

Sequential Sampling of Noisy Information: Attention, Stopping, and Risk-Taking*

Ferdinand M. Vieider¹

¹*RISL $\alpha\beta$, Department of Economics, Ghent University*

7 September 2026

Abstract

I present a model of decision making in which canonical risk-taking regularities and attention allocation emerge jointly from a common cognitive process. A decision maker chooses between two options but has access only to noisy samples of the underlying choice attributes. She must therefore decide where to direct attention and when to stop sampling and choose. The Bayesian decision maker tracks posterior uncertainty over the sampled attributes. This uncertainty serves a functional role: it governs the value of further sampling, directs attention, determines stopping, and generates stochastic choice. The model yields a distinctive prediction. Exogenously increasing the noise of one decision attribute raises the value of sampling that attribute and therefore attracts additional attention to it. This compensation is incomplete, however, so that greater attention coincides with lower behavioural sensitivity to the noisier attribute. Although this appears to conflict with recent evidence that individuals who attend relatively more to probabilities are also more probability-sensitive, the two patterns can coexist because they arise from different cognitive primitives. I test this contrast in a mouse-tracking experiment and find the predicted increase in probability attention alongside reduced probability sensitivity, while the positive cross-sectional attention–sensitivity association remains present within conditions.

[...]a wealth of information creates a poverty of attention and a need to allocate that attention efficiently among the overabundance of information sources that might consume it.

—Herbert A. Simon ([1971](#), pp. 40–41)

1 Introduction

This paper examines the risky-choice behaviour of a decision maker who must choose between two options, but only has access to noisy samples of their attributes. The decision maker then confronts two related problems. The first is how to allocate information search across the different attributes. The second is when to stop collecting information

*I gratefully acknowledge funding by the Research Foundation–Flanders under grant nr. G064826N: “Noisy Cognition as a Unified Driver of Economic Decisions”. I am indebted to Rava Azeredo da Silveira and Rafael Polanía for helpful comments. All errors remain my own.

and make a choice, even if the available evidence remains incomplete. I develop the framework for risky choice, where probabilities and outcomes provide natural targets of information search, although the underlying problem is more general. The model predicts a two-way causal relationship between information search and the evolving probability of choice: residual uncertainty about the choice attributes shapes how long the decision maker samples and where attention is directed, while the resulting allocation of attention shapes the probability with which each option is chosen.

Statistical origins. At face value, this appears to be a standard problem of sequential statistical inference. In the sequential probability ratio test (*SPRT*) developed by Abraham Wald (1945)—and in related procedures used by Alan Turing during World War II (see Gold and Shadlen, 2002)—evidence for two competing hypotheses is accumulated sequentially in a likelihood ratio. This evidence may be combined with prior odds and with the relative losses associated with different decisions. Sampling stops, and a decision is made, once the accumulated evidence crosses one of two decision boundaries. Closely related mechanisms underlie many evidence-accumulation models in psychology (Ratcliff, 1978; Busemeyer and Townsend, 1993; Bogacz et al., 2006; Krajbich et al., 2010).

Here I argue that, while these models capture important features of sequential decision making, they leave out a feature of cognition that becomes central once information can be sampled selectively. Decision makers appear not only to accumulate evidence, but also to keep track of how uncertain they remain about the different attributes of the choice options. The *SPRT* does not represent this uncertainty separately: the accumulated likelihood ratio is sufficient for deciding when to stop and which hypothesis to select. Once information can be sampled selectively, however, uncertainty about the sampled attributes becomes decision-relevant in its own right. Attribute-specific uncertainty can direct attention towards those attributes for which further evidence is most valuable, in the spirit of Chernoff (1959), and determine when the expected benefit of additional information no longer justifies its cost, in the spirit of Arrow et al. (1949). Attention allocation, stopping, and choice thus emerge jointly from a common inferential process.

Information acquisition and stopping. I build the framework on the static noisy-cognition model of Khaw et al. (2021) and its generalization by Vieider (2024). Khaw et al. (2021) model Bayesian inference from a single noisy sample of lottery outcomes, while as-

suming probabilities to be perceived objectively. They show that small-stakes risk aversion can emerge from Bayesian inference itself, without any need to invoke a utility function. [Vieider \(2024\)](#) generalizes this framework to accommodate probability-dependent risk taking as well as small-stakes risk aversion; [Oprea and Vieider \(2026\)](#) provide experimental evidence for its underlying information-processing mechanism, showcasing parallel effects for risk and time. Closely related Bayesian models of noisy cognition have been proposed in psychology to explain distorted belief updating ([Zhu et al., 2020; 2024](#)).

Here, I extend this framework to a setting in which information is acquired sequentially. In the two-dimensional setting of [Vieider \(2024\)](#), additional samples can be drawn either from probabilities or from outcomes, so that the decision maker must determine not only how much information to acquire, but also where to acquire it. As information accumulates, posterior uncertainty falls and the marginal value of an additional sample declines. Stopping is governed by trading off the marginal value of the next sample against its cognitive cost, while attention allocation depends on where additional information is currently most valuable. The model therefore generates joint predictions for information acquisition and choice that can be tested using process-tracing measures.

I deliberately keep the underlying inference and choice architecture close to the static models. In particular, I preserve their scale-invariant representation of the choice problem rather than introducing an explicit weighting of mistakes by their economic consequences, and I retain their posterior-draw choice rule: an option is chosen with the posterior probability that it is superior. These assumptions are not claimed to provide a uniquely optimal or normative architecture. The remit here is instead to ask what this noisy-cognition framework implies once information acquisition becomes sequential and endogenous.

Information search is governed by the reduction in the probability of mistakes: additional sampling is valuable insofar as it lowers the probability of choosing the inferior option ([Robson, 2001b;a](#)). I implement this criterion using a one-step-lookahead rule: sampling continues when the expected reduction in error from one additional sample exceeds its cognitive cost, and attention is directed toward the information source for which this reduction is greatest. This differs from standard optimal-sampling applications, which typically solve a dynamic stopping problem over a lower-dimensional state and select an option deterministically once sampling ends.¹ Here, uncertainty remains separately

¹Prominent examples include [Drugowitsch et al. \(2012\)](#) and [Fudenberg et al. \(2018\)](#). In these models,

represented across choice attributes because its distribution determines where information should be sampled, while residual posterior uncertainty continues to affect terminal choice through the inherited posterior-draw rule.

Canonical risk-taking regularities from noisy inference. An important attraction of the noisy-inference framework is that risk-taking regularities such as probability insensitivity, decreasing sensitivity to outcomes, and excess sensitivity to losses—as summarized in prospect theory (*PT*; [Kahneman and Tversky, 1979](#))—can emerge from the inference process itself, without requiring stable tastes for risk ([Khaw et al., 2021; 2023; Vieider, 2024](#)). The same mechanism also generates stochasticity in observed choice, avoiding problems that arise when the stochastic choice rule is specified independently of the underlying preference model ([Wilcox, 2011; Bhatia and Loomes, 2017](#)).

Endogenizing information acquisition adds a process dimension to this account. Noisy representations determine where additional information is valuable and therefore shape the search that precedes choice; the resulting allocation of attention, in turn, changes the evolution of posterior uncertainty and the choice probabilities that obtain when sampling stops. Attention is consequently neither merely an input into choice nor simply a reflection of the importance of an attribute, but an adaptive response to residual uncertainty about the attributes on which choice depends.

Adaptive cognition. In the sequential noisy sampling model (*SNSM*) I develop here, familiar departures from benchmark choice models such as expected-utility maximization are consequences of cognition adapted to the structure and costs of the underlying inference problem. Decision attributes must be inferred from noisy evidence, while the precision of those inferences depends endogenously on where attention is directed and how long information is sampled. Behavioural regularities can therefore arise not from fixed distortions applied to objectively given quantities, but from an information-processing system that allocates limited cognitive resources in response to the demands of the decision environment ([Simon, 1956](#)).

The perspective presented here is complementary to efficient coding models, which ask how representational precision should adapt to the statistical distribution of quantities

the dynamic stopping problem is reduced to a one-dimensional latent state, with terminal choice deterministic conditional on that state. Such a reduction is less natural in the present setting, where the research question concerns how uncertainty over distinct choice attributes governs where information is acquired.

encountered in the environment (Laughlin, 1981; Robson, 2001a; Netzer, 2009; Polania et al., 2019; Frydman and Jin, 2022; Prat-Carrabin and Woodford, 2022; Polania and Hare, 2026). Efficient-coding models explain how limited representational resources generate stimulus-dependent precision, so that atypical quantities may be encoded less precisely. The present model takes the representational system as given and asks how prior uncertainty and coding noise shape the subsequent acquisition of information. A distinctive implication is that difficult or atypical stimuli can induce compensatory increases in sampling that nevertheless remain incomplete, in the sense that the resulting representation remains less precise despite the additional attention.

Contrasting causal and cross-sectional attention effects. The model generates a seemingly paradoxical prediction. An exogenous increase in the representational noise of an attribute raises the value of sampling that attribute and therefore attracts additional attention towards it. This compensatory reallocation is nevertheless incomplete: sensitivity to the noisier attribute still declines. Experimentally increasing probability noise is therefore predicted to lead decision makers to inspect probabilities more while allowing those probabilities to exert less systematic influence on choice.

This prediction appears to conflict with the positive cross-sectional association between probability attention and probability sensitivity documented by Pachur et al. (2018). In the SNSM, however, the two patterns arise from different cognitive primitives. Stable differences in prior uncertainty can generate a positive association between attention and sensitivity across individuals, whereas an exogenous increase in coding noise attracts compensatory attention while reducing the precision ultimately achieved. Precisely because the two effects run in opposite directions, their joint occurrence provides a diagnostic test of the model. I therefore test this prediction using mouse-tracking data.

Attention: effects and determinants. Attention has become an increasingly important explanatory construct in theories of decision making. A growing body of work incorporates attentional processes to explain choice behaviour (e.g., Krajbich et al., 2010; Krajbich and Rangel, 2011; Bordalo et al., 2012; Bhatia, 2014; Li and Camerer, 2022), while experimental studies show that exogenous manipulations of attention can causally influence choice (e.g., Armel et al., 2008; Pachur et al., 2018; Chung et al., 2025). Existing approaches differ considerably, however, in whether attention is treated primarily as an

input into choice or as an endogenous outcome of the decision problem itself.

In attentional evidence-accumulation models, fixation affects the rate at which evidence is accumulated, while attention allocation itself is typically not modelled (e.g., [Krajbich et al., 2010](#); [Krajbich and Rangel, 2011](#)). Related approaches derive attentional weighting from bottom-up properties of the visual stimulus or from economic salience (e.g., [Koch and Ullman, 1985](#); [Bordalo et al., 2012](#); [Li and Camerer, 2022](#)). These models allow attention to respond systematically to features of the choice environment, but do not generally model attention as the outcome of a sequential information-acquisition problem.

A closely related literature models attention as an endogenous choice over information acquisition, allowing decision makers to determine what information to acquire subject to cognitive or economic costs (e.g., [Sims, 2003](#); [Gabaix et al., 2006](#); [Callaway et al., 2021](#); [Jang et al., 2021](#); [Liang et al., 2022](#); [Yang and Krajbich, 2023](#); [Hébert and Woodford, 2023](#); [Gluth et al., 2026](#)). The SNSM, however, uses a different mechanism and asks a complementary question: how evolving uncertainty over separately represented choice attributes can jointly govern where attention is directed, when sampling stops, and how the resulting information shapes choice. This yields a process-level account that links observed risk taking to measures of information acquisition, in line with recent calls in the literature ([Pachur et al., 2018](#); [Zilker and Pachur, 2022](#)). Attention allocation, stopping, and stochastic choice therefore arise from a common cognitive mechanism rather than being specified separately.

2 Sequential Accumulation of Noisy Information

I model choices between binary lotteries yielding a prize x with probability p and an outcome $y < x$ otherwise, and sure amounts of money c intermediate between x and y . The model naturally extends to losses and mixed lotteries, but I develop the main equations below for gains (i.e., $x > c > y \geq 0$). While the key intuitions extend more broadly, the cognitive perspective developed here implies that different choice environments may induce different information-processing regimes ([Bouchouicha et al., 2024](#)).

2.1 Sequential sampling and discriminability of choice options

The choice problem. Let us begin from the benchmark problem of maximizing expected value: choosing the lottery over the sure amount whenever $px + (1 - p)y > c$. This

rule maximizes expected payoffs in the long run, and is therefore normatively appealing at least for low-stakes, repeated decisions (Wakker, 2010; Woodford, 2020; Khaw et al., 2021). More importantly for present purposes, it provides a benchmark in the absence of primitive risk attitudes or preference-based risk motives, allowing canonical patterns of risk-taking to emerge solely from the sequential accumulation of noisy information.

Rearranging terms and taking natural logarithms of both sides yields

$$\ln\left(\frac{p}{1-p}\right) > \ln\left(\frac{c-y}{x-c}\right).$$

This representation has several advantages. First, it is equivalent to choosing the lottery whenever the expected benefit of choosing the lottery over the sure amount, $p(x-c)$, exceeds the expected cost, $(1-p)(c-y)$. Second, the comparative formulation avoids assigning absolute utility indices to outcomes, which are often psychologically and empirically difficult to identify (Tversky, 1969). Third, the resulting attribute-wise trade-off creates parallels with decision making over time delays (Scholten et al., 2014; Oprea and Vieider, 2026), has received direct support from eye-tracking data (Arieli et al., 2011), and is consistent with evidence that the brain represents the relative value of competing choice options (Fitzgerald et al., 2009; Lim et al., 2011).²

Bayesian priors about choice attributes. The mechanism is straightforward: the decision maker combines noisy samples with prior expectations to form Bayesian posterior distributions over the relevant choice attributes.³ The priors take the following form:

$$\ln\left(\frac{p}{1-p}\right) \sim \mathcal{N}\left(\ln\left(\frac{p_0}{1-p_0}\right), \sigma_p^2\right), \quad \ln\left(\frac{c-y}{x-c}\right) \sim \mathcal{N}\left(\ln\left(\frac{k_0}{b_0}\right), \sigma_o^2\right).$$

The parameters governing these priors, p_0 , b_0 , k_0 , σ_p , and σ_o , are taken to be invariant over the duration of the ‘experiment’ studied below. This isolates the effects of sequential

²Taking logarithms is immaterial for the benchmark choice rule itself, since it preserves the comparison between expected benefits and expected costs. It is nevertheless computationally convenient because it transforms multiplicative relationships into additive ones, allowing noisy information about probabilities and cost-benefit quantities to accumulate sequentially over time. The key conclusions derived below, however, remain intact without log-transforming the choice rule.

³It is important to note that I do not assume that the decision maker carries out these calculations consciously, or indeed that she needs to be able to do so. Rather, Bayesian information processing is a key hypothesis about brain functioning in neuroscience (Knill and Pouget, 2004; Ma et al., 2006; 2023), and is generally thought to operate subconsciously.

sampling from slower changes in the underlying prior representations.⁴

The prior means need not coincide with the true distribution of stimuli encountered in the environment. In the present context, I argue that $p_0 < 0.5$ and $k_0 > b_0$ provide psychologically plausible learned starting points (although they may differ at the individual level). Together, these assumptions express caution in the absence of reliable information: before sampling, the decision maker expects favourable outcomes to be relatively unlikely and the downside of the lottery to outweigh its upside. As we shall see, these priors create a higher evidentiary burden for choosing the riskier alternative than the safer one. They therefore do not impose preferences, but define the initial state from which noisy evidence is accumulated. The prior variances σ_p^2 and σ_o^2 capture uncertainty about these initial expectations. Larger prior variance implies weaker confidence in the prior and gives new evidence greater scope to influence the resulting posterior beliefs.

Sequential integration of noisy signals. When presented with a choice problem, decision makers sequentially sample noisy signals about the log-odds and the log cost-benefit ratio.⁵ Noise arises in both channels. Probability signals may be noisy because described probabilities are difficult to represent precisely, because some descriptive formats are harder to parse than others, or both. Outcome signals may likewise be noisy because the relative costs and benefits of the available options must be evaluated and compared internally. The model therefore treats chance assessment and comparative valuation as distinct noisy information channels that can be sampled with different intensity. The treatment of comparative value as an internally represented quantity connects naturally to evidence that relative values are represented in neural populations (Padoa-Schioppa and Assad, 2006; Padoa-Schioppa, 2007; Fitzgerald et al., 2009; Kable and Glimcher, 2009; Lim et al., 2011; Levy and Glimcher, 2012; Knudsen and Wallis, 2022).

One possible implementation is that encoded numerical outcomes are translated into comparative value signals by evaluating value differences relative to representative differences retrieved from memory, in the spirit of Decision-by-Sampling (Stewart et al., 2006). Such a mechanism can implement efficient value coding (Bhui and Gershman, 2018) and fits the

⁴Normality on the logarithmic scale provides a useful approximation while keeping Bayesian updating analytically tractable.

⁵In principle, costs and benefits could be sampled separately, as could the individual quantities entering each comparison. I abstract from these possibilities in order to isolate the central mechanism of attribute-wise attention allocation between probability and outcome information.

present framework particularly well because it naturally generates sequential samples of comparative value.⁶ Other computational implementations of efficient value coding, such as divisive normalization (Glimcher, 2010; Rustichini et al., 2017; Webb et al., 2020), are equally compatible with the framework. The precise computational mechanism is therefore left open; what matters for the present model is that the translation from numerical outcomes into comparative value can itself be noisy.

This comparative-value account is complementary to models in which individual numerical quantities are represented with noise (Khaw et al., 2021). In efficient-coding models, such noise reflects the optimal allocation of limited perceptual resources across numerical magnitudes (Heng et al., 2020; Frydman and Jin, 2022; Polania and Hare, 2026). Efficient numerical coding therefore provides a natural interpretation of the probability channel. For outcomes, the two mechanisms may operate sequentially: numerical magnitudes may first be efficiently encoded with noise and subsequently translated into comparative values through an efficient value-coding mechanism. Recent neural evidence suggests that both efficient numerical and efficient value representations can contribute to decision making (Bedi et al., 2026). This two-stage view also provides a natural route through which unfamiliar numerical units may increase noise in comparative valuation (Garagnani and Vieider, 2025; Oprea and Vieider, 2026). To keep the present model tractable, I abstract from the additional noise arising in the numerical encoding of outcomes and place the entire outcome noise directly on the comparative value signal.

Formally, a single sample i from the log-odds and log cost-benefit ratio takes the form

$$r_{p,i} \sim \mathcal{N}\left(\ln \frac{p}{1-p}, \nu_p^2\right), \quad r_{o,i} \sim \mathcal{N}\left(\ln \frac{c-y}{x-c}, \nu_o^2\right).$$

The samples are unbiased: in expectation, they recover the underlying choice quantities generating the signals. Any individual draw, however, may deviate from the truth because of noise, characterized by the standard deviations ν_p and ν_o .

Bayesian inference on choice quantities. The decision problem now becomes one of Bayesian inference: noisy observations are treated as evidence generated by latent states of the world. Here, these states are the choice quantities entering the expected-value

⁶Here, Decision-by-Sampling is applied to value *differences* rather than individual values: the relative upside and downside of the lottery are compared with representative value differences stored in memory and evaluated according to their ordinal rank.

comparison—specifically, the log-odds and log cost-benefit ratio that are unknown to the decision maker and generate the noisy signals. The task is therefore to infer these latent quantities from the evidence accumulated over time.

Assume there is a series of samples $\mathbf{r}_{p,n} = \{r_{p,1}, \dots, r_{p,n}\}$ and $\mathbf{r}_{o,m} = \{r_{o,1}, \dots, r_{o,m}\}$, which we can take as given for the time being. Aggregating these noisy observations has two important consequences. First, the average signal becomes increasingly precise as additional samples are accumulated. By the law of large numbers, the sample means, $\bar{r}_{p,n} \triangleq \frac{1}{n} \sum_i r_{p,i}$ and $\bar{r}_{o,m} \triangleq \frac{1}{m} \sum_j r_{o,j}$, converge to the underlying latent quantities, while their sampling variances, $\bar{\nu}_{p,n}^2 \triangleq \nu_p^2/n$ and $\bar{\nu}_{o,m}^2 \triangleq \nu_o^2/m$, shrink toward zero as n, m grow large. Second, this increasing precision reduces posterior uncertainty, so that posterior precision increases with accumulated evidence. Because sampling variance falls only with the inverse number of observations—and standard deviation with the inverse square root—uncertainty becomes progressively harder to reduce as additional samples are collected. Each new observation therefore produces a progressively smaller reduction in uncertainty.

Applying standard Bayesian updating under the normal-normal conjugate model yields the following posterior expectations for the log-odds and the log cost-benefit ratio:

$$\mathbb{E} \left[\ln \frac{p}{1-p} \mid \mathbf{r}_{p,n} \right] = \frac{\bar{\nu}_{p,n}^2}{\sigma_p^2 + \bar{\nu}_{p,n}^2} \ln \frac{p_0}{1-p_0} + \frac{\sigma_p^2}{\sigma_p^2 + \bar{\nu}_{p,n}^2} \bar{r}_{p,n},$$

$$\mathbb{E} \left[\ln \frac{c-y}{x-c} \mid \mathbf{r}_{o,m} \right] = \frac{\bar{\nu}_{o,m}^2}{\sigma_o^2 + \bar{\nu}_{o,m}^2} \ln \frac{k_0}{b_0} + \frac{\sigma_o^2}{\sigma_o^2 + \bar{\nu}_{o,m}^2} \bar{r}_{o,m}.$$

The posterior expectation is therefore a precision-weighted average of prior beliefs and sampled evidence. Early in the sampling process, posterior beliefs remain strongly influenced by prior expectations because the accumulated evidence is still noisy. As additional samples are acquired, the sample mean becomes increasingly precise while the sample variance shrinks, causing posterior beliefs to place progressively greater weight on the observed evidence than on the prior. Importantly, this form of probabilistic inference is compatible with biologically plausible implementations, either through probabilistic population codes (Ma et al., 2006) or through sampling-based representations (Sanborn and Chater, 2016).

While the posterior expectations summarize the decision maker’s current beliefs about the latent choice quantities, the posterior variances summarize the uncertainty surrounding

those beliefs. Under Bayesian updating, the posterior variances are given by

$$s_{p,n}^2 \triangleq \mathbb{V}\left[\log \frac{p}{1-p} \middle| \mathbf{r}_{p,n}\right] = \frac{\bar{\nu}_{p,n}^2 \sigma_p^2}{\bar{\nu}_{p,n}^2 + \sigma_p^2}, \quad s_{o,m}^2 \triangleq \mathbb{V}\left[\log \frac{c-y}{x-c} \middle| \mathbf{r}_{o,m}\right] = \frac{\bar{\nu}_{o,m}^2 \sigma_o^2}{\bar{\nu}_{o,m}^2 + \sigma_o^2},$$

where $\bar{\nu}_{p,n}^2 \triangleq \nu_p^2/n$ and $\bar{\nu}_{o,m}^2 \triangleq \nu_o^2/m$.

Posterior uncertainty therefore remains high whenever prior beliefs are diffuse, sampled evidence is noisy, or only limited evidence has been accumulated. Conversely, accumulating additional evidence reduces posterior uncertainty by averaging out coding noise. This reduction, however, occurs at a diminishing rate: each additional sample contributes progressively less information than the previous one. This diminishing marginal reduction in posterior uncertainty plays a central role in the model developed below. As uncertainty falls, the expected informational value of acquiring another sample eventually becomes too small to justify its cognitive cost, giving rise to endogenous stopping.

2.2 Accumulated evidence and choice discriminability

Here, I illustrate exogenous effects of evidence accumulation. This serves to put some key intuitions in place, before moving to endogenous attention allocation.

Trading off inferred quantities. Having formed posterior beliefs about the log-odds and the log cost-benefit ratio, the decision maker compares the two inferred quantities. The posterior comparison favours the lottery whenever its inferred log-odds exceed its inferred log cost-benefit ratio:

$$\mathbb{E}\left[\log \frac{p}{1-p} \middle| \mathbf{r}_{p,n}\right] > \mathbb{E}\left[\log \frac{c-y}{x-c} \middle| \mathbf{r}_{o,m}\right].$$

Substituting the posterior expectations and rearranging, the posterior comparison favours the lottery whenever

$$\gamma_n \bar{r}_{p,n} - \beta_m \bar{r}_{o,m} > \theta_{n,m}, \tag{1}$$

where

$$\gamma_n \triangleq \frac{\sigma_p^2}{\sigma_p^2 + \bar{\nu}_{p,n}^2}, \quad \beta_m \triangleq \frac{\sigma_o^2}{\sigma_o^2 + \bar{\nu}_{o,m}^2}.$$

The quantities γ_n and β_m are Bayesian evidence weights after n probability samples and m cost-benefit samples, respectively. They determine how strongly accumulated evidence

influences posterior inference relative to prior expectations. Because both weights increase as evidence accumulates, posterior beliefs place progressively greater weight on sampled information. The term $\theta_{n,m}$ represents a prior-induced evidentiary threshold that must be exceeded before the lottery is chosen. Its interpretation is developed below. Equation 1 determines which option is currently favoured on the basis of the inferred quantities. Whether this comparison also supports a reliable decision depends, however, on the uncertainty surrounding those inferences. The implications of this distinction are easiest to understand by first considering the special case of neutral prior expectations.

Evidence accumulation under neutral priors. Suppose that the prior means are neutral, with $p_0 = .5$ and $k_0 = b_0$, so that the prior log-odds as well as the prior log cost-benefit ratio are both zero and the prior-induced threshold vanishes: $\theta_{n,m} = 0$. This special case isolates the effect of noisy evidence accumulation on inference and choice.

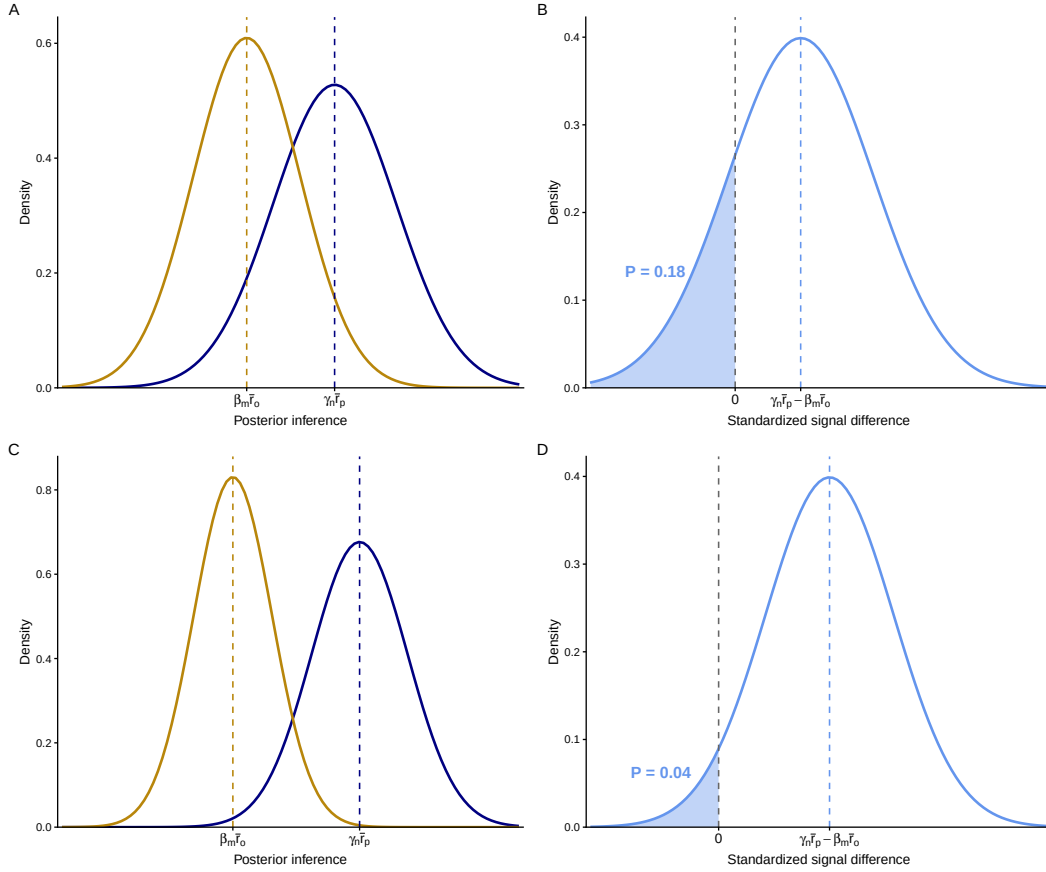


Figure 1: Discriminability between options as a function of accumulated evidence.

Panels A and C illustrate how posterior inference changes as evidence accumulates. With only limited evidence (Panel A), the posterior means remain relatively close because the

evidence weights γ_n and β_m are still well below one, while posterior uncertainty remains substantial. The log-odds posterior is wider than the log cost-benefit posterior because probability signals are assumed to be noisier ($\nu_p > \nu_o$).⁷ After additional samples have been acquired (Panel C), the posterior means move farther apart as γ_n and β_m increase, while both posterior distributions become narrower as posterior uncertainty declines. Together, these changes make the inferred quantities increasingly distinct.

Panels B and D show the implications for choice. With limited evidence (Panel B), discriminability between the two options remains low, resulting in an 18% probability of choosing the inferior sure amount despite the lottery's higher expected value. After additional evidence has been accumulated (Panel D), the standardized evidence difference shifts to the right and becomes more concentrated, reducing this probability to 4%. The superior option is therefore chosen more consistently as evidence accumulates.

The changes illustrated in Figure 1 can be summarized by the discriminability of the two options. Assuming that the posterior quantities are independent, the posterior variance of their difference is given by the sum of the two posterior variances, $s_{p,n}^2 + s_{o,m}^2$. Discriminability therefore takes the form

$$d'_{n,m} = \frac{\gamma_n \bar{r}_{p,n} - \beta_m \bar{r}_{o,m} - \theta_{n,m}}{\sqrt{s_{p,n}^2 + s_{o,m}^2}}. \quad (2)$$

The numerator of Equation 2 determines which option is currently favoured, while the denominator measures the uncertainty surrounding that comparison. Discriminability therefore increases when the inferred quantities move farther apart, when posterior uncertainty declines, or when both changes occur simultaneously.

Under the normal posterior, the probability of the lottery being chosen is

$$P[(x, p; y) \succ c \mid \mathbf{r}_{p,n}, \mathbf{r}_{o,m}] = \Phi(d'_{n,m}), \quad (3)$$

where Φ denotes the standard normal cumulative distribution function. Although Equation 3 has the mathematical form of a probit choice rule, its interpretation differs funda-

⁷This captures the intuition that precise numerical probabilities may be less familiar and therefore harder to represent than monetary outcomes, which people encounter routinely in everyday life. Described probabilities, by contrast, may constitute a relatively unfamiliar representational format for many decision makers (Gigerenzer and Hoffrage, 1995; Hertwig et al., 2004).

mentally from that of a conventional random utility model. Stochasticity does not arise from an additive error term appended to final utility, but instead reflects posterior uncertainty about the quantities on which choice depends. The direction and variability of choice are therefore jointly determined by the same process of noisy inference. Risk-taking and stochasticity are consequently intimately linked.

Inference under cautious priors. The neutral-prior case isolates the role of evidence accumulation. Prior expectations may also systematically favour one option over the other. In full generality, the prior-induced threshold takes the following form:

$$\theta_{n,m} \triangleq (1 - \beta_m) \log \frac{k_0}{b_0} - (1 - \gamma_n) \log \frac{p_0}{1 - p_0}.$$

Cautious priors can thus be represented by $p_0 < .5$, which assigns prior probability mass to relatively unfavourable odds; and by $k_0 > b_0$, which indicates that costs are expected to exceed benefits. Together, these prior means generate a positive threshold $\theta_{n,m}$, so that stronger evidence is required before the lottery is favoured.

Figure 2 repeats the inference process shown in Panels C and D of Figure 1, using the same relatively large evidence base but replacing the neutral prior means with cautious ones. Panel A shows the posterior distributions under cautious priors as solid lines and the corresponding distributions under neutral priors as dotted lines. Relative to the neutral-prior case, posterior inference about the probability shifts to the left (since $p_0 < 0.5$), while posterior inference about the cost-benefit ratio shifts to the right (since $k_0 > b_0$). The two inferred quantities are therefore closer together for any fixed amount of evidence. Their widths remain unchanged because the evidence bases are held fixed; changing the prior means shifts posterior location but not posterior precision.

Panel B shows the implication for stochastic choice. Under cautious priors, the comparison is made against the positive threshold $\theta_{n,m}$ rather than against zero. Equivalently, the threshold reduces the posterior mean difference in the numerator of Equation 2. Because the posterior variances are unchanged, discriminability is consequently lower. In the example, the probability of choosing the inferior sure amount increases from approximately 4% under neutral priors to approximately 18% under cautious priors.

Cautious prior expectations therefore require more evidence to reach any given level of

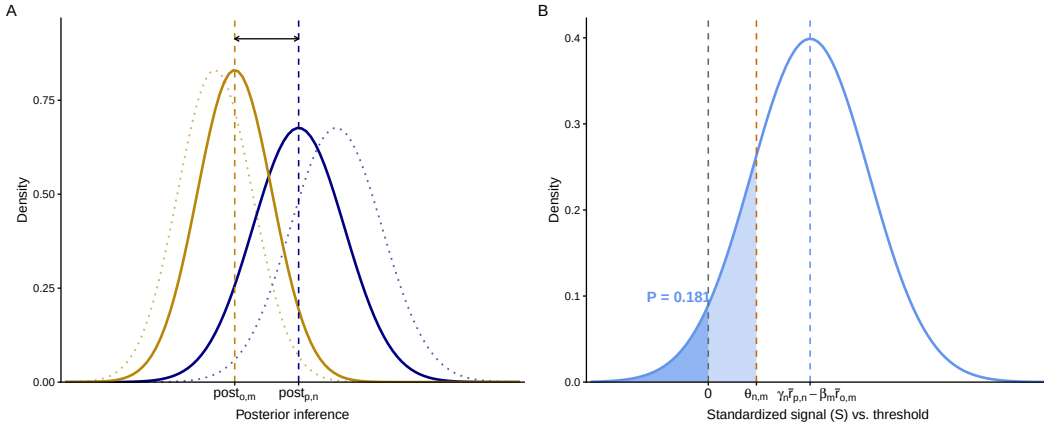


Figure 2: Discriminability between options under neutral and cautious prior means.

discriminability. This can prevent premature lottery choice when evidence is sparse or noisy, but it can also delay choice of a lottery that is in fact superior. Because prior means enter the comparison with weights $1 - \gamma_n$ and $1 - \beta_m$, their influence declines as evidence accumulates. That is, $\gamma_n, \beta_m \rightarrow 1$ implies $\theta_{n,m} \rightarrow 0$. Prior-induced caution is therefore strongest when information is sparse or unreliable.

Prior variances affect inference in a different way. Whereas prior means shift the evidentiary threshold, prior variances determine both how strongly sampled evidence is weighted and how much posterior uncertainty remains after any given amount of evidence has been accumulated. Their effect on discriminability therefore operates through both the numerator and denominator of Equation 2. Once information acquisition becomes endogenous, prior variances acquire an additional role by affecting the value of further evidence.

2.3 A generative model of risk-taking

The discriminability patterns illustrated above generate average risk-taking patterns that resemble those usually described by prospect-theory functionals. Here, I show that the model can produce stylized choice patterns such as risk aversion increasing in the probability of winning (captured by probability weighting in PT) and choices of the sure amount even in cases when the lottery is objectively superior (captured by utility curvature in PT), without treating these functionals as taste-based primitives.

Coding noise and probability distortions. Under fixed and equal sampling, the canonical inverse-S probability-weighting pattern documented in PT—i.e. risk aversion increasing in the probability of winning a prize—obtains if and only if probability infor-

mation is represented less precisely relative to prior dispersion than outcome information, $\nu_p/\sigma_p > \nu_o/\sigma_o$. I derive this result under the benchmark $n = m$, which isolates the role of relative discriminability. As shown below, this simple characterization no longer holds once sampling is chosen endogenously: differences in prior uncertainty can alter attention allocation sufficiently to reinforce or even overturn the effect of relative coding precision.

To see how the model generates probability distortions, replace the accumulated signals by their means. At the point of indifference, the numerator of Equation 2 implies

$$\ln \left(\frac{c-y}{x-c} \right) = \frac{\gamma_n}{\beta_m} \ln \left(\frac{p}{1-p} \right) - \frac{1}{\beta_m} \theta_{n,m}.$$

Exponentiating gives the odds form

$$\frac{c-y}{x-c} = \delta_{n,m} \left(\frac{p}{1-p} \right)^{\gamma_n/\beta_m}, \text{ where } \delta_{n,m} \triangleq \exp \left(-\frac{\theta_{n,m}}{\beta_m} \right).$$

Equivalently, defining the normalized certainty-equivalent weight by $w(p) \triangleq (c-y)/(x-y)$, so that $w(p)/(1-w(p)) = (c-y)/(x-c)$, we obtain the following as-if decision-weighting function generated by the model:

$$w(p) = \frac{\delta_{n,m} p^{\gamma_n/\beta_m}}{\delta_{n,m} p^{\gamma_n/\beta_m} + (1-p)^{\gamma_n/\beta_m}}.$$

Substituting for $\theta_{n,m}$ gives

$$\delta_{n,m} = \left(\frac{p_0}{1-p_0} \right)^{(1-\gamma_n)/\beta_m} \left(\frac{k_0}{b_0} \right)^{(\beta_m-1)/\beta_m}.$$

Formally, this expression has the linear-in-log-odds form of a popular probability-weighting function used in PT (Goldstein and Einhorn, 1987; Gonzalez and Wu, 1999; Zhang and Maloney, 2012). Its interpretation is different, however. Rather than representing a primitive transformation of probabilities, the function describes how risk taking varies jointly with probability and outcome information. The ratio γ_n/β_m governs relative probability sensitivity, with values $\gamma_n/\beta_m < 1$ generating probability insensitivity, while $\delta_{n,m}$ governs the overall elevation of risk taking. Both parameters arise from the underlying inference process and can therefore change when either probability or outcome representations

change. The familiar probability-weighting shape thus emerges here as an implication of noisy inference rather than as a primitive taste-based distortion of probabilities.

Panel A in Figure 3 plots the implied pattern of risk taking across the probability interval. When probability information is represented less precisely than cost-benefit information relative to prior dispersion, $\nu_p/\sigma_p > \nu_o/\sigma_o$, risk taking follows the familiar inverse-S pattern known from PT estimations because $\gamma_n/\beta_m < 1$. When the two sources of information are represented with equal precision relative to prior dispersion, $\nu_p/\sigma_p = \nu_o/\sigma_o$, the resulting pattern is uniformly convex, with the downward shift relative to the 45-degree line reflecting the influence of cautious priors. When cost-benefit information is represented less precisely than probability information, $\nu_p/\sigma_p < \nu_o/\sigma_o$, the pattern instead becomes S-shaped. Thus, the shape of risk taking across probabilities is governed by the relative discriminability of probability and cost-benefit information, rather than by a stable probability-weighting primitive grounded in tastes for risk.

Outcome coding and risk aversion. Panel B complements the probability-dependent patterns by showing how average risk taking changes as cost-benefit discriminability declines, plotted as a function of $1 - \beta_m$. As cost-benefit information becomes less reliable, choices are informed increasingly by prior expectations. In other words, the weight $1 - \beta_m$ placed on the prior mean increases, which for a cautious or pessimistic prior expectation $k_0/b_0 > 1$ yields monotonic increases in risk aversion. Thus, the same Bayesian updating mechanism that generates probability-dependent patterns also generates variation in average risk taking: reduced cost-benefit discriminability increases reliance on cautious cost-benefit priors. Consequently, two decision makers with identical prior beliefs but different levels of outcome coding noise can display different levels of average risk aversion and would therefore be described by different “as-if utility curvature” in a conventional preference model. The prediction for average choice behaviour is therefore intrinsically linked to stochasticity in choice (see [Vieider, 2025](#), for an empirical characterization).

More generally, the mapping from the sequential noisy sampling model (*SNSM*) into risk taking is determined jointly by prior means and posterior evidence weights. The relevant objects are the prior probability odds $p_0/(1 - p_0)$ and the prior cost-benefit ratio k_0/b_0 , weighted respectively by $(1 - \gamma_n)/\beta_m$ and $(\beta_m - 1)/\beta_m$. Neither γ_n nor β_m is itself a probability-weighting or utility-curvature parameter. Rather, they govern the extent to which inferred attributes are pulled toward their prior means. The parameter

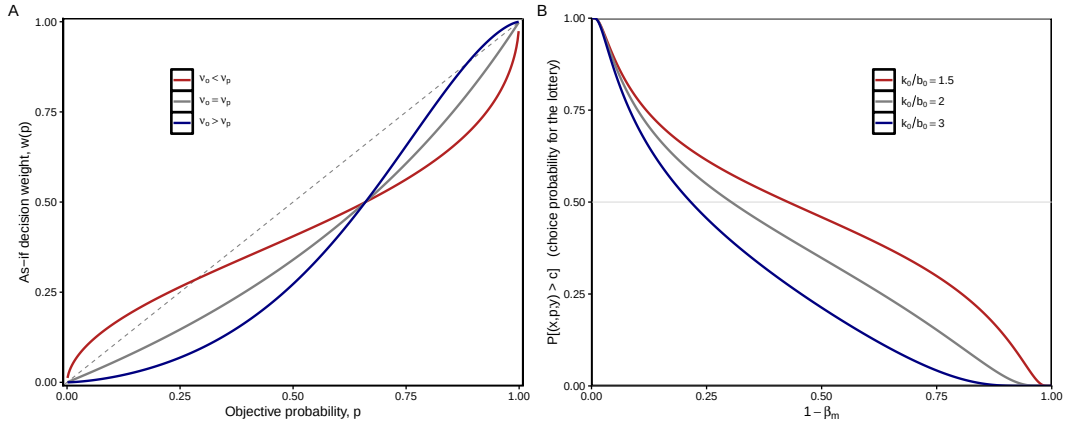


Figure 3: Prospect-theory-like choice patterns generated by noisy Bayesian inference

γ_n therefore determines sensitivity to probability information and the influence of prior probability beliefs, whereas β_m determines sensitivity to cost-benefit information and the influence of prior cost-benefit beliefs. When these weights vary systematically with the underlying attributes, the resulting prior-induced shifts in risk taking can appear, under a conventional preference-based interpretation, as probability-weighting or utility-curvature effects. The next section examines how such variation arises when sample acquisition is determined endogenously.

Together, Panels A and B therefore show how PT-like average choice patterns can emerge as descriptive consequences of Bayesian noisy inference rather than from primitive preference distortions. Importantly, the same inference process also determines stochastic choice probabilities. Average risk taking and stochasticity are therefore generated jointly, avoiding problems that can arise when deterministic preferences and stochastic choice are specified separately (Wilcox, 2011; Bhatia and Loomes, 2017; Apesteguia and Ballester, 2018). Conditional on a fixed number of samples, reliable evidence receives substantial weight and choices closely track the sampled signals. As representational noise increases, the evidence is discounted more strongly and choices are pulled toward prior expectations. High noise therefore need not imply convergence toward random choice: with non-neutral priors, prior beliefs can stabilize behaviour precisely when current evidence becomes unreliable (Vieider, 2025).

3 Endogenous stopping and attention allocation

I have so far treated the numbers of probability and cost-benefit samples, n and m , as exogenous. I now allow the decision maker to determine both how much information to acquire and which choice quantity to sample. Sampling continues when the expected reduction in choice error exceeds its cost, and attention is directed toward the attribute for which an additional sample has the greatest net value.⁸

3.1 Error reduction, attention allocation, and stopping

Expected error reduction. A key quantity is the expected reduction in the probability of making a mistake produced by one additional sample. This criterion has a natural precedent in evolutionary accounts of decision making. Robson (2001b) motivates approximate utility representations by their ability to reduce the probability of a mistake when discrimination is imperfect. I adopt this probability-of-error criterion here because it closely tracks the logic of the static noisy-cognition models on which I build.⁹

Let $q_{n,m} \triangleq \Phi(d'_{n,m})$ denote the posterior probability that the lottery is superior conditional on the current evidence. Following the posterior-draw choice rule inherited from the static noisy-cognition framework, this same probability governs terminal choice: the lottery is chosen with probability $q_{n,m}$ and the sure amount with probability $1 - q_{n,m}$. Thus, even conditional on an identical evidence history, residual posterior uncertainty continues to generate stochasticity in choice. A mistake occurs either when the lottery is chosen although the sure amount is superior, or when the sure amount is chosen although the lottery is superior. Under the posterior-draw rule, each event has posterior probability $q_{n,m}(1 - q_{n,m})$, so that the total posterior probability of making a mistake is

$$M_{n,m} \triangleq 2q_{n,m}(1 - q_{n,m}). \quad (4)$$

Although this expression is symmetric in $q_{n,m}$, the underlying posterior need not favour

⁸The model is related to the multi-Attribute Search and Choice model of Gluth et al. (2026), which also represents option attributes as uncertain beliefs updated through sequential search. The approaches differ mainly in how search and termination are determined. The multi-Attribute Search and Choice model uses an exogenously parameterized posterior-overlap criterion for stopping. Here, attention allocation and stopping are jointly determined by the expected reduction in choice error relative to sampling costs.

⁹Netzer (2009) instead emphasizes the consequences of errors, so that mistakes matter in proportion to the fitness losses they generate. Incorporating this logic would require weighting errors by their economic consequences rather than merely by whether the inferior option is chosen. I leave this extension aside here and retain the scale-invariant architecture of Khaw et al. (2021), under which the magnitude of the mistake does not enter the stopping criterion.

the two options symmetrically. Prior expectations, signal noise, and accumulated evidence may affect the inferred log-odds and log cost-benefit ratio differently, shifting $q_{n,m}$ towards one option or the other.

The posterior-draw rule has a close analogue in Thompson sampling (Thompson, 1933; Russo et al., 2018), where actions are selected according to their posterior probability of being optimal. In changing environments, such posterior sampling can serve an adaptive exploration function: actions that are currently inferred to be inferior may still be chosen when posterior uncertainty leaves a non-negligible probability that they are in fact superior, thereby generating informative feedback. This provides an adaptive rationale for retaining stochastic choice when internal evidence remains inconclusive.¹⁰ I do not model this feedback-based learning process here. I therefore retain the posterior-draw rule as a consequence of the static choice setup on which I build, without claiming any optimality.¹¹

Information acquisition and stopping. Conditional on the current information set $\mathcal{F}_{n,m}$, the decision maker can stop, sample one additional probability signal, or sample one additional cost-benefit signal. Let $a \in \{p, o\}$ denote the sampled quantity, with $(n', m') = (n + 1, m)$ following a probability sample and $(n', m') = (n, m + 1)$ following a cost-benefit sample. The gross value of action a is the expected reduction in the posterior probability of making a mistake:

$$\begin{aligned} \Delta M_a(n, m) &\triangleq M_{n,m} - \mathbb{E}[M_{n',m'} \mid \mathcal{F}_{n,m}] \\ &= 2\mathbb{V}[q_{n',m'} \mid \mathcal{F}_{n,m}], \end{aligned} \tag{5}$$

where the second equality follows from the martingale property of posterior beliefs. Conditional on the current information set, the next posterior probability may move up or down

¹⁰The adaptive value of such stochastic choice is particularly transparent when internal representations remain very noisy. Suppose, for instance, that probability information is sampled with very high noise and that a cautious prior consequently leaves the sure amount favoured at stopping even when the lottery is in fact substantially superior. A deterministic terminal rule would then repeatedly select the sure amount. Posterior sampling instead preserves a positive probability of choosing the lottery whenever residual uncertainty leaves some probability that it is superior. These choices, in turn, can generate information that eventually corrects the prior-induced ranking.

¹¹An alternative would be to make terminal choice deterministic conditional on the accumulated evidence, selecting the posterior-best option at stopping. The Bayesian inference architecture would remain unchanged, but the mapping from posterior uncertainty into terminal error would then involve a max/min rule rather than the posterior-draw mapping used here. This would in turn alter the value assigned to additional information and potentially the resulting stopping and attention policies. Characterizing this alternative lies beyond the present analysis.

after an additional signal, but its expected value remains equal to the current posterior probability: $\mathbb{E}[q_{n',m'} \mid \mathcal{F}_{n,m}] = q_{n,m}$. The expected reduction in error therefore comes not from a systematic expected shift in the posterior mean, but from the dispersion of possible posterior updates. This dispersion is exactly $\mathbb{V}[q_{n',m'} \mid \mathcal{F}_{n,m}]$.¹² The sampling value of action a is the expected reduction in the probability of a mistake, net of the cognitive cost of acquiring the additional signal:

$$SV_a(n, m) = \Delta M_a(n, m) - \kappa, \quad (6)$$

where κ denotes the cognitive cost of acquiring one additional signal.¹³ Additional sampling is worthwhile only when the expected reduction in choice errors exceeds this cost. The rule is myopic in the dynamic-programming sense: it evaluates the expected benefit of one additional sample relative to stopping thereafter, without recursively optimizing over subsequent sampling opportunities.

A key difference from existing optimal-sampling models concerns terminal choice. In models such as [Tajima et al. \(2016\)](#) and [Fudenberg et al. \(2018\)](#), the accumulated evidence determines the selected option once sampling stops, whereas in the SNSM residual

¹²Using $M_{n,m} = 2q_{n,m}(1 - q_{n,m})$, the expected reduction in error is

$$\Delta M_a(n, m) = 2q_{n,m}(1 - q_{n,m}) - 2\mathbb{E}[q_{n',m'}(1 - q_{n',m'}) \mid \mathcal{F}_{n,m}].$$

Expanding both products gives

$$\begin{aligned} \Delta M_a(n, m) &= 2q_{n,m} - 2q_{n,m}^2 \\ &\quad - 2\mathbb{E}[q_{n',m'} \mid \mathcal{F}_{n,m}] + 2\mathbb{E}[q_{n',m'}^2 \mid \mathcal{F}_{n,m}]. \end{aligned}$$

The martingale property means that, conditional on the information currently available, the expected posterior probability after one further signal is equal to the current posterior probability: $\mathbb{E}[q_{n',m'} \mid \mathcal{F}_{n,m}] = q_{n,m}$. The posterior may therefore move up or down after the next signal, but there is no expected directional change in its level. Substituting this equality causes the two first-moment terms, $2q_{n,m}$ and $-2\mathbb{E}[q_{n',m'} \mid \mathcal{F}_{n,m}]$, to cancel, leaving

$$\Delta M_a(n, m) = 2(\mathbb{E}[q_{n',m'}^2 \mid \mathcal{F}_{n,m}] - q_{n,m}^2).$$

The martingale property does not imply that $\mathbb{E}[q_{n',m'}^2 \mid \mathcal{F}_{n,m}] = q_{n,m}^2$: although the expected posterior level is unchanged, the posterior itself varies across possible future signals. Using $q_{n,m}^2 = (\mathbb{E}[q_{n',m'} \mid \mathcal{F}_{n,m}])^2$, the remaining difference is therefore exactly the conditional variance of the future posterior,

$$\mathbb{V}[q_{n',m'} \mid \mathcal{F}_{n,m}] = \mathbb{E}[q_{n',m'}^2 \mid \mathcal{F}_{n,m}] - q_{n,m}^2.$$

Hence,

$$\Delta M_a(n, m) = 2\mathbb{V}[q_{n',m'} \mid \mathcal{F}_{n,m}].$$

¹³The qualitative results do not depend on a particular functional form for sampling costs. In the simulations below, I assume constant per-sample costs for simplicity. Increasing marginal costs are also plausible. In that case, unusually intensive sampling of one attribute raises the marginal cost of further information acquisition more generally, so that attention drawn disproportionately toward one channel tends to be accompanied by relatively less sampling of the other.

posterior uncertainty continues to affect choice even conditional on an identical evidence history. This same attribute-specific posterior uncertainty also determines where an additional sample is most valuable, providing the link from stopping to endogenous attention allocation. I turn to this next.

Goal-directed attention. The sampling values defined above determine not only whether sampling continues, but also which quantity is sampled conditional on continued information acquisition. The problem is therefore one of sequentially choosing the most valuable experiment (Chernoff, 1959): at each point, the decision maker can sample probability information, sample cost-benefit information, or stop.

Let $a \in \{p, o, s\}$ denote the available actions, where p corresponds to sampling an additional probability signal, o to sampling an additional cost-benefit signal, and s to stopping. The sampling values $SV_p(n, m)$ and $SV_o(n, m)$ are defined as in Equation 6. The value of stopping is normalized to $SV_s(n, m) = 0$. The resulting action is then

$$a^*(n, m) = \arg \max_{a \in \{p, o, s\}} SV_a(n, m). \quad (7)$$

Thus, probability information is sampled whenever $SV_p(n, m) > \max\{SV_o(n, m), 0\}$, cost-benefit information is sampled whenever $SV_o(n, m) > \max\{SV_p(n, m), 0\}$, and sampling stops whenever $\max\{SV_p(n, m), SV_o(n, m)\} \leq 0$.

This rule makes attention endogenous: attention is allocated to the quantity for which an additional sample has the highest expected value. Attributes therefore attract attention when resolving their remaining uncertainty is expected to improve the decision.¹⁴

3.2 Two types of uncertainty and relative attention

Uncertainty about each choice attribute enters the model in one of two ways, both of which affect attention in different ways.

Relative coding noise, attention, and probability sensitivity. I begin by characterizing the comparative statics that arise when coding noise ν_p and ν_o change while prior variances are held fixed. To obtain a clean benchmark, let $\sigma_p = \sigma_o = 1$. Although the

¹⁴Although attention is allocated by attribute in the model, this can induce systematic patterns in option-level attention in empirical data. Attribute-wise and option-wise attention can therefore capture complementary dimensions of the information-acquisition process, as emphasized by Zilker and Pachur (2023).

absolute level of prior uncertainty will turn out to matter in the more general case, this normalization isolates the role of relative coding noise.

Under this benchmark, the ordering of coding noise at the single-sample level continues to determine relative sensitivity at the endogenous stopping point. Conditional on equal prior variances, $\nu_p/\sigma_p > \nu_o/\sigma_o$ is necessary and sufficient for insensitivity to probabilities, $\gamma^*/\beta^* < 1$. Endogenous sampling on its own does not overturn this result, as the simulations in Figure 4 illustrate. Panel A establishes the key result for attention allocation. In the symmetric benchmark, where probability and cost-benefit information are equally noisy, attention is allocated symmetrically and $n^* = m^*$. When coding noise for probabilities exceeds coding noise for outcomes, $\nu_p/\sigma_p > \nu_o/\sigma_o$, the probability channel attracts more—not less—attention, so that $n^*/m^* > 1$. This establishes a central testable prediction: making an information channel noisier increases the attention allocated to it.¹⁵

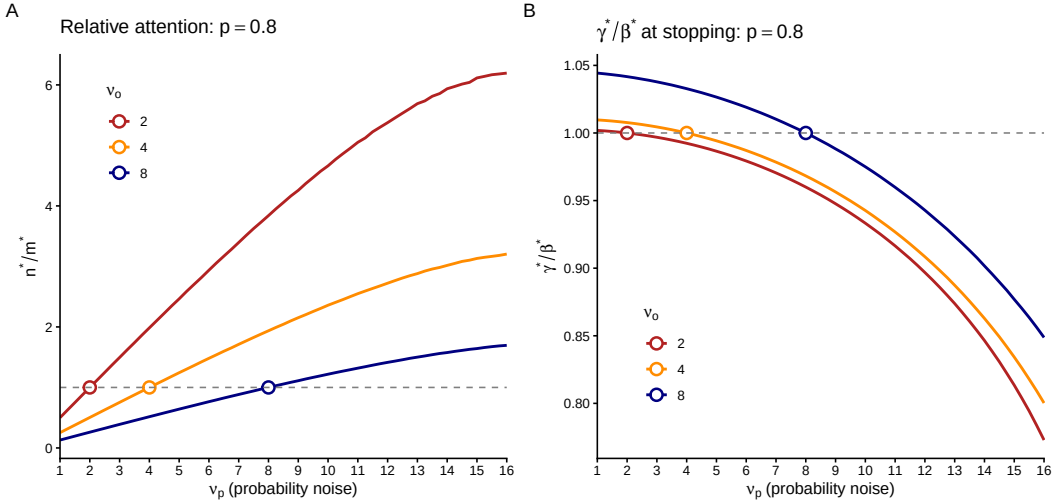


Figure 4: Relative coding noise, attention allocation, and probability sensitivity

Panel B illustrates the central implication for choice. Even though the noisier probability channel attracts additional attention, the resulting increase in sampling is insufficient to eliminate its initial precision disadvantage. The probability-sensitivity ratio γ_n^*/β_m^* therefore *decreases* as relative attention to probabilities n^*/m^* increases. In this sense, increased attention allocation when a channel increases in noisiness is compensatory but

¹⁵This comparative static need not hold globally. If coding noise becomes sufficiently large, an additional sample may convey so little information that its sampling value eventually falls, causing attention to the corresponding channel to decline. The predicted relation between coding noise and attention can therefore be nonmonotonic at extreme noise levels.

incomplete—the increased samples do not fully make up for the reduction in signal fidelity. This yields the corresponding testable prediction for choice behaviour: as probability coding noise increases, probability sensitivity is predicted to decrease.

Taken together, these findings generate a seemingly paradoxical prediction: greater attention to probabilities can be accompanied by lower sensitivity to them in choice. This prediction can be tested cleanly by increasing probability noise exogenously and jointly examining the resulting changes in attention allocation and choice behaviour. At face value, it appears to contradict the positive correlation between relative attention to probabilities and probability sensitivity documented by [Pachur et al. \(2018\)](#). As I show next, however, the contradiction is only apparent and is itself anticipated by the model.

Prior variances modulate information search. So far, I have largely abstracted from the parameters describing the prior by fixing them at normalized values. Prior variances, however, play a crucial role in shaping attention allocation and choice at the individual level.¹⁶

The key comparative statics are easiest to illustrate through simulation. Let $\nu_o/\sigma_o = 2$. Panel A in Figure 5 shows attention allocation for fixed ratios $\nu_p/\sigma_p = \{4, 8, 12\}$, while multiplying both ν_p and σ_p by a common factor v . This leaves relative coding noise unchanged but increases their absolute scale. As v , and hence σ_p , increases, progressively more attention is allocated to the probability channel. Thus, across individuals with different cognitive primitives ν_p and σ_p , relative attention to probabilities can increase even while ν_p/σ_p remains constant. At the same time, for any given value of v , greater relative probability noise continues to increase attention to the probability channel.

Panel B shows the complementary implications for choice. Individuals with a larger multiplier v —and hence higher ν_p and σ_p relative to ν_o and σ_o —are more sensitive to probabilities. At sufficiently high values, this can generate probability *over*-sensitivity, corresponding to S-shaped probability weighting. The mechanism is interesting in its own right. As v increases, attention shifts toward probability information, so that n^*/m^* rises. The resulting increase in γ^* relative to β^* raises γ^*/β^* . Thus, greater relative attention to probabilities is accompanied by greater probability sensitivity.

¹⁶Prior variances are held fixed within the experiment for parsimony and to isolate comparative statics across their initial levels. Allowing them to evolve would require additional learning dynamics, and is left for future work.

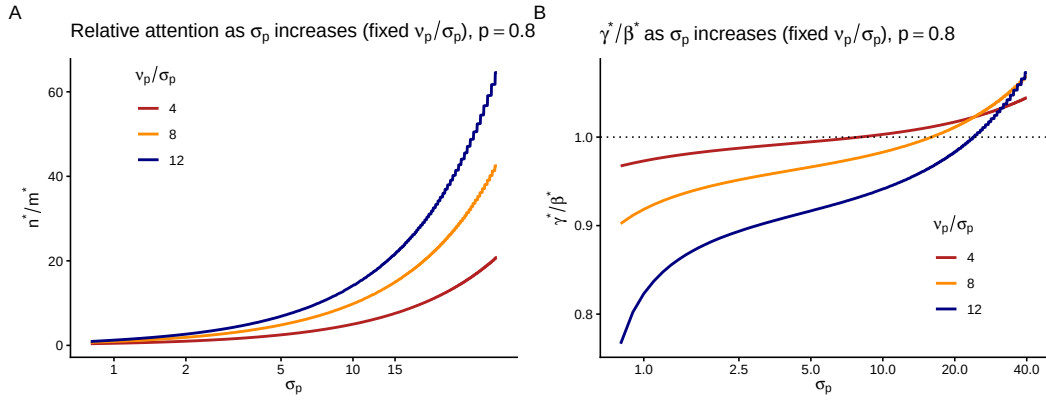


Figure 5: Prior dispersion, attention allocation, and probability sensitivity

This contrasts sharply with an exogenous increase in ν_p/σ_p . In that case, n^* increases, but the compensatory increase in probability sampling is insufficient to offset the deterioration in signal fidelity. Hence, even though relative attention to probabilities increases, γ^*/β^* falls. The contrast has a simple intuition. Coding noise and prior variance enter posterior uncertainty in closely related ways, but they affect the value of additional samples differently. Greater coding noise makes the existing representation less precise while simultaneously reducing the diagnostic value of each subsequent sample, producing compensatory but incomplete sampling. Greater prior variance, by contrast, raises the value of acquiring information without degrading the fidelity of the signals themselves. Additional samples therefore remain highly informative, allowing increased sampling to translate into greater behavioural sensitivity.

Individuals who are less certain about probability information may have both greater coding noise and more diffuse priors over probabilities. The latter generates an adaptive response: greater prior uncertainty raises the value of resolving probability information and therefore induces more intensive sampling.¹⁷ When coding noise and prior dispersion increase proportionally, so that ν_p/σ_p remains stable, individuals with greater uncertainty will allocate more attention to probabilities and may consequently achieve greater probability sensitivity. Prat-Carrabin and Woodford (2026) provide evidence that coding noise increases sublinearly with prior dispersion. In that case, increases in σ_p are accompanied by declines in ν_p/σ_p , which may further reinforce the positive between-person association between attention to probabilities and probability sensitivity.

¹⁷Liang et al. (2022) likewise study dynamic attention allocation across multiple information sources, but in a setting where attention acquires external information about unknown quantities rather than improving noisy internal representations of given choice information.

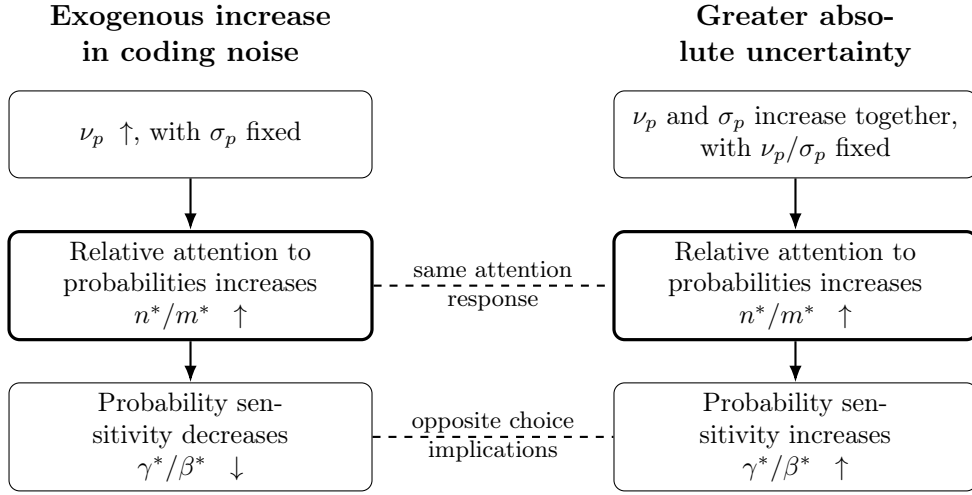


Figure 6: Different sources of increased attention to probabilities. An exogenous increase in probability coding noise and an increase in absolute uncertainty both raise relative attention to probabilities, but have opposite implications for probability sensitivity.

An important anchor for these results is the fully symmetric case. When the two channels have equal prior dispersion and equal coding noise, $\sigma_p = \sigma_o$ and $\nu_p = \nu_o$, endogenous sampling is symmetric and $\gamma^*/\beta^* = 1$. Departures from this benchmark can then be generated while holding the coding-noise-to-prior-dispersion ratio constant within each channel. Increasing prior dispersion σ_p in the probability channel, together with a proportional increase in ν_p , raises the value of probability sampling and increases γ^*/β^* . Conversely, increasing σ_o together with a proportional increase in ν_o shifts sampling toward outcome information and lowers γ^*/β^* . Thus, once prior dispersion is allowed to differ across channels, relative coding noise alone no longer determines relative sensitivity at the endogenous stopping point. Sufficiently large prior uncertainty can attract enough compensatory sampling to offset, and eventually overturn, a disadvantage in relative coding precision.

The model thus accommodates the positive association between relative attention to probabilities and probability sensitivity documented by [Pachur et al. \(2018\)](#). What matters is therefore not only the total amount of information acquired, but also how attention is distributed across the constituent attributes of the choice problem. This parallels evidence that attentional imbalance across dimensions carries information beyond overall search intensity ([Zilker and Pachur, 2023](#)). At the same time, the model predicts the opposite association when probability noise is increased exogenously: attention to probabilities rises while probability sensitivity falls. Figure 6 summarizes this contrast. The same increase in

relative attention can therefore reflect different underlying cognitive primitives and have opposite implications for choice. This apparent paradox provides a particularly stringent empirical test of the mechanism proposed here.

3.3 Cautious prior means direct search and regularize inference.

So far, the analysis has focused on comparative statics describing differences across individuals or experimental conditions rather than variation across tasks. Prior means can, however, generate systematic task-level variation in attention allocation and choice. This provides a further illustration of the adaptive role of prior expectations: prior means not only regularize noisy inference, but also direct information search toward those attributes for which current evidence departs most consequentially from prior expectations.

Prior information, sampling, and choice. Figure 7 shows how prior means affect sampling and choice. Panel A plots the evolution of lottery choice probabilities over the sampling process for a neutral probability prior, $p_0 = 0.5$ (solid lines), and a cautious prior, $p_0 = 0.25$ (dashed lines). Costs and benefits are kept symmetric, so that the lottery is attractive when the winning probability is high, $p = 0.8$, and unattractive when the winning probability is low, $p = 0.2$. The cost-benefit prior is neutral, $b_0 = k_0$, so the figure isolates the role of probability priors. Under the neutral prior, the problem is symmetric across the two probabilities. As evidence accumulates, the lottery is chosen increasingly reliably when $p = 0.8$, and rejected with similar reliability when $p = 0.2$. The cautious prior breaks this symmetry. It shifts the posterior comparison against the lottery, requiring more evidence to favour the lottery when $p = 0.8$, while accelerating rejection when $p = 0.2$. Prior means therefore act as evidentiary burdens, determining how much evidence must be accumulated before the posterior comparison moves decisively in one direction. In doing so, they also shape where additional sampling is most valuable.

The asymmetric prior produces two distinct patterns. At $p = 0.2$, sampled evidence largely confirms the prior, so sampling stops quickly and the sure option is chosen more consistently than under the neutral prior. At $p = 0.8$, prior expectations and sampled evidence conflict, making additional probability samples valuable because they reduce the weight placed on the pessimistic prior (see [Oprea and Vieider, 2024](#), for evidence of this in a decision-from-experience setting). Sampling therefore continues longer, but does not fully eliminate the prior’s influence. At stopping, the high-probability lottery is chosen

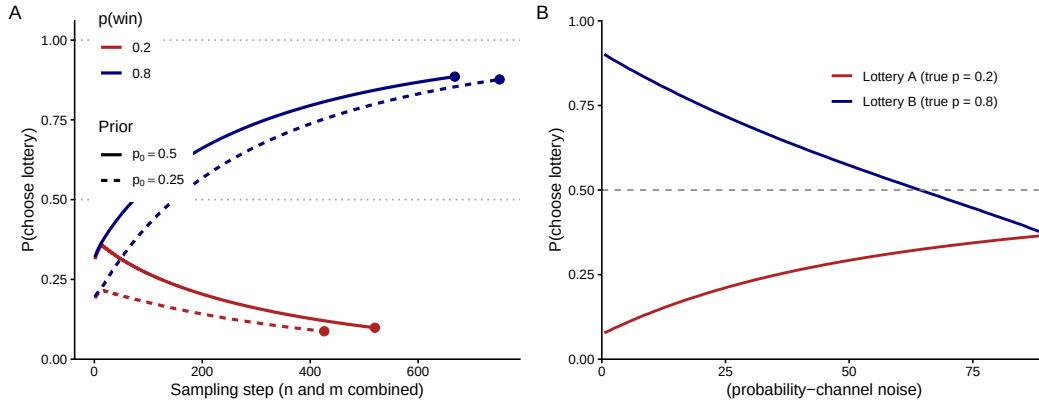


Figure 7: Prior means, information acquisition, and choice

slightly less often than under the neutral prior, and less reliably than the sure option in the corresponding low-probability case. The cautious prior can thus generate asymmetric information search and leave a residual degree of risk aversion when evidence remains sufficiently noisy.

Panel B depicts choice probabilities at stopping as a function of increasing log-odds noise, ν_p , for fixed outcome noise, ν_o , and $p_0 = 0.25$. At relatively low noise levels, the superior option is chosen fairly reliably: the lottery is chosen when $p = 0.8$, and the sure option when $p = 0.2$.¹⁸ For the low-probability lottery, lottery choice increases slightly as log-odds noise rises, but then stabilizes at a low level. Here, the prior is already close to the true probability, so little is gained from further sampling. For the high-probability lottery, by contrast, the prior and the true probability are far apart, initially making additional sampling particularly valuable. As probability signals become increasingly noisy, however, each additional sample becomes less diagnostic, and the marginal value of further sampling eventually declines. Beyond some level of noise, the acquired evidence is no longer reliable enough to overcome cautious prior expectations, and the superior lottery is chosen less than half the time. The prior therefore stabilizes choice as signals become highly unreliable: rather than converging toward random choice, behaviour becomes increasingly anchored by prior expectations (Vieider, 2025).

Taken together, these results show that prior means adaptively direct information search toward tasks in which current evidence conflicts with prior expectations, while allowing early stopping when the evidence confirms them. Prior expectations thus help economize

¹⁸Choice probabilities do not fully converge to 0 or 1 at low probability noise because of the fixed outcome noise assumed in the simulation.

on scarce attentional resources. When signals become sufficiently noisy, however, further sampling loses value and choice becomes increasingly guided by the prior. Cautious priors can therefore generate both asymmetric attention and residual risk aversion.

3.4 Relation to evidence-accumulation models.

The difference between standard evidence-accumulation models and the SNSM may seem surprising in light of their common connection to Wald’s (1945) sequential probability ratio test (*SPRT*) and to signal-detection theory (Green et al., 1966; Gold and Shadlen, 2001). A closer look at this connection clarifies both the similarities and the key departure.

A mapping between the SNSM and the SPRT. To see the connection to Wald, let H_0 denote the hypothesis that the sure option is superior to the lottery, $V_S \geq V_L$, and let H_1 denote the hypothesis that the lottery is superior, $V_L > V_S$. A type-I error occurs when the lottery is chosen although H_0 is true. A type-II error occurs when the sure option is chosen although H_1 is true. The SPRT accumulates evidence in the form of a likelihood ratio between these two hypotheses and stops once the evidence is strong enough to select one of them, given the desired tradeoff between the two errors. In the Wald–Wolfowitz (1948) formulation, this tradeoff depends on prior probabilities, accumulated evidence, and the relative losses associated with the two possible mistakes.

The SNSM has an analogous evidence-balance term. Let $N_{n,m}$ denote the numerator of the posterior discriminability index $d'_{n,m}$:

$$N_{n,m} = \gamma_n \bar{r}_{p,n} - \beta_m \bar{r}_{o,m} - \theta_{n,m}.$$

This term summarizes the current posterior balance of evidence in favour of the lottery over the sure option. In this sense, $N_{n,m}$ is analogous to the accumulated log-likelihood ratio in the SPRT. Priors shift the evidentiary burden through $\theta_{n,m}$, while accumulated samples shift the balance of evidence through the posterior-weighted probability and cost-benefit signals.

The correspondence is useful, but it is not exact. In the SPRT, the evidence variable is a likelihood ratio over external hypotheses and is sufficient for the accumulated evidence. In the SNSM, by contrast, the state of the decision process consists of uncertain internal representations of noisy choice quantities. The relative costs and benefits that determine

which option is superior are also not simply imposed from outside, but must themselves be inferred from noisy signals. Finally, sampling can be allocated selectively across distinct sources of uncertainty. The decision maker can acquire additional probability information or additional cost-benefit information, and these channels may differ in coding noise, prior uncertainty, and marginal sampling value.

Internal representation of posterior uncertainty. The most important departure is the denominator of d' . The SNSM does not base choice only on the signed evidence balance $N_{n,m}$, but on the discriminability of that balance:

$$d'_{n,m} = \frac{N_{n,m}}{\sqrt{s_{p,n}^2 + s_{o,m}^2}}.$$

The denominator has no direct counterpart as a separately represented posterior-uncertainty term in the classical SPRT or in standard DDM choice equations ([Drugowitsch and Pouget, 2012](#)). It represents the posterior uncertainty that remains after the current sequence of samples. This residual uncertainty is not merely noise around the evidence variable; it is part of the cognitive state on which choice is based. This is the key representational departure. Standard accumulation models typically assume that, conditional on the accumulated evidence reaching a given state, choice is determined by that state relative to an evidentiary threshold or boundary. In the SNSM, the same signed evidence balance can generate different choice probabilities depending on how uncertain the underlying cognitive representations remain.

Closely related approaches. An important precursor to the SNSM is Decision Field Theory ([Busemeyer and Townsend, 1993](#)). Decision Field Theory likewise constructs risky choice dynamically from repeated comparisons of the consequences associated with different events. The probabilities and outcome utilities entering these comparisons, however, are treated as given. Stochasticity arises from noisy momentary comparisons and from the sequence in which event-contingent comparative evaluations are sampled and accumulated over time. [Bhatia \(2014\)](#) extends this framework by allowing event sampling itself to be stochastic, with noise in attention allocation introduced through a reduced-form mixture process. This substantially expands the explanatory reach of the sequential framework.

The SNSM differs at the level of representation. Probabilities and comparative outcome values are not taken as known inputs into the sampling process, but must themselves be inferred from noisy internal signals. Residual uncertainty therefore concerns the underlying choice attributes themselves, rather than only the evolving preference balance generated by sampling across events. This also changes the role of attention. In Decision Field Theory and its extensions, attention governs which event enters the momentary comparison. In the SNSM, attention is allocated across choice attributes according to the expected value of reducing their remaining uncertainty. Probability and cost-benefit information may therefore attract different amounts of attention depending on their current precision and decision relevance. In principle, attribute-level information acquisition could coexist with sampling across events, but the two mechanisms are conceptually distinct.

Finally, the model bears some resemblance to attentional drift-diffusion models (*aDDMs*; [Krajbich et al., 2010](#)). In the aDDM, not only dwell time but also final fixation may be related to choice because the evidence accumulated during that fixation can push the decision variable across the choice boundary. A relation between final fixation and choice is also fully compatible with the SNSM, albeit for subtly different reasons. Because attention is endogenous, the final sampled attribute reflects posterior uncertainty immediately before stopping: an attribute may be sampled last because uncertainty about it remains the most decision-relevant unresolved quantity. Final fixation can therefore predict subsequent choice both because the final sample updates the evidence and because the posterior state that made this sample valuable may already favour the option ultimately chosen.

4 Experiment: Shocking probability noise

The pre-registered experiment tests a key qualitative prediction of the model: that making probability representations noisier should increase relative attention to probabilities while nevertheless reducing probability sensitivity.¹⁹ The seeming contradiction with the opposite between-subject correlation documented by [Pachur et al. \(2018\)](#)—which is nevertheless naturally accommodated by the model—makes for a particularly sharp test case.

¹⁹The experiment and hypotheses were pre-registered on OSF and are accessible at the following link: https://osf.io/xj9wv/overview?view_only=97698049c9134a79b166c2ec7159f264

4.1 Experimental design and identification

Participants. I recruited 422 participants via Prolific UK, who completed a total of 137 choice trials each. Participants were randomly assigned to the baseline ($n = 211$) or treatment ($n = 211$) condition (see below). All participants were required to enable their webcam for the duration of the study to screen for automated or inattentive responding (Yang and Krajbich, 2021). One trial was randomly selected and paid out with a 10% probability as a bonus in addition to a fixed base payment (see Online Appendix A.2 for the instructions). After applying the preregistered exclusion criteria, 389 participants remain for the main analysis. I report the corresponding analyses for the full sample as a robustness check in the Online Appendix.

Design and treatment. Each trial presented a binary choice between a two-outcome lottery and a sure amount of money. The lottery paid £18 with probability p and £0 otherwise. Probabilities were drawn from $p \in \{.1, .2, .3, .5, .7, .8, .9\}$, and for each probability level the sure amount took every integer value from £1 to £17, yielding a full factorial bracketing design of $7 \times 17 = 119$ primary trials. An additional 18 trials, drawn from two auxiliary outcome pairs (£20-or-£0 and £22-or-£4, each at $p \in \{.1, .5, .9\}$), were interspersed to prevent participants from disengaging from outcome information. The treatment was administered between-subjects and concerned how probabilities were displayed. In the *baseline* condition, probabilities were shown as fractions out of 10 (e.g., 3/10). In the *treatment* condition, both the numerator and denominator were multiplied by an integer drawn independently from 3–17, excluding 10 to avoid round out-of-100 fractions, so that the same underlying probability was displayed in a harder-to-parse form (e.g., 27/90 rather than 3/10) that varied unpredictably in scale from trial to trial. This manipulation was designed to increase representational noise in the probability signal without altering the objective probabilities or the decision problem itself.

Process-tracing measures. Trial information was concealed in a Mouselab-style interface programmed in jsPsych (De Leeuw, 2015): the probability and outcome associated with each option were hidden inside labelled boxes that revealed their value while the participant’s cursor hovered over them and re-concealed the value once the cursor moved away. Participants could inspect any box, in any order, as many times as they wished, before indicating their choice. For each trial I recorded the cumulative dwell time (in millisec-

onds) spent on each box, which I aggregate into a probability-information channel and an outcome-information channel for analysis. The mapping is deliberately made at the level of information channels rather than individual mathematical objects. Participants observe the constituent monetary outcomes x , y , and c , from which the comparative quantities entering the model can be constructed internally. This aggregation maps the process-tracing measures directly onto the two information channels distinguished in the model, reflecting the probability-versus-outcome attention contrast studied by [Pachur et al. \(2018\)](#).

4.2 Results

I first examine treatment effects on attention allocation and choice behaviour. I then turn to between-subject associations between attention and behaviour.

Numerical probability ratios and attention allocation. I begin with the treatment effects on information acquisition. The primary measures are dwell times on probability and outcome information. I interpret dwell time as a continuous proxy for the number of noisy samples accumulated from each information channel: longer inspection corresponds to more draws and hence, on average, greater posterior precision. To limit the influence of unusually long observations, which likely reflect interruptions or breaks, I winsorize dwell times at the 99th percentile. I then take their square roots, corresponding to the theoretically relevant transformation in the model, and average the transformed measures across tasks to obtain participant-level indices. Unless stated otherwise, statistical comparisons reported in the text are based on two-sided Wilcoxon rank-sum tests.

Panel A of Figure 8 shows the empirical cumulative distribution functions of participants' mean square-root-transformed dwell time on probability information, separately by treatment condition. Probability dwell time increases markedly when probabilities are represented using more complex numerical ratios ($p < 0.001$), consistent with the model's prediction that greater probability-coding noise induces additional sampling of probability information. Panel B shows the corresponding effect on outcome attention. Outcome dwell time also increases ($p = 0.002$), although more moderately, and total attention consequently rises in the treatment condition ($p < 0.001$). These results illustrate compensatory information acquisition: when probability signals become less reliable, participants devote more attention to reducing the resulting uncertainty. The increase in outcome attention further indicates that the adjustment is not confined to the directly

manipulated information channel, which aligns with model predictions.

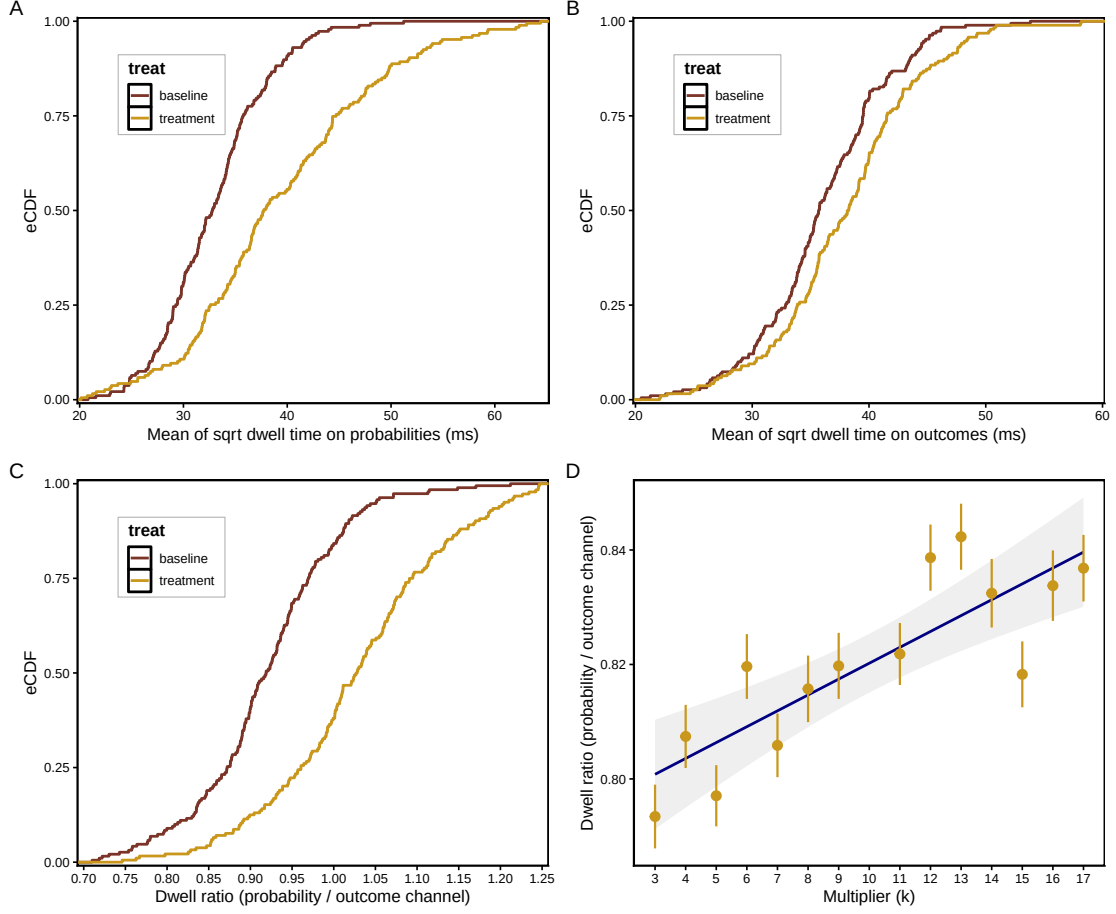


Figure 8: Attention allocation by treatment condition

Panel C shows the quantity most directly relevant to the model: attention to probability information relative to attention to outcome information. The treatment produces a clear shift in this ratio. Making probabilities harder to parse substantially increases the relative attention devoted to probabilities ($p < 0.001$), consistent with the model's prediction that noisier probability signals raise the sampling value of the probability channel.

Panel D replaces the binary treatment indicator with the numerical multiplier applied to the probability ratio. Relative probability attention increases with the multiplier, suggesting that processing difficulty rises with the numerical scale in which the numerator and denominator are expressed. The relationship is not perfectly monotonic, however: multipliers of 5 and 15 produce smaller increases than some immediately neighbouring values. Moreover, even the smallest multiplier yields a relative-attention ratio well above the baseline-condition average of 0.73.

Numerical probability ratios and choice behaviour. I next examine risky-choice behaviour. For each probability level, I use the observed choices to construct a normalized certainty-equivalent measure, $dw(p) = \frac{c-y}{x-y}$, where c denotes the estimated certainty equivalent of the lottery paying x with probability p and y otherwise. The normalization expresses the certainty equivalent as a proportion of the distance between the two lottery outcomes. Under dual expected utility (Preston and Baratta, 1948; Yaari, 1987), this measure can be interpreted as a decision weight; equivalently, it corresponds to the probability weight implied by prospect theory under linear utility. The construction summarizes choice proportions at each probability level and therefore abstracts from possible nonmonotonicities and other trial-level irregularities. Below, I include probit regressions that do not require this assumption.

Panel A of Figure 9 plots the aggregate numerator of Equation 2 in log-odds form. The elevation of the fitted function captures average risk taking through $\delta_{n,m}$, whereas its slope captures probability sensitivity through γ_n/β_m . The fitted function is markedly flatter in the treatment than in the baseline condition. Thus, although the treatment increases relative attention to probability information, it reduces behavioural sensitivity to probabilities. Taken together, these findings support the model’s prediction of incomplete compensatory sampling: participants acquire additional probability information, but the increase in attention is insufficient to offset the lower fidelity of individual probability signals.

Panel B reinforces this result by showing the empirical cumulative distribution functions of participant-level probability-sensitivity indices. For each symmetric probability pair around 0.5, I compute $\frac{dw(0.5+s)-dw(0.5-s)}{2s}$, which measures the change in decision weights relative to the corresponding change in objective probability. An index of 1 therefore corresponds to linear treatment of probabilities, values below 1 indicate probability insensitivity, and values above 1 indicate oversensitivity. The distribution shifts to the left in the treatment condition, indicating lower probability sensitivity ($p = 0.040$).

The second key pattern concerns the elevation of the fitted function in Panel A. At even odds, the fitted function is more elevated in the treatment condition, indicating greater rather than lower average risk taking. Panel C of Figure 9 shows the corresponding empirical cumulative distribution functions of participants’ overall proportions of safe choices. Although the treatment difference narrowly misses conventional significance lev-

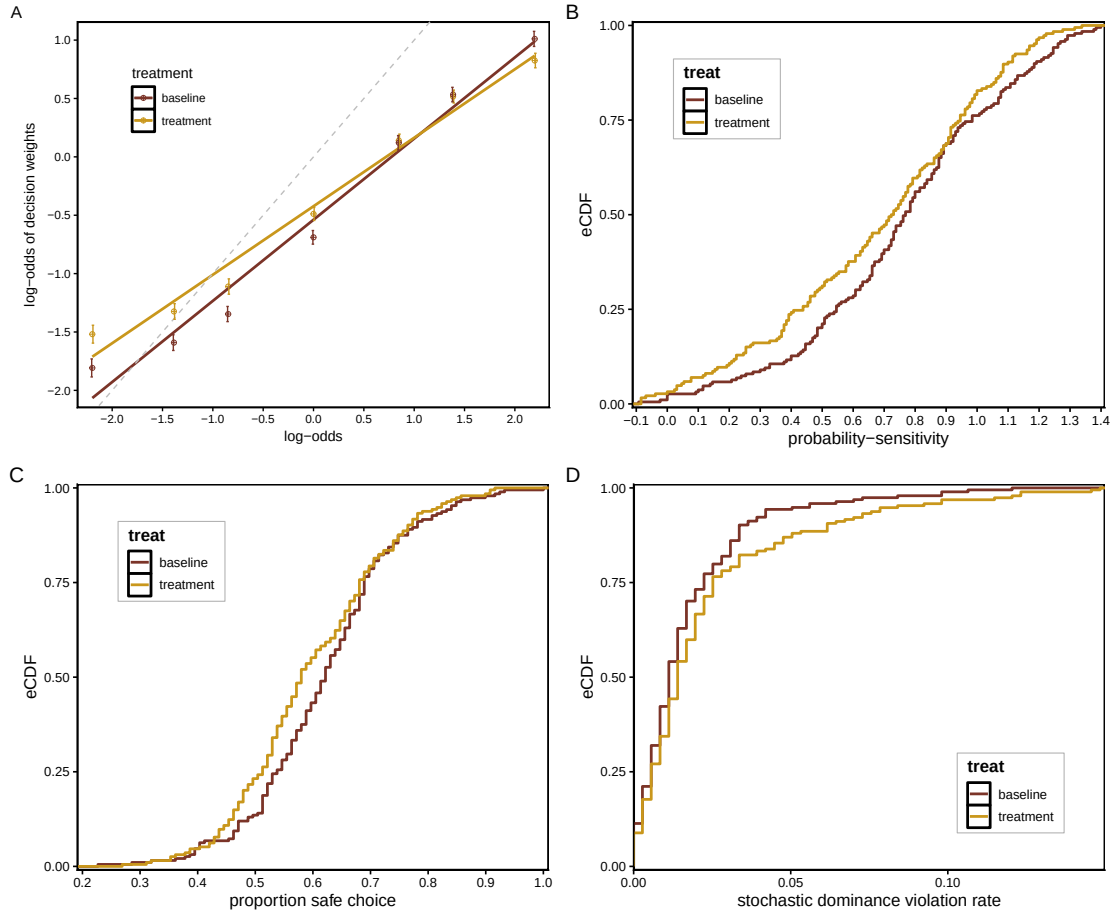


Figure 9: Choice behaviour by treatment condition

els ($p = 0.068$), its direction is opposite to the predicted increase in risk aversion. This result also contrasts with [Oprea and Vieider \(2026\)](#), who document increased risk aversion when either probability or outcome information becomes noisier. In their probability manipulation, replacing visually displayed urns with percentages reduces probability sensitivity, as in the present experiment, but also increases risk aversion. Here, by contrast, lower probability sensitivity is accompanied by slightly greater average risk taking.

Although this explanation was not preregistered, one possibility is the ratio-bias effect: scaling up both the numerator and denominator of a probability ratio can increase its perceived magnitude even when the underlying probability remains unchanged ([Miller et al., 1989](#); [Kirkpatrick and Epstein, 1992](#); [Denes-Raj et al., 1995](#); [Alonso and Fernandez-Berrocal, 2003](#)). If larger multipliers make the represented probability appear larger, they may shift the elevation of the risk-taking function upward and counteract the increase in risk aversion otherwise predicted by reduced probability discriminability. Given the post-

hoc nature of this explanation, however, it needs to be subjected to additional empirical scrutiny before being accepted.

Finally, Panel D examines a complementary prediction arising from the denominator of Equation 2: increasing the noisiness of one attribute should increase stochasticity in choice. I measure stochasticity nonparametrically using first-order stochastic-dominance violations.²⁰ Such violations are more frequent in the treatment condition ($p = 0.024$). This provides a further signature of incomplete compensation: the additional attention devoted to probability information does not fully offset the reduction in signal fidelity, leading both to lower probability sensitivity and to more frequent violations of consistency precepts.

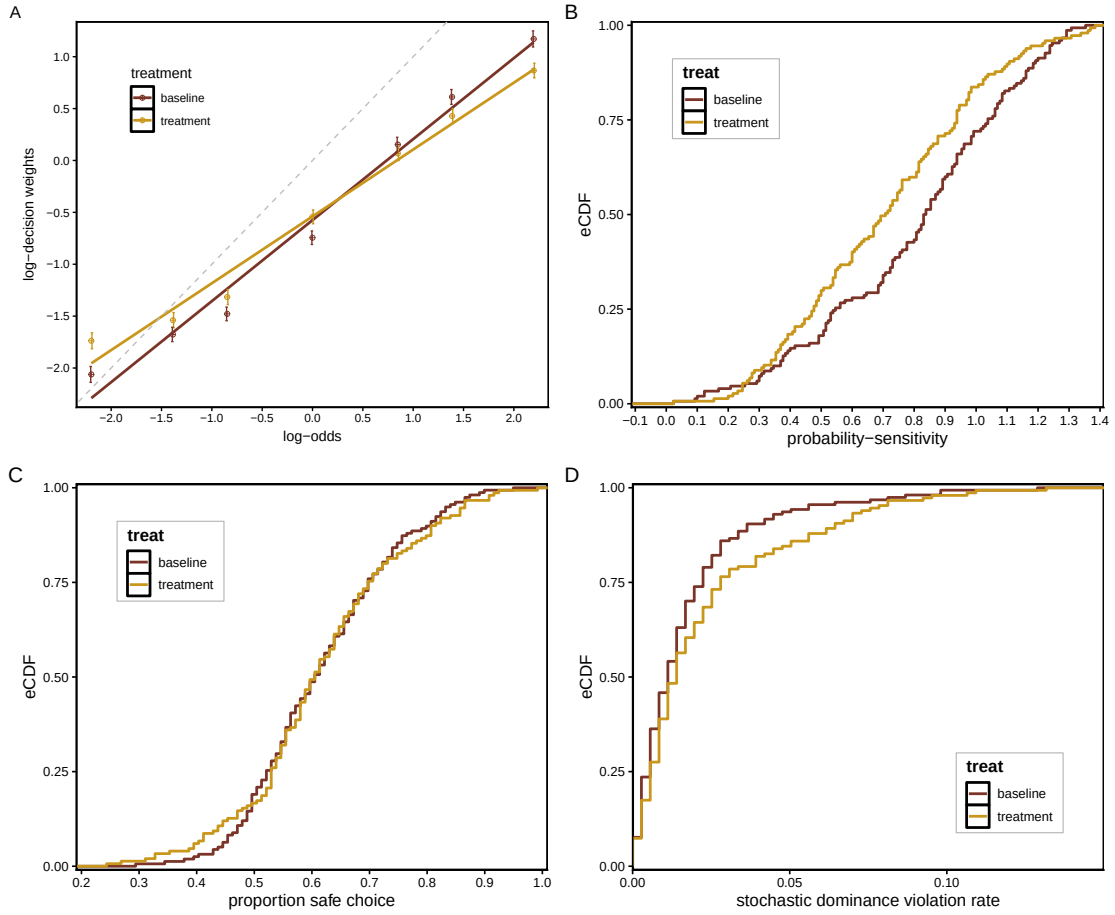


Figure 10: Choice behaviour in the behavioural benchmark experiment

²⁰Specifically, for each participant I compute the proportion of probability comparisons in which the lottery is chosen at a lower probability but the sure option is chosen at a higher probability.

A behavioural benchmark experiment. One potential concern is that requiring participants to actively reveal information may itself affect choice behaviour. I therefore conducted a behavioural benchmark experiment using the identical interface, except that all decision-relevant information was visible from the outset. All other design features, recruitment procedures, and sample-size targets were held constant, although the fully visible interface naturally provides no attention measures.

Figure 10 reproduces the four choice panels from Figure 9. The central treatment effects are qualitatively aligned across the two implementations. Probability sensitivity again decreases in the treatment condition ($p = 0.001$), while first-order stochastic-dominance violations again become more frequent, although the latter difference narrowly misses conventional significance levels ($p = 0.055$). There is no treatment difference in average risk aversion ($p = 0.933$). Thus, the failure to observe the predicted increase in risk aversion does not appear to be caused by the mouse-contingent interface.

Individual-level correlations. The treatment effects reported above differ sharply from the cross-sectional association documented by Pachur et al. (2018). Experimentally increasing the noisiness of probability representations raises relative attention to probabilities while reducing probability sensitivity. By contrast, in the absence of experimentally induced variation, Pachur et al. (2018) find that individuals who devote relatively more attention to probabilities also exhibit greater probability sensitivity. The model predicts that these patterns can coexist because they arise from different sources of variation. A causal increase in coding noise induces compensatory but incomplete sampling, whereas stable individual differences in prior uncertainty or signal precision can generate a positive association between relative attention and achieved sensitivity.

Panel A of Figure 11 examines the participant-level association between relative probability attention and probability sensitivity within each treatment condition. The treatment effects are clearly visible: relative probability attention shifts upward, whereas probability sensitivity shifts downward when probabilities are represented using larger numbers. Within each condition, however, the association between the two variables remains positive. The correlation is $r = 0.295$ ($p < 0.001$) in the baseline condition and $r = 0.474$ ($p < 0.001$) in the treatment condition. Thus, the experimentally induced movement across conditions runs in the opposite direction from the cross-sectional association observed within conditions. This is the pattern predicted by the model: greater coding noise

causes compensatory but incomplete attention, whereas stable individual differences can generate a positive relation between relative attention and achieved sensitivity.

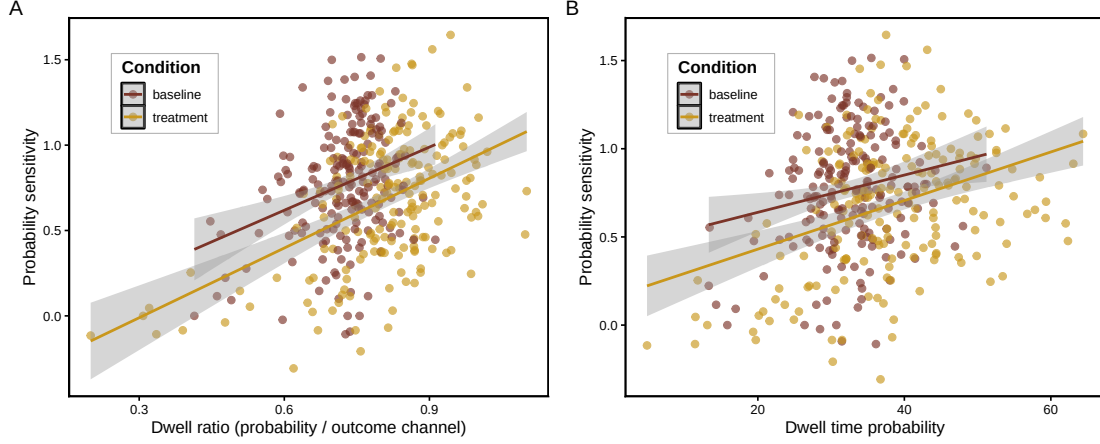


Figure 11: Attention and probability sensitivity within treatment conditions

Panel B examines the same association using a measure closer to that employed by [Pachur et al. \(2018\)](#): absolute attention to probability information, measured here as square-root-transformed probability dwell time in line with the model. The association is again positive in both conditions, with $r = 0.186$ ($p = 0.009$) in the baseline condition and $r = 0.370$ ($p < 0.0001$) in the treatment condition.

Taken together, Panels A and B resolve the apparent contradiction. Within each treatment condition, individuals who devote more attention to probability information exhibit greater probability sensitivity. Experimentally making probabilities harder to parse, however, has the opposite effect: it increases both relative and absolute attention to probabilities while reducing probability sensitivity. This joint pattern provides a direct test of the model’s central mechanism of compensatory but incomplete sampling.

Regression analysis. To assess whether the preceding results are robust to the aggregation used in the participant-level analyses, and to characterize the underlying choice patterns more fully, Table 1 reports reduced-form regressions of risky choice on the treatment manipulation, task characteristics, and attention measures. I estimate Bayesian hierarchical probit models with participant-specific random intercepts, thereby accounting for repeated choices within participants and persistent heterogeneity in baseline risk taking. The models are estimated in Stan ([Carpenter et al., 2017](#)). Reported coefficients are posterior means, with posterior standard deviations in parentheses.

The reduced-form probit specification maps closely onto the stochastic choice equation of the model. For trial t , let $\ell_{p,t} \triangleq \ln[p_t/(1 - p_t)]$ denote the log-odds and $\ell_{o,t} \triangleq \ln[(c_t - y_t)/(x_t - c_t)]$ denote the log cost-benefit ratio introduced above. For the primary trials, where $x = 18$ and $y = 0$, the latter reduces to $\ell_{o,t} = \ln[c_t/(18 - c_t)]$. Dividing the posterior evidence balance by β_m gives

$$\frac{\gamma_n}{\beta_m} \ell_{p,t} - \ell_{o,t} - \frac{\theta_{n,m}}{\beta_m}.$$

I therefore enter the log cost-benefit ratio as an offset, fixing its coefficient at -1 . Under this normalization, the coefficient on log-odds provides a reduced-form proxy for the model’s relative probability sensitivity, γ_n/β_m , while the intercept captures the elevation of the choice function induced by prior expectations. The treatment interaction with log-odds consequently tests whether increasing probability-representation noise reduces relative probability sensitivity, as predicted by the model.

Column (1) estimates the treatment effects on the elevation and probability sensitivity of the risk-taking function. Column (2) replaces the binary treatment indicator with the numerical multiplier applied to the probability ratio. Column (3) decomposes relative probability attention and total attention into between-participant differences and within-participant deviations from each participant’s mean. The regressions confirm the results already shown in the figures above. In addition, they show that within-subject task-level differences are directionally aligned with between-subject effects, while not necessarily being of the same magnitude.

5 Conclusion

This paper has developed a Bayesian account of risky choice in which probabilities and outcomes are represented through noisy internal samples. Posterior uncertainty determines both how attention is allocated and when information acquisition stops. The model therefore links noisy representation, endogenous attention, and systematic distortions in risky choice within a common framework.

The experiment provides a direct test of the model’s central mechanism. Making probability information harder to parse increases both absolute and relative attention to probabilities, yet reduces behavioural sensitivity to probability differences. At the same time, a positive cross-sectional association between probability attention and probability sensitivity remains present within treatment conditions. These findings reconcile the experi-

Table 1: Reduced-form probit regressions of risky choice

	Dependent variable: risky choice		
	(1) Treatment	(2) Multiplier	(3) Attention decomposition
Treatment	0.084 (0.048)		0.090 (0.050)
Multiplier		0.001 (0.002)	
Log-odds	0.507 (0.007)	0.509 (0.007)	0.548 (0.007)
<i>Interactions with log-odds</i>			
Treatment $\times$ log-odds	-0.086 (0.010)		-0.152 (0.010)
Multiplier $\times$ log-odds		-0.008 (0.001)	
Relative probability attention $\times$ log-odds, between			0.204 (0.012)
Relative probability attention $\times$ log-odds, within			0.096 (0.006)
Total attention $\times$ log-odds, between			-0.020 (0.010)
Total attention $\times$ log-odds, within			-0.065 (0.007)

Note. Entries are posterior means from Bayesian hierarchical probit regressions; posterior standard deviations are reported in parentheses. All models include participant-specific random intercepts. Risky choice equals one when the participant selects the lottery. Log-odds is centered at even odds, so the treatment and multiplier main effects describe changes in risky choice at $p = 0.5$. Column (2) uses the raw numerical multiplier applied to the probability ratio. In Column (3), relative probability attention and total attention are decomposed into participant means (*between*) and deviations from those means (*within*). The corresponding attention main effects are included in the model but omitted from the table for compactness. Attention coefficients describe conditional associations and should not be interpreted as causal mediation effects.

mentally induced effect of coding noise with the individual-level pattern documented in earlier work: additional attention can be compensatory without being sufficient to restore the influence of degraded information.

The present analysis focuses on binary risky choices in the gain domain and on a targeted manipulation of probability representation. Extensions to outcome manipulations, visual salience, and to other decision domains such as mixed gambles, delay-discounting, or unknown or experienced probabilities are highly desirable to probe the scope of the model predictions. The general predictions are nuanced: Greater attention may signal more precise representations, but it may also be a response to information that is intrinsically more difficult to encode. This makes exogenous manipulations of the cognitive primitives particularly well-suited to studying joint effects on attention allocation and choice.

References

- Alonso, D. and Fernandez-Berrocal, P. (2003). Irrational decisions: Attending to numbers rather than ratios. *Personality and Individual Differences*, 35(7):1537–1547.
- Apestequia, J. and Ballester, M. A. (2018). Monotone stochastic choice models: The case of risk and time preferences. *Journal of Political Economy*, 126(1):74–106.
- Arieli, A., Ben-Ami, Y., and Rubinstein, A. (2011). Tracking decision makers under uncertainty. *American Economic Journal: Microeconomics*, 3(4):68–76.
- Armel, K. C., Beaumel, A., and Rangel, A. (2008). Biasing simple choices by manipulating relative visual attention. *Judgment and Decision Making*, 3(5):396–403.
- Arrow, K. J., Blackwell, D., and Girshick, M. A. (1949). Bayes and minimax solutions of sequential decision problems. *Econometrica*, pages 213–244.
- Bedi, S., de Hollander, G., Harl, M., and Ruff, C. C. (2026). Multi-stage efficient coding of perception and value in goal-directed behavior. Technical report, Preprint.
- Bhatia, S. (2014). Sequential sampling and paradoxes of risky choice. *Psychonomic Bulletin & Review*, 21(5):1095–1111.
- Bhatia, S. and Loomes, G. (2017). Noisy preferences in risky choice: a cautionary note. *Psychological Review*, 124(5):678–687.
- Bhui, R. and Gershman, S. J. (2018). Decision by sampling implements efficient coding of psychoeconomic functions. *Psychological Review*, 125(6):985.
- Bogacz, R., Brown, E., Moehlis, J., Holmes, P., and Cohen, J. D. (2006). The physics of optimal decision making: a formal analysis of models of performance in two-alternative forced-choice tasks. *Psychological Review*, 113(4):700.
- Bordalo, P., Gennaioli, N., and Shleifer, A. (2012). Salience theory of choice under risk. *The Quarterly Journal of Economics*, 127(3):1243–1285.

- Bouchouicha, R., Oprea, R., Vieider, F. M., and Wu, J. (2024). Is prospect theory really a theory of choice? Technical report, Ghent University Discussion Papers.
- Busemeyer, J. R. and Townsend, J. T. (1993). Decision field theory: a dynamic-cognitive approach to decision making in an uncertain environment. *Psychological Review*, 100(3):432.
- Callaway, F., Rangel, A., and Griffiths, T. L. (2021). Fixation patterns in simple choice reflect optimal information sampling. *PLoS Computational Biology*, 17(3):e1008863.
- Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., and Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, 76(1):1–32.
- Chernoff, H. (1959). Sequential design of experiments. *The Annals of Mathematical Statistics*, 30(3):755–770.
- Chung, H.-K., Doren, N., Mononen, L., Lu, M., Grueschow, M., Hayward Könnecke, H., Jetter, A., Quednow, B. B., Netzer, N., and Tobler, P. N. (2025). Improving rationality by increasing attention. Technical report, Working Paper.
- De Leeuw, J. R. (2015). jspsych: A javascript library for creating behavioral experiments in a web browser. *Behavior Research Methods*, 47(1):1–12.
- Denes-Raj, V., Epstein, S., and Cole, J. (1995). The generality of the ratio-bias phenomenon. *Personality and Social Psychology Bulletin*, 21(10):1083–1092.
- Drugowitsch, J., Moreno-Bote, R., Churchland, A. K., Shadlen, M. N., and Pouget, A. (2012). The cost of accumulating evidence in perceptual decision making. *Journal of Neuroscience*, 32(11):3612–3628.
- Drugowitsch, J. and Pouget, A. (2012). Probabilistic vs. non-probabilistic approaches to the neurobiology of perceptual decision-making. *Current Opinion in Neurobiology*, 22(6):963–969.
- Fitzgerald, T. H., Seymour, B., and Dolan, R. J. (2009). The role of human orbitofrontal cortex in value comparison for incommensurable objects. *The Journal of Neuroscience*, 29(26):8388–8395.
- Frydman, C. and Jin, L. J. (2022). Efficient coding and risky choice. *Quarterly Journal of Economics*, 137:161–213.
- Fudenberg, D., Strack, P., and Strzalecki, T. (2018). Speed, accuracy, and the optimal timing of choices. *American Economic Review*, 108(12):3651–3684.
- Gabaix, X., Laibson, D., Moloche, G., and Weinberg, S. (2006). Costly information acquisition: Experimental analysis of a boundedly rational model. *American Economic Review*, 96(4):1043–1068.
- Garagnani, M. and Vieider, F. M. (2025). Economic consequences of numerical adaptation. *Psychological Science*, 36(6):407–420.
- Gigerenzer, G. and Hoffrage, U. (1995). How to improve bayesian reasoning without instruction: Frequency formats. *Psychological review*, 102(4):684.

- Glimcher, P. W. (2010). *Foundations of neuroeconomic analysis*. Oxford University Press.
- Gluth, S., Deakin, J., and Rieskamp, J. (2026). A theory of multiattribute search and choice. *Psychological Review*.
- Gold, J. I. and Shadlen, M. N. (2001). Neural computations that underlie decisions about sensory stimuli. *Trends in Cognitive Sciences*, 5(1):10–16.
- Gold, J. I. and Shadlen, M. N. (2002). Banburismus and the brain: decoding the relationship between sensory stimuli, decisions, and reward. *Neuron*, 36(2):299–308.
- Goldstein, W. M. and Einhorn, H. J. (1987). Expression theory and the preference reversal phenomena. *Psychological Review*, 94(2):236–254.
- Gonzalez, R. and Wu, G. (1999). On the Shape of the Probability Weighting Function. *Cognitive Psychology*, 38:129–166.
- Green, D. M., Swets, J. A., et al. (1966). *Signal detection theory and psychophysics*, volume 1. Wiley New York.
- Hébert, B. and Woodford, M. (2023). Rational inattention when decisions take time. *Journal of Economic Theory*, 208:105612.
- Heng, J. A., Woodford, M., and Polania, R. (2020). Efficient sampling and noisy decisions. *Elife*, 9:e54962.
- Hertwig, R., Barron, G., Weber, E. U., and Erev, I. (2004). Decisions from experience and the effect of rare events in risky choice. *Psychological Science*, 15:534–539.
- Jang, A., Sharma, R., and Drugowitsch, J. (2021). Optimal policy for attention-modulated decisions explains human fixation behavior. *eLife*, 10.
- Kable, J. W. and Glimcher, P. W. (2009). The neurobiology of decision: consensus and controversy. *Neuron*, 63(6):733–745.
- Kahneman, D. and Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, 47(2):263 – 291.
- Khaw, M. W., Li, Z., and Woodford, M. (2021). Cognitive imprecision and small-stakes risk aversion. *The Review of Economic Studies*, 88(4):1979–2013.
- Khaw, M. W., Li, Z., and Woodford, M. (2023). Cognitive imprecision and stake-dependent risk attitudes. Technical report, NBER Working Paper 30417.
- Kirkpatrick, L. A. and Epstein, S. (1992). Cognitive-experiential self-theory and subjective probability: further evidence for two conceptual systems. *Journal of Personality and Social Psychology*, 63(4):534.
- Knill, D. C. and Pouget, A. (2004). The bayesian brain: the role of uncertainty in neural coding and computation. *TRENDS in Neurosciences*, 27(12):712–719.
- Knudsen, E. B. and Wallis, J. D. (2022). Taking stock of value in the orbitofrontal cortex. *Nature Reviews Neuroscience*, 23(7):428–438.
- Koch, C. and Ullman, S. (1985). Shifts in selective visual attention: towards the underlying neural circuitry. *Human Neurobiology*, 4(4):219–227.

- Krajbich, I., Armel, C., and Rangel, A. (2010). Visual fixations and the computation and comparison of value in simple choice. *Nature Neuroscience*, 13(10):1292–1298.
- Krajbich, I. and Rangel, A. (2011). Multialternative drift-diffusion model predicts the relationship between visual fixations and choice in value-based decisions. *Proceedings of the National Academy of Sciences*, 108(33):13852–13857.
- Laughlin, S. (1981). A simple coding procedure enhances a neuron’s information capacity. *Zeitschrift für Naturforschung c*, 36(9-10):910–912.
- Levy, D. J. and Glimcher, P. W. (2012). The root of all value: a neural common currency for choice. *Current opinion in neurobiology*, 22(6):1027–1038.
- Li, X. and Camerer, C. F. (2022). Predictable effects of visual salience in experimental decisions and games. *The Quarterly Journal of Economics*, 137(3):1849–1900.
- Liang, A., Mu, X., and Syrgkanis, V. (2022). Dynamically aggregating diverse information. *Econometrica*, 90(1):47–80.
- Lim, S.-L., O’Doherty, J. P., and Rangel, A. (2011). The decision value computations in the vmPFC and striatum use a relative value code that is guided by visual attention. *Journal of Neuroscience*, 31(37):13214–13223.
- Ma, W. J., Beck, J. M., Latham, P. E., and Pouget, A. (2006). Bayesian inference with probabilistic population codes. *Nature Neuroscience*, 9(11):1432–1438.
- Ma, W. J., Kording, K. P., and Goldreich, D. (2023). *Bayesian Models of Perception and Action: An Introduction*. MIT press.
- Miller, D. T., Turnbull, W., and McFarland, C. (1989). When a coincidence is suspicious: The role of mental simulation. *Journal of Personality and Social Psychology*, 57(4):581.
- Netzer, N. (2009). Evolution of time preferences and attitudes toward risk. *American Economic Review*, 99(3):937–55.
- Oprea, R. and Vieider, F. M. (2024). Minding the gap: On the origins of the description-experience gap. Technical report, Working Paper.
- Oprea, R. and Vieider, F. M. (2026). Information processing and the anomalies of risk and time. Technical report, Working Paper.
- Pachur, T., Schulte-Mecklenbeck, M., Murphy, R. O., and Hertwig, R. (2018). Prospect theory reflects selective allocation of attention. *Journal of Experimental Psychology: General*, 147(2):147–169.
- Padoa-Schioppa, C. (2007). Orbitofrontal cortex and the computation of economic value. *Annals of the New York Academy of Sciences*, 1121(1):232–253.
- Padoa-Schioppa, C. and Assad, J. A. (2006). Neurons in the orbitofrontal cortex encode economic value. *Nature*, 441(7090):223–226.
- Polania, R. and Hare, T. (2026). Similarity and comparison complexity. *Preprint*.
- Polania, R., Woodford, M., and Ruff, C. C. (2019). Efficient coding of subjective value. *Nature neuroscience*, 22(1):134–142.

- Prat-Carrabin, A. and Woodford, M. (2022). Efficient coding of numbers explains decision bias and noise. *Nature Human Behaviour*, 6(8):1142–1152.
- Prat-Carrabin, A. and Woodford, M. (2026). Endogenous precision of the number sense. *Elife*, 13:RP101277.
- Preston, M. G. and Baratta, P. (1948). An experimental study of the auction-value of an uncertain outcome. *The American Journal of Psychology*, 61:183–193.
- Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, 85(2):59.
- Robson, A. J. (2001a). The biological basis of economic behavior. *Journal of Economic Literature*, 39(1):11–33.
- Robson, A. J. (2001b). Why would nature give individuals utility functions? *Journal of Political Economy*, 109(4):900–914.
- Russo, D. J., Van Roy, B., Kazerouni, A., Osband, I., and Wen, Z. (2018). A tutorial on thompson sampling. *Foundations and trends® in machine learning*, 11(1):1–99.
- Rustichini, A., Conen, K. E., Cai, X., and Padoa-Schioppa, C. (2017). Optimal coding and neuronal adaptation in economic decisions. *Nature Communications*, 8(1):1208.
- Sanborn, A. N. and Chater, N. (2016). Bayesian brains without probabilities. *Trends in Cognitive Sciences*, 20(12):883–893.
- Scholten, M., Read, D., and Sanborn, A. (2014). Weighing outcomes by time or against time? Evaluation rules in intertemporal choice. *Cognitive Science*, 38(3):399–438.
- Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review*, 63(2):129.
- Simon, H. A. (1971). Designing organizations for an information-rich world. In Greenberger, M., editor, *Computers, Communications, and the Public Interest*, pages 38–72. Johns Hopkins Press, Baltimore.
- Sims, C. A. (2003). Implications of rational inattention. *Journal of Monetary Economics*, 50(3):665–690.
- Stewart, N., Chater, N., and Brown, G. D. (2006). Decision by sampling. *Cognitive Psychology*, 53(1):1–26.
- Tajima, S., Drugowitsch, J., and Pouget, A. (2016). Optimal policy for value-based decision-making. *Nature Communications*, 7(1):12400.
- Thompson, W. R. (1933). On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*, 25(3/4):285–294.
- Tversky, A. (1969). Intransitivity of preferences. *Psychological Review*, 76(1):31.
- Vieider, F. M. (2024). Decisions under uncertainty as bayesian inference on choice options. *Management Science*, 70(12):9014–9030.
- Vieider, F. M. (2025). The Structure of Stochastic Choice. *Working Paper*.

- Wakker, P. P. (2010). *Prospect Theory for Risk and Ambiguity*. Cambridge University Press, Cambridge.
- Wald, A. (1945). Sequential tests of statistical hypotheses. *Annals of Mathematical Statistics*, 16(2):117–186.
- Wald, A. and Wolfowitz, J. (1948). Optimum character of the sequential probability ratio test. *The Annals of Mathematical Statistics*, 19:326–339.
- Webb, R., Glimcher, P. W., and Louie, K. (2020). Divisive normalization does influence decisions with multiple alternatives. *Nature Human Behaviour*, 4(11):1118–1120.
- Wilcox, N. T. (2011). ‘Stochastically more risk averse:’ A contextual theory of stochastic discrete choice under risk. *Journal of Econometrics*, 162(1):89–104.
- Woodford, M. (2020). Modeling imprecision in perception, valuation, and choice. *Annual Review of Economics*, 12:579–601.
- Yaari, M. E. (1987). The Dual Theory of Choice under Risk. *Econometrica*, 55(1):95–115.
- Yang, X. and Krajbich, I. (2021). Webcam-based online eye-tracking for behavioral research. *Judgment and Decision Making*, 16(6):1485–1505.
- Yang, X. and Krajbich, I. (2023). A dynamic computational model of gaze and choice in multi-attribute decisions. *Psychological Review*, 130(1):52–70.
- Zhang, H. and Maloney, L. T. (2012). Ubiquitous log odds: a common representation of probability and frequency distortion in perception, action, and cognition. *Frontiers in Neuroscience*, 6:1.
- Zhu, J.-Q., Sanborn, A. N., and Chater, N. (2020). The bayesian sampler: Generic bayesian inference causes incoherence in human probability judgments. *Psychological review*, 127(5):719.
- Zhu, J.-Q., Sundh, J., Spicer, J., Chater, N., and Sanborn, A. N. (2024). The auto-correlated bayesian sampler: A rational process for probability judgments, estimates, confidence intervals, choices, confidence judgments, and response times. *Psychological Review*, 131(2):456.
- Zilker, V. and Pachur, T. (2022). Nonlinear probability weighting can reflect attentional biases in sequential sampling. *Psychological Review*, 129(5):949–975.
- Zilker, V. and Pachur, T. (2023). Attribute attention and option attention in risky choice. *Cognition*, 236:105441.

ONLINE APPENDIX

A Additional details on experiment

A.1 Analysis including all subjects

Here, I reproduce the analysis from the main text while including all 422 subjects—also those who clicked through without opening any boxes.

Numerical probability ratios and attention allocation. Panel A of Figure 12 shows the empirical cumulative distribution functions of participants’ mean square-root-transformed dwell time on probability information, separately by treatment condition. Probability dwell time increases markedly when probabilities are represented using more complex numerical ratios ($p < 0.001$), consistent with the model’s prediction that greater probability-coding noise induces additional sampling of probability information. Panel B shows the corresponding effect on outcome attention. Outcome dwell time also increases ($p = 0.002$), although more moderately, and total attention consequently rises in the treatment condition ($p < 0.001$). These results illustrate compensatory information acquisition: when probability signals become less reliable, participants devote more attention to reducing the resulting uncertainty. The increase in outcome attention further indicates that the adjustment is not confined to the directly manipulated information channel, which aligns with model predictions.

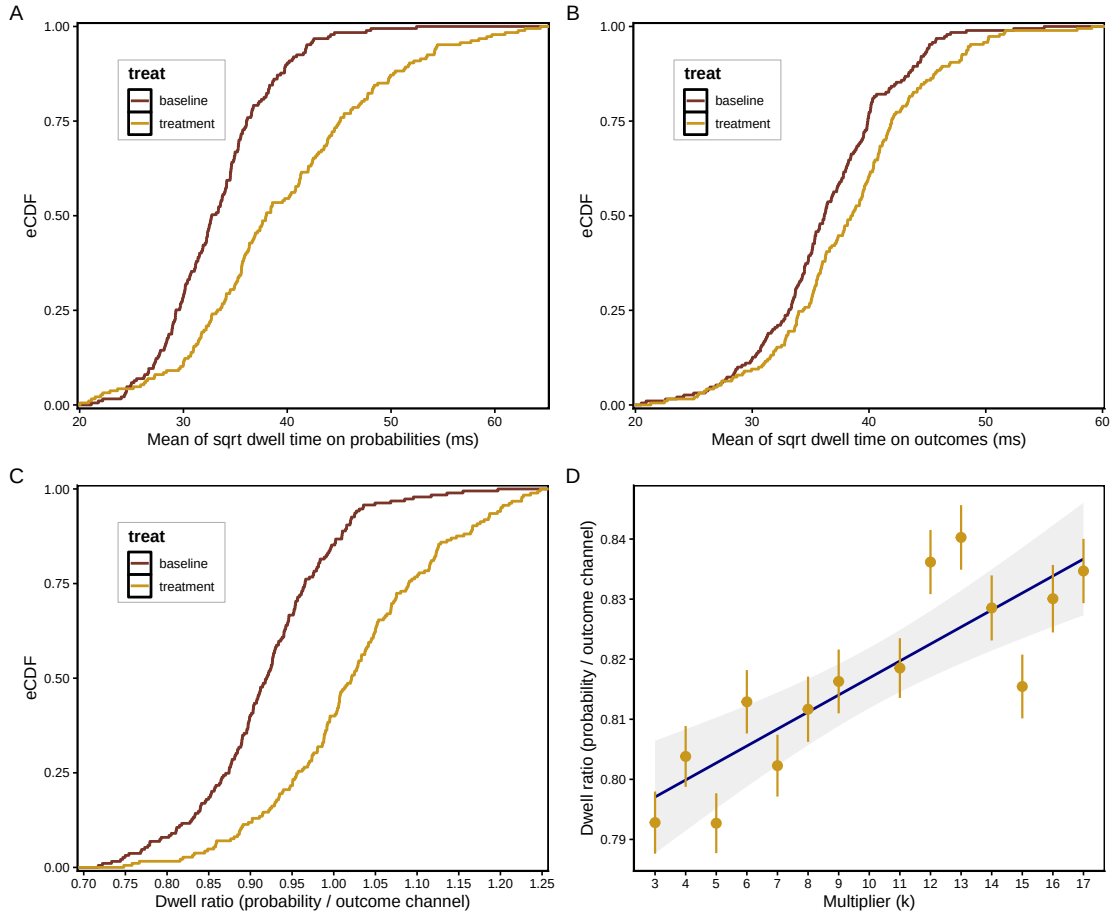


Figure 12: Attention allocation by treatment condition

Panel C shows the quantity most directly relevant to the model: attention to probability information relative to attention to outcome information. The treatment produces a clear shift in this ratio. Making probabilities harder to parse substantially increases the relative attention devoted to probabilities ($p < 0.001$), consistent with the model’s prediction that noisier probability signals raise the sampling value of the probability channel.

Panel D replaces the binary treatment indicator with the numerical multiplier applied to the probability ratio. Relative probability attention increases with the multiplier, suggesting that processing difficulty rises with the numerical scale in which the numerator and denominator are expressed. The relationship is not perfectly monotonic, however: multipliers of 5 and 15 produce smaller increases than some immediately neighbouring values. Moreover, even the smallest multiplier yields a relative-attention ratio well above the baseline-condition average of 0.73.

Numerical probability ratios and choice behaviour. I next examine risky-choice behaviour. For each probability level, I use the observed choices to construct a normalized certainty-equivalent measure, $dw(p) = \frac{c-y}{x-y}$, where c denotes the estimated certainty equivalent of the lottery paying x with probability p and y otherwise. The normalization expresses the certainty equivalent as a proportion of the distance between the two lottery outcomes. Under dual expected utility (Preston and Baratta, 1948; Yaari, 1987), this measure can be interpreted as a decision weight; equivalently, it corresponds to the probability weight implied by prospect theory under linear utility. The construction summarizes choice proportions at each probability level and therefore abstracts from possible nonmonotonicities and other trial-level irregularities. Below, I include probit regressions that do not require this assumption.

Panel A of Figure 13 plots the aggregate numerator of Equation 2 in log-odds form. The elevation of the fitted function captures average risk taking through $\delta_{n,m}$, whereas its slope captures probability sensitivity through γ_n/β_m . The fitted function is markedly flatter in the treatment than in the baseline condition. Thus, although the treatment increases relative attention to probability information, it reduces behavioural sensitivity to probabilities. Taken together, these findings support the model’s prediction of incomplete compensatory sampling: participants acquire additional probability information, but the increase in attention is insufficient to offset the lower fidelity of individual probability signals.

Panel B reinforces this result by showing the empirical cumulative distribution functions of participant-level probability-sensitivity indices. For each symmetric probability pair around 0.5, I compute $\frac{dw(0.5+s)-dw(0.5-s)}{2s}$, which measures the change in decision weights relative to the corresponding change in objective probability. An index of 1 therefore corresponds to linear treatment of probabilities, values below 1 indicate probability insensitivity, and values above 1 indicate oversensitivity. The distribution shifts to the left in the treatment condition, indicating lower probability sensitivity ($p = 0.040$).

The second key pattern concerns the elevation of the fitted function in Panel A. At even odds, the fitted function is more elevated in the treatment condition, indicating greater rather than lower average risk taking. Panel C of Figure 9 shows the corresponding empirical cumulative distribution functions of participants’ overall proportions of safe choices. Although the treatment difference narrowly misses conventional significance levels ($p = 0.068$), its direction is opposite to the predicted increase in risk aversion. This result also contrasts with Oprea and Vieider (2026), who document increased risk aver-

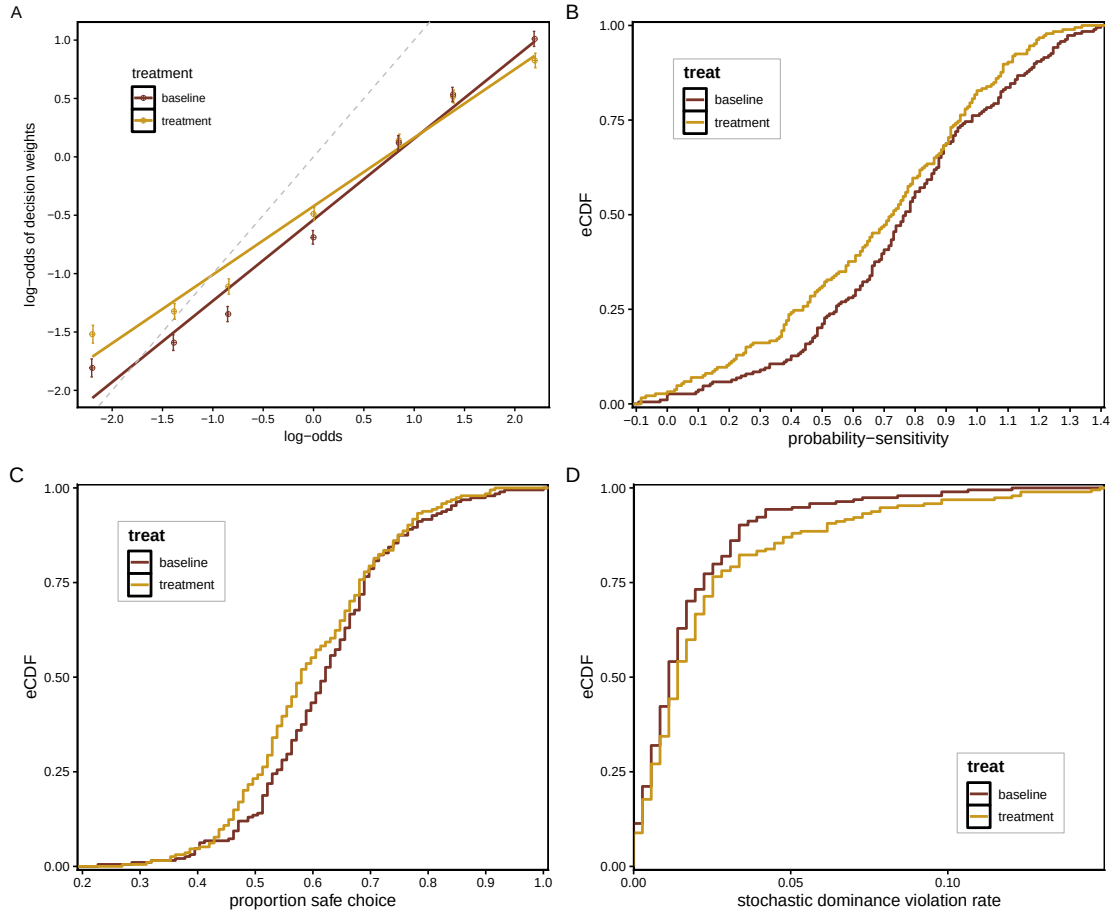


Figure 13: Choice behaviour by treatment condition

sion when either probability or outcome information becomes noisier. In their probability manipulation, replacing visually displayed urns with percentages reduces probability sensitivity, as in the present experiment, but also increases risk aversion. Here, by contrast, lower probability sensitivity is accompanied by slightly greater average risk taking.

Although this explanation was not preregistered, one possibility is the ratio-bias effect: scaling up both the numerator and denominator of a probability ratio can increase its perceived magnitude even when the underlying probability remains unchanged (Miller et al., 1989; Kirkpatrick and Epstein, 1992; Denes-Raj et al., 1995; Alonso and Fernandez-Berrocal, 2003). If larger multipliers make the represented probability appear larger, they may shift the elevation of the risk-taking function upward and counteract the increase in risk aversion otherwise predicted by reduced probability discriminability. Given the post-hoc nature of this explanation, however, it needs to be subjected to additional empirical scrutiny before being accepted.

Finally, Panel D examines a complementary prediction arising from the denominator of Equation 2: increasing the noisiness of one attribute should increase stochasticity in choice. I measure stochasticity nonparametrically using first-order stochastic-dominance violations.²¹ Such violations are more frequent in the treatment condition ($p = 0.024$). This

²¹Specifically, for each participant I compute the proportion of probability comparisons in which the lottery is chosen at a lower probability but the sure option is chosen at a higher probability.

provides a further signature of incomplete compensation: the additional attention devoted to probability information does not fully offset the reduction in signal fidelity, leading both to lower probability sensitivity and to more frequent violations of consistency precepts.

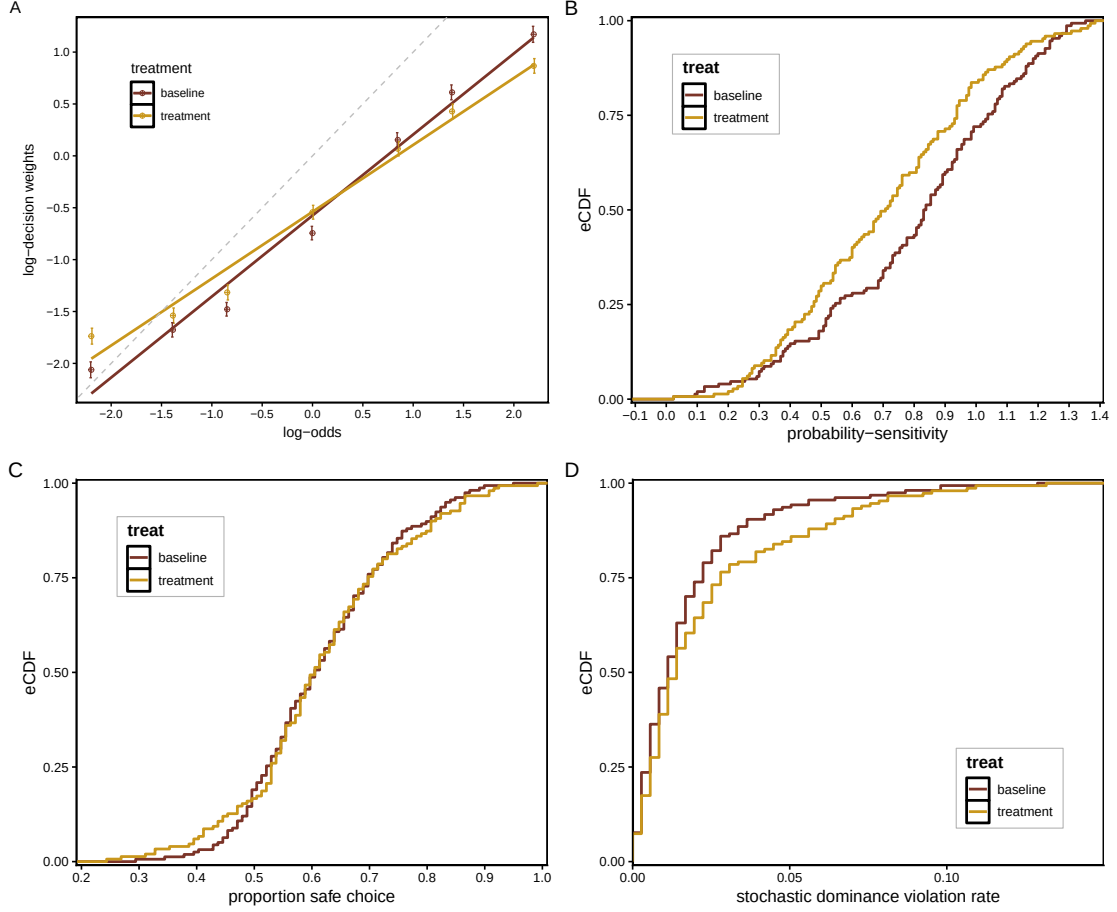


Figure 14: Choice behaviour in the behavioural benchmark experiment

Individual-level correlations. The treatment effects reported above differ sharply from the cross-sectional association documented by Pachur et al. (2018). Experimentally increasing the noisiness of probability representations raises relative attention to probabilities while reducing probability sensitivity. By contrast, in the absence of experimentally induced variation, Pachur et al. (2018) find that individuals who devote relatively more attention to probabilities also exhibit greater probability sensitivity. The model predicts that these patterns can coexist because they arise from different sources of variation. A causal increase in coding noise induces compensatory but incomplete sampling, whereas stable individual differences in prior uncertainty or signal precision can generate a positive association between relative attention and achieved sensitivity.

Panel A of Figure 11 examines the participant-level association between relative probability attention and probability sensitivity within each treatment condition. The treatment effects are clearly visible: relative probability attention shifts upward, whereas probability sensitivity shifts downward when probabilities are represented using larger numbers. Within each condition, however, the association between the two variables remains pos-

itive. The correlation is $r = 0.295$ ($p < 0.001$) in the baseline condition and $r = 0.474$ ($p < 0.001$) in the treatment condition. Thus, the experimentally induced movement across conditions runs in the opposite direction from the cross-sectional association observed within conditions. This is the pattern predicted by the model: greater coding noise causes compensatory but incomplete attention, whereas stable individual differences can generate a positive relation between relative attention and achieved sensitivity.

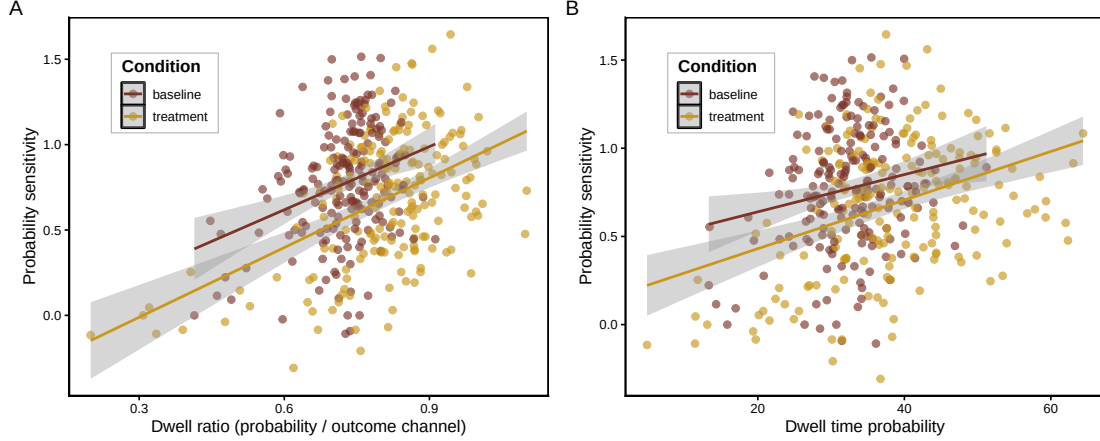


Figure 15: Attention and probability sensitivity within treatment conditions

Panel B examines the same association using a measure closer to that employed by [Pachur et al. \(2018\)](#): absolute attention to probability information, measured here as square-root-transformed probability dwell time in line with the model. The association is again positive in both conditions, with $r = 0.186$ ($p = 0.009$) in the baseline condition and $r = 0.370$ ($p < 0.0001$) in the treatment condition.

Taken together, Panels A and B resolve the apparent contradiction. Within each treatment condition, individuals who devote more attention to probability information exhibit greater probability sensitivity. Experimentally making probabilities harder to parse, however, has the opposite effect: it increases both relative and absolute attention to probabilities while reducing probability sensitivity. This joint pattern provides a direct test of the model's central mechanism of compensatory but incomplete sampling.

A.2 Experimental instructions and screenshots

Information & Consent

Thank you for taking part in this study. In the course of the study we will ask you to make choices between options entailing monetary payments with certain probabilities. In the end, you will have a 1/10 probability of being selected for a bonus. If selected, the bonus will be determined by your choice in one randomly selected decision. The precise procedure will be explained in the instructions.

During the study, you will need to use a mouse or trackpad. This is required because you will need a cursor to reveal information about the choice options. It is thus only possible to take part in this study using a laptop or desktop computer. If you are reading this on a tablet or phone, please leave this study and join again from a computer.

We require you to turn on your camera for the duration of the study. The camera serves to verify that a real person is present, by periodically checking that a face is detected in view. No other data will be captured or stored, and eye movements or gaze position are not tracked or recorded. In particular, at no time will we take or store pictures of you or your surroundings. On the following page, you will be asked for your explicit consent to turn on your camera for this purpose. If you do not consent to turning on your camera, you cannot participate in this study.

The current study follows ethical guidelines, and all information you are given is true. The study is conducted by Dr. Ferdinand Vieider at Ghent University, Belgium. In case of any questions or concerns, you can reach out via email at ferdinand.vieider@ugent.be, or alternatively, by anonymous chat on Prolific.

You can proceed by giving your informed consent to these conditions below. If you do not consent you may indicate so, and you will automatically leave the study.

I consent to take part

I do not consent

Webcam check

This study uses your device's webcam to verify that a real person is present, by periodically checking that a face is detected in view. **No video is recorded, stored, or uploaded** — all processing happens locally in your browser, and eye movements or gaze position are not tracked. We only store whether a face was detected at intervals during the task, not any images or video.

On the next screen your browser will ask for camera permission. Please click "Allow" to continue.

If you would prefer not to grant camera access, you can return your submission via the Prolific study page without penalty.

Enable camera and continue

Instructions

We will ask you to choose between binary lotteries and sure amounts of money. The lotteries offer a prize that is larger than the sure amount, but also have a downside that is smaller than the sure amount. The amounts involved are described in Pounds, and the probabilities are indicated as fractions throughout.

When you first see the problem, various amounts and probabilities are hidden from view. Move your mouse over a box to reveal its value; move it away to hide the value again. You can reveal the different values and probabilities in any order and as often and for as long as you wish. Once you have enough information, you can indicate your decision.

The positioning of the sure amount and lottery (left or right) is randomized. The position of the high and low outcome of the lottery (top or bottom) is also randomized. Please consider all the attributes of the choice problem carefully before indicating your choice: once a choice is made, you cannot go back.

Please carefully consider the information provided before indicating a choice: any choice could ultimately determine your bonus. If a given choice is selected to count towards your bonus payment, the computer will automatically determine your payment. If you have chosen the sure option, you will simply be paid that amount. If you have chosen the lottery, a ball will be extracted from an urn representing the winning probability. If a winning ball is extracted you will be paid the high outcome in the lottery; if not, you will be paid the lower outcome.

Once you continue from this page, you will see two practice trials. After these practice trials, the main trials start.

Start practice

Move your mouse over a box to reveal its value; move away to hide it again. Look at boxes in any order, as many times as you like. When you are ready, choose Option A or Option B.

Option A

£11

Choose A

Option B

Probability 1

Outcome 1

Probability 2

Outcome 2

Choose B