boundary (inset, dashed line) and shaded (red) distributions show the difference in BIC values when decisions are simulated from integration to a decreasing boundary (inset, solid line). Three different slopes are used for decreasing boundaries (a) 15 degrees, (b) 30 degrees and (c) 60 degrees.

- 1 using both evidence as time as predictors. Furthermore, when the slope of the decision
- 2 boundaries is increased for the simulated decisions, the difference in BIC values increases
- 3 (compare the shaded region in the three plots), showing that the BIC of the time-varying
- 4 bounds model decreases with increase in the slope used to generate the decisions.