The Structure of Stochastic Choice*

Ferdinand M. Vieider[1]

[1]*RISLαβ, Department of Economics, Ghent University*

April 27, 2026

Abstract

Volatility in choices creates substantial identification problems for recovering stable preferences from risky choices. Recent work in economics has therefore developed process-based models that derive both average choice behaviour and stochastic variation from a coherent underlying decision process. This paper shows that choice variability can be used to discriminate between two broad classes of such models. In experimental tests, I find a pronounced inverse-U relationship between how easily choice options can be distinguished and choice variability. This pattern supports models based on Bayesian inference from noisy signals, and rejects models in which greater difficulty necessarily leads to more random choice.

Keywords: stochastic choice; Bayesian inference; risk-taking

JEL-classification: D81; C91

*Address: Sint-Pietersplein 6, 9000 Ghent, Belgium; `ferdinand.vieider@ugent.be`. A previous version of this manuscript was circulated under the name "Back to Thurstone? The Psychophysics of Stochastic Choice". I gratefully acknowledge funding by the Research Foundation–Flanders under grant nr. G008021N: "Causal determinants of preferences". This paper has benefited greatly from discussions with Larbi Alaoui, Miguel Ballester, Colin Camerer, Graham Loomes and Songfa Zhong. All errors remain my own.

1 Motivation

Behaviour is intrinsically variable. When individuals face the same choice repeatedly, their decisions often differ from one occasion to the next. To recover stable taste parameters from such data, researchers therefore typically augment deterministic decision models with an additive noise term that captures unexplained variation in choice. The mapping from stable tastes to stochastic choice frequencies implied by these models, however, is not generally unique, creating well-known identification problems for the underlying taste parameters. As a result, the same stochastic choice patterns can admit sharply different interpretations.

One way to address the disconnect arising from separate specification of the deterministic decision model and the stochastic choice rule is to model the decision process itself. A growing literature takes this approach by developing process-based models of choice. A central feature of these models is that both systematic choice patterns and choice variability arise from constraints on the information-processing architecture, rather than from separate preference and error components. Prominent examples include models in which limited representational resources are optimally allocated across expected outcomes (Robson, 2001b; Stewart et al., 2006; Netzer, 2009; Frydman and Jin, 2022), sequential evidence accumulation models (Ratcliff, 1978; Ratcliff and McKoon, 2008; Fudenberg et al., 2018), and Bayesian models of noisy cognition (Natenzon, 2019; Khaw et al., 2021; Vieider, 2024b; 2023).

I show that these models can be grouped into two broad classes based on the mechanisms that drive their stochastic choice predictions. These mechanisms imply sharply different relationships between choice variability and the fidelity of internal representations of choice-relevant quantities, holding economic value constant. This allows me to use systematic changes in choice variability across decision environments to distinguish between them. Far from being a nuisance, stochastic choice thus becomes a key diagnostic tool for discriminating between competing accounts of the decision process.

Random utility and preference identification. A central difficulty in random utility models is that deterministic preference parameters need not map uniquely into observed choice behaviour (Wilcox, 2011; Apesteguia and Ballester, 2018). From the opposite vantage point, the same observed stochastic choice patterns can be compatible with

multiple underlying taste parameters (e.g., different levels of risk aversion). Figure 1 illustrates this identification problem in a standard expected-utility (EU) framework with additive noise. Following Apesteguia and Ballester (2018), the figure plots the probability of choosing a lottery—assumed to exceed the sure option in expected value—as a function of the coefficient of relative risk aversion.[1]

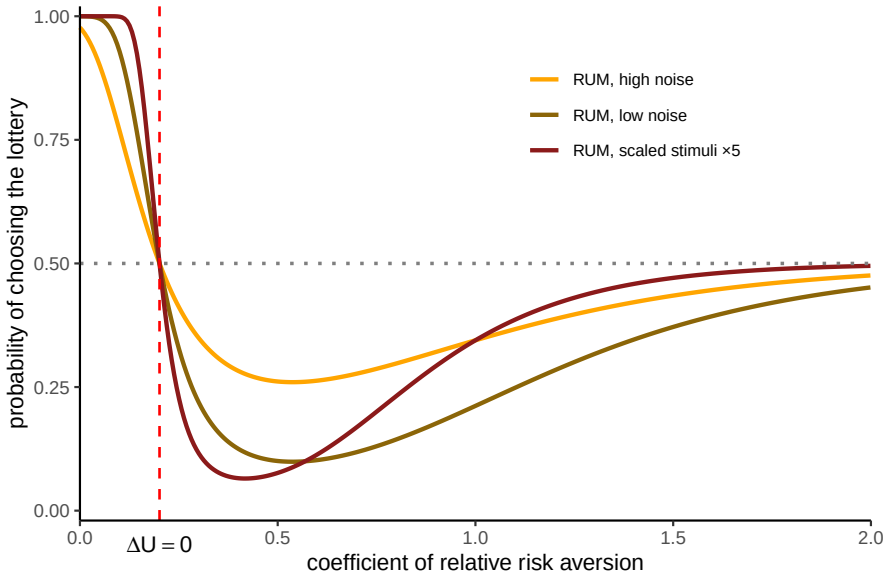


Figure 1: Choice proportion of lottery as a function of risk aversion

The simulations are based on a constant relative risk aversion (CRRA) utility function $u(x) = x^{1-r}$ for $r \neq 1$, and $u(x) = \ln x$ for $r = 1$. Stochastic choice is governed by the difference in random utilities $\widetilde{U}(x,p) = p\,u(x) + \epsilon_x$ and $\widetilde{U}(c,1) = u(c) + \epsilon_c$, where ϵ_x and ϵ_c capture unobserved disturbances. Assuming $\epsilon_x, \epsilon_c \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,\sigma^2)$ implies $\epsilon \triangleq \epsilon_x - \epsilon_c \sim \mathcal{N}(0, 2\sigma^2)$, yielding the Probit choice rule $\Pr\big[(x,p) \succ c\big] = \Phi\Big(\frac{p\,u(x)-u(c)}{\sqrt{2}\,\sigma}\Big)$. Choices are generated for a baseline comparison between $(60, 0.2)$ and $(8, 1)$. In the "low noise" simulation, $\sigma = 1$, and in the "high noise" simulation $\sigma = 2$. The scaled-stimuli simulation uses the same noise level as the high-noise case, but multiplies all outcomes by a factor of five, corresponding to a choice between $(300, 0.2)$ and $(40, 1)$. ΔU indicates the difference in deterministic utilities, $U(x,p) - U(c,1) = pu(x) - u(c)$. The qualitative non-monotonicity does not depend on either the EU or the normality assumptions, but reflects a general feature of the interaction between utility curvature and additive noise.

A risk-neutral decision-maker will prefer the lottery to the sure option, since the former has higher expected value. As risk aversion increases, outcomes become compressed and the utility difference between the options declines, yielding an initial decline in the choice probability of the risky option. As risk aversion continues to rise, however, the deterministic expected-utility difference between the lottery and the sure option continues to shrink, eventually becoming small relative to the noise term. Noise starts to dominate

[1] One interpretation of this plot is as representing a population of decision-makers who face the same objective choice problem and share the same variance of internal noise, but differ in the curvature of their utility functions. While EU and a Probit specification are adopted here for concreteness in the simulation, the qualitative patterns illustrated in the figure are a general feature of random utility models.

the deterministic utility difference, the choice options become harder to discriminate, and choice probabilities drift back toward stochastic indifference. Importantly, the same non-monotonicity in stochastic choice can arise even when preferences and noise are held fixed (i.e. for one and the same decision-maker) but choice primitives are scaled.[2] This illustrates a general implication of the random utility framework: as options become harder to discriminate, choice variability increases monotonically.[3]

This raises a natural question: how should one interpret the relationship between discriminability and choice variability highlighted above? In particular, should the resulting patterns be viewed as behavioural predictions of the model, or as artefacts of how stochastic choice is specified? Random utility models are typically interpreted as reduced-form representations of choice behaviour: they are meant to capture statistical regularities in observed choices without, by themselves, pinning down a unique behavioural or structural interpretation. Because the deterministic decision model and the stochastic choice specification are chosen independently, the relationship between discriminability and variability is not structurally grounded within this framework. It therefore provides no disciplined account of whether and how noise should vary as options become harder to distinguish—the one cannot inform the other without imposing additional assumptions that are not themselves implied by the model.

Stochastic Choice from Noisy Representations. Process-based models address this ambiguity by deriving both average behaviour and choice variability from a common underlying decision process, rather than specifying them independently. Studying stochastic choice through the lens of such models thus has two key advantages. First, these models provide stylized representations of the mechanisms generating choice behaviour, and therefore admit a generative—i.e. causal—interpretation. As a result, their

[2]For illustration, Figure 1 also plots choice probabilities for a rescaled version of the same choice problem, in which all outcomes are multiplied by a constant factor. Under constant relative risk aversion, such rescaling leaves deterministic preferences unchanged. Nevertheless, it alters the discriminability of the options relative to the fixed noise level, leading to markedly different stochastic choice behaviour. Analogous reversions toward stochastic indifference can also arise at very small stakes, which may be hard to discriminate even at low levels of risk aversion.

[3]Several contributions address the non-monotonic relationship between risk aversion and stochastic choice probabilities in standard random utility specifications. In particular, Wilcox (2011) and Apesteguia and Ballester (2018) propose approaches that restore monotonicity of choice probabilities under expected utility. These contributions are complementary to the present discussion. Importantly, even when such monotonicity is imposed, the qualitative implication that choice variability increases as options become harder to discriminate remains a feature of those stochastic choice frameworks. A more detailed discussion is provided in Online Appendix B.

predictions are behavioural in nature, rather than reduced-form, as in standard random utility models. Second, because both systematic choice patterns and stochastic variation originate from the same cognitive frictions in information processing, these models impose nontrivial structure on stochastic choice itself. This structure will be central to the analysis that follows.

Models proposed within this general framework fall into two sharply distinct classes, which differ in how noisy internal representations of choice attributes are transformed into choices. One class of models addresses noise primarily at the *encoding* stage, by optimally shaping internal representations given environmental constraints (usually given by the statistical distribution of choice-relevant quantities). A second class of models also exploits information about environmental distributions, but in a different way: prior statistical knowledge is used to optimally *decode* noisy representations in order to infer the underlying choice primitives that generated them. Although these two approaches to managing noise in internal representations are not mutually exclusive—and may well coexist in practice—existing models typically emphasize one or the other mechanism. As a result, the two model classes imply sharply different relationships between discriminability (i.e. how easily choice options can be distinguished) and choice variability.

Encoding-based models address representational noise at the encoding stage, by optimally shaping internal representations given the structure of the environment. This class includes evolutionary models such as those proposed by Robson (2001a;b) and Netzer (2009), the Decision-by-Sampling model of Stewart et al. (2006), as well as the efficient coding approach of Frydman and Jin (2022). They also include a broad class of evidence accumulation models (Ratcliff, 1978; Fudenberg et al., 2018). Although these models differ in their assumed sources of internal noise and in the optimality criteria used to derive representational codes, they share a common implication for stochastic choice: As the discriminability between options declines, choices are predicted to become increasingly variable, approaching stochastic indifference in the limit. In this respect, encoding-based models share a qualitative implication with standard random utility models, but these can now be unambiguously be interpreted as behavioural predictions. This follows directly from the emphasis these models place on the informativeness of internal signals: when internal representations of choice options carry relatively little signal about their relative attractiveness, responses necessarily become more random.

Decoding-based models, by contrast, rely on Bayesian inference to mitigate the impact of representational noise. Noisy internal signals are weighted in proportion to their informativeness: unreliable or extreme signals are discounted and pulled toward prior expectations, yielding inferences that optimally trade off signal and prior information. This class includes, for example, the model of small-stakes risk aversion proposed by Khaw et al. (2021), as well as related Bayesian approaches to decision-making under uncertainty (e.g., Natenzon, 2019). A central implication of this process is regression to the mean of the prior. As signals become less reliable, they receive progressively less weight in decisions, which are instead increasingly guided by prior expectations. This stabilizing property has sharp implications for stochastic choice. Rather than increasing monotonically, choice variability is predicted to peak at intermediate levels of discriminability and to decline again when noise overwhelms the informational content of the signal. In other words, Bayesian decoding predicts an inverse-U shaped relationship between discriminability and stochastic choice—the opposite of the prediction made by encoding-based models.

Testing Predictions of Encoding and Decoding Models. I test the contrasting predictions of encoding- and decoding-based models by systematically manipulating outcome discriminability in a lottery choice experiment. The key manipulation consists in varying the numerical scale in which monetary outcomes are denominated, by mapping payoffs into experimental currency units. As shown by Garagnani and Vieider (2025), such scale transformations can systematically shift decision noise in predictable ways. Oprea and Vieider (2025) use an analogous manipulation to demonstrate how changes in numerical scale affect sensitivity to probabilities and time delays. By employing a wide range of exchange rates between GBP and experimental currency units—specifically, mappings of 1:1, 1:100, and 1:10,000—I generate substantial variation in outcome discriminability while holding underlying preferences fixed. This design induces controlled changes in the fidelity of internal outcome representations, allowing me to directly test how choice variability responds as discriminability deteriorates.

The experimental manipulation yields a clear and robust pattern. Outcome discriminability varies systematically across the different payoff scales, confirming that the manipulation is successful and—importantly—that it generates substantial variation in the signal-to-noise ratio of internally represented choice primitives. As outcome discrim-

inability is reduced by rescaling monetary payoffs, choice variability initially increases. Once outcome representations become dominated by noise, however, choice variability begins to decline again. In other words, stochastic choice exhibits a pronounced inverse-U shaped relationship with discriminability. Importantly, I show that this pattern emerges both in econometric tests and in nonparametric analyses designed to provide an assessment without imposing strong structural assumptions.

Contribution and Relation to the Literature. This paper builds on a growing literature that models decision-making as arising from noisy internal representations of choice-relevant quantities. A closely related contribution is Natenzon (2019), who proposes a Bayesian inference model of stochastic choice to explain attraction and compromise effects. Khaw et al. (2021) propose and test a model in which small-stake risk aversion emerges from noise in the representation of choice primitives. Several papers independently propose explanations of probability weighting as an outgrowth of noisy representations (Zhang et al., 2020; Enke and Graeber, 2023; Frydman and Jin, 2023; Oprea, 2024; Oprea and Vieider, 2024; Vieider, 2024b). Similar principles have been applied to explain choices over mixed gain-loss gambles (Bouchouicha et al., 2025), and deviations from exponential discounting benchmarks (Vieider, 2023; Enke et al., 2024; Oprea and Vieider, 2025). While several of these papers propose choice rules grounded in Bayesian inference, their empirical focus lies on average choice patterns or specific anomalies, rather than on the structure of stochastic choice itself.

The paper is also closely related to a literature emphasizing the limits of inferring preferences from observed choice behaviour (Ballinger and Wilcox, 1997; Loomes, 2005). Nielsen and Rehbeck (2022) allow subjects to choose whether to commit to rational-choice axioms and study how deviations from these principles are reconciled in subsequent choices, with the aim of distinguishing tastes from mistakes. McGranaghan et al. (2024) examine common ratio violations using both binary choices and certainty equivalents, and show that such violations in binary choice can arise even under expected utility when stochastic choice is taken into account. de Clippel et al. (2024) compare behaviour in settings with experimentally induced demand to decision-making under risk, finding strong correlations across contexts that suggest a prominent role for heuristics or as-if rationality rather than stable utility maximization. My results are closely related to this literature in that they likewise support the view that observed choices may reflect mis-

takes. The contribution here, however, is to focus on the structure of choice variability itself as a diagnostic tool.

Finally, the paper is related to work that studies stochastic choice using additional observables beyond choice itself. A literature in psychology emphasizes the role of decision times and their relationship to processes of evidence accumulation (Krajbich et al., 2010), with related applications to decision-making under risk (Busemeyer and Townsend, 1993; Zilker and Pachur, 2022). Recent work in economics has similarly leveraged response times to improve the identification of choice models and stochastic components (Alós-Ferrer et al., 2021), or to enhance out-of-sample prediction of risky choices (Alós-Ferrer and Garagnani, 2024). The present contribution is complementary to this literature, focusing on the structure of choice variability across tasks rather than on auxiliary process measures such as response times.

2 Encoding-based models of stochastic choice

In a random utility model (RUM), discriminability between options and stochasticity in choice are modelled independently, and the model remains silent about the cognitive origin of the noise. As a result, the same observed choice behaviour can be attributed either to changes in preferences or to changes in unobserved disturbances, with no principled way to adjudicate between the two. This ambiguity suggests examining stochastic choice from the perspective of explicitly generative models. In such models, both discriminability and stochastic choice arise endogenously from assumptions about how noise enters internal representations of choice stimuli, creating a direct link between discrimination and stochasticity.

2.1 The psychophysics of noisy internal representations

The origin of the random utility model can be traced to Thurstone (1927a;b). Thurstone developed his framework to model discrimination as the outcome of noisy internal representations, building on earlier work by Weber and Fechner. I describe this framework here not for historical reasons, but because its core equations provide a common blueprint for the stochastic choice predictions of the models examined in this paper.

Consider two objects, denoted x and c. In Thurstone's original examples these might

correspond to two physical weights, with the task being to judge which object is heavier. For concreteness, suppose that $x > c$. Thurstone posits that physical magnitudes are mapped into *internal psychophysical representations*, and that these representations are inherently noisy. Let the internal representations of x and c be denoted by r_x and r_c, respectively. His framework assumes

$$r_x = f(x) + \varepsilon_x, \qquad r_c = f(c) + \varepsilon_c,$$

where $f(\cdot)$ is a deterministic psychophysical mapping and ε_x and ε_c are stochastic disturbances. Representational noise is Gaussian and is assumed independent across objects,[4] with $\varepsilon_x \sim \mathcal{N}(0, \sigma_x^2)$ and $\varepsilon_c \sim \mathcal{N}(0, \sigma_c^2)$.[5]

Given this structure, Thurstone's *law of comparative judgment* takes the form

$$\Pr(r_x - r_c > 0) = \Phi\left(\frac{f(x) - f(c)}{\sqrt{\sigma_x^2 + \sigma_c^2}}\right),$$

where $\Pr(r_x - r_c > 0)$ denotes the probability of correctly identifying the heavier object (in the running example) on any given trial, and $\Phi(\cdot)$ is the standard normal cumulative distribution function. From a purely formal point of view, the standard Probit specification of a random utility model is obtained by assuming homoscedastic noise ($\sigma_x = \sigma_c \triangleq \sigma$) and by replacing the psychophysical mapping with a utility-based one, $f(x) = U(x, p)$ and $f(c) = U(c, 1)$.

Notwithstanding this formal similarity, the key difference is that in Thurstone's framework both discriminability in the numerator and judgment noise in the denominator arise jointly from noisy internal representations. Judgment accuracy is therefore governed by two components: the difference between the internal codes, $f(x) - f(c)$, and the precision of these codes, captured by the inverse of $\sigma_x^2 + \sigma_c^2$. Larger differences in physical attributes (as mapped through $f(\cdot)$) and more precise internal representations yield more accurate judgments. Conversely, stimuli that lie close together in psychophysical

[4] I maintain this assumption of independence here and throughout. Note, however, that independence is not required: Thurstone (1927a) explicitly discusses a version of the model with correlated noise. Such correlations may improve predictions in certain contexts; see Natenzon (2019) for a prominent example.

[5] A key conceptual point—often misunderstood in later economic applications—concerns the role of the normality assumption. For Thurstone, the Gaussian distribution does *not* describe the distribution of external physical stimuli such as x and c. Rather, it reflects an assumption about the distribution of *internal representational noise*. Normality is therefore a property of the encoding process, not of the environment itself.

space will be discriminated unreliably: even when $x > c$, mistakes occur with positive probability.

2.2 Encoding-based models of stochastic choice

Thurstone's work was concerned with the discrimination of physical attributes. Extending this framework to economic choice problems—such as decisions under risk—requires comparing objects that are themselves composed of multiple elements, including outcomes and probabilities. This introduces an additional layer of structure. To fix ideas, I focus on models in which the allocation of cognitive resources—and hence the structure of representational noise—is endogenously tied to the decision environment. In particular, I consider two prominent and closely related approaches developed for decisions under risk: the Decision-by-Sampling model of Stewart et al. (2006), and the efficient-coding account of Frydman and Jin (2022). The underlying logic, however, is more general and extends beyond these specific formulations.

Decision-by-Sampling (DbS) models subjective value as arising from comparisons of the stimulus to a finite sample of draws from memory. For continuity with the previous discussion, I focus only on outcome encoding (probability encoding is analogous). Let the environmental distribution of outcomes stored in memory be F. When the decision maker evaluates an outcome x, she compares it to N samples from memory, $Y_1, \ldots, Y_N$, drawn i.i.d. from F. The internal representation of x is then its empirical rank,

$$r_x \;=\; \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}\{x > Y_i\}.$$

This construction provides a sampling-based microfoundation for Thurstone's internal representations. Conditional on x, the indicators $\mathbf{1}\{x > Y_i\}$ are Bernoulli random variables with success probability $F(x)$. By the law of large numbers, r_x converges to $F(x)$ as N goes to infinity. For finite N, sampling variability implies that

$$r_x = F(x) + \varepsilon_x, \qquad \varepsilon_x \sim \mathcal{N}\left(0,\; \frac{F(x)\big[1 - F(x)\big]}{N}\right),$$

where the approximate normality of ε_x follows from the central limit theorem applied to the sum of N Bernoulli trials (cf. Casella and Berger, 2024, ch. 5).

Evaluating a lottery (x, p) against a sure payoff c yields a stochastic choice rule of the familiar Probit form:

$$\Pr\big[(x,p) \succ c\big] = \Phi\left(\frac{p\,F(x) - F(c)}{\sqrt{\sigma_x^2 + \sigma_c^2}}\right),$$

where $\sigma_x^2 \triangleq \frac{F(x)\big[1-F(x)\big]}{N}$ and $\sigma_c^2 \triangleq \frac{F(c)\big[1-F(c)\big]}{N}$. This yields the Thurstone discriminability model as a special case, but with rank in memory replacing physical intensity, and with binomial sampling variance replacing the arbitrary Gaussian noise term of the random utility Probit.

The stochastic-choice implications of this representation follow directly from the properties of the sampling variance. When outcomes fall into sparsely populated regions of the environmental distribution F, they become difficult to discriminate: small differences in stimulus magnitude translate into only weak differences in expected rank. At the same time, the binomial sampling variance remains substantial. As a result, the signal-to-noise ratio in the internal representations deteriorates, and sampling noise increasingly dominates the comparison of internal values. Because sampling noise is the sole source of stochasticity in the model, this implies that choice probabilities drift back toward stochastic indifference as discriminability declines. The predicted pattern mirrors the non-monotonic stochastic-choice implications of the RUM discussed above.

Efficient coding and stochastic choice. Efficient coding is a foundational concept in neuroscience, describing how scarce neural resources optimally adapt to the statistical structure of stimuli in the environment (Barlow, 1961; Laughlin, 1981). Related ideas have been formalized in cognitive science through rate–distortion approaches to optimal encoding under capacity constraints (e.g., Sims, 2016). More recently, efficient-coding principles have been formalized in models of risky choice, most prominently in the efficient-coding account developed by Heng et al. (2020) and adapted to decision making under risk by Frydman and Jin (2022). Here, I focus on this latter application.

Attributes of a lottery—such as the prize x in (x, p) or the sure amount c—are internally represented by a noisy population code. In the efficient-coding framework of Frydman and Jin (2022), a stimulus x is encoded by a population of n binary neurons, each firing independently with probability $\theta_X(x)$; similarly, c elicits firing probability $\theta_C(c)$.

The internal codes are the resulting spike counts, $r_x = \sum_{i=1}^{n} s_i$ and $r_c = \sum_{i=1}^{n} t_i$, with $s_i \sim \text{Bernoulli}(\theta_X(x))$ and $t_i \sim \text{Bernoulli}(\theta_C(c))$. This, in turn, implies that r_x and r_c will follow binomial distributions governed by probabilities $\theta_X(x)$ and $\theta_C(c)$.

Efficient coding determines the functions $\theta_X(\cdot)$ and $\theta_C(\cdot)$ by optimally allocating coding precision across the environmental distributions F_X and F_C. This induces a smooth, non-linear compression of the stimulus magnitudes. The functions $v_X(x)$ and $v_C(c)$ denote the deterministic posterior means implied by this coding scheme.[6] Because the underlying spike counts are binomial, the decoded values inherit approximately Gaussian noise via a standard delta-method argument (cf. Casella and Berger, 2024, ch. 5). The internal codes can therefore be written in the same form as in DbS and Thurstone:

$$r_x = v_X(x) + \varepsilon_x, \qquad r_c = v_C(c) + \varepsilon_c,$$

with $\varepsilon_x \sim \mathcal{N}\big(0, \sigma_X^2(x)\big)$ and $\varepsilon_c \sim \mathcal{N}\big(0, \sigma_C^2(c)\big)$, where $\sigma_X^2(x) \triangleq n\,\theta_X(x)\big[1 - \theta_X(x)\big]$ and $\sigma_C^2(c) \triangleq n\,\theta_C(c)\big[1 - \theta_C(c)\big]$.[7] Defining the associated decision-stage values as $V(x, p) = p\,v_X(x)$ and $V(c, 1) = v_C(c)$, yields a choice probability of the Thurstone-RUM form:

$$\Pr\big[(x, p) \succ c\big] \approx \Phi\left(\frac{p\,v_X(x) - v_C(c)}{\sqrt{\sigma_X^2(x) + \sigma_C^2(c)}}\right),$$

where $\sigma_X^2(x)$ and $\sigma_C^2(c)$ reflect the efficient-coding-induced neural noise.

The efficient-coding model thus fits naturally into the classical random-utility framework and shares the qualitative stochastic-choice implications of Decision-by-Sampling. As outcomes move into sparsely populated regions of the environmental distribution, optimal representational compression maps given objective differences into very small differences in decoded values. Discriminability therefore deteriorates: changes in stimulus magnitude translate into only weak differences in internal representations, while

[6] Although framed in Bayesian terms, the efficient-coding model of Frydman and Jin (2022) effectively behaves as a likelihood-based model at the level of choice. The assumed uniform prior is an improper prior in Bayesian statistics, and on a bounded stimulus domain it implies that the posterior is proportional to the likelihood. As a result, posterior uncertainty is not propagated into the decision stage, and Bayesian regression to the mean does not impact the stochastic choice rule.

[7] The variance expressions in Decision-by-Sampling and efficient-coding models differ only by normalization. In Decision-by-Sampling, internal representations correspond to normalized ranks, yielding a variance proportional to $F(x)\big[1 - F(x)\big]/N$. In efficient-coding models, representations are typically written in terms of unnormalized sample counts or neural activity, leading to a variance proportional to $n\,\theta_X(x)\big[1 - \theta_X(x)\big]$. Both formulations arise from Binomial sampling and are equivalent up to scale.

representational noise remains substantial. As a result, the signal-to-noise ratio in internal representations falls, and stochastic fluctuations increasingly dominate the comparison of values. This implies that choice probabilities drift toward stochastic indifference as discriminability declines, mirroring the implications of Decision-by-Sampling and the random-utility model.

Other encoding-based accounts. The models discussed above do not exhaust the class of encoding-based models. For instance, the evolutionary accounts of Robson (2001a;b) and Netzer (2009) provide prominent examples of other encoding-based approaches. In these models, decision-makers face potentially infinitely many consumption levels but can assign utility only in discrete steps, reflecting cognitive limits akin to Weber's just-noticeable differences. From an evolutionary perspective, optimality then requires allocating utility jumps in proportion to the probability density of consumption opportunities in the environment, with the precise criterion depending on whether one models the minimization of decision errors (Robson, 2001a;b) or the maximization of evolutionary fitness (Netzer, 2009). Although formulated at a more abstract level, these models yield the same qualitative stochastic-choice implications as the accounts considered above: as discriminability deteriorates in low-density regions of the environment, choice probabilities revert toward stochastic indifference.

There exists, furthermore, a close connection to evidence-accumulation models. Decision field theory (Busemeyer and Townsend, 1993), in particular, provides an early application to risky choice that microfounds a random utility model through sequential evidence accumulation. More generally, Webb (2019) shows that a broad class of such models—including the standard drift-diffusion model (Ratcliff, 1978) and optimal stopping models such as Fudenberg et al. (2018)—map into heteroscedastic random utility specifications at the level of choice. As the difference in value between options shrinks, the drift of the accumulation process declines, slowing decisions and increasing the influence of noise. Because stopping times are endogenous, low discriminability leads both to longer decision times and to a higher likelihood of mistakes, as choices become increasingly noise-driven. Consequently, these models again imply a monotone relationship between discriminability and choice variability, consistent with standard random utility models.

3 Decoding-based models of stochastic choice

I next examine Bayesian inference models. In contrast to encoding-based accounts, stochastic choice in this class arises primarily at the *decoding* stage: noisy internal representations are not acted upon directly, but are combined with prior knowledge to form posterior beliefs. A central implication of Bayesian decoding is regression toward prior expectations. As internal signals become less informative, posterior beliefs place increasing weight on the prior, leading behaviour to stabilize rather than becoming arbitrarily noisy. To isolate this mechanism as transparently as possible, I will assume homoscedastic coding errors and abstract from optimal adaptation of encoding noise across the stimulus space. This restriction is not inherent to Bayesian models; it serves to highlight the distinctive stochastic-choice implications generated by Bayesian inference itself.

3.1 Bayesian inference restores stochastic monotonicity

3.2 Bayesian inference restores stochastic monotonicity

Khaw, Li and Woodford (2021) (henceforth: *KLW*) develop a neural model of lottery evaluation in which discriminability between outcomes and stochasticity in choice arise jointly from noisy numerical magnitude perception combined with Bayesian inference. The model builds on the idea that numerical magnitudes are represented in a compressed manner, reflecting efficient neural coding under biological constraints. Outcomes x are encoded as

$$r_x \;=\; \ln(x) + \varepsilon_x, \qquad \varepsilon_x \sim \mathcal{N}(0, \nu^2).$$

This fits the Thurstone template $r_x = f(x) + \varepsilon_x$ with $f(x) = \ln(x)$, which can be viewed as a neural implementation of Fechner's law.[8]

The model diverges from the approaches considered so far by introducing a second stage that leverages Bayesian decoding to mitigate perceptual noise. The decision maker is assumed to hold a prior belief over log magnitudes, $\ln(x) \sim \mathcal{N}(\mu, \tau^2)$,[9] and to decode the noisy internal code r_x using Bayes' rule. The resulting posterior mean serves as the

[8]The Gaussian noise term is motivated by the tuning curves of number neurons, which exhibit approximately constant variance in log space (Dehaene, 2003).

[9]Under Gaussian noise in log space, a normal prior over $\ln(x)$ is the conjugate representation of beliefs learned from repeated noisy inference. Importantly, this does not require the true environmental distribution of x to be log-normal. Rather, it reflects a reduced-form characterization of subjective beliefs that arise when a decision-maker learns about magnitudes through noisy internal representations.

internal value representation of x and is given by

$$m_x \triangleq \mathbb{E}[\ln(x) \mid r_x] = \alpha\, r_x + (1-\alpha)\,\mu, \qquad \alpha \triangleq \frac{\tau^2}{\tau^2+\nu^2}.$$

The posterior mean is a weighted average of the noisy signal and the prior mean, where the weight on the signal, α, depends on the relative reliability of the two sources of information. When coding noise is low ($\nu^2 \to 0$), $\alpha \to 1$ and the posterior closely tracks the true stimulus (since $\mathbb{E}[r_x] = \ln(x)$). Conversely, when coding noise is high ($\nu^2 \to \infty$), $\alpha \to 0$ and the posterior collapses toward the prior mean. This Bayesian decoding rule is optimal in the sense that it minimizes the mean squared error across repeated trials (see, e.g., Bishop, 2006; Ma et al., 2023).

Because m_x is a linear transformation of the noisy code r_x, it remains stochastic even after Bayesian decoding. Conditional on x, the expectation and variance of the posterior mean m_x across repeated presentations of the same outcome take the form

$$\mathbb{E}[m_x \mid x] = \alpha\, \ln(x) + (1-\alpha)\,\mu, \qquad \mathbb{V}[m_x \mid x] = \alpha^2\nu^2 = \tau^2\,\alpha\,(1-\alpha),$$

where the second equality on the right follows from the definition of α. Thus, while Bayesian decoding attenuates perceptual noise, it does not eliminate variability in internal value representations. Importantly, the variance of the decoded representation depends non-monotonically on discriminability: it is small when α is close to 1 (the signal dominates), also small when α is close to 0 (the prior dominates), and maximized at the intermediate point $\alpha = \frac{1}{2}$, where prior and signal receive equal weight. This inverted-U relationship between representational variability and discriminability is the key property that distinguishes Bayesian decoding from the encoding-based accounts discussed above.

To connect this representational structure to stochastic choice, it is useful to write the implied decision rule for a simple binary lottery. Following Khaw et al. (2021), consider a lottery (x, p) evaluated against a sure payoff c. The decision maker compares the posterior mean of the sure amount, m_c, to the sum of the log-probability and posterior log-value of the risky payoff, $\ln p + m_x$. Under the assumptions above (and treating

probabilities as encoded without noise), this yields the probit choice rule

$$\Pr[(x,p) \succ c] = \Phi\left(\frac{\ln p + \alpha \ln\left(\frac{x}{c}\right)}{\sqrt{2}\,\alpha\nu}\right). \tag{1}$$

Equation (1) has the overall form of a probit random-utility model, but with a crucial difference: both the mean perceptual difference in the numerator and the variability in the denominator are jointly determined by coding noise ν^2 and prior variance τ^2 through the discriminability parameter $\alpha = \tau^2/(\tau^2 + \nu^2)$.

This structure has two key implications for our present purposes. First, as emphasized by Khaw et al. (2021), outcome discriminability α plays the role of an as-if risk-attitude parameter: holding (μ, τ^2) fixed, increasing coding noise ν^2 lowers α, thereby increasing apparent risk aversion.[10] Second—and more importantly for this paper—the model makes a non-trivial prediction about stochastic choice. Holding prior variance τ^2 fixed, increasing coding noise ν^2 initially raises choice variability by pushing α toward $\frac{1}{2}$ (when $\nu = \tau$), where representational variance is maximized. Beyond this point, further increases in coding noise reduce choice variability, as posterior beliefs become increasingly dominated by the prior. In the limit as $\alpha \to 0$, the posterior collapses onto the prior mean and choices become more, not less, coherent.

To make the relationship between discriminability and decision noise explicit, Figure 2 plots the implied decision-noise term $\tau\sqrt{2\,\alpha\,(1-\alpha)}$ as a function of outcome discriminability α. Panel A displays the characteristic inverted-U shape: decision noise is small when outcomes are either highly discriminable ($\alpha \approx 1$) or strongly dominated by prior expectations ($\alpha \approx 0$), and is maximized at the intermediate point $\alpha = \frac{1}{2}$ where prior and signal receive equal weight.

Panel B illustrates the corresponding implications for lottery choice proportions. In contrast to the random-utility case examined above, the predicted choice probability of the risky option now decreases monotonically in coding noise ν (which, for fixed τ, directly governs as-if utility curvature through α). This monotonicity arises precisely because decision noise declines again at low discriminability: Bayesian regression toward the prior prevents noise from overwhelming the as-if utility difference in the numerator.

[10] To see this, note that the expression $\ln(p) + \alpha\ln(x) - \alpha\ln(c)$ is a monotonic transformation of $px^\alpha - c^\alpha$, so that α governs the curvature of the implied utility function in a manner analogous to constant relative risk aversion.

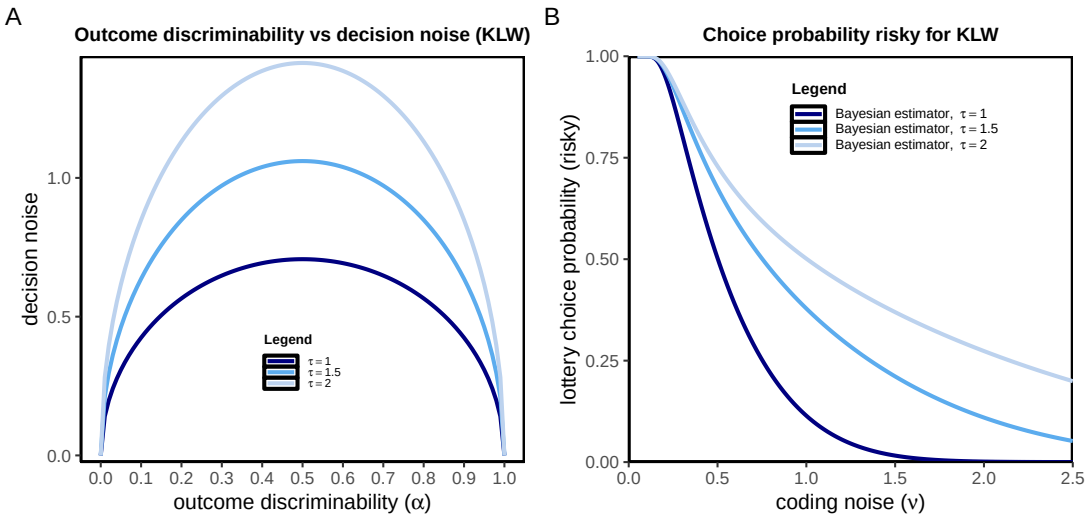


Figure 2: Stochastic choice predictions under Bayesian inference (KLW)

The figure illustrates the effect of Bayesian inference on stochastic choice behaviour. Panel A shows the relationship between discriminability α on the horizontal axis, and decision noise $\sqrt{2}\nu\alpha$ on the vertical axis. Decision noise is highest at $\alpha = \frac{1}{2}$, and declines for smaller and larger values of α. Panel B plots coding noise ν on the horizonal axis against predicted choice proportions of the superior risky option on the vertical axis. As ν increases as-if risk aversion $1 - \alpha$ also increases; nevertheless, the choice probability of the higher-EV risky option declines monotonically.

As a result, the predicted choice probability of the lottery declines monotonically with ν. The slope of this decline depends on the prior standard deviation τ, with larger values of τ producing both higher decision noise (panel A) and a slower decrease in lottery choice probabilities (panel B).

This stands in sharp contrast to the likelihood-based models discussed above. In the random utility model, discriminability depends solely on the distance between deterministic utilities: choices become noisy when utility differences are small, and converge toward indifference as curvature compresses those differences. In Decision-by-Sampling and efficient-coding models, by contrast, discriminability is shaped by the mapping from objective magnitudes to internal representations, which reflects the statistical structure of the environment. Regions of the stimulus space in which internal codes are densely packed—or in which the mapping flattens—exhibit low discriminability, even when objective differences are substantial. The Bayesian inference model departs from both frameworks. Because Bayesian regression to the mean suppresses noise when signals are uninformative, choice behaviour stabilizes as α becomes small, restoring a form of stochastic monotonicity in choice proportions that the other generative models examined here do not predict.

A subtle but important point concerns the behaviour of the prior variance τ^2. Within the logic of the KLW framework the prior is not externally fixed: it must itself be learned from noisy posterior inferences. Because the organism infers environmental statistics through the same noisy perceptual channel used for individual decisions, the precision of the learned prior will naturally depend on the coding noise ν. For the behavioural predictions developed above, however, only the relative magnitude of signal and noise matters, as captured by the outcome discriminability parameter $\alpha = \tau^2/(\tau^2 + \nu^2)$. Thus, whenever coding noise increases more rapidly than the learned prior variance (thereby reducing α)—or when individuals differ in their signal-to-noise ratios α—the model robustly yields maximal decision variability at $\alpha = \frac{1}{2}$ and stabilization of choices in the high-noise regime. This observation will be important for interpreting the econometric flexibility of the model in the next subsection.

3.3 Empirical test strategy

Figure 2 highlights a second feature of the KLW framework that informs the model-based tests proposed below (which are complemented by nonparametric tests). For a given prior standard deviation τ, variation in coding noise ν jointly determines as-if utility curvature (through $\alpha = \tau^2/(\tau^2 + \nu^2)$) and decision noise $\tau\sqrt{2\,\alpha\,(1-\alpha)}$, generating the characteristic inverted-U relation between them. Thus, within the KLW model with fixed τ, risk attitudes and choice variability are tightly linked: changing ν necessarily affects both discriminability and decision noise in a specific way.

From a purely econometric perspective, however, ν and τ are free parameters. This flexibility allows τ to co-move with ν in such a way that α remains approximately constant while decision noise, $\tau\sqrt{2\,\alpha\,(1-\alpha)}$, increases or decreases. More generally, by allowing both ν and τ to vary freely across individuals or treatments, an econometric estimation of the KLW probit specification can dissociate as-if risk attitudes (captured by α) from response variability (captured by $\tau\sqrt{2\,\alpha\,(1-\alpha)}$). In particular, the probit specification may yield estimates of decision noise that increase monotonically as discriminability falls, even though this is *not a behavioural prediction* of the model. In this sense, the econometric application of the model can reveal empirical patterns consistent with either encoding- or decoding-based accounts, even though its behavioural predictions are fundamentally decoding-based, reflecting the assumption that τ does not increase as rapidly

as ν.

This limiting case has a natural interpretation. The first stage of the KLW framework corresponds to a homoscedastic version of the Thurstone model. As $\tau \to \infty$, the prior variance becomes unbounded, effectively yielding an improper (flat) prior. In this limit, the posterior becomes proportional to the likelihood, and Bayesian decoding collapses to a purely encoding-based estimator.

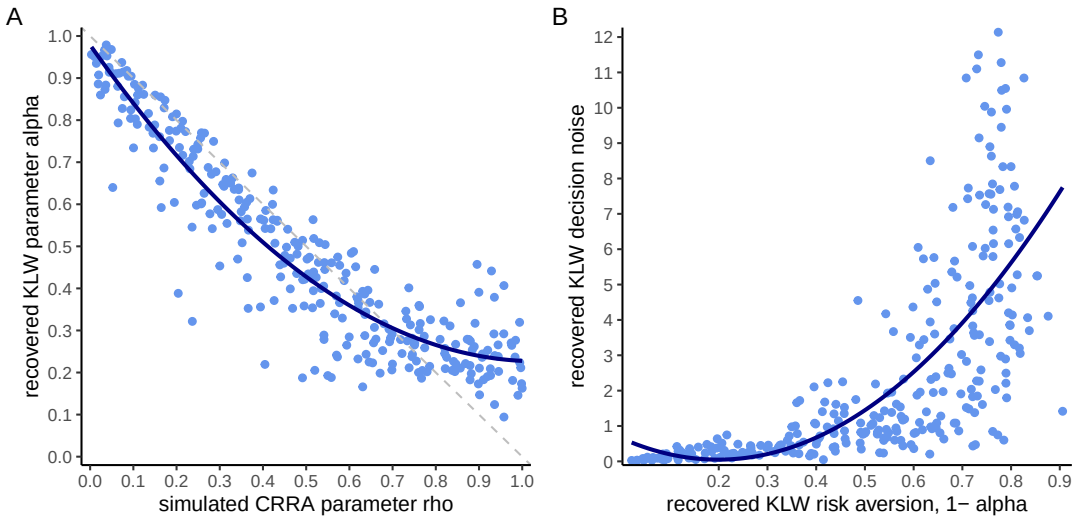


Figure 3: Effects of visual display for rewards in pounds versus pence

Simulations shown in the figure are based on an expected utility specification with CRRA utility such as described in the notes to figure 1. Choices are simulated based on a random utility specification, such as to exhibit the non-monotonicities described in the introductory figure. Panel A plots simulated CRRA parameters against the KLW discriminability parameter α recovered from the simulated EU plus RUM choices. Panel B plots decision noise in the KLW probit as a function of $1 - \alpha$, i.e. of as-if risk aversion in KLW.

Figure 3 reports a simulation that highlights this flexibility. Choices are generated using the CRRA-based random utility model, with true CRRA coefficients ρ drawn uniformly from $[0, 1]$ (see Online Appendix A for details). The simulated data are then fitted using the KLW specification in equation (1). Panel A plots the true, choice-generating parameter ρ on the horizontal axis against the recovered KLW discriminability parameter α on the vertical axis. For moderate levels of risk aversion, α closely tracks the underlying ρ. However, once ρ exceeds about 0.75, the recovered α plateaus rather than continuing to fall. This flattening arises because utility differences become so small relative to the noise level that further increases in curvature no longer generate systematically different stochastic choice patterns: RUM-simulated choices drift toward random responding and cease to be informative about curvature—precisely the identification problem illustrated

in figure 1, and shared by encoding-based models.

Panel B shows the recovered decision-noise term $\tau\sqrt{2\,\alpha\,(1-\alpha)}$. In stark contrast to the inverse-U relationship *predicted behaviourally* by KLW, the noise *econometrically recovered from the RUM simulation* increases monotonically as $1-\alpha$ increases (i.e. as outcome-discriminability α decreases). In other words, although the KLW model predicts a specific non-monotonic pattern between as-if risk aversion and decision noise, it is nevertheless econometrically flexible enough to accommodate the monotonic relationship between discriminability and choice variability generated by a standard RUM (or its generative counterparts).

This econometric flexibility—together with the behavioural specificity of the KLW predictions—is the key feature exploited by the model-based tests developed below. Because the inverse-U pattern arises as a behavioural implication of Bayesian decoding rather than as a restriction imposed on the econometric mapping, it provides a sharp diagnostic for distinguishing underlying choice-generating mechanisms.

4 Empirical evidence

The theoretical analysis above generates sharply contrasting predictions about how discriminability affects stochastic choice. Encoding-based models predict that as outcomes become harder to discriminate, decision noise increases monotonically, pushing choices toward stochastic indifference in low-discriminability regions. In contrast, the Bayesian decoding framework predicts a non-monotonic relationship: choice variability peaks when discriminability is intermediate and declines again when noise dominates entirely. This implies that Bayesian inference reintroduces a form of stochastic monotonicity in choice patterns.

4.1 Experimental setup

To test these predictions, I design an experiment that generates systematic variation in discriminability and coding noise. Subjects repeatedly choose between risky lotteries (x, p) and a sure amount c. The probability p varies across nine levels $\{0.1, \ldots, 0.9\}$, while the sure amount c spans a range centered on the expected value px, with a somewhat wider range on the risk-averse side (see Online Appendix F for stimuli and instructions).

This design induces a rich set of perceptual tradeoffs across both outcome and probability dimensions. Each subject completes 175 binary choices, including a 10% fraction of repeated trials that allow me to assess within-subject choice consistency.

Numerical magnitude manipulation. The key experimental manipulation alters the numerical scaling of outcomes while keeping their economic value fixed. Subjects are randomly assigned to one of three between-subjects *magnitude-scaling treatments*, which differ only in the numerical units used to represent payoffs:

- **Low-magnitude treatment**: £1 = 1 ECU
- **Medium-magnitude treatment**: £1 = 100 ECU
- **High-magnitude treatment**: £1 = 10,000 ECU

Across treatments, the underlying choice problems are economically identical; only the numerical representation of outcomes differs. Motivated by a broad class of models, Garagnani and Vieider (2025) show that representing quantities outside an adapted numerical range alters the precision of internal magnitude representations. The key point of the treatments is to generate sufficient individual variation in coding noise to allow for meaningful individual-level tests of hoy variability in choice changes with outcome-discriminability. Importantly, the manipulation targets the fidelity of internal representations and is thus well-defined under both model classes considered here. This variation allows us to test whether decision noise increases monotonically as discriminability falls—as predicted by encoding-based models—or instead follows the inverse-U pattern implied by Bayesian decoding models.

The experiment was conducted online using Prolific (UK), with a target sample of 200 subjects per treatment (total $N = 600$). The median completion time was 14 minutes. Subjects were compensated for their time in accordance with Prolific's payment guidelines. In addition, one out of every ten subjects was randomly selected to receive an additional performance-based payment determined by the outcome of one randomly chosen decision.

4.2 Analysis

I estimate the main behavioural parameters using the Bayesian-inference model of Vieider (2024b), which generalizes the KLW framework to settings in which both outcome magnitudes and probabilities vary.[11] This generalization is well suited to the present experiment, as the richer stimulus space induces substantial variation in discriminability—particularly in low-signal regions that are central to the empirical test. Importantly, all core qualitative features of the KLW model are preserved in this specification, including (i) the inverse-U-shaped relationship between outcome-discriminability and decision noise, and (ii) the fact that this pattern is a behavioural prediction rather than a mechanical consequence of the econometric specification.

The key idea of the model is that the *log-odds* of winning, $\ln\left(\frac{p}{1-p}\right)$, and the *relative log cost–benefit ratio*, $\ln\left(\frac{c-y}{x-c}\right)$, are inferred from noisy internal representations. Bayesian decoding of these noisy percepts yields the following probit choice equation, rescaled by dividing through $1/\beta$ (see Online Appendix C for the derivation):

$$\Pr[(x,p;y) \succ c] \;=\; \Phi\left(\frac{\frac{\gamma}{\beta}\ln\left(\frac{p}{1-p}\right) - \ln\left(\frac{c-y}{x-c}\right) - \frac{1}{\beta}\ln(\theta)}{\sqrt{\nu_o^2 + \nu_p^2\left(\frac{\gamma}{\beta}\right)^2}}\right), \tag{2}$$

where $\ln(\theta) \triangleq (1-\beta)\ln(\kappa) + (\gamma-1)\ln(\eta)$ captures the weighted contribution of the two priors for the log-cost benefits and the log-odds. The parameters $\beta \triangleq \frac{\tau_o^2}{\tau_o^2+\nu_o^2}$ and $\gamma \triangleq \frac{\tau_p^2}{\tau_p^2+\nu_p^2}$ measure outcome- and likelihood-discriminability, respectively (with subscripts o and p denoting outcomes and probabilities).

The relation between outcome-discriminability and choice variability. Equation 2 comprises three key elements for our present purposes:

1. The ratio γ/β captures the sensitivity of average choice proportions to the log-odds of winning. The primary prediction is that increasing numerical magnitudes will

[11] A prominent related account of probability weighting is Frydman and Jin (2023), who derive nonlinear probability weights from an efficient-coding framework in which representational precision adapts optimally to the environmental distribution of probabilities. Bayesian inference is used to characterize optimal decoding under this coding scheme and to derive the implied certainty-equivalent mapping. However, Bayesian inference does not govern stochastic choice itself: the model does not generate a Bayesian stochastic choice equation linking posterior uncertainty to choice variability. The present analysis instead focuses on models in which Bayesian decoding directly determines the structure of stochastic choice, yielding distinct predictions for how decision noise varies with discriminability.

raise outcome coding noise ν_o, thereby reducing outcome-discriminability β and increasing sensitivity to probabilities.

2. The factor $1/\beta$ modulates the influence of prior expectations on choice. A decline in outcome-discriminability β increases the weight placed on prior beliefs. Under the assumption of a risk-averse prior, $\theta < 1$, this shifts average choices toward the safe option.[12] This is a secondary prediction, inasmuch as it depends on the latent prior means κ and η. Nonetheless, it mirrors the increase in as-if risk aversion α in the KLW model, illustrating how the present framework nests similar comparative statics within a richer generative structure.

3. The denominator captures decision noise. Bayesian decoding again implies that variability is not monotone in outcome-discriminability: as discriminability starts falls, choice variability initially rises, but once signals become sufficiently unreliable, choices are increasingly governed by prior beliefs and variability declines again. This is the same qualitative prediction as in the KLW setting.

This yields a set of clear and testable predictions. First, choice variability—captured by the scale of the noise term in equation (2)—is predicted to be inverse-U shaped in outcome-discriminability β, mirroring the structural logic underlying the KLW predictions discussed above. Second, likelihood-sensitivity γ/β, which can be approximated nonparametrically, should exhibit the same inverse-U relationship with nonparametric proxies of choice variability, such as stochastic dominance violations. Finally, the model implies that average choice patterns are also linked to discriminability: the influence of prior beliefs, captured by $(1/\beta)\ln(\theta)$, increases as β declines, implying an inverse-U relationship between empirical proxies of risk aversion and choice variability which closely mimics the prediction of the KLW model. Taken together, these predictions establish a common structure across structural parameters and nonparametric indices, and guide the empirical analysis that follows.

Identification and estimation. Following Vieider (2024b), I place outcome and probability discriminability on a common representational scale by setting $\nu_o = \nu_p = \nu$ in estimations. This restriction reflects the assumption that numerical magnitude and

[12] This assumption is based on commonly observed risk aversion in binary choice experiments. Within the logic of the model, such risk aversion arises from pessimistic expectations about the log-odds and/or the log-cost benefits of the lottery (cfr. Oprea and Vieider, 2024; 2025).

probability are encoded with comparable baseline noise, while allowing discriminability to differ across the probability and outcome dimensions through the prior variances τ_o^2 and τ_p^2. In the presence of independent variation in p, x, and c, this normalization ensures separate identification of β and γ. Online Appendix E reports alternative specifications that place β and γ on a common prior-variance scale while allowing coding noise ν_o and ν_p to differ, and provides simulation evidence of full parameter recoverability.

I estimate equation (2) using hierarchical Bayesian methods in `Stan` (Carpenter et al., 2017). The hierarchical structure regularizes noisy individual-level estimates in a principled way while preserving genuine heterogeneity. Population-level parameters receive weakly informative but proper hyperpriors. Convergence diagnostics include $\widehat{R}$ statistics, checks for divergent transitions, and inspection of posterior correlation structures. All estimation details are provided in Online Appendix D, and Vieider (2024a) offers a tutorial introduction to hierarchical Bayesian estimation of decision models in `Stan`.

4.3 Results

I begin by examining how the experimental manipulation affects the parameters recovered from the econometric model. Unless otherwise noted, all reported p-values refer to two-sided Wilcoxon rank-sum tests applied to the posterior means of the individual-level parameters obtained from the hierarchical Bayesian estimation.

Manipulation checks and treatment effects. Panel A of Figure 4 displays empirical cumulative distribution functions (eCDFs) of the estimated outcome-discriminability parameter β across the three ECU-multiplier conditions. The distributions exhibit a strict monotonic ordering: discriminability is highest in the low-multiplier treatment, intermediate in the medium treatment, and lowest in the high-multiplier treatment. Pairwise comparisons confirm that all differences are statistically significant ($p < 0.001$ in each case). We also observe substantial heterogeneity within each treatment, and importantly, a non-negligible mass of observations with $\beta < \frac{1}{2}$, corresponding to regimes in which internal noise outweighs the signal in Bayesian decoding. This confirms that increasing the ECU multiplier systematically reduces outcome discriminability and generates the exogenous variation required for the main empirical test.

Panel B turns to the distribution of the estimated decision-noise parameter, $\omega \triangleq$

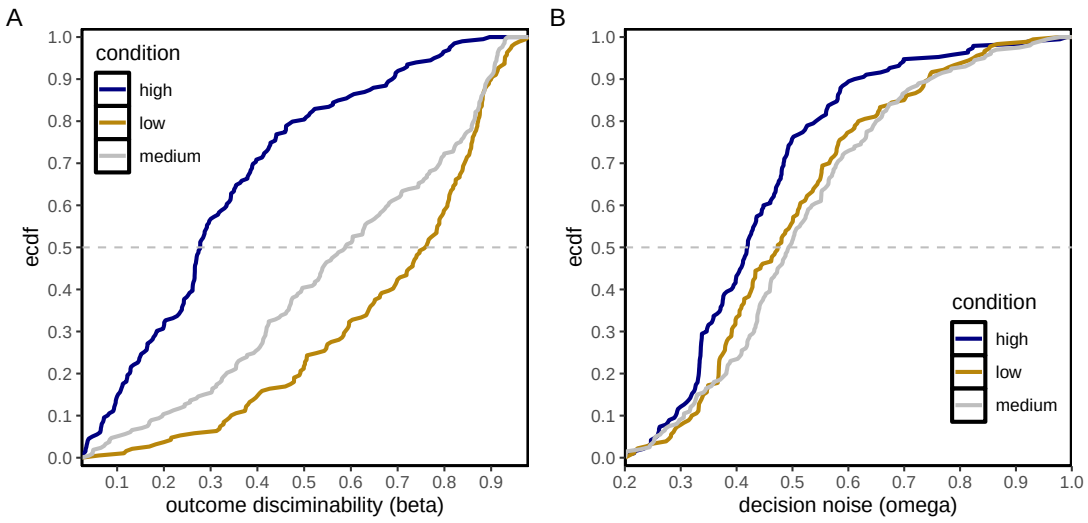


Figure 4: Treatment effects on outcome-discriminability and decision noise

Treatment level distributions of key parameters. Panel A shows the empirical cumulative distribution functions per treatment of outcome-discriminability β. Panel B shows the empirical cumulative distribution functions of decision noise ω.

$\nu\sqrt{\beta^2+\gamma^2}$, which corresponds to the composite standard deviation of the probit error term in equation (2) under the maintained restriction $\nu_o = \nu_p = \nu$. The ordering across treatments is now markedly different. The low-ECU treatment exhibits *intermediate* levels of decision noise, whereas the high-ECU treatment—despite producing the lowest outcome discriminability—shows the *lowest* levels of decision noise ($p < 0.001$ in both pairwise comparisons). The medium-ECU condition displays a slight but statistically insignificant increase in decision noise relative to the low-ECU condition ($p = 0.282$).

Taken together, Panels A and B already point to a clear departure from the monotonic predictions of encoding-based models. Reducing stimulus discriminability does not lead to a monotonic increase in decision noise. Instead, decision noise declines again in the high-noise regime, providing initial evidence for the inverse-U relationship between discriminability and stochastic choice implied by Bayesian decoding.

Bayesian regression to the mean reins in choice stochasticity. Our key interest lies in the individual-level relationship between outcome discriminability β and decision noise $\omega \triangleq \nu\sqrt{\beta^2+\gamma^2}$. Bayesian decoding predicts that decision noise should be *non-monotonic* and inverse-U shaped in outcome discriminability: as internal signals become increasingly noisy, posterior beliefs place greater weight on the prior mean, which eventually stabilizes behaviour and reduces stochasticity.

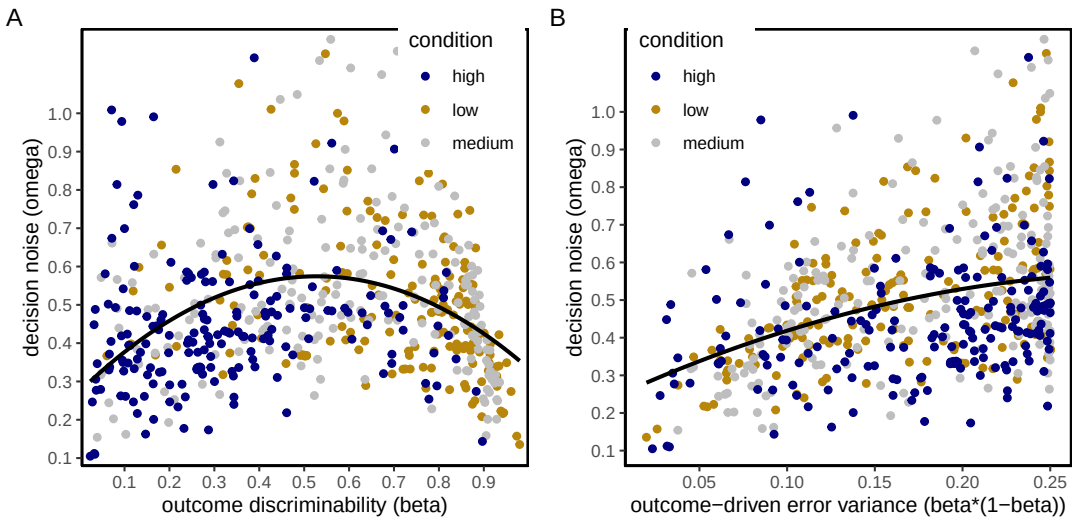


Figure 5: Relationship between outcome discriminabiity and decision noise

Relationship between outcome discriminability and decision noise. Panel A shows a scatter plot of outcome discriminability β on the horizontal axis, against decision noise ω on the vertical axis. Panel B plots the portion of error variance driven by the outcome dimension, $\beta\,(1-\beta)$ on the horizontal axis against decision noise ω on the vertical axis. In both cases the points have been fit by a quadratic regression line. Some extreme observations have been removed to improve the visual experience.

Panel A of Figure 5 plots outcome discriminability β on the horizontal axis against decision noise ω on the vertical axis. The scatterplot reveals a pronounced inverse-U pattern: decision noise is highest at intermediate levels of discriminability and declines both when discriminability is very low and when it is very high. A quadratic regression of decision noise on outcome discriminability confirms this pattern: the linear term is strongly positive ($\hat{\beta}_1 = 1.15$, $p < 0.001$), the quadratic term strongly negative ($\hat{\beta}_2 = -1.10$, $p < 0.001$), and the implied maximum occurs at $\hat{\beta} \approx 0.52$.

Panel B presents the same relationship using the canonical inverse-U transformation $\beta\,(1-\beta)$ that governs the variance term in the Bayesian posterior. When plotted on this scale, decision noise increases monotonically in $\beta\,(1-\beta)$, which attains its maximum at $\beta = 0.5$. This pattern follows directly from the structure of Bayesian decoding: noise is smallest when internal signals are either highly reliable or entirely uninformative (in which case the posterior collapses onto the prior), that is, when $\beta \to 0$ or $\beta \to 1$ and hence $\beta\,(1-\beta) \to 0$. Decision noise is largest when noisy signals and heterogeneous priors interact most strongly, namely at $\beta \approx 0.5$, where $\beta\,(1-\beta)$ is maximized.

In Appendix E.1, I re-estimate all parameters using the KLW model in equation (1). Despite the richer probability variation in the present experiment—which the KLW spec-

ification was not designed to accommodate—the key qualitative patterns are unchanged: (i) outcome discriminability α varies systematically across treatments, and (ii) decision noise declines again at low discriminability, consistent with Bayesian regression toward the prior mean. Together, these results provide strong empirical support for Bayesian decoding and contradict the monotonic stochastic-choice predictions of encoding-based models.

Nonparametric results. The analysis so far has relied on parameters recovered from a structural model to contrast the predictions of encoding-based accounts with those of Bayesian inference. I now show that the central findings replicate using entirely nonparametric indices that capture the same underlying concepts of discriminability and stochasticity, without imposing any functional-form restrictions or structural assumptions. This serves two purposes. First, it rules out that the inverse-U pattern documented above is an artefact of the particular econometric specification. Second, it shows that the key behavioural prediction of Bayesian decoding emerges directly in the raw choice data, independently of how decision noise is parameterized, and of the specific functional and distributional assumptions in the model.

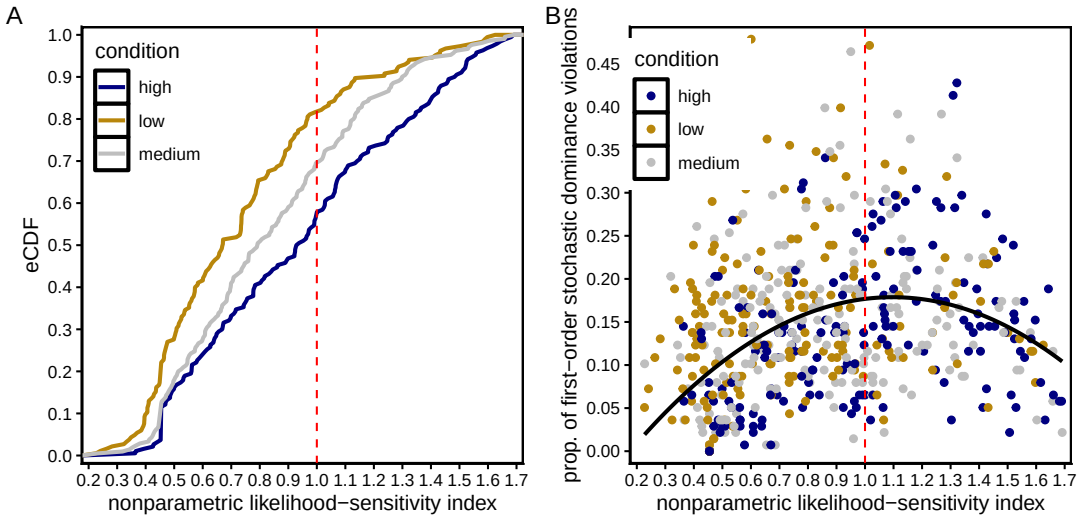


Figure 6: Nonparametric analysis of likelihood-sensitivity and stochastic dominance violations

The figure shows a nonparametric analysis of the relationship between likelihood-sensitivity in risky choice and first-order stochastic dominance violations, measured as a percentage of all possible violations. Panel A shows eCDFs for likelihood-sensitivity across treatments and subjects. Panel B shows a scatter plot of the nonparametric likelihood sensitivity index on the horizontal axis, against the nonparametric index of the proportion of stochastic dominance violations on the vertical axis.

Panel A of Figure 6 displays empirical cumulative distribution functions (eCDFs) of a

nonparametric index of likelihood sensitivity. This index captures how sharply a subject's probability of choosing the risky option responds to changes in the probability of receiving a fixed prize of £22. It is constructed from normalized differences in stochastic certainty-equivalent approximations taken symmetrically around $p = 0.5$.[13] The index equals 1 under perfect likelihood sensitivity (as in EU), values below 1 indicate insensitivity, and values above 1 indicate oversensitivity. Conceptually, this index serves as a nonparametric analogue of the γ/β ratio governing likelihood sensitivity in the structural model in equation (2). Decreasing outcome-discriminability β is then predicted to *increase* likelihood-sensitivity γ/β.

The eCDFs exhibit a striking monotonic ordering that mirrors the ordering of outcome discriminability β across treatments. Likelihood sensitivity is lowest in the 1 ECU condition, higher in the 100 ECU condition, and highest in the 10,000 ECU condition ($p < 0.001$ in all comparisons). In other words, as outcome representations become noisier, subjects become *more* sensitive to probability differences. These nonparametric findings closely parallel those reported by Oprea and Vieider (2025), who examine the effects of experimentally induced noise in the outcome versus probability dimension on average choice patterns using a a closely related experimental manipulation.

Panel B plots the likelihood-sensitivity index (horizontal axis) against a nonparametric measure of individual-level violations of first-order stochastic dominance (vertical axis), expressed as a proportion of all possible violations. I here focus on a *dimension-specific* notion of dominance violations, holding probabilities fixed and varying outcomes. This choice aligns the nonparametric measure directly with the experimental manipulation, which increases noise in the encoding of outcomes rather than probabilities. Specifically, a violation is recorded whenever, for a given probability p, choices are non-monotonic in the sure amount c. A violation therefore corresponds to *multiple switching* (e.g., choosing the lottery, then the sure amount, and then the lottery again). The three other measures of stochastic inconsistency that can be recovered from the data—including stochastic dominance violations across probabilities, stochastic dominance violations in

[13] For each probability pair $\{p, 1-p\}$ in $\{0.9{:}0.1,\ 0.8{:}0.2,\ 0.7{:}0.3\}$, I construct a stochastic certainty-equivalent proxy from the subject's choice proportion for the risky option. The likelihood-sensitivity measure for each pair is the normalized difference in choice proportions between the high- and low-probability presentations of the same lottery; normalization divides the difference in choice proportions by the corresponding difference in objective probabilities. The individual-level index is obtained by averaging these normalized differences across the three probability pairs.

the outcome-dominated choices with a prize of £20, and test-retest variability in repeated choice—all yield qualitatively identical patterns, shown in Online Appendix E.3. The resulting relationship forms—once again—a pronounced inverse-U shape.[14]

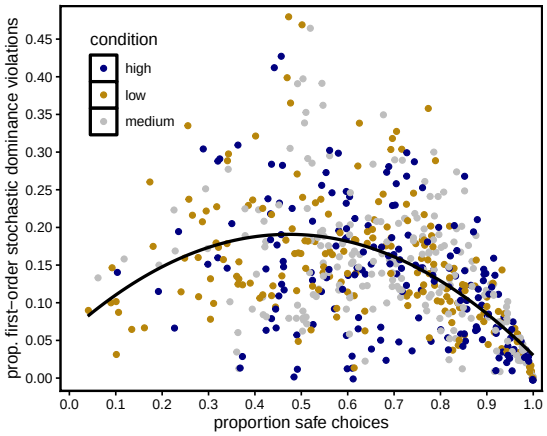


Figure 7: Nonparametric analysis of likelihood-sensitivity and stochastic dominance violations

The figure shows a nonparametric analysis of the relationship between the proportion of safe choices and first-order stochastic dominance violations, measured as a percentage of all possible violations.

Figure 7 further examines how choice volatility relates to risk aversion, capturing the secondary prediction arising from the effect of $1/\beta$ on the joint prior means in θ. The figure plots the proportion of choices of the sure alternative across all decisions on the horizontal axis against the same measure of within-list stochastic dominance violations used above on the vertical axis. This relationship is also of independent interest, because it provides a direct test of the association between risk aversion and discriminability in KLW. The resulting pattern is again inverse-U shaped, consistent with the predictions of the model. The decline in variability at the upper end of the distribution should be interpreted with caution, since the experimental design mechanically constrains the observable measure of risk aversion. Importantly, however, the inverse-U pattern is not confined to these boundary regions: variability changes smoothly across the interior of the support, suggesting that the overall relationship is not driven solely by censoring at the endpoints.

Discussion: A doubly counter-intuitive nonparametric pattern. The inverse-U relationship between likelihood-sensitivity and choice variability is *doubly counter-intuitive*. First, from a standard noise-based perspective, one would expect more stochastic-

[14] This is confirmed by a quadratic regression: both the linear term ($b_1 = 65.4$, $p < 2 \times 10^{-16}$) and the negative quadratic term ($b_2 = -30.8$, $p < 2 \times 10^{-16}$) are strongly significant, indicating that violations rise with sensitivity at low values of the index but decline again at higher values.

dominance violations in the high-noise regime. Yet violations are lowest precisely when outcome-coding noise is greatest (and hence likelihood oversensitivity is most pronounced). Within the logic of the Bayesian-decoding framework, this occurs because regression toward the prior mean stabilizes behaviour when the signal is highly uninformative. Second, one might expect low likelihood sensitivity to mechanically generate more dominance violations, as choices become less responsive to probability differences. Yet violations are remarkably rare among subjects with very low sensitivity to changes in probabilities—including stochastic dominance violations *across probabilities*, shown in Online Appendix E.3. Instead, the largest number of violations arises for subjects with intermediate discriminability, where signals and prior expectations receive equal weight.

From the perspective of economic rationality, this pattern appears paradoxical. It is precisely the subjects with *near-perfectly calibrated likelihood sensitivity*—those whose average behaviour most closely resembles the expected utility benchmark—who exhibit the largest number of stochastic-dominance violations. In other words, behaviour that appears most "rational" from the perspective of the deterministic normative benchmark model coincides with the greatest instability in choice. Understanding this pattern requires a different conception of rationality. Conditional on noisy internal representations, Bayesian decoding constitutes an optimal response to uncertainty. It is precisely this optimality that generates the coexistence of near-perfect probability sensitivity with heightened stochasticity in choice. Randomness in behaviour, rather than reflecting cognitive failure, can therefore be a direct consequence of optimal inference under noise.

5 Conclusion

This paper asked what stochastic choice reveals about the processes that generate behaviour. Rather than treating randomness as an exogenous disturbance, I examined generative models in which both average discriminability between choice options and stochasticity across choices arise endogenously from noise in internal representations. Encoding-based models predict that choice variability should increase monotonically as options become harder to discriminate. Bayesian decoding, by contrast, yields a sharply different implication: regression toward prior expectations stabilizes behaviour when signals become uninformative, suppressing stochasticity in the high-noise regime.

The experiment was designed to adjudicate between these competing accounts. By rescaling outcomes into increasingly coarse numerical units, I generated exogenous variation in outcome discriminability while holding the economic environment fixed. Structural and nonparametric analyses converge on the same conclusion. Although discriminability declines monotonically as outcomes are expressed on coarser numerical scales, choice variability does not. Instead, the relationship between discriminability and stochastic choice follows a pronounced inverse-U pattern: variability peaks at intermediate discriminability and declines again when internal signals become very noisy. At the same time, lower outcome discriminability is associated with greater sensitivity to probabilities and fewer violations of stochastic dominance—patterns that run counter to encoding-based accounts but follow directly from Bayesian decoding.

Taken together, these findings imply that stochastic choice bears the distinctive signature of Bayesian inference. Apparent randomness in behaviour is not simply the result of imprecise encoding or independent errors, but reflects optimal decoding of noisy internal signals using prior information about the environment. Efficient-coding considerations may still shape the fidelity of internal representations, but they operate within a decoding architecture that constrains how noise is translated into behaviour. More broadly, the results show that patterns of stochastic choice—far from being a nuisance—contain systematic information that can be used to distinguish between competing models of decision-making under risk.

From an economic perspective, these results have direct implications for how we think about the relationship between tastes and choice variability. Deviations from expected-utility benchmarks need not reflect stable non-standard preferences, nor independent errors of the kind assumed in random utility models, but may instead arise from optimal inference under representational noise. Conversely, this also implies—perhaps paradoxically—that behaviour consistent with expected utility may be observed precisely in high-noise environments, and need not correspond to stable underlying tastes at all. Adjudicating whether observed behaviour truly reflects preferences, rather than the structured consequences of noise, therefore requires careful attention to the structure of stochastic choice data—the central insight of this paper.

References

Alós-Ferrer, Carlos and Michele Garagnani, "Improving risky-choice predictions using response times," *Journal of Political Economy Microeconomics*, 2024, *2* (2), 335–354.

_ , Ernst Fehr, and Nick Netzer, "Time will tell: Recovering preferences when choices are noisy," *Journal of Political Economy*, 2021, *129* (6), 1828–1877.

Apesteguia, Jose and Miguel A Ballester, "Monotone stochastic choice models: The case of risk and time preferences," *Journal of Political Economy*, 2018, *126* (1), 74–106.

Ballinger, T. Parker and Nathaniel T. Wilcox, "Decisions, Error and Heterogeneity," *The Economic Journal*, 1997, *107* (443), 1090–1105.

Barlow, Horace B., "Possible Principles Underlying the Transformation of Sensory Messages," in Walter A. Rosenblith, ed., *Sensory Communication*, Cambridge, MA: MIT Press, 1961, pp. 217–234.

Bishop, Christopher M., *Pattern recognition and machine learning*, Vol. 4, Springer, 2006.

Bouchouicha, Ranoua, Yuchi Li, and Ferdinand M. Vieider, "Loss-sensitivity versus Loss-aversion," Technical Report, Ghent University Discussion Papers 2025.

Busemeyer, Jerome R and James T Townsend, "Decision field theory: a dynamic-cognitive approach to decision making in an uncertain environment.," *Psychological Review*, 1993, *100* (3), 432.

Carpenter, Bob, Andrew Gelman, Matthew D Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Marcus Brubaker, Jiqiang Guo, Peter Li, and Allen Riddell, "Stan: A probabilistic programming language," *Journal of Statistical Software*, 2017, *76* (1), 1–32.

Casella, George and Roger Berger, *Statistical inference*, Chapman and Hall/CRC, 2024.

de Clippel, Geoffroy, Ryan Oprea, and Kareen Rozen, "As-If," *Working Paper*, 2024.

Dehaene, Stanislas, "The neural basis of the Weber–Fechner law: a logarithmic mental number line," *Trends in Cognitive Sciences*, 2003, *7* (4), 145–147.

Enke, Benjamin and Thomas Graeber, "Cognitive uncertainty," *Quarterly Journal of Economics*, 2023, *138* (4), 2021–2067.

_ , _ , and Ryan Oprea, "Complexity and Time," *Journal of the European Economic Association*, 2024, *23* (5), 1838–1867.

Frydman, Cary and Lawrence J Jin, "Efficient coding and risky choice," *Quarterly Journal of Economics*, 2022, *136*, 161–213.

_ and Lawrence J. Jin, "On the Source and instability of probability weighting," Working Paper 2023.

Fudenberg, Drew, Philipp Strack, and Tomasz Strzalecki, "Speed, accuracy, and the optimal timing of choices," *American Economic Review*, 2018, *108* (12), 3651–3684.

Garagnani, Michele and Ferdinand M. Vieider, "Economic Consequences of Numerical Adaptation," *Psychological Science*, 2025, *36* (6), 407–420.

Heng, Joseph A, Michael Woodford, and Rafael Polania, "Efficient sampling and noisy decisions," *Elife*, 2020, *9*, e54962.

Khaw, Mel Win, Ziang Li, and Michael Woodford, "Cognitive imprecision and small-stakes risk aversion," *The Review of Economic Studies*, 2021, *88* (4), 1979–2013.

Krajbich, Ian, Carrie Armel, and Antonio Rangel, "Visual fixations and the computation and comparison of value in simple choice," *Nature Neuroscience*, 2010, *13* (10), 1292–1298.

Laughlin, Simon, "A simple coding procedure enhances a neuron's information capacity," *Zeitschrift für Naturforschung c*, 1981, *36* (9-10), 910–912.

Loomes, Graham, "Modelling the stochastic component of behaviour in experiments: Some issues for the interpretation of data," *Experimental Economics*, 2005, *8*, 301–323.

Ma, Wei Ji, Konrad Paul Kording, and Daniel Goldreich, *Bayesian Models of Perception and Action: An Introduction*, MIT press, 2023.

McGranaghan, Christina, Kirby Nielsen, Ted O'Donoghue, Jason Somerville, and Charles D Sprenger, "Distinguishing common ratio preferences from common ratio effects using paired valuation tasks," *American Economic Review*, 2024, *114* (2), 307–347.

Natenzon, Paulo, "Random choice and learning," *Journal of Political Economy*, 2019, *127* (1), 419–457.

Netzer, Nick, "Evolution of time preferences and attitudes toward risk," *American Economic Review*, 2009, *99* (3), 937–55.

Nielsen, Kirby and John Rehbeck, "When choices are mistakes," *American Economic Review*, 2022, *112* (7), 2237–2268.

Oprea, Ryan, "Decisions Under Risk are Decisions Under Complexity," *American Economic Review*, 2024, *114* (12), 3789—-3811.

_ **and Ferdinand M. Vieider**, "Minding the Gap: On the Origins of the Description-Experience Gap," Technical Report, Working Paper 2024.

_ **and Ferdinand M Vieider**, "The New Psychophysics of Risk and Time," *Working Paper*, 2025.

Ratcliff, Roger, "A theory of memory retrieval," *Psychological Review*, 1978, *85* (2), 59.

_ **and Gail McKoon**, "The diffusion decision model: theory and data for two-choice decision tasks," *Neural computation*, 2008, *20* (4), 873–922.

Robson, Arthur J, "The biological basis of economic behavior," *Journal of Economic Literature*, 2001, *39* (1), 11–33.

_ , "Why would nature give individuals utility functions?," *Journal of Political Economy*, 2001, *109* (4), 900–914.

Sims, Chris R, "Rate–distortion theory and human perception," *Cognition*, 2016, *152*, 181–198.

Stewart, Neil, Nick Chater, and Gordon DA Brown, "Decision by sampling," *Cognitive Psychology*, 2006, *53* (1), 1–26.

Thurstone, Louis L, "A law of comparative judgment.," *Psychological Review*, 1927, *34* (4), 273.

_ , "Psychophysical analysis," *The American Journal of Psychology*, 1927, *38* (3), 368–389.

Vieider, Ferdinand M., "Cognitive Foundations of Delay-Discounting," *Working Paper*, 2023.

_ , "Bayesian Estimation of Decision Models," Technical Report, RISLab 2024.

_ , "Decisions under Uncertainty as Bayesian Inference on Choice Options," *Management Science*, 2024, *70* (12), 9014–9030.

Webb, Ryan, "The (neural) dynamics of stochastic choice," *Management Science*, 2019, *65* (1), 230–255.

Wilcox, Nathaniel T., "'Stochastically more risk averse:' A contextual theory of stochastic discrete choice under risk," *Journal of Econometrics*, May 2011, *162* (1), 89–104.

Zhang, Hang, Xiangjuan Ren, and Laurence T Maloney, "The bounded rationality of probability distortion," *Proceedings of the National Academy of Sciences*, 2020, *117* (36), 22024–22034.

Zilker, Veronika and Thorsten Pachur, "Nonlinear probability weighting can reflect attentional biases in sequential sampling.," *Psychological Review*, 2022, *129* (5), 949.

ONLINE APPENDIX

A Simulation details for Figure 3

To illustrate the qualitative predictions of a heteroscedastic random-utility model, I simulate binary risky choices under a probit specification with independent additive noise. The simulation proceeds in four steps.

(i) Lottery construction. Risky lotteries take the form (x, p), where the prize x ranges from 60 to 200 in steps of 20, and the probability takes one of two values $p \in \{0.40, 0.55\}$. The sure amounts c range from 8 to 160 in steps of 8. I retain only dominated pairs with $c < x$, yielding the choice set

$$\mathcal{S} = \{(x, p), c : x \in \{60, \ldots, 200\},\ c \in \{8, \ldots, 160\},\ p \in \{0.40, 0.55\},\ c < x\}.$$

(ii) Heterogeneous preference and noise parameters. I generate $N = 300$ simulated individuals. Each individual is endowed with: (i) a CRRA curvature parameter

$$\rho_i \sim \text{Uniform}(0, 1),$$

and (ii) a noise parameter

$$\sigma_i \sim \text{Lognormal}(\log 5,\ 0.9^2),$$

which governs the magnitude of the random disturbance in the probit model.

(iii) Sampling of choice problems. For each individual i, I draw 21 choice problems at random (with replacement) from $\mathcal{S}$ and combine these with the full set of problems once, producing a mixed design with both common and idiosyncratic choice sets. The resulting dataset contains multiple observations per individual on a diverse set of lotteries.

(iv) Stochastic choice generation. Utility for each lottery is given by the CRRA specification $u(x) = x^{1-\rho_i}$ (with the usual logarithmic limit case). The deterministic

utility difference for individual i on choice problem j is therefore

$$\Delta_{ij} = p_j\, x_j^{1-\rho_i} - c_j^{1-\rho_i}.$$

Choices follow a probit random-utility specification with heteroscedastic noise (i.e. the noise term is individual-specific):

$$\Pr((x_j, p_j) \succ c_j \mid i) = \Phi\left(\frac{\Delta_{ij}}{\sqrt{2}\,\sigma_i}\right),$$

and the observed binary choice is generated as

$$y_{ij} \sim \text{Bernoulli}\left(\Phi\left(\frac{\Delta_{ij}}{\sqrt{2}\,\sigma_i}\right)\right).$$

This simulated dataset is then analyzed using the same estimation procedures as for the empirical data, allowing me to compare the nonparametric patterns implied by a standard heteroscedastic RUM with those observed in the experiment.

Hierarchical Bayesian estimation of the KLW model

For each choice problem i for individual n, the risky option is a lottery (x_i, p_i) and the sure option is c_i. As in the main text, I rewrite the deterministic component of the KLW decision index in terms of outcome ratios and probabilities,

$$\ell_i = \log\left(\frac{x_i}{c_i}\right), \qquad \log p_i = \log(p_i).$$

At the individual level, the KLW model is parameterised by a coding-noise parameter $\nu_n > 0$ and a response-noise parameter $\sigma_n > 0$. These are mapped into the signal weight

$$\alpha_n = \frac{\sigma_n^2}{\sigma_n^2 + \nu_n^2}$$

and the effective probit scale

$$\omega_n = \sqrt{2}\,\nu_n \alpha_n = \sqrt{2}\,\frac{\sigma_n^2 \nu_n}{\sigma_n^2 + \nu_n^2},$$

as derived in Section 3. Conditional on (α_n, ω_n), the choice probability for individual n on trial i is

$$\Pr\big((x_i, p_i) \succ c_i \mid n\big) = \Phi\left(\frac{\alpha_n \, \ell_i + \log p_i}{\omega_n}\right),$$

and the observed choice is

$$y_{in} \sim \text{Bernoulli}\left(\Phi\left(\frac{\alpha_n \, \ell_i + \log p_i}{\omega_n}\right)\right).$$

To allow for heterogeneity across subjects, I place a bivariate normal prior on the log-parameters $(\log \nu_n, \log \sigma_n)$:

$$\begin{pmatrix} \log \nu_n \\ \log \sigma_n \end{pmatrix} \sim \mathcal{N}_2\big(\boldsymbol{\mu},\, \Sigma\big), \qquad \Sigma = \text{diag}(\boldsymbol{\tau})\, R\, \text{diag}(\boldsymbol{\tau}),$$

where $\boldsymbol{\mu} = (\mu_1, \mu_2)^\top$ is a vector of population means, $\boldsymbol{\tau} = (\tau_1, \tau_2)^\top$ is a vector of population standard deviations, and R is a 2×2 correlation matrix.

Priors are chosen to be weakly informative on the log scale:

$$\mu_j \sim \mathcal{N}(0, 5), \quad j = 1, 2; \qquad \tau_j \sim \text{Exponential}(5), \quad j = 1, 2; \qquad R \sim \text{LKJ}(4).$$

The normal priors on μ_j imply that ν_n and σ_n are a priori spread over several orders of magnitude (since $\log \nu_n$ and $\log \sigma_n$ are typically between -10 and 10 with overwhelming probability). The exponential priors on τ_j favour moderate between-subject heterogeneity while still allowing large values, and the LKJ prior with shape parameter 4 puts mild weight on correlations near zero but does not rule out strong correlations. Overall, these priors are deliberately diffuse: their role is to stabilise inference without imposing tight a priori restrictions on the population distribution of (ν_n, σ_n).

B Contextual Utility and Random Preferences

This section examines whether two fixes of the random-utility pathology described in the literature—contextual utility and random-preference specifications—alter the qualitative predictions on the relationship between discriminability and decision noise documented in the main text.

B.1 Contextual utility

Contextual-utility correction was developed by Wilcox (2011) to address a well-known issue in estimating expected-utility (EU) models with random-utility error. In a standard EU–RUM specification,

$$U(x,p) = p\,u(x), \qquad U(c,1) = u(c),$$

the curvature of $u(\cdot)$ interacts mechanically with the range of outcomes: changing the scale of the stimulus space (for example by multiplying all outcomes by a constant) alters the effective steepness of the deterministic index and thereby the sensitivity of choice probabilities to payoff differences. Contextual utility introduces a simple normalization that rescales $u(\cdot)$ within each display or context.

Let k index the context (here corresponding to the ECU scaling level). Each context k includes a lower and upper outcome bound, $\underline{x}_k$ and $\overline{x}_k$. The contextual transformation replaces $u(\cdot)$ with an affine normalization,

$$u_k(x) = \frac{u(x) - u(\underline{x}_k)}{u(\overline{x}_k) - u(\underline{x}_k)},$$

which maps the context-specific minimum to 0 and the maximum to 1. Preferences remain EU, but the scale of the deterministic index depends on the range of outcomes shown in context k.

In each context, choices are generated from the transformed deterministic utilities

$$U_k(x,p) = p\,u_k(x), \qquad U_k(c,1) = u_k(c),$$

combined with a random-utility error term. Under a probit specification,

$$\Pr\big((x,p) \succ c \mid k\big) = \Phi\left(\frac{U_k(x,p) - U_k(c,1)}{\sigma_k}\right),$$

where σ_k is the (context-specific or common) noise scale. Because the contextual transformation is affine, it preserves the ratio-scale structure of EU but simply rescales the deterministic difference $U_k(x,p) - U_k(c,1)$.

Importantly, noise remains an exogenous additive term appended at the final stage,

and the model retains the core implication of the standard RUM: as discriminability declines, the denominator dominates and choice probabilities collapse monotonically toward 1/2. Incorporating contextual utility does not generate the inverse-U pattern in decision noise predicted by Bayesian decoding; the qualitative monotonicity of RUM is unchanged.

B.2 Random-preference specifications

Random-preference models replace a fixed utility function with a distribution over utility indices, typically by assuming that each choice instance draws a parameter ψ that governs risk attitudes or utility curvature. This relocates randomness from an additive error term (as in classical RUM) to the underlying preferences themselves. As Apesteguia and Ballester (2018) stress, such models can repair several comparative-statics pathologies of RUM applied to EUT: choice probabilities can be made monotone in attributes even when the standard additive-error specification would violate monotonicity. In that sense, random preferences improve the link between economic structure and observed choice probabilities.

However, this change in where randomness enters the model does not introduce a mechanism akin to Bayesian regression to the mean. Let U_ψ denote a von Neumann–Morgenstern utility function indexed by a random parameter ψ, and let A and B be two lotteries. A random-preference model implies

$$\Pr(A \succ B) = \Pr(U_\theta(A) - U_\theta(B) > 0) = \Pr\big(\Delta(A, B|\psi) > 0\big)\,,$$

where $\Delta(A, B|\psi)$ is a deterministic function of the stimuli for each ψ. Changing the discriminability of the lotteries affects Δ through the payoffs and probabilities, and hence shifts the measure of θ for which A is preferred to B. But there is no internal prior over outcomes or probabilities, and no encoding–decoding stage in which the variance of the decision index can shrink when the signal becomes uninformative. All stochasticity is driven by variation in ψ.

From an econometric perspective, this implies that random-preference models and Bayesian-decoding models make fundamentally different predictions for how decision noise should behave as discriminability varies. In Bayesian models, noise arises from corrupted inter-

nal representations, and when the signal is very weak the decoder regresses toward the prior, *reducing* the dispersion of the decision index and stabilizing choice. In random-preference models, by contrast, the distribution of ψ is fixed, and there is no structural link between low discriminability and a reduction in effective noise. Any inverse-U relationship between discriminability and stochastic choice would therefore have to be imposed indirectly via the assumed distribution of preferences or the design of the stimulus space, rather than emerging as a generic implication of the model. In particular, random preferences do not predict the specific pattern highlighted in this paper: low decision noise when discriminability is either very high or very low, and maximal noise at intermediate levels of discriminability.

C Bayesian Inference Model derivation

I first describe the inference process for a generic choice attribute z. The decision maker receives an unbiased but noisy signal $r_z \sim \mathcal{N}(ln(z), \nu_z^2)$, where ν_z^2 is the error variance of the signal. The decision-maker holds prior beliefs $ln(z) \sim \mathcal{N}(\mu_z, \sigma_z^2)$ about the distribution of choice quantities in the environment. She subsequently decodes the noisy signal by combining it with the prior in a Bayesian inference process, yielding the posterior mean

$$\mathbb{E}[ln(z)|r_z] = \frac{\sigma_z^2}{\sigma_z^2 + \nu_z^2} r_z + \frac{\nu_z^2}{\sigma_z^2 + \nu_z^2} \mu_z.$$

This posterior mean is not directly observable to the experimenter. We can, however, average over many repeated trials involving the same choice attributes. Given that r_z is unbiased, this yields the following expectation of the response distribution:

$$\mathbb{E}\left[\mathbb{E}[ln(z)|r_z]\middle|z\right] = \frac{\lambda_z}{\lambda_z + \xi_z} ln(z) + \frac{\xi_z}{\lambda_z + \xi_z} \mu_z,$$

where the signal has been replaced by its mean. The proof follows trivially from the property of the normal distribution whereby $n \sim \mathcal{N}(m, s)$ implies $a + bn \sim \mathcal{N}(a + bm, b^2 s^2)$.

The final step is to substitute the log-odds and log cost-benefits for z and to trade them off. Define $\gamma \triangleq \frac{\sigma_p^2}{\sigma_p^2 + \nu_p^2}$ as the evidence weight put on the true log-odds in the Bayesian inference process. Equivalently, define $\beta \triangleq \frac{\sigma_o^2}{\sigma_o^2 + \nu_o^2}$, where o stands for outcomes, i.e. the

cost-benefit ratio. Trading off the attributes involves choosing the risky wager whenever $\gamma r_p + (1-\gamma) ln(\eta) > \beta r_o + (1-\beta) ln(\kappa)$, with the prior means defined as in the main text. Rewriting the choice equation as $\gamma r_p - \beta r_o - [(1-\beta) ln(\kappa) + (\gamma -) ln(\eta)] > 0$ and defining $[(1-\beta) ln(\kappa) + (\gamma - 1) ln(\eta)] \triangleq ln(\theta)$, we can again apply the properties of the normal distribution to arrive at the response distribution:

$$\gamma\, r_p - \beta\, r_o - ln(\theta) \sim \mathcal{N}\left(\gamma\, ln\left(\frac{p}{1-p}\right) - \beta\, ln\left(\frac{c-y}{x-c}\right) - ln(\theta)\ ,\ \gamma^2 \nu_p^2 + \beta^2 \nu_o^2\right).$$

The version in the main Probit equation in the main text obtains equivalently, by first dividing everything by β in $\gamma r_p - \beta r_o - [(1-\beta) ln(\kappa) + (\gamma -) ln(\eta)] > 0$ and then following the same steps.

D Bayesian estimation details

In this section I describe the hierarchical Bayesian estimation procedure used to recover individual-level psychophysical parameters from the experimental data. Each choice problem i for individual n consists of a risky lottery (x_i, p_i) with lower outcome ℓ_i and higher outcome h_i, evaluated against a sure amount c_i. Two transformed regressors are constructed:

$$\text{lcb}_i = \log\left(\frac{c_i - \ell_i}{h_i - c_i}\right), \qquad \text{llr}_i = \log\left(\frac{p_i}{1-p_i}\right),$$

corresponding respectively to the log-relative position of the sure outcome within the risky payoff range, and the log-likelihood ratio of the probability p_i.

Individual-level parameters. Each individual n is characterised by five positive psychophysical parameters:

$$\nu_n, \quad \sigma_{o,n}, \quad \sigma_{p,n}, \quad \eta_n, \quad \xi_n.$$

The parameter ν_n governs the overall noise in the internal code ('coding noise'). The parameters $\sigma_{o,n}$ and $\sigma_{p,n}$ control the noise in the outcome and probability channels, respectively. The parameters η_n and ξ_n capture the idiosyncratic prior means for outcomes and probabilities in the Bayesian decoding stage. All parameters are defined on the log-scale for estimation convenience:

$$(\log \nu_n, \log \sigma_{o,n}, \log \sigma_{p,n}, \log \eta_n, \log \xi_n) \sim \mathcal{N}_5(\boldsymbol{\mu}, \Sigma).$$

Signal weights and effective scale. Following the psychophysical derivations in Section 3, the model implies separate signal weights for outcomes and probabilities:

$$\alpha_n = \frac{\sigma_{o,n}^2}{\sigma_{o,n}^2 + \nu_n^2}, \qquad \gamma_n = \frac{\sigma_{p,n}^2}{\sigma_{p,n}^2 + \nu_n^2}.$$

These govern the extent to which decoded internal values depend on current stimuli versus prior expectations. The effective probit scale is

$$\omega_n = \nu_n \sqrt{\alpha_n^2 + \gamma_n^2},$$

which reflects the joint contribution of coding noise and Bayesian regression effects.

Choice probability. For individual n on trial i, the decision index implied by the Bayesian decoder is

$$d_{in} = \frac{\gamma_n \, \mathrm{llr}_i + (1 - \gamma_n) \, \log(\eta_n) - \big[\alpha_n \, \mathrm{lcb}_i + (1 - \alpha_n) \, \log(\xi_n)\big]}{\omega_n}.$$

The associated choice probability for the risky option is

$$\Pr\big((\ell_i, h_i, p_i) \succ c_i \mid n\big) = \Phi(d_{in}),$$

and the observed choice is generated as

$$y_{in} \sim \mathrm{Bernoulli}\big(\Phi(d_{in})\big).$$

Hierarchical structure and priors. To allow for flexible heterogeneity across individuals, the five log-parameters are given a population-level multivariate normal prior,

$$\theta_n = \begin{pmatrix} \log \nu_n \\ \log \sigma_{o,n} \\ \log \sigma_{p,n} \\ \log \eta_n \\ \log \xi_n \end{pmatrix} \sim \mathcal{N}_5(\boldsymbol{\mu}, \, \Sigma), \qquad \Sigma = \mathrm{diag}(\boldsymbol{\tau}) \, R \, \mathrm{diag}(\boldsymbol{\tau}),$$

with weakly informative priors:

$$\mu_j \sim \mathcal{N}(0, 10), \qquad \tau_j \sim \text{Exponential}(1), \qquad R \sim \text{LKJ}(1).$$

The normal priors on the population means are deliberately diffuse, corresponding to several orders of magnitude variation in the implied scale parameters. The exponential priors on the standard deviations favour moderate heterogeneity while allowing large variation in the population. The LKJ(1) prior places a uniform density over correlation matrices, avoiding any a priori bias toward independence.

These priors are chosen to be minimally informative. Their purpose is not to regularise behaviour toward any particular psychophysical structure, but simply to stabilise estimation while allowing the posterior to be driven almost entirely by the data. In practice, posterior inferences are highly robust to reasonable alternative choices of prior scale.

Estimation workflow

All models are estimated in Stan using Hamiltonian Monte Carlo (HMC) as implemented in the No-U-Turn Sampler (NUTS). For each specification, I run four parallel Markov chains with 2,000 iterations each, discarding the first 2,000 iterations of each chain as warm-up. This yields 4,000 post warm-up draws for posterior inference.

Priors are specified as described above and are deliberately weakly informative, ensuring that posterior inferences are driven primarily by the likelihood rather than by prior regularisation. All continuous parameters are sampled on an unconstrained scale and transformed to the positive reals where appropriate, which improves sampler stability.

Convergence is assessed using standard diagnostics. I report the potential scale reduction factor $\widehat{R}$ for all parameters and generated quantities, requiring values below 1.02 as evidence of satisfactory mixing. Effective sample sizes (both bulk and tail) are examined to ensure that posterior quantities are estimated with sufficient precision. Trace plots and rank plots are inspected to verify that the chains explore the posterior without divergent transitions or pathological behaviour.

Posterior summaries reported in the main text and online appendix are based on the aggregated post warm-up draws from all four chains. Point estimates are posterior means unless otherwise stated, and uncertainty is summarised using central 95% posterior credible intervals.

D.1 Parameter recoverability and identification

To assess whether the Bayesian decoder can recover the underlying psychophysical parameters from finite binary choice data, I conduct a recoverability exercise based on fully simulated choices. The simulation proceeds in three steps: (i) construction of the choice set; (ii) generation of individual-level psychophysical parameters; and (iii) stochastic choice generation under the full Bayesian model.

(i) Choice set. Each simulated choice consists of a risky lottery (h, p) that pays a high outcome $h = 140$ with probability $p \in \mathcal{P}$ and a low outcome $y = 0$ otherwise, evaluated against a sure amount $c \in \mathcal{C}$. The probability set is

$$\mathcal{P} = \{0.05,\ 0.10,\ 0.15,\ 0.30,\ 0.45,\ 0.50,\ 0.55,\ 0.60,\ 0.70,\ 0.85,\ 0.90,\ 0.95\},$$

and the set of sure amounts is

$$\mathcal{C} = \{10, 15, 20, \ldots, 130\}.$$

The Cartesian product $\mathcal{P} \times \mathcal{C}$ yields the choice set $\mathcal{S}$ containing all combinations of (p, c) paired with the fixed lottery $(h = 140, y = 0)$.

For realism, each simulated individual is later presented with the full choice set together with a small random sample of additional repeated trials (six per individual) to mimic the uneven design of the empirical dataset.

(ii) Psychophysical parameters. I simulate $N = 100$ heterogeneous individuals. Each individual is endowed with four positive psychophysical parameters:

$$\nu_n > 0, \qquad \sigma_{o,n} > 0, \qquad \sigma_{p,n} > 0, \qquad \xi_n > 0.$$

These govern the coding noise, the outcome-channel noise, the probability-channel noise, and the prior mean on outcomes, respectively. Parameters are drawn from diffuse distributions chosen to span a wide region of the parameter space:

$$\nu_n \sim |\mathcal{N}(0.6,\, 0.3^2)|,$$
$$\sigma_{o,n} \sim |\mathcal{N}(0.9,\, 0.4^2)|,$$
$$\sigma_{p,n} \sim |\mathcal{N}(0.7,\, 0.35^2)|,$$
$$\xi_n \sim |\mathcal{N}(0.7,\, 0.3^2)|,$$

where the absolute value ensures positivity. The Bayesian decoder implies two signal weights,

$$\alpha_n = \frac{\sigma_{o,n}^2}{\sigma_{o,n}^2 + \nu_n^2}, \qquad \gamma_n = \frac{\sigma_{p,n}^2}{\sigma_{p,n}^2 + \nu_n^2},$$

and a prior-adjusted outcome factor,

$$\delta_n = \xi_n^{1-\gamma_n}.$$

The effective probit scale is

$$\omega_n = \nu_n \sqrt{\alpha_n^2 + \gamma_n^2}.$$

(iii) Stochastic choice generation. For individual n facing choice problem (p_i, c_i), the Bayesian decoder implies the decision index

$$d_{in} = \frac{\gamma_n \, \log \frac{p_i}{1-p_i} \; - \; \alpha_n \, \log\!\left(\frac{c_i - y}{h - c_i}\right) \; + \; \log(\delta_n)}{\omega_n}.$$

The corresponding choice probability for selecting the risky option is

$$\Pr\big((h, p_i) \succ c_i \mid n\big) = \Phi(d_{in}),$$

and the observed choice is drawn as

$$y_{in} \sim \text{Bernoulli}(\Phi(d_{in}))\,.$$

(iv) Resulting dataset. The final simulated dataset contains all choice pairs (p_i, c_i) crossed with all simulated individuals, plus six additional randomly sampled repeated problems per individual. This construction yields a dataset closely mirroring the structure and heterogeneity of the empirical data, allowing a direct test of whether the hierarchical Bayesian estimator can recover the individual-level psychophysical parameters from finite, realistically noisy choice data.

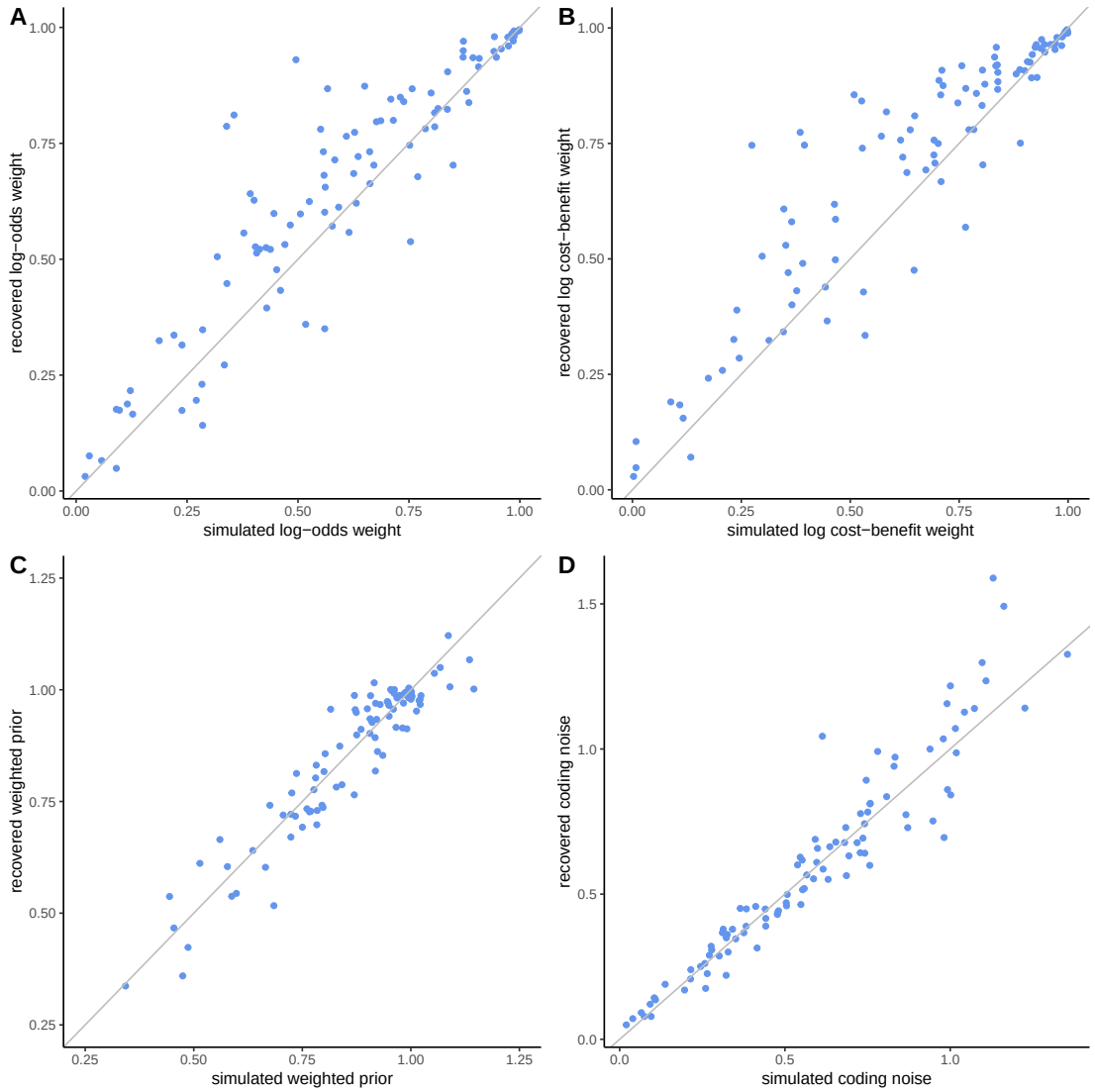


Figure 8: Recoverability of Bayesian Inference Model parameters

Recoverability of Bayesian Inference Model parameters

Figure 8 shows that parameter recoverability is indeed excellent. All spearman correlations between simulated and recovered parameters fall above 0.9.

E Analysis: Robustness checks

I here present various robustness checks for the main analysis described in the body of the paper. In particular, I 1) re-estimate the data using the KLW model instead of the generalization including probability transformations; 2) check robustness to putting β and γ on a common variance scale instead of a common noise scale; and 3) I compare the nonparametric likelihood-sensitivity result to an equivalent result based on the parametric estimates.

E.1 Re-estimation using the KLW model

Figure 9 displays the relationship between individual-level discriminability (α) and the estimated decision-noise parameter $\sqrt{2}\,\nu\alpha$ obtained from fitting the KLW model to the experimental data. With the horizontal axis reversed, movement from left to right corresponds to a transition from high to low discriminability. The key pattern predicted by the Bayesian inference mechanism is clearly present: as discriminability declines, the estimated decision noise falls steadily. This is the precise *opposite* of the qualitative prediction shared by likelihood-based models such as Decision-by-Sampling, or likelihood-only efficient coding, all of which imply that noise should rise monotonically as discriminability worsens. The distinctive Bayesian signature is therefore readily apparent in the experimental data.

We do not, however, observe uniformly low noise for the highest levels of discriminability at the extreme left of the graph. Although most observations in this region come from the low-ECU treatment, the estimated decision noise exhibits substantial vertical dispersion: subjects with very similar discriminability display widely different noise levels. This dispersion is readily explained. When outcome discriminability is high, the KLW noise term $\sqrt{2}\,\nu\alpha$ becomes increasingly sensitive to heterogeneity in how subjects react to *probabilities*. In the present experiment, probability sensitivity varies markedly across individuals, and the as-if expected-utility representation embedded in KLW does not model such heterogeneity explicitly. The residual vertical spread at high discriminability therefore reflects cross-subject variation in probability processing, not a failure of the Bayesian-decoding mechanism.

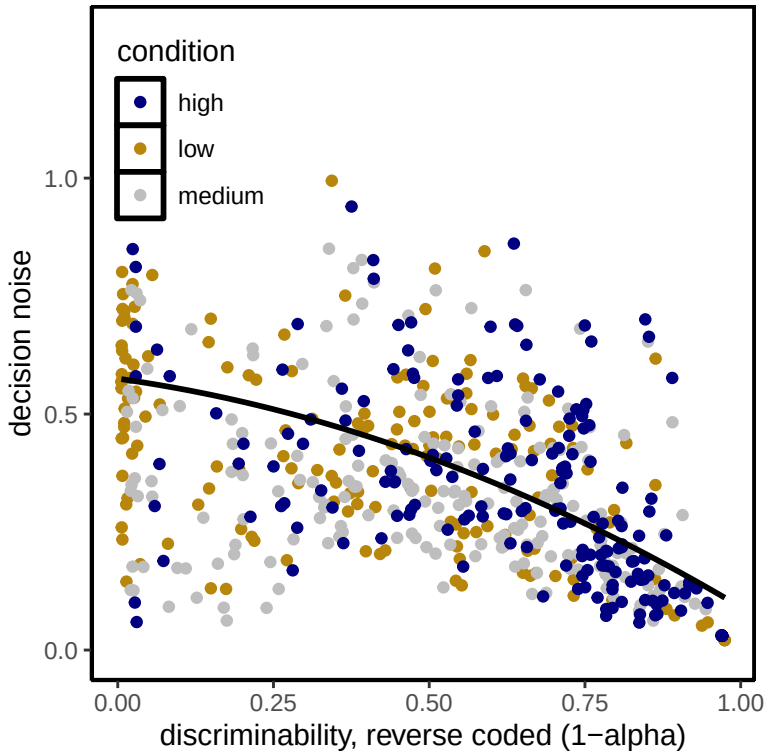


Figure 9: Outcome-discriminability and decision noise in the KLW model

E.2 Alternative prior-variance parameterization

Next, we examine the robustness of the inverse-U relationship to the specific econometric assumptions adopted in the main text. In the benchmark estimation in the main text, I imposed a single coding-noise parameter ν shared across the outcome and probability channels, i.e. $\nu = \nu_o = \nu_p$. Here, I relax this restriction and allow the coding-noise parameters to differ, $\nu_o \neq \nu_p$. To ensure that the two signal weights β (for outcomes) and γ (for probabilities) remain on a common scale—and that they remain identifiable from choice data alone—I impose a common prior variance, $\sigma \triangleq \sigma_o = \sigma_p$, in the Bayesian decoder.

Figure 10 plots the resulting relationship between outcome discriminability β (horizontal axis) and decision noise $\omega \triangleq \sqrt{\nu_o^2\beta^2 + \nu_p^2\gamma^2}$ (vertical axis). The inverse-U pattern is reproduced cleanly: decision noise is lowest when the signal is either very strong or very weak, and reaches its maximum at intermediate discriminability. Allowing ν_o and ν_p to differ therefore does not attenuate the key qualitative signature of Bayesian inference. If anything, the flexibility introduced by heterogeneous coding noise sharpens the curvature

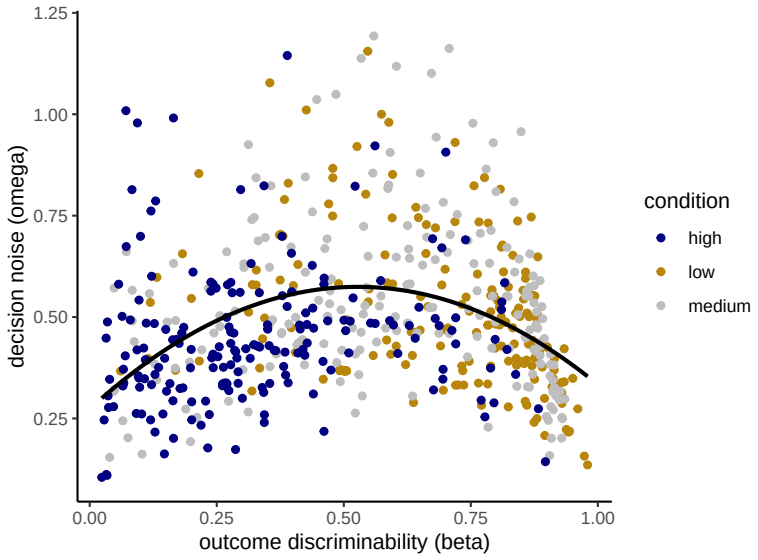


Figure 10: Outcome discriminability and decision noise under a common prior-variance scale.

of the inverse-U, reinforcing the view that the pattern is intrinsic to Bayesian regression to the prior rather than an artifact of the baseline parameterization.

E.3 Nonparametric noise measures

In the main text, I correlated the nonparametric likelihood-sensitivity index with a measure of stochastic dominance violations calculated conditioning on the probability p. The reason was that this index most cleanly reflects variation across outcomes, which is the dimension for which coding noise is being manipulated in the experiment.

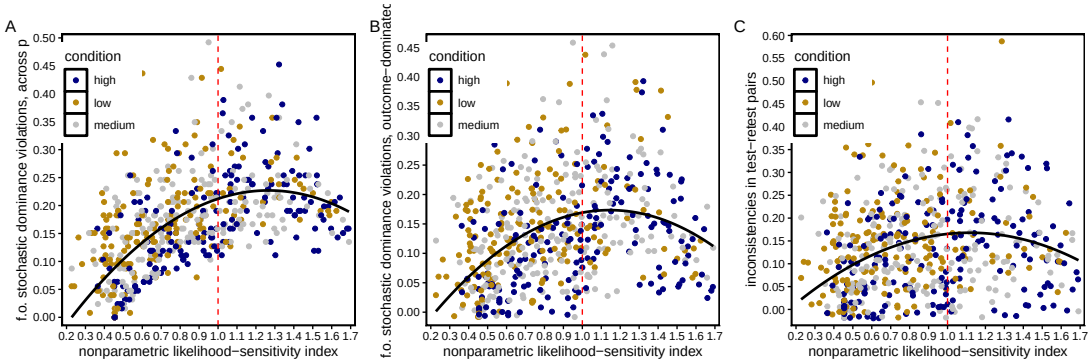


Figure 11: Likelihood-sensitivity and between-probability stochastic dominance violations

Here, I show that the key result of the inverse-U relation is robust to using a different measures of stochastic dominance violations. There are three such alternative measures: (i) a measure of stochastic dominance violations calculated *across* p, i.e. holding all

outcomes constant and registering violations as a switch from choosing the lottery to choosing the sure outcome as p increases; (ii) a measure using specifically outcome-dominated lists, where the lottery is chosen for prize x, but the sure outcome is chosen for prize $x - 2$, holding p, c, y constant; and (iii) a measure of whether choices were inconsistent across repetitions of an identical choice task.

Figure 11 shows correlations of these three measures with the nonparametric likelihood-sensitivity index. In all three cases, the relationship is clearly inverse-U shaped. While the inverse-U relationship appears somewhat weaker in panel C, which plots the relationship with inconsistent choices in repeated choice tasks, that is to be expected: This is indeed the measure that yields the lowest statistical power, since it is based on relatively few binary comparisons. Overall, the pattern is remarkably robust regardless of the specific measure of variability that is used.

F Experimental materials and stimuli

All choice stimuli were constructed from a simple generative rule designed to ensure (i) wide but controlled variation in likelihood discriminability, (ii) balanced coverage of probabilities and sure amounts, and (iii) naturalistic spacing of risky and safe options.

The basic stimulus was a binary lottery (x, p) paying $x > 0$ with probability p and 0 otherwise, evaluated against a sure amount c. The construction proceeded in four steps.

1. Base probability grid. The main set of risky options used a coarse grid of probabilities,

$$p \in \{0.1, 0.2, \ldots, 0.9\}.$$

2. Sure-amount bands centred on the expected value. For each probability p, sure amounts were chosen from a band centred on the expected value of the lottery, $\mathrm{EV}(p) = px$. The sure amount c was drawn from the integer range

$$c \in \{\lfloor \mathrm{EV}(p) - 10 \rfloor, \ldots, \lceil \mathrm{EV}(p) + 6 \rceil\},$$

truncated to lie in the global domain $c \in \{1, \ldots, 21\}$. This rule keeps the risky and sure options economically comparable while providing systematic variation in outcome discriminability.

3. Two auxiliary lists with a lower prize. To broaden the range of risk-return combinations, two additional blocks of trials were created in which the prize was set to $x = 20$ instead of $x = 22$. These blocks used only $p = 0.3$ and $p = 0.7$, with sure amounts generated by the same EV-centred rule as above. This yields two "matched" lists differing only in the lottery prize.

4. A small set of repeated trials. To enable nonparametric reliability checks, a random 15% subsample of the main $x = 22$ block was duplicated and flagged as repeat trials. These repetitions were presented at different points in the sequence.

5. Final trial set. The three components—the full $x = 22$ block, the two $x = 20$ blocks, and the repeat subsample—were combined and sorted by (p, x, c) to generate the final trial list. Each trial was also assigned a unique identifier. Probabilities were displayed to subjects as integer percentages $\text{pcp} = 100 \times p$.

The resulting stimulus set provides a dense, well-controlled design that varies discriminability through probability and outcome scaling while keeping the overall structure of the task intuitive for participants.

The Instructions read as follows:

INSTRUCTIONS

Thank you for taking part in this study.

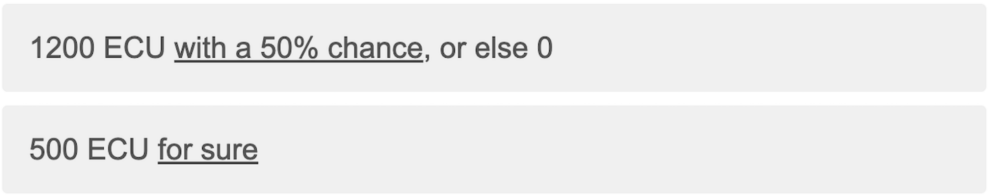
We will ask you to take repeated decisions involving lotteries. On each screen, you will be asked to choose between **a lottery** and **a sure amount** of money.

Chances of winning the prize in the lottery are **always indicated in percentages**.

Monetary payments are **always indicated in Economic Currency Units (ECU)**. At the end, you may be paid a bonus based on one of your choices: **the conversion factor for that bonus is 100 ECU to 1 pound**.

There are no right or wrong answers—we are purely interested in your preferences.

Here is an example of a choice task:



In the example above, you are asked to choose between a lottery paying 1200 ECU with a 50% chance (or else nothing), and a sure payment of 500 ECU.

If this is the randomly selected choice paid for real:

- If you selected the **sure amount**, we will pay you that amount

- If you selected the **lottery**, we will draw a ball from an urn containing 100 sequentially numbered balls. If the ball extracted bears a number between 1 and 50 inclusive, we will pay you the prize. If the ball contains a number between 51 and 100 inclusive, we will pay you nothing.

You will be presented repeatedly with such tasks, and you are asked to indicate your choice for each one of those tasks. Notice that **both the amounts and the chances involved may change from screen to screen**. Please consider the information carefully and choose your preferred option.