Disentangling Suboptimal Updating: Task Difficulty, Structure, and Sequencing*

Marina Agranov† Pëllumb Reshidi ‡

August 2026

Abstract

We study the underlying reasons for the failure of individuals to adhere to Bayes' rule and decompose this departure into three elements: (i) task difficulty, (ii) information structure, and (iii) timing of information arrival. In a series of controlled experiments, we systematically alter all three elements and quantify their magnitude. We link task difficulty with the degree of nonlinearity embedded in Bayesian updating: participants make their largest mistakes precisely where the posterior responds sharply to small changes in signal accuracies.

*Agranov gratefully acknowledges the support of NSF grant SES-2214040. We thank Larbi Alaoui, Dan Benjamin, Ben Enke, Alessandro Lizzeri, Kristof Madarasz, Kirby Nielsen, Pietro Ortoleva, Leeat Yariv, Sevgi Yuksel, and participants at many seminars and conferences for helpful comments and suggestions.

†Department of Economics, California Institute of Technology and NBER; marina.agranov@gmail.com

‡Department of Economics, Florida State University; preshidi@fsu.edu

1 Introduction

Across various contexts, updating beliefs is critical in the decision-making process. In the last few decades, economists and psychologists have made significant progress in understanding how people engage with new information. In many situations, the standard Bayesian updating framework accurately describes the evolution of beliefs (Grether, 1978; Camerer, 1987; Charness and Levine, 2005).[1] At the same time, there are frequent and consistent deviations from Bayesian updating. These deviations persist when people have ample opportunities to learn (Esponda et al., 2023) and when stakes are high (Enke et al., 2023). Moreover, these deviations are prevalent even among professionals who frequently deal with such issues (Eddy, 1982). What makes some belief-updating tasks more difficult and, as a result, more mistake prone than others? This is the focus of our paper.

Belief-updating environments often vary along several dimensions that can influence how individuals incorporate information. In particular, the intrinsic difficulty of the updating problem, the way information is structured, and the timing with which information arrives can all affect belief formation. Because these features frequently vary jointly across environments, isolating their individual contributions is not straightforward. In this paper, we focus on these factors and study their independent effects within a unified experimental framework.

We present findings from a series of experiments involving belief-updating tasks. Starting with the simplest environment, we consider a setting where the state is binary and both states are equally likely. A decision-maker receives two simultaneous binary signals and reports her posterior about the state. Our analysis, conducted across various parameters, empirically documents how deviations from Bayesian predictions, termed mistakes, depend on signals' precisions. We observe two empirical regularities: first, participants make larger mistakes as the accuracy of both signals increases, even when the gap between signal accuracies remains fixed; second, participants make larger mistakes as the gap between signal accuracies widens. We show these regularities closely track nonlinearities in

[1] Not only humans employ Bayesian updating! Valone (2006) reviews experiments involving animals and concludes that a variety of them, across different ecological contexts, behave in a manner consistent with Bayesian updating. Furthermore, a large literature in cognitive science models perception and decision-making as probabilistic inference under uncertainty, often using Bayesian frameworks (Knill and Pouget, 2004; Gold and Shadlen, 2007; Pouget et al., 2013).

Bayesian updating. The difficulty—or, equivalently, the complexity—of integrating information from signals with different accuracies increases when posteriors change significantly with small changes in signal accuracies.[2] Regions in which these nonlinearities are more pronounced are precisely regions in which participants make larger mistakes, unable to fully incorporate the extent to which one signal is more informative than the other.

We formalize this notion of task difficulty in three ways. The first is a model-free approach that links difficulty to the nonlinearity of the Bayesian posterior, measured directly from task primitives. The other two are behavioral models: one adapts the widely used Grether (1980) model to capture how mistakes scale with nonlinearity, and the other proposes an alternative model where agents fail to fully respond to changes in the second derivative of the posterior. All three approaches predict the empirical regularities described above.

We further show that our notion of task difficulty is distinct from several alternatives recently proposed in the literature. The first alternative suggests that mistakes are driven by proximity to corner beliefs (0% or 100%). However, when participants start with an uninformative prior and receive a single signal, the vast majority report the exact Bayesian posterior regardless of proximity to the corner—difficulty arises instead when multiple sources of information must be integrated. A second explanation is the compression effect and its sensitivity to cognitive uncertainty (Enke and Graeber, 2023a). Enke and Graeber (2023a) document that compression in probability judgments is related to cognitive uncertainty, with higher uncertainty associated with greater overweighting of low-probability events and underweighting of high-probability events, leading to compression of beliefs toward 50/50.[3] Using their dataset, we show that even after controlling for cognitive uncertainty, both the level of signal precision and the gap between signal precisions remain significant drivers of updating mistakes. We further demonstrate that our findings differ from predictions based on cognitive noise (Augenblick et al., 2025) and from approaches emphasizing the cardinality of the state space (Ba et al., 2024). Overall, we show that task difficulty from nonlinearities in Bayesian updating is distinct from other proposed sources of mistakes.

[2] Throughout the paper, we treat the terms difficulty and complexity as interchangeable and use them as such.

[3] The authors measure cognitive uncertainty as one minus confidence in the reported belief.

We next investigate how the mistakes identified in scenarios with simultaneous signals translate to more typical belief-updating tasks, where the decision-maker receives information sequentially. To do so, we manipulate two features of the environment: information structure (whether the same information arrives via an informative prior plus one signal, or an uninformative prior plus two signals) and information sequencing (whether signals arrive simultaneously or sequentially, and in what order). Although the two information structures are mathematically equivalent, they differ conceptually, and our experiment provides the first empirical test of this equivalence.

We document several findings. First, compared to what we term the Baseline treatment—where participants begin with an informative prior and receive a signal—providing the same information through two simultaneous signals significantly reduces mistakes. Strikingly, in our low-difficulty parametrization, which uses the canonical Kahneman and Tversky (1973) parameters that have served as a leading example of base-rate neglect for decades (Esponda et al., 2023; Gneezy et al., 2013), this manipulation eliminates base-rate neglect entirely.[4] Second, we find no aggregate effect of information structure on reported beliefs, but we document a sizable recency bias that operates regardless of task difficulty. The optimal information design therefore depends on the environment: in low-difficulty tasks, simultaneous release performs best; in high-difficulty tasks, presenting the high-accuracy signal last leverages recency bias to counteract the difficulty-induced underweighting. We close by decomposing base-rate neglect, showing that it arises primarily from sequencing and task difficulty.

Taken together, our paper contributes on two fronts. We propose a notion of task difficulty rooted in the nonlinearities of Bayesian updating and provide empirical evidence supporting it. More broadly, by varying task difficulty, information structure, and information sequencing while keeping the underlying Bayesian problem unchanged, we isolate each factor's independent effect and quantify its relative importance in driving mistakes.

The remainder of the paper is structured as follows. We survey the literature in Section 1.1. In Section 2, we lay out the conceptual framework. We dedicate Section 3 to the experimental design and procedures. In Section 4, we present the

[4]Base-rate neglect refers to a bias whereby individuals underweight or disregard the information contained in their prior when evaluating conditional likelihoods.

aggregate results of our experiment. Individual-level analyses are presented in the Online Appendix (Agranov and Reshidi, 2026). We conclude in Section 5.

1.1 Literature Review

Our paper builds on a large body of research documenting errors in probabilistic reasoning that produce beliefs misaligned with Bayesian predictions; Benjamin (2019) provides the most recent survey.

A central question in this literature is understanding when we should expect the largest deviations from Bayesian predictions. Three recent papers make important progress in this direction. Enke and Graeber (2023a) link updating mistakes to a measure of cognitive uncertainty, which captures how confident people are in their reported beliefs. The authors show that greater cognitive uncertainty is associated with a compression of reported beliefs toward the 50–50 point. By manipulating the computational complexity of the task, they further show that cognitive uncertainty partly reflects subjective perceptions of task difficulty. Augenblick et al. (2025) document that individuals overinfer from weak signals and underinfer from strong signals. To account for these patterns, the authors develop a model of cognitive imprecision about signal informativeness, in which individuals know the direction of updating but not its magnitude. Ba et al. (2024) hypothesize that difficulty depends on the size of the state space and develop a model of noisy cognition and representativeness that predicts underreaction to new information in simple environments (two states) and overreaction in more complex ones.

Like these papers, we characterize features of the environment that determine when and to what extent people report suboptimal beliefs, interpreting such deviations as a proxy for task difficulty. However, our notion of task difficulty differs from those discussed above and is rooted in nonlinearities in Bayesian updating. In Section 4.3, we show that cognitive uncertainty, cognitive imprecision, and the cardinality of the state space cannot account for the mistakes we observe in our treatments. We therefore view our work as identifying a distinct and complementary dimension of task difficulty.

Our notion of task difficulty connects to a literature on suboptimal decisions in other environments with nonlinear features. These include exponential growth bias—the tendency to underestimate compound growth processes in financial decisions (Wagenaar and Sagaria, 1975; Stango and Zinman, 2009; Levy and Tasoff,

2016, 2017); the schmeduling heuristic, which refers to simplified ways of constructing mental representations of nonlinear incentive schemes (Rees-Jones and Taubinsky, 2020); the MPG illusion, which refers to mistaken beliefs that gasoline consumption decreases linearly with a car's MPG (Larrick and Soll, 2008); and mistakes in evaluating atemporal payments that involve a large number of steps (Enke et al., 2025).[5] To our knowledge, our study is the first to extend this concept to belief-updating tasks and to demonstrate its usefulness in organizing errors in the formation of posterior beliefs.

Our notion of task difficulty also relates to complexity notions in lottery choices. Recent papers by Shubatt and Yang (2025) and Enke and Shubatt (2023) emphasize the role of dissimilarity among lotteries as a key source of decision complexity. In both belief-updating tasks and lottery choice problems, decision-makers must integrate and trade off multiple components of the problem. Shubatt and Yang (2025) argue that choices are easier when options are similar because less aggregation of tradeoffs is required to compare them, whereas greater dissimilarity increases complexity. Enke and Shubatt (2023) document that state-by-state dissimilarity among lotteries in the choice set is the leading empirical determinant of choice complexity, which they term 'tradeoff complexity'.

Our results point to a related mechanism in belief-updating tasks. We show that individuals make more mistakes when they must aggregate information from two opposing signals with different levels of precision. This pattern is consistent with the broader idea that the integration of information becomes more difficult when the components to be combined are more dissimilar. In this sense, the difficulty we identify in belief updating is conceptually related to the notion of tradeoff complexity studied in lottery choice.[6]

Finally, our paper also relates to the literature on base-rate neglect, one of the most persistent biases in belief updating. First documented by Kahneman and Tversky (1972) and Bar-Hillel (1980), this bias has since been studied extensively, both empirically and theoretically (Benjamin, 2019; Benjamin et al., 2019). Recent

[5] See also papers that use the number of kinks in piecewise-linear schedules to study mistakes (Shaffer, 2020; Goodman and Puri, 2025).

[6] Dissimilarity is not the only notion of complexity studied in the lottery choice literature. Other work links complexity to the number of distinct outcomes in the lottery support (Bernheim and Sprenger, 2020; Puri, 2022; Fudenberg and Puri, 2022) or the difficulty of aggregating outcomes and probabilities into a single value (Oprea, 2024b). More broadly, a growing literature studies complexity beyond lottery choice; see Oprea (2024a) for a recent survey.

work shows that it can persist even after extensive opportunities to learn from feedback (Esponda et al., 2023).[7] The bias is largely insensitive to financial incentives (Enke et al., 2023) but can be attenuated by presenting information as frequencies (Koehler, 1996; Barbey and Sloman, 2007), using intuitive rather than abstract framings (Cheng and Holyoak, 1985; Gigerenzer and Hoffrage, 1995; Enke et al., 2023; Ganguly et al., 2000), or asking for forecasts rather than state beliefs (Fan et al., 2022). While these interventions can attenuate the bias, they rarely eliminate it. We study a different manipulation and show that, for parameter values commonly used in the literature, it eliminates base-rate neglect from the outset. Although we use base-rate neglect as a case study, our insights apply more broadly to belief-updating environments involving probabilistic information.[8]

2 Conceptual Framework

2.1 Setup

Consider a standard belief-updating task. The state is binary $\omega \in \{F, S\}$, denoting, for example, whether a project is a Failure or a Success. A decision-maker does not know the state but holds prior belief $P(F) = p_0$. They observe the realization of a signal s, which can take either a negative ($s = n$) or positive ($s = p$) value. The signal has accuracy θ_s, which summarizes the probability that it correctly reveals the state, $P(n|F) = P(p|S) = \theta_s$. Upon observing a signal, Bayes' rule implies that the posterior beliefs are given by

$$P(F|s) = \frac{P(s|F)p_0}{P(s|F)p_0 + P(s|S)(1-p_0)}. \tag{1}$$

The posterior reflects that both the prior and the signal contain valuable information, which the decision-maker uses to update their beliefs.

[7] See also the "nudge" literature, which studies interventions that improve decision-making (Thaler and Sunstein, 2008; Thaler and Benartzi, 2004; Madrian and Shea, 2001).

[8] See also a more distantly related literature on the mechanisms that generate biased updating. In some of these models individuals understand Bayesian updating but distort probabilities—for example due to misspecified sampling processes (Rabin, 2002; Rabin and Vayanos, 2010), noisy perception or rational inattention (Woodford, 2020; Enke and Graeber, 2023a), while in others attention distortions are driven by salience (Bordalo et al., 2026). Our paper differs from this literature by focusing on which updating environments generate large errors, rather than on the psychological mechanisms behind them.

2.2 Information Structure

For our analysis, it is useful to distinguish between informative and uninformative priors. In the binary case, the uninformative prior assigns equal probability to both states, $p_0 = 1/2$. With this prior, once an agent receives a signal, the posterior is fully determined by the signal's realization and accuracy.[9] In this sense, the prior does not interfere with the information conveyed by the signal. Henceforth, we refer to a prior as *uninformative* if it assigns probability $1/2$ to each state and as *informative* otherwise.

Consider an agent with an informative prior $P(F) = p_0 > 1/2$. This prior can be interpreted as the posterior resulting from an initial uninformative prior $P(F) = \tilde{p}_0 = 1/2$ and a negative signal with accuracy $\theta_s = p_0$.[10] In this case,

$$P(F|s = n) = \frac{\theta_s 1/2}{\theta_s 1/2 + (1 - \theta_s)1/2} = \theta_s = p_0.$$

Thus, having a prior $p_0 > 1/2$ is equivalent to starting from an uninformative prior and observing a negative signal with accuracy p_0. This equivalence extends to settings with multiple signals. Consider an agent with an uninformative prior $\tilde{p}_0 = 1/2$ who receives two signals with accuracies θ_1 and θ_2, where $\theta_2 = \theta_s$ matches the accuracy of the single signal in the original setup. Conditional on $s_1 = n$, the posterior is

$$P(F|s_2, s_1 = n) = \frac{P(s_2|F)P(F|s_1 = n)}{P(s_2|F)P(F|s_1 = n) + P(s_2|S)(1 - P(F|s_1 = n))}.$$

If the first signal has accuracy $\theta_1 = p_0$, this becomes

$$P(F|s_2, s_1 = n) = \frac{P(s_2|F)p_0}{P(s_2|F)p_0 + P(s_2|S)(1 - p_0)}, \tag{2}$$

which coincides with the posterior obtained when an agent with prior p_0 receives a signal with accuracy θ_s (compare equation (2) and equation (1)). Thus, an information structure with an informative prior and one signal is equivalent to one with an uninformative prior and two signals. Importantly, this equivalence holds regardless of whether the signals arrive simultaneously or sequentially. This is critical for our setup because, as we move from one treatment to another, we alter

[9] Specifically, if the signal accuracy is θ_s and the realization is negative (positive), Bayes' rule implies that the project is a Failure (Success) with probability θ_s.

[10] Assuming $p_0 > 1/2$ is without loss of generality; if $p_0 < 1/2$, let the realized signal be $s = p$ and signal accuracy be $\theta_s = 1 - p_0$.

the structure and sequencing of information, yet the problem remains unchanged.

2.3 Task Difficulty

Motivating Example. To illustrate the general idea behind our notion of task difficulty, consider the following example. There are two information structures, both with an uninformative prior and two signals, the accuracy of which differs by five percentage points. In the first case, signal accuracies are $\theta_1 = 0.85$ and $\theta_2 = 0.80$, while in the second, they are $\theta_1 = 0.97$ and $\theta_2 = 0.92$. Consider an individual who observes two signals, $s_1 = n$ and $s_2 = p$, and updates posterior beliefs. If both signals had identical accuracy, say both equal to 0.80, or both equal to 0.92, then it seems natural to place equal weight on both signals and form a posterior equal to the original prior of 50%. However, because the first signal is more accurate than the second, individuals may argue they should weigh the first signal slightly more and, thus, compute a posterior that leans more toward Failure than Success.

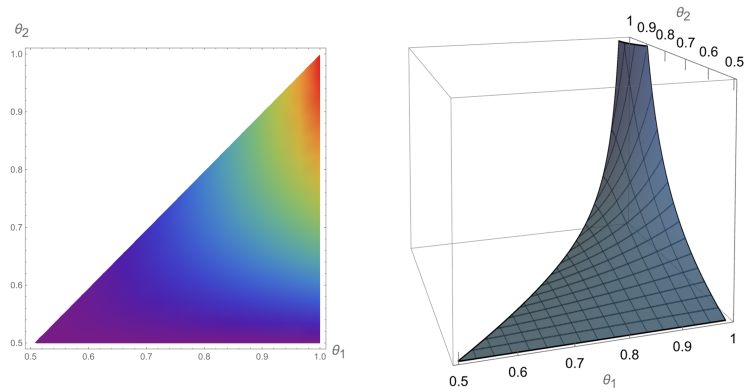
Although this is true in both cases, the extent to which the first signal is more informative than the second is quite different between them. The Bayesian posterior probability of Success is 41.38% in the first case, seemingly a plausible value near which individuals' average posteriors may lie. In contrast, the Bayesian posterior is 26.2% in the second case, despite the difference in signal accuracies remaining only five percentage points. This pattern becomes even more pronounced with higher accuracy signals or as the disparity between signal accuracies increases.

Task Difficulty and Nonlinearities. The difficulty illustrated above stems from the non-triviality of incorporating information from signals with different accuracies. The challenge is to know how much more one needs to react to the higher accuracy signal compared to the lower accuracy one. The answer to this question depends on the accuracy of both signals. We argue individuals may struggle to fully internalize the extent to which Bayesian updating requires nonlinear thinking. Figure 1 depicts the derivative of the posterior with respect to the high-accuracy signal.[11] The marginal increase is unchanged only when facing a single signal ($\theta_2 = 1/2$ reduces the low-accuracy signal to an uninformative one). In all other cases, the marginal change in the posterior depends not only on the signal's own accuracy but on the accuracy of the other signal as well, i.e., the Bayesian

[11] Predictions are unchanged if alternatively, we focused on the low-accuracy signal.

posterior is no longer linear. These nonlinearities generate what we term task difficulty—the intrinsic complexity of the updating problem.

Figure 1: Bayesian Posterior Derivative w.r.t high-accuracy signal, θ_1



Notes: The left graph depicts a heatmap (warmer colors represent higher values), and the right graph is a 3D plot. In both graphs, we focus on $\theta_1 \geq \theta_2$, since θ_1 denotes the high-accuracy signal.

We argue that when individuals do not fully account for nonlinearities, they perform better in regions where these nonlinearities are mild and worse where they are strong. When small changes in signal precision lead to large shifts in posteriors, mistakes are larger. For example, the case with $\theta_1 = 0.85$ and $\theta_2 = 0.80$ lies in a relatively linear region, whereas $\theta_1 = 0.97$ and $\theta_2 = 0.92$ lies in a more nonlinear one; accordingly, we expect larger errors in the latter.

By task difficulty, we do not mean perceived difficulty. From the participant's perspective, the task changes little across parameter values. Rather, difficulty arises from the nonlinear structure of Bayesian updating itself. If individuals use rules that do not fully capture these nonlinearities, discrepancies will be small in some regions and large in others. We define regions with larger discrepancies—and thus more mistake-prone—as more difficult, reflecting the inherent complexity of the updating problem.

Concrete Formulations. We formalize this idea in three ways in the Appendix. Section A.1 proposes a simple approach based on Bayesian posteriors, in which difficulty increases with posterior nonlinearity. Section A.2 adapts the Grether (1980) model—widely used in empirical work—by imposing minimal restrictions that capture these features; the resulting model is more inert than the Bayesian benchmark, especially where small changes in signal accuracy imply large posterior shifts. Section A.3 models agents as only partially responding to curvature,

reflecting limited sensitivity to second derivatives. These are three possible ways to capture how nonlinearities in Bayesian updating can hinder belief formation, and we expect future work to develop additional specifications of this mechanism.

Testable Implications. The preceding discussion yields two testable implications. First, task difficulty increases with the level of signal accuracies, holding the gap fixed. Second, task difficulty increases with the gap between signal accuracies, holding the level fixed. It is these testable implications that we take to the data. The empirical footprint of these implications is the wedge between posteriors reported by participants and Bayesian predictions. If our notion of task difficulty holds water, we expect to see greater mistakes in more difficult belief-updating problems—higher levels of and gaps between signal accuracies.[12]

3 Experiment Design

Our experiment varies three features: task difficulty, information structure, and information sequencing. Task difficulty is manipulated through parameter values. Information structure varies whether participants receive an informative prior and one signal or an uninformative prior and two signals. Information sequencing varies whether the two signals arrive simultaneously or sequentially. To isolate the interaction between timing and signal accuracy, we include two sequential treatments that differ only in the order of accuracies (high–low vs. low–high).

Main Task. Participants face a standard belief-updating task with a binary state and binary signal(s). In each round, the state is represented by a project drawn from a pool; it is a Failure with probability p and a Success with probability $1-p$. Participants know p but not the realized state. Depending on the treatment, they receive one or two conditionally independent signals with known accuracies.[13] When two signals are present, their accuracies differ: θ_1 (θ_2) denotes the higher (lower) accuracy signal. Using the strategy method, we elicit posterior be-

[12]While our experiment employs a discrete two-state, two-signal setting, the underlying intuition can generalize more broadly. For instance, under normal updating with a diffuse prior, the posterior mean is a precision-weighted average of the signals. When signals are initially imprecise, small changes in one signal's variance produce only modest adjustments in the weight assigned to each signal. However, when signals are already highly precise, even a small change in variance can induce a large shift in weights—paralleling the mechanism identified in our discrete framework.

[13]Signal accuracy is the probability that a signal correctly identifies the true state, i.e., $\theta_s = P(s = n \mid F) = P(s = p \mid S)$.

liefs for every possible signal realization.[14] This ensures a balanced dataset. Each participant is assigned to one treatment and completes 20 rounds.[15]

Feedback. At the end of each round, participants observe the realized signal(s) (positive or negative) and the state (Success or Failure). Outcomes from all previous rounds are displayed in a summary table at the bottom of the screen. We provide detailed feedback to mitigate memory issues that could disrupt learning and confound results (see Online Appendix screenshots).

Treatments. In the **Baseline** treatment, participants start with an informative prior p and receive one signal with accuracy θ_2.[16] This setup mirrors the classic design of Kahneman and Tversky (1972) and the base-rate neglect literature.

In the **Simultaneous** treatment, participants have an uninformative prior and receive two conditionally independent signals with accuracies θ_1 and θ_2. With two binary signals, four signal combinations are possible, but under an uninformative prior these reduce to two cases by symmetry: aligned signals (both positive or both negative) and misaligned signals (one of each). To maintain comparability, we randomly draw the first signal and use the strategy method to elicit posteriors for both realizations of the second signal. Details are in the Online Appendix.[17]

[14]The strategy method is widely used in experiments, including belief-updating tasks (Gneezy et al., 2013; Esponda et al., 2023). For a comparison with the direct-response method, see Brandts and Charness (2011).

[15]Repeated tasks are standard and allow participants to familiarize themselves with the interface and approach optimal responses; we address learning in the data analysis section.

[16]For each participant, we randomly assign whether p refers to Success or Failure (with equal probability) at the start of the experiment, and this remains fixed throughout.

[17]This design ensures a balanced dataset and that participants encounter all signal combinations across rounds.

Table 1: Sessions, Treatments, and Parameter Values

Session	# Participants	Treatment	Parameter	Prior	Low-Accuracy Signal	High-Accuracy Signal
1	101	Baseline	$\tilde{A}$	p=0.85	$\theta_2 = 0.80$	—
2	101	Simultaneous	A	p=0.50		$\theta_1 = 0.85$
3	101	Sequential High-Low				
4	100	Sequential Low-High				
5	99	Baseline	$\tilde{B}$	p=0.95	$\theta_2 = 0.85$	—
6	99	Simultaneous	B	p=0.50		$\theta_1 = 0.95$
7	102	Sequential High-Low				
8	100	Sequential Low-High				
9	100	Simultaneous	C	p=0.50	$\theta_2 = 0.75$	$\theta_1 = 0.85$
10	100		D		$\theta_2 = 0.80$	$\theta_1 = 0.90$
11	100		E		$\theta_2 = 0.85$	$\theta_1 = 0.90$
12	99		F		$\theta_2 = 0.90$	$\theta_1 = 0.95$

In the **Sequential High-Low** treatment, participants start with an uninformative prior and receive two signals sequentially, with the higher-accuracy signal first. After observing its realization, they report their updated belief, and then use the strategy method to report beliefs for both possible realizations of the second signal.[18] The **Sequential Low-High** treatment is identical, except the lower-accuracy signal arrives first.

In summary, in each treatment and round, we elicit two beliefs: one when the signal(s) align with the prior and one when they do not.[19]

Parameters. Table 1 summarizes treatments and parameter values. We use two main parametrizations, A and B. Under A, when two signals are present, accuracies are $(\theta_2, \theta_1) = (0.80, 0.85)$; in the Baseline (denoted $\tilde{A}$), the prior is $p = 0.85$ and the signal has accuracy $\theta_2 = 0.80$, matching standard values in the base-rate neglect literature (Kahneman and Tversky, 1972; Esponda et al., 2023). Under B, signal accuracies are $(\theta_2, \theta_1) = (0.85, 0.95)$, and in the Baseline (denoted $\tilde{B}$), $p = 0.95$ and $\theta_2 = 0.85$. As noted in Section 2.2, Bayesian predictions are identical across treatments within each parametrization. Parametrization B serves to test the task difficulty mechanism and the robustness of results from A.

[18] This mirrors the Simultaneous treatment, where the first signal is drawn and beliefs are elicited for both realizations of the second signal.

[19] A signal aligns with the prior if the prior leans toward Failure (Success) and the realized signal is negative (positive). With two signals, alignment means both signals have the same realization; misalignment means one is positive and the other negative.

Four additional parametrizations, C–F, allow us to decompose how the level and gap in signal accuracies affect task difficulty. Based on Section 2.3, we rank all parametrizations from easiest (A) to most difficult (B). The ordering follows two principles, described above, with C_j denoting the difficulty of parametrization j:

1. *Higher accuracy levels increase difficulty, holding the gap fixed:*
   - $C_A(0.80, 0.85) < C_E(0.85, 0.90) < C_F(0.90, 0.95)$
   - $C_C(0.75, 0.85) < C_D(0.80, 0.90) < C_B(0.85, 0.95)$
2. *Larger accuracy gaps increase difficulty, holding the level fixed:*
   - $C_A(0.80, 0.85) < C_D(0.80, 0.90)$
   - $C_E(0.85, 0.90) < C_B(0.85, 0.95)$

Interface. Figure 11 in Section B in the Appendix shows the interface for the Baseline treatment. Detailed descriptions of the interface and instructions for each treatment can be found in the Online Appendix.

Subject Pool. We ran the experiment on Prolific with about 100 participants per treatment (12 treatments; N = 1,202). Participants were U.S.-based, aged 18–70, fluent in English, and had high approval ratings. Each treatment had equal numbers of men and women. The main experiment was conducted in October–December 2022, with two additional treatments (D and E) run in July 2023.

Participants' Payments. Participants received a $5 completion payment in all treatments. In addition, each had a 20% chance of being selected for a bonus: one round was randomly chosen, and responses in that round determined an additional $20 payment. Beliefs were incentivized using the standard BDM method.[20] The experiment lasted about 20 minutes, with average earnings of $7.97.

Implementation. The experiment was approved by Caltech (IR22-1237) and Duke University IRB (2023-0033) and preregistered on aspredicted.org.[21] The ex-

[20] BDM is incentive-compatible regardless of risk preferences (Becker et al., 1964). We also informed participants that truth-telling maximizes expected payoffs; Danz et al. (2022) show that announcing that truth-telling is optimal is an effective way to elicit true beliefs.

[21] The experiment was conducted in three waves: the initial wave with parametrization A and B in October 2022; decomposing task difficulty treatments, parametrization C and F, in December 2022; additional task difficulty treatments, parametrization D and E, in July 2023. Each wave was separately preregistered on aspredicted.org; see preregistrations at aspredicted.org/blind.php?x=833_2RK, ...x=QWN_L6G, and ...x=638_8K7.

perimental software was programmed in oTree (Chen et al., 2016). Instructions and screenshots of the interface are presented in the Online Appendix.

4 Results

Approach to data analysis. Our main analysis focuses on cases in which participants receive misaligned information.[22] When information is aligned, the Bayesian posterior is very close to zero.[23] With predicted values near the boundary, implementation errors are likely to be asymmetric, making these observations less informative for our main analysis. Accordingly, as stated in our preregistration, we focus on elicitations from misaligned signals.[24] We nevertheless use all elicitations in individual-level analyses reported in the Online Appendix.

To simplify the presentation, we normalize the prior and the high-accuracy signal to be negative. Since we focus on misaligned signals, this normalization implies a positive realization for the low-accuracy signal.

4.1 Updating from One Signal

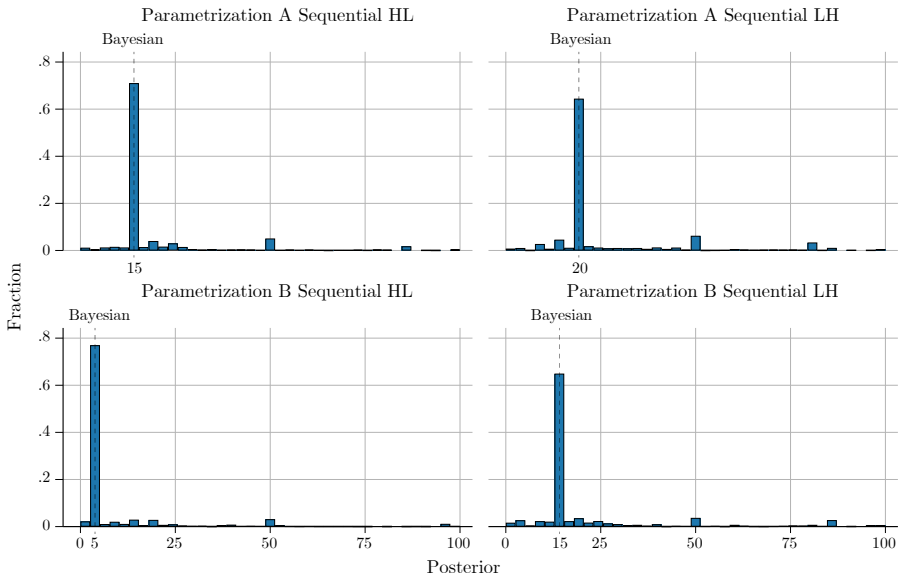
Our notion of task difficulty emphasizes that difficulty arises when people integrate signals with different accuracies. When only one signal is observed, such difficulty is absent by definition. We therefore begin by examining how participants update their beliefs after observing a single signal. This data is generated from the Sequential treatments, where beliefs are elicited twice: once after the first signal and again via the strategy method for both realizations of the second signal.

[22] In the Baseline treatment, information is aligned (misaligned) if the signal's realization agrees (disagrees) with the direction of the prior. In all other treatments, information is aligned (misaligned) if the realized signals are the same (different).

[23] These probabilities are 0.042 for parametrization A, 0.009 for B, 0.055 for C, 0.027 for D, 0.019 for E, and 0.006 for F.

[24] An alternative approach would treat these observations as truncated, but doing so requires assumptions about the truncation process and the distribution of implementation errors.

Figure 2: Sequential Treatments: Posteriors after the First Signal



Notes: The histograms of participants' posteriors are reported with the fraction of choices on the vertical axis and posteriors on the horizontal axis. The dashed lines indicate the Bayesian posterior.

Our data show that beliefs after the first signal closely track Bayesian posteriors (Figure 2). The upper (lower) panels correspond to parametrization A (B), with dashed lines at 15%, 20%, 5%, and 15%. The vast majority of choices cluster around these values. Because noise to the left is bounded at 0 while the right tail is unconstrained, mean responses sit a few percentage points above the Bayesian posterior, while the median matches it exactly in all four cases.[25] This is consistent with our framework: with one informative signal, the Bayesian posterior is linear in signal accuracy, so mistakes are minimal.

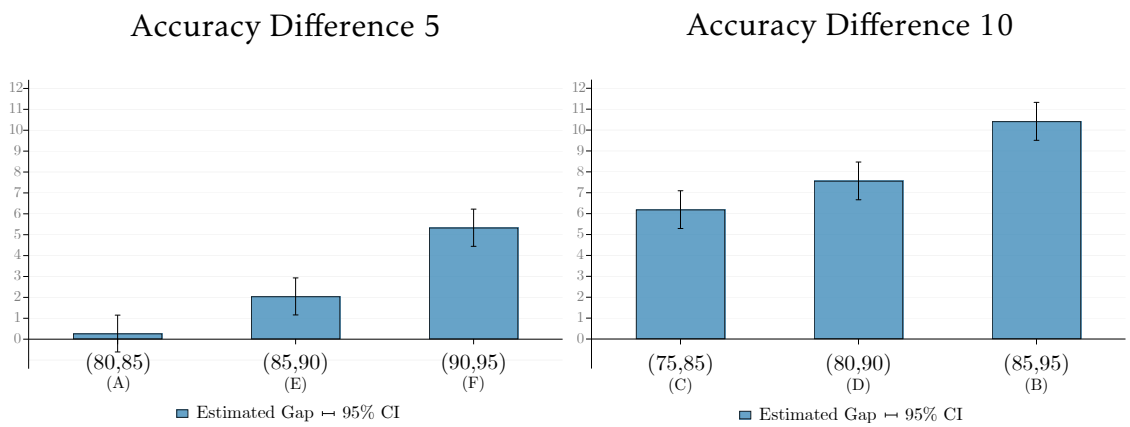
Result 1 (Updating with One Signal)**.** *When starting from an uninformative prior and observing a single signal, participants' beliefs are broadly consistent with Bayesian updating.*

[25]Two minor deviations appear in the data. A small share of participants report a posterior of 50, and a few report the inverse of the correct value (85 instead of 15, 80 instead of 20, 95 instead of 5, and 85 instead of 15). Inspection of individual responses indicates these are occasional mistakes by a few participants rather than systematic deviations.

4.2 Task Difficulty

The updating task becomes more difficult when it requires combining two signals with different precisions. To study this, we focus on the Simultaneous treatments. Figure 3 shows how the gap between reported and Bayesian posteriors varies with the level of signal accuracies, holding their difference fixed (5 in the left panel, 10 in the right). In both cases, higher accuracy levels lead to larger mistakes.

Figure 3: The Impact of Signal Accuracy Levels on Task Difficulty

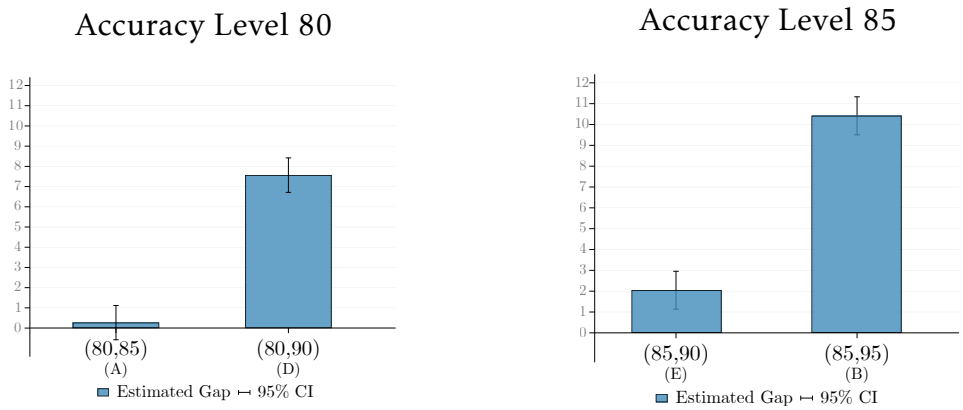


Notes: We report the difference between observed and Bayesian posteriors, averaged across participants and rounds, with 95% confidence intervals.

Figure 4 shows how mistakes respond to larger gaps in signal accuracies, holding the level fixed (80 in the left panel, 85 in the right). In both cases, increasing the gap leads to larger mistakes.[26]

[26] Using an overreaction metric—normalizing the difference by the distance between the Bayesian posterior and the prior, as in Ba et al. (2024)—yields similar results: mistakes increase with both the gap and the level of signal accuracies. The only exception is when the gap is 10, where mistakes remain relatively stable across levels.

Figure 4: The Impact of Signal Accuracy Difference on Task Difficulty



Notes: We report the difference between observed and Bayesian posteriors, averaged across participants and rounds, with 95% confidence intervals.

We collect these findings in Table 2, where we regress the gap between reported and Bayesian posteriors on the level (θ_2) and the difference between signal accuracies ($\theta_1 - \theta_2$), both centered at our easiest parametrization A ($\theta_2 = 80$, $\theta_1 - \theta_2 = 5$). The first column pools all 20 rounds; the second restricts to the last five. The constant is statistically indistinguishable from zero in both columns, indicating that at the easiest parametrization beliefs align with Bayesian predictions on average. Beyond that, each percentage-point increase in the level widens the gap by roughly 0.46 pp, and each percentage-point increase in the difference widens it by 1.57 pp—both effects are stable across all rounds and the last five, suggesting limited learning.

Table 2: Accuracy Level and Difference: Impact on Observed Gap

	Gap	
	All Rounds	Last 5 Rounds
Level	0.464***	0.421***
	(0.124)	(0.152)
Difference	1.566***	1.555***
	(0.240)	(0.289)
Constant	0.228	-1.253
	(0.854)	(1.063)
N (observations)	11,980	2,995
K (individuals)	599	599

Individual-level clustered errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Level centered at $\theta_2 = 80$; Difference centered at 5.

Result 2 (Mistakes: Level and Difference)**.** *Both the level and the difference between signal accuracies increase the gap between observed and Bayesian posteriors.*

Linear Thinking Estimation. The reduced-form regression above documents that mistakes scale with both the level and the difference in signal accuracies. We now ask how well a structural model with a single parameter—capturing limited ability to incorporate nonlinearities—fits these patterns. As we show in detail in Section A.3 in the Appendix, the posterior of such an agent can be decomposed into two parts: the Bayesian posterior and a fully linear posterior:

$$\tilde{P}(S|s_2 = p, s_1 = n) = \alpha \underbrace{\frac{\theta_2(1-\theta_1)}{\theta_2 + \theta_1 - 2\theta_2\theta_1}}_{\text{Bayesian Posterior}} + (1-\alpha)\underbrace{\left(\frac{1}{2} - \theta_1 + \theta_2\right)}_{\text{Fully Linear}}.$$

The α parameter quantifies the agent's ability to incorporate nonlinearities: $\alpha = 0$ indicates a complete failure to do so, while $\alpha = 1$ corresponds to Bayesian updating. We estimate α via linear regression and present the results in Table 3, with the estimated model shown in Figure 5. The estimated value is approximately 0.42 across all rounds, rising to 0.56 in the last five.[27] Participants thus adjust toward Bayesian behavior with experience, but the adjustment is modest and a sizable gap remains. The model that best fits the data places participants between Bayesian and linear updaters: they exhibit some nonlinear thinking but only partially internalize the nonlinearities required for Bayesian updating.[28]

Result 3 (Updating: Bayesian vs. Linear)**.** *Participants only partially incorporate nonlinearities involved in Bayesian updating. Their behavior is best described by a model that lies between Bayesian and fully linear updating.*

4.3 Task Difficulty Robustness Checks

Task Difficulty and Cognitive Uncertainty. It is useful to compare the implications of our notion of task difficulty with those predicted by cognitive uncertainty (Enke and Graeber, 2023a). Cognitive uncertainty predicts a compression effect: when individuals are less confident in their answers, they report beliefs closer to 50. Enke and Graeber (2023a) provide evidence for this pattern using a dataset

[27] In the Online Appendix we examine whether the estimated α varies across demographic groups, including sex, age, and ethnicity. Of these, only age exhibits a statistically significant association, with older participants tending to have lower estimated values of α.

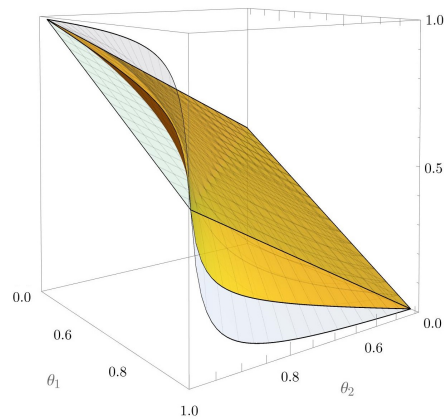
[28] Applying a similar exercise with a nonlinear regression on the alternative modified Grether model in Section A.2 yields approximately 0.46 across all rounds and 0.59 in the last five.

Table 3: Estimated α

	All Rounds	Last 5 Rounds
$\hat{\alpha}$	0.417***	0.564***
	(0.0555)	(0.0663)
N (observations)	11,980	2,995
K (individuals)	599	599

Individual-level clustered errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure 5: Estimated Model



Notes: The figure illustrates the Bayesian ($\alpha = 1$) and fully linear ($\alpha = 0$) models through transparent graphs, along with the estimated model from the last five rounds ($\alpha = 0.56$) shown in yellow.

that includes belief elicitations and individual-level measures of cognitive uncertainty.[29]

Table 4: Robustness Check

	Gap					
				Low CU	Mid CU	High CU
Difference	0.167***	0.173***	0.178***	0.118***	0.247***	0.222***
	(0.0154)	(0.0153)	(0.0155)	(0.0217)	(0.0245)	(0.0398)
Level	0.235***	0.233***	0.228***	0.204***	0.245***	0.326***
	(0.0509)	(0.0508)	(0.0501)	(0.0750)	(0.0735)	(0.109)
Cognitive Unc		0.0572***	0.0529***	0.0618	0.141**	0.0221
		(0.0179)	(0.0180)	(0.0584)	(0.0593)	(0.126)
Other Controls	No	No	Yes	No	No	No
N	2,866	2,866	2,866	1,490	1,007	369

Individual-level clustered errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: Each column represents a separate regression of the gap between reported and Bayesian posteriors. *Other Controls* include participants' age, education, Raven scores, and gender. We renormalize $Difference = \frac{Difference}{\max(Difference)} \cdot 100$ and $Level = \frac{Level}{\max(Level)} \cdot 100$ to ensure they are in the same order of magnitude as Cognitive Uncertainty, which can take values between 0 and 100. The *Low*, *Mid*, and *High CU* columns group observations with $CU \leq 33$, $33 < CU \leq 66$ and $66 < CU$ respectively.

We use their data, as published in their replication package (Enke and Graeber, 2023b).[30] To ensure comparability with our setting, we focus on cases with an

[29] Enke and Graeber (2023a) measure cognitive uncertainty by asking participants how certain they are that the Bayesian posterior lies within a 2-percentage-point window around their stated posterior. This question appears after participants report their updated posteriors.

[30] We are grateful to the authors for sharing their data.

informative prior and a single signal.[31] Regression analysis in Table 4 shows that both the level of and the difference between signal accuracies—our measure of task difficulty—significantly affect the gap between reported and Bayesian posteriors even after controlling for cognitive uncertainty.[32]

Task Difficulty and Proximity to Corner Beliefs. The analysis of updating with one signal presented in Section 4.1 shows that participants can make correct choices even far from the midpoint (50%) when the problem is linear. For example, in the most extreme case—when the correct response is 5 in the Sequential B HL treatment after the first signal—roughly 75% of participants choose this option. After the second signal, the most extreme posterior occurs when $\theta_1 = 0.95$ and $\theta_2 = 0.85$, yielding a posterior of 22.97, which is much farther from the boundary than the earlier case. Hence, the increase in mistakes as signal precision rises cannot be explained by difficulty of making correct choices near the extremes.

Task Difficulty and Cognitive Noise. Augenblick et al. (2025) show that individuals may under- or over-infer from signals in ways consistent with a cognitive-noise model in which noise interacts with signal precision. While related to our findings, our framework emphasizes that behavior depends not only on the precision of a given signal but also on the number of information sources and their precisions. We estimate a simple model that captures their framework

$$\log\left(\frac{p_1}{1-p_1}\right) = \log\left(\frac{p_0}{1-p_0}\right) + k\left[\log\left(\frac{\theta_s}{1-\theta_s}\right)\right]^{\beta}$$

where $p_1 = P(F|s = n)$. The Baseline treatment (parameters A and B) yields $\hat{k} = 1.35$ and $\hat{\beta} = 1.78$, whereas both should equal 1 under Bayesian updating. These estimates imply that participants over-infer from the signal, especially when precision is high (see Section 4.4). Re-estimating the model in the Sequential treatment—focusing on beliefs after the first signal to match a prior-plus-one-signal environment—gives $\hat{k} = 0.79$ and $\hat{\beta} = 1.09$, much closer to Bayesian values. The

[31] The parameters include priors (50, 70, 90, 95, 99) and signal accuracies (65, 70, 75, 90). Our findings in Section 4.5 suggest that beliefs in this setting should resemble those in treatments with sequential signals and an uninformative prior. While sequencing may also affect beliefs, a back-of-the-envelope calculation shows that, for these parameters, it explains only a small share of the observed gap relative to task difficulty.

[32] In the last three columns of Table 4, we split the data by low, medium, and high cognitive uncertainty and estimate separate regressions for each group. Within these subsets, cognitive uncertainty has limited explanatory power. Nevertheless, our main parameters of interest remain statistically significant and economically large across all specifications.

additional weight from $\hat{\beta} > 1$ is mainly offset by the lower weight from $\hat{k} < 1$, making participants' choices remarkably close to the optimal posteriors, as discussed in Section 4.1.

The stark differences across settings imply that parameters estimated in one environment poorly predict behavior in another. This suggests that updating depends on the full learning environment, not just the precision of a single signal. In our framework, the number of information sources and the precisions of *all* signals jointly determine how effectively individuals incorporate information.

Task Difficulty and the Cardinality of the State Space. Relatedly, Ba et al. (2024) show that how people incorporate signals depends on the complexity of the state space, finding strong evidence that the number of signal realizations affects updating. Our results differ by showing that even when the number of information sources and the cardinality of the state space are fixed, performance in belief updating varies substantially. Thus, while the cardinality matters, as shown by Ba et al. (2024), substantial behavioral variation can arise even when cardinality is held constant.

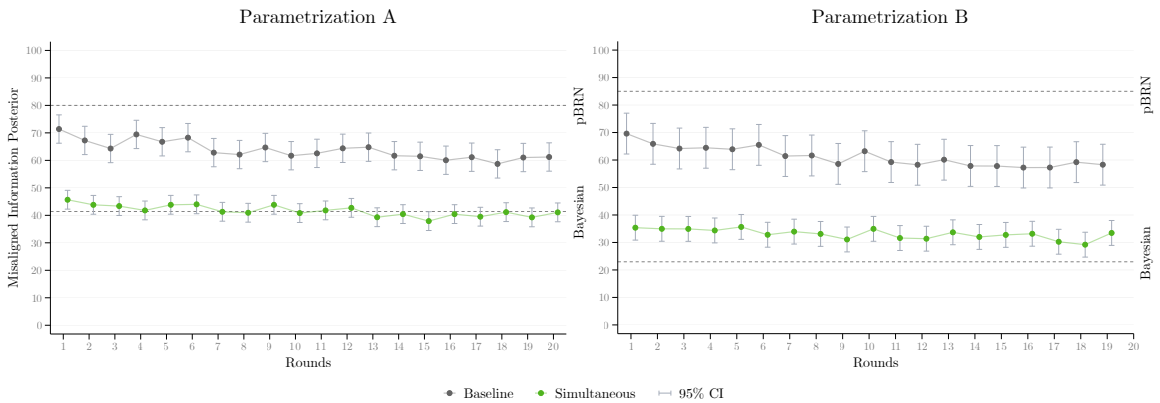
Further Robustness Checks. In Appendix Section C, we compare our results to three additional concepts: *Accuracy Ratios*, *Preferences for Simplicity*, and *Bayesian Posterior levels*. We show that these concepts either explain our findings only partially or are inconsistent with the data.

To summarize, we view these papers as complementary to our work, as they also seek to identify features that shape belief updating. We show that our notion of task difficulty is distinct from cognitive uncertainty and cognitive noise, is not driven by proximity to corner beliefs, and affects updating even when state-space cardinality is fixed.

4.4 Baseline vs. Simultaneous Treatments

Figure 6 plots round-by-round average posteriors under misaligned information. The dashed lines represent the Bayesian benchmark (41.38 in A, 22.97 in B) and the posterior under perfect base-rate neglect (80 in A, 85 in B), where the prior is ignored and only the signal is used.

Figure 6: Posteriors in Baseline and Simultaneous Treatments



Notes: We report the round-by-round average posteriors alongside the 95% confidence intervals, clustered at the individual level. The lower horizontal dashed line depicts the Bayesian posterior, while the top dashed line depicts the perfect base-rate neglect posterior.

We begin with low-difficulty parametrization A, which is also the benchmark used in the base-rate neglect literature. In the Simultaneous treatment, both signals arrive at once, eliminating sequencing effects, and all information is conveyed through signals, minimizing the role of information structure. This makes Simultaneous under parametrization A a natural comparison to Baseline. In the Baseline, we observe little learning across rounds, with an average posterior of 63.79, consistent with prior work.[33] This lack of learning holds across treatments. In contrast, in the Simultaneous treatment, the average posterior is 41.65 and statistically indistinguishable from the Bayesian benchmark ($p = 0.785$; Table 5) and it occurs right from the outset.

This result is striking given that the underlying inference problem is identical across treatments (Section 2.2). It also contrasts with extensive evidence that base-rate neglect is persistent and robust to higher incentives (Enke et al., 2023), feedback (Esponda et al., 2023), and more intuitive framings (Gigerenzer and Hoffrage, 1995). Here, appropriate information structure and timing alone fully restore Bayesian updating—without averaging across rounds or other aggregation.

Result 4 (Belief Updating Immediate Correction)**.** *In low-difficulty tasks, delivering information through two simultaneous signals yields posteriors statistically indistinguishable from the Bayesian benchmark.*

[33]See Esponda et al. (2023), which uses the same baseline and parameters; initial beliefs are around 64 and decline only slowly with learning.

The right panel of Figure 6 shows that the Simultaneous treatment no longer attains the Bayesian benchmark in the high-difficulty parametrization B. The average posterior is 33.39, compared to the Bayesian level of 22.97. Since information structure and sequencing are held constant, the difference from Simultaneous under A is driven solely by task difficulty. Consistent with earlier evidence, higher difficulty creates a wedge between observed and optimal behavior.

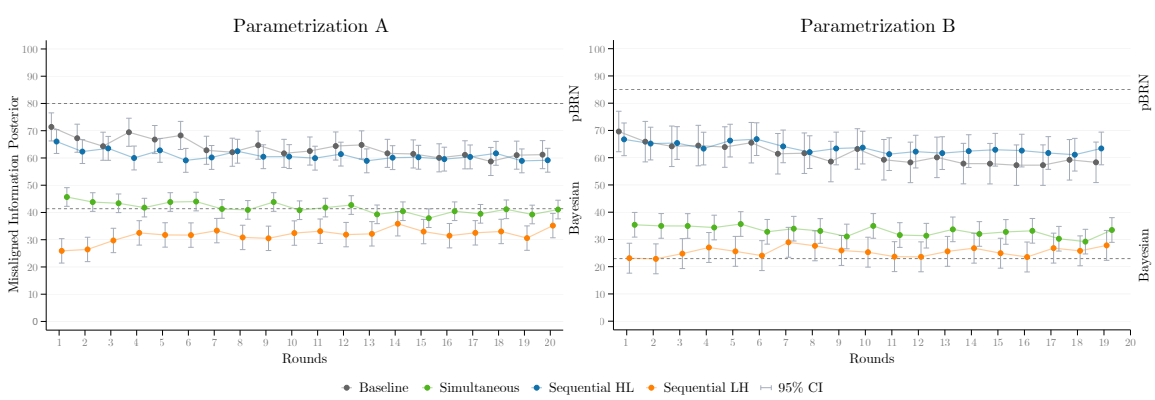
Result 5 (Belief Correction in Difficult Tasks)**.** *In high-difficulty tasks, simultaneous signals reduce but do not eliminate the gap between observed and Bayesian beliefs.*

4.5 Information Structure and Sequencing

Having established the differences between the Baseline and Simultaneous treatments, we now examine the factors driving this gap. Two features distinguish the treatments: (i) information structure—Baseline uses an informative prior and one signal, while Simultaneous uses an uninformative prior and two signals conveying equivalent information; and (ii) timing—information arrives sequentially in Baseline but all at once in Simultaneous. Either or both may account for differences in behavior across treatments.

We use the Sequential treatments to isolate these channels. Figure 7 shows round-by-round average posteriors for all four treatments under parametrization A, and Table 5 reports average posteriors pooled across rounds.

Figure 7: Posteriors in All A and B Treatments



Notes: We report the round-by-round average posteriors alongside the 95% confidence intervals, clustered at the individual level. The lower horizontal dashed line depicts the Bayesian posterior, while the top dashed line depicts the perfect base-rate neglect posterior.

The Effect of Information Structure per se. The informative prior in the Baseline conveys the same information as the first signal in Sequential HL, with identical second-signal accuracy and timing. Thus, any difference between these two treatments comes exclusively from the way information is communicated—information structure.

We assess this in two steps. First, we compare beliefs after the first signal in Sequential HL with the induced prior in Baseline. Theory predicts equivalence, and the data confirm it: participants update correctly from an uninformative prior and a single signal (see Figure 2 and Section 4.1). Second, we compare final beliefs in Sequential HL and Baseline. We find no aggregate differences (p-values: 0.28 under A, 0.62 under B; see Figure 7 and Table 5). This masks some individual-level variation, explored in the Online Appendix.[34]

Result 6 (Information Structure Effect)**.** *We do not find a statistically significant effect of information structure on the average reported beliefs.*

The Effect of Information Sequencing. The Sequential HL and Sequential LH treatments differ only in the order of high- and low-accuracy signals, allowing us to isolate ordering effects from differential weighting of signal accuracy. Figure 7 shows that sequencing has a large impact on belief updating. In both treatments, participants overweight the most recent signal. In Sequential HL, they overweight the second (positive) signal, yielding posteriors around 60.89 under A and 63.70 under B. In Sequential LH, they overweight the second (negative) signal, producing lower-than-Bayesian posteriors of 31.70 under A and 25.04 under B.[35]

Result 7 (Recency Bias)**.** *We document a sizable recency bias independent of signal accuracy and task difficulty.*

Our results are consistent with prior evidence of recency bias (Pitz and Reinhold, 1968; Edenborough, 1975; Grether, 1992), but our framework isolates and quantifies recency bias separately from other factors and tests its robustness across parametrizations. Notably, the bias is sizable despite the minimal delay between

[34]In particular, we show that information structure mostly affects participants who rely on all information they receive: their beliefs in Sequential HL are less extreme than in Baseline.

[35]We normalize the high-accuracy signal to be negative and the low-accuracy signal to be positive.

signals—participants report beliefs after the first signal and immediately update after the second, with the interface reminding them of the first realization.

Robustness Across Parameters. Comparing the left and right panels of Figure 7 shows that the ranking of estimated means across treatments is identical under parametrizations A and B. The Baseline and Sequential HL treatments yield statistically indistinguishable average posteriors, followed by a significantly lower posterior in the Simultaneous treatment and an even lower one in Sequential LH. Thus, we do not detect a statistically significant aggregate effect of information structure—whether information is delivered via an informative prior and one signal or via an uninformative prior and two signals—on reported beliefs. In contrast, sequencing matters: we observe a large recency bias.

Result 8 (Robustness)**.** *The ranking between all four treatments is preserved across different parametrizations, as are results regarding information structure and sequencing.*

4.6 Countering Biases

In Section 4.4, we showed that simultaneous information release yields beliefs indistinguishable from Bayesian under the low-difficulty parametrization A, but only partially closes the gap under the high-difficulty parametrization B. Two forces drive these patterns: in difficult environments, participants underweight the high-accuracy signal under simultaneous information, while a strong recency bias leads them to overweight the most recent signal. This suggests that presenting the high-accuracy signal last may offset the difficulty bias. Figure 7 supports this: in high-difficulty settings, Sequential LH yields posteriors closer to the Bayesian benchmark than Simultaneous, and Table 5 confirms they are not statistically different from Bayesian ($p = 0.325$). The optimal information structure thus depends on the environment: simultaneous release performs well when tasks are easy, while sequencing helps when tasks are difficult by leveraging recency bias.

Result 9 (Countering Biases)**.** *Sequential information delivery can mitigate difficulty bias by exploiting recency bias.*

4.7 Drivers of Base-Rate Neglect

In Table 5, we present average reported beliefs for all treatments, both for the entire dataset and for the last five rounds separately. In both parametrizations, the Baseline treatment exhibits comparable relative levels of base-rate neglect.

Table 5: Estimated Means

	Parameters A ($\tilde{A}$)		Parameters B ($\tilde{B}$)	
	All Rounds	Last 5 Rounds	All Rounds	Last 5 Rounds
Baseline	63.79	60.43	61.93	57.97
	(1.967)	(2.423)	(2.854)	(3.447)
Simultaneous	41.65	40.29	33.39	31.77
	(0.985)	(1.293)	(1.435)	(1.678)
Sequential HL	60.89	59.95	63.70	62.35
	(1.785)	(1.966)	(2.205)	(2.538)
Sequential LH	31.70	32.56	25.04	25.79
	(1.633)	(1.849)	(2.093)	(2.464)
N	8,060	2,015	8,000	2,000

Individual-level clustered errors in parentheses

$$\frac{\mu^A_{Bench} - \mu^A_{Bayes}}{\mu^A_{pBRN} - \mu^A_{Bayes}} = \frac{63.79 - 41.38}{80 - 41.38} \approx 0.58, \qquad \frac{\mu^B_{Bench} - \mu^B_{Bayes}}{\mu^B_{pBRN} - \mu^B_{Bayes}} = \frac{61.93 - 22.97}{85 - 22.97} \approx 0.63.$$

In the calculations above, a score of 0 implies that, on average, participants choose the Bayesian posterior, while a score of 1 implies that, on average, participants choose the perfect base-rate neglect (pBRN) posterior. Recall that pBRN agents disregard the initial information and solely follow the signal.

Next, we turn to a decomposition of the observed level of base-rate neglect, attributing it to information structure, sequencing, and task difficulty. As discussed before, information structure has no effect on beliefs at the aggregate level, while the task difficulty and the sequencing do. The extent to which base-rate neglect is influenced by sequencing can be computed as follows

$$Sequencing^A = \frac{\mu^A_{SeqHL} - \mu^A_{Sim}}{\mu^A_{Bench} - \mu^A_{Bayes}} = \frac{60.89 - 41.65}{63.79 - 41.38} \approx 0.86.$$

$$Sequencing^B = \frac{\mu^B_{SeqHL} - \mu^B_{Sim}}{\mu^B_{Bench} - \mu^B_{Bayes}} = \frac{63.70 - 33.39}{61.93 - 22.97} \approx 0.78.$$

In addition, task difficulty plays a role in parametrization B and its contribution

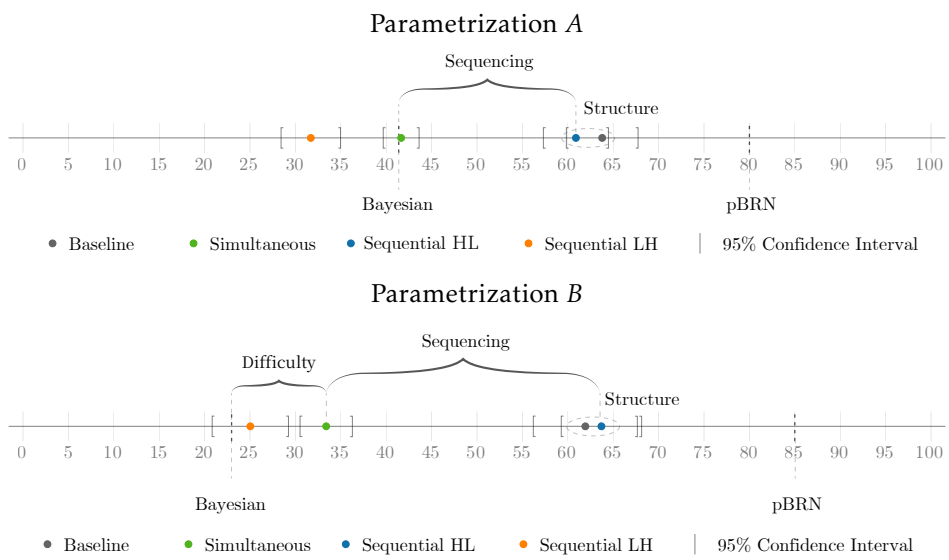
to the overall misspecified beliefs is

$$Difficulty^B = \frac{\mu^B_{Sim} - \mu^B_{Bayes}}{\mu^B_{Bench} - \mu^B_{Bayes}} = \frac{33.39 - 22.97}{61.93 - 22.97} \approx 0.27.$$

The remaining portion of base-rate neglect, although not statistically significant, is due to information structure. We visually summarize these effects in Figure 8, which illustrates the y-axis of Figure 7 after aggregating data across rounds.

Result 10 (Drivers of Base-Rate Neglect)**.** *Information sequencing is the main catalyst of base-rate neglect, with task difficulty also playing a significant role.*

Figure 8: Average posteriors in all treatments



5 Conclusions

Through a series of experiments, we examine how task difficulty, information structure, and information sequencing influence belief updating. While previous work has documented many deviations from Bayesian updating, the factors generating such deviations are often intertwined. By varying these three factors while holding the underlying Bayesian problem constant, we disentangle the independent role of each element and quantify their relative importance in driving mistakes. Adjusting these factors shifts observed behavior from closely matching the Bayesian benchmark to exhibiting sizable deviations, with each element having a distinct impact at the aggregate or individual level.

Through a range of treatments and supplementary data, we provide a comprehensive test of a notion of task difficulty rooted in the nonlinearities of the underlying problem. The evidence suggests that nonlinearities play an important role in belief updating. We believe people's limited ability to account for nonlinearities likely extends beyond belief updating, and we see this as a promising direction for future research.

A Concrete Specifications of Task Difficulty

A.1 Parsimonious Approach

When comparing two signals with different accuracies, the region of interest becomes the region from the juncture where both signals have the same accuracy to the point characterized by the accuracies of the signals. A Bayesian agent would be able to fully follow these changes and figure out how much more they need to weigh the high-precision signal compared to the lower-precision one. However, an agent struggling to follow such nonlinearities might make errors proportional to these cumulative nonlinearities. To link the difficulty of a task with these cumulative nonlinearities, we integrate the absolute value of the second derivative of the Bayesian posterior, starting from the juncture where two signals have equal accuracy up to the point of interest

$$C(\theta_2, \theta_1) = \int_{\theta_2}^{\theta_1} \left| \frac{\partial^2 P(\tilde{\theta}_1, \theta_2 \mid s_1, s_2)}{\partial \tilde{\theta}_1^2} \right| d\tilde{\theta}_1. \tag{3}$$

Above, $P(\tilde{\theta}_1, \theta_2 | s_1, s_2)$ represents the Bayesian posterior given signal accuracies $\tilde{\theta}_1$, θ_2, and signal realizations s_1 and s_2.[36] Low-difficulty (low-complexity) environments will be those in which the Bayesian posterior is rather linear, and thus, neglecting nonlinearities will not matter much, resulting in low $C(\theta_2, \theta_1)$ values. In such environments, a somewhat linear approximated understanding of the environment proves effective. High-difficulty environments will be those in which the Bayesian posterior is rather nonlinear. In these environments, aggregate nonlinearities are large, leading to large $C(\theta_2, \theta_1)$ values. In these environments, relying on a linear approximated understanding of the environment leads to sizable discrepancies.

For ease of exposition, we express the accuracy of a more accurate signal θ_1 in terms of the less accurate signal θ_2, i.e., $\theta_1 = \theta_2 + \eta$, where η is a positive constant.[37] The above definition of task difficulty leads to two main testable implications.

1. Task Difficulty increases with signal accuracy: $\frac{\partial C(\theta_2, \theta_2 + \eta)}{\partial \theta_2} > 0$.

[36] This is one of many ways to capture the nonlinearities of the environment. Another measure that gives almost identical predictions in this setup is the Gini coefficient (Lorenz curve), which looks at the deviation of a graph of interest (the distribution of wealth) from the 45-degree line (a fully linear function).

[37] Naturally, the value of η is restricted to be $\eta \in (0, 1 - \theta_2)$.

2. Task Difficulty increases with the signal gap: $\frac{\partial C(\theta_2, \theta_2+\eta)}{\partial \eta} > 0$.

As the level and/or the gap between signal accuracies increases, the updating task takes place on a more nonlinear region, which, according to our predictions, may lead participants to make larger mistakes.[38]

A.2 Modified Grether Model

Grether (1980) proposed a generalization of Bayesian updating that accommodated a variety of deviations commonly observed in experiments. This generalization is extensively used in the literature on beliefs (Benjamin, 2019). In this section, we explore a modification of the Grether (1980) model that yields comparable qualitative predictions to our task difficulty concept discussed in Section 2.3.

Grether (1980) writes the posterior-odds ratio in the following form

$$\frac{\pi(S|s_2, s_1)}{\pi(F|s_2, s_1)} = \left(\frac{P(s_2|S)}{P(s_2|F)} \frac{P(s_1|S)}{P(s_1|F)}\right)^{\alpha} \left(\frac{P(S)}{P(F)}\right)^{\beta}$$

where $\pi(\cdot)$ captures an agent's possibly biased beliefs. If $\alpha = \beta = 1$, the model reduces to Bayesian updating, whereas $\alpha < 1$ ($\beta < 1$) implies underinference, extracting less information from the signal (prior) than prescribed by Bayes' rule. For comparability purposes, we focus on the uninformative prior case with $\frac{P(S)}{P(F)} = 1$.

We make two modifications to this formulation, given an uninformative prior: (i) agents properly update beliefs when only one signal is informative, and (ii) in the presence of more than one signal, agents under-follow signals more, the more accurate signals are. The first modification relates to the idea that task difficulty, or high nonlinearities, only appear when there is more than one source of information. The second modification captures the idea that under Bayesian updating, agents are expected to react more and more to small changes in signals' accuracies as these accuracies grow larger. It is in such cases that, we believe, the aforementioned sluggishness is emphasized, and individuals fail to properly Bayesian update.

[38] When both signals have the same realizations, there are parameter values that make the effect of the level non-monotonic. However, our study focuses on cases with misaligned information, as declared in the preregistration and described further in Section 3. For misaligned signals, the above predictions hold true for all parameter values.

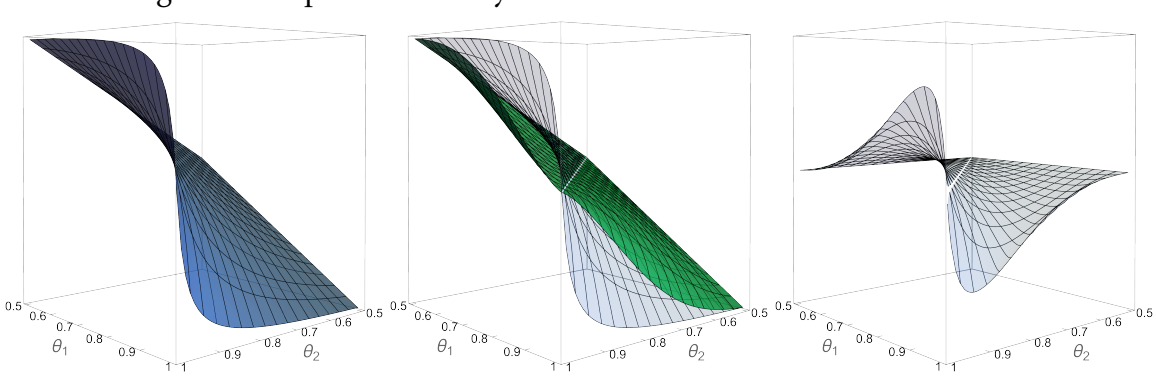
One way to accommodate these two modifications in the Grether (1980) model is to replace parameter α with a function $\gamma(\theta_1, \theta_2)$ that depends on the signals' accuracies:

$$\gamma(\theta_1,\theta_2) = \alpha \cdot \underbrace{(2min(\theta_1,\theta_2)-1)}_{\in[0,1],\text{ weight on }\alpha} + 1 \cdot \underbrace{(2-2min(\theta_1,\theta_2))}_{\in[0,1],\text{ weight on }1} \qquad 0 \leq \alpha \leq 1 \qquad (4)$$

Note that the above functional form reflects the two proposed modifications. Indeed, if $\min(\theta_1, \theta_2) = 1/2$ then $\gamma(\theta_1, \theta_2) = 1$ in line with the first modification and as $\min(\theta_1, \theta_2)$ increases $\gamma(\theta_1, \theta_2)$ decreases, in line with the second modification.

To illustrate, Figure 9 focuses on the case of two misaligned signals and compares the Bayesian posterior with the one induced by the modified Grether model across different θ_1 and θ_2 values.[39] By design, the functions are in agreement when one signal is uninformative (θ_1 or θ_2 is 0.5). The functions are also in agreement if both signals have equal accuracy ($\theta_1 = \theta_2$) because, in such cases, both models assign equal weight to each signal.[40]

Figure 9: Gap Between Bayesian Posterior and Modified Grether



Notes: The figures above show the Bayesian posterior (left), a transparent Bayesian posterior, and the posterior of the modified Grether model (middle), as well as their difference (right). The graphs are plotted for values θ_1 and $\theta_2 \in [0.5, 0.99]$. The value of α is set to 0.

[39]The posterior belief implied by the modified Grether model for two misaligned signals is

$$\pi(S|s_2 = p, s_1 = n) = \frac{\left(\frac{\theta_2}{1-\theta_2}\frac{1-\theta_1}{\theta_1}\right)^{\gamma(\theta_1,\theta_2)}}{\left(\frac{\theta_2}{1-\theta_2}\frac{1-\theta_1}{\theta_1}\right)^{\gamma(\theta_1,\theta_2)}+1}$$

[40]The two functions are also in agreement if one of the signals is fully informative (θ_1 or θ_2 is 1). To make displaying the graphs clearer, this region is clipped off. Figure 9 only displays posteriors for θ_1 and $\theta_2 \in [0.5, 0.99]$.

In all other cases, the two functions differ. If signal accuracies are low or if the gap between the signal accuracies is small, the modified Grether model results in posteriors close to Bayesian ones. However, when both signals are highly accurate—especially when their accuracies also differ substantially—the modified Grether model produces noticeably more inert updating than the Bayesian one. This is the region in which the sluggishness of our model plays a large role, leading to large differences. These predictions mimic those described in Section 2.3.

A.3 Linear Thinking Model

In this section, we provide an alternative model that captures the difficulty of incorporating nonlinearities in belief-updating tasks. We assume that agents struggle to fully appreciate rapid changes required to form correct posterior beliefs in response to small changes in fundamentals and, instead, are only able to adapt to these changes partially.

Focusing on the case of misaligned signals, let $\tilde{P}(S|s_2 = p, s_1 = n)$ be the posterior of an agent who struggles to incorporate nonlinearities. We define the derivative of this posterior to be a weighted average—with weight $\alpha \in [0,1]$ on the derivative of the Bayesian posterior and weight $1-\alpha$ on a constant. A constant is chosen to respect the idea that agents update beliefs correctly when there is only one signal, i.e.,

$$\frac{\partial \tilde{P}(S|s_2 = p, s_1 = n)}{\partial \theta_1} = \alpha \frac{\partial P(S|s_2 = p, s_1 = n)}{\partial \theta_1} + (1-\alpha) \overbrace{\frac{\partial P(S|s_2 = p, s_1 = n)}{\partial \theta_1}\Big|_{\theta_2 = \frac{1}{2}}}^{\text{a constant}}$$

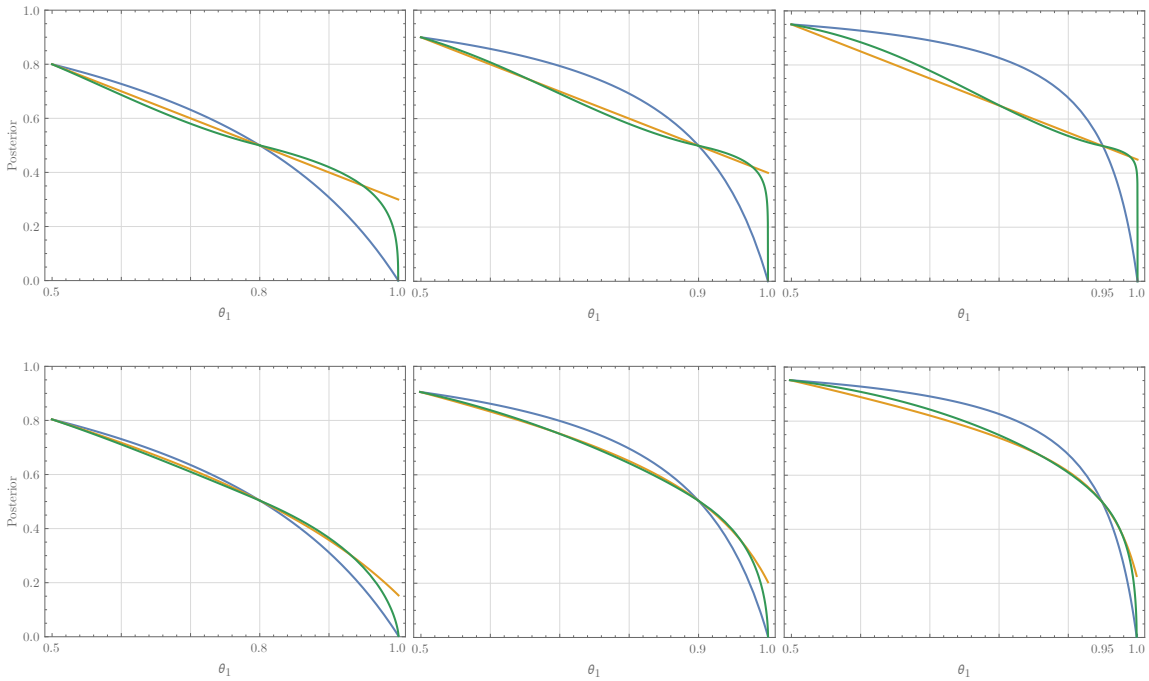
The derivative with respect to the other signal accuracy is similar. Solving the system of partial differential equations and imposing the boundary condition that when both signals are uninformative, $\theta_1 = \theta_2 = \frac{1}{2}$, the posterior coincides with the prior, i.e., $\tilde{P}(S \mid s_2 = p, s_1 = n)\big|_{\theta_1 = \theta_2 = 1/2} = 1/2$, yields

$$\tilde{P}(S|s_2 = p, s_1 = n) = \alpha \underbrace{\frac{\theta_2(1-\theta_1)}{\theta_2 + \theta_1 - 2\theta_2\theta_1}}_{\text{Bayesian Posterior}} + (1-\alpha) \underbrace{\left(\frac{1}{2} - \theta_1 + \theta_2\right)}_{\text{Fully Linear}}$$

Note that since the first derivative is a convex combination of the Bayesian first

derivative and a constant, the second derivative will always be lower in magnitude than the Bayesian second derivative.

Figure 10: Model Prediction Comparison



Notes: In both rows, from left to right, we use $\theta_2 = \{0.80, 0.90, 0.95\}$. The top row uses $\alpha = 0$ and the bottom row uses $\alpha = 1/2$. The blue, green, and orange graphs represent the Bayesian, modified Grether, and Linear Thinking posteriors, respectively.

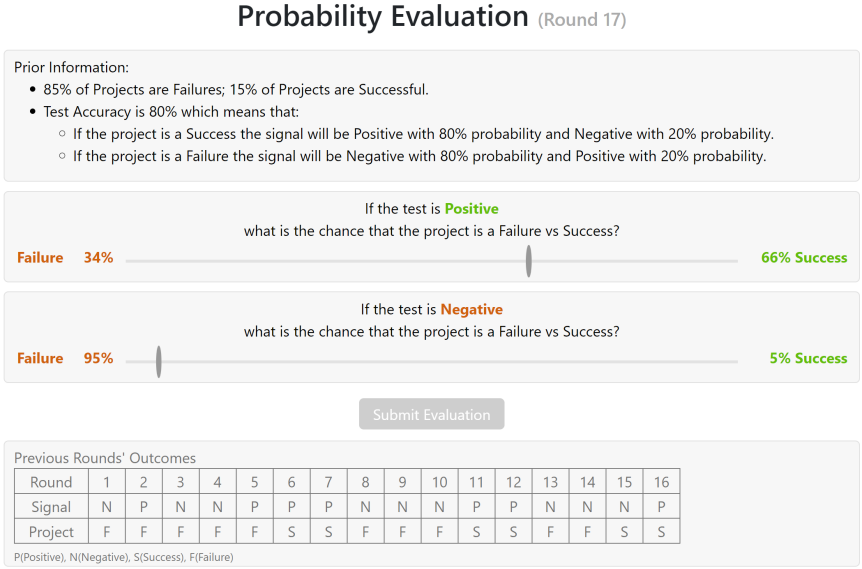
As a final step, we compare the Bayesian posterior $P(S|s_2 = p, s_1 = n)$ with the posterior from the modified Grether model $\pi(S|s_2 = p, s_1 = n)$ discussed in Section A.2, and the linear thinking posterior derived in this section $\tilde{P}(S|s_2 = p, s_1 = n)$. We show these posteriors for various levels of θ_1 with $\alpha = 0$ in the first row and $\alpha = 1/2$ in the second row of Figure 10. The modified Grether and the linear thinking model produce similar predictions, both in the direction of departure from Bayesian updating as well as in magnitude. The main difference between the two models emerges when one of the signals becomes fully informative ($\theta_1 = 1$). Near this region, the modified Grether model quickly converges to 0, while the linear thinking model does not. Again, since the fully informative case is not the focus of this study, for the relevant regions (away from $\theta_1 = 1$), the two models produce qualitatively similar predictions. Importantly, under both models, when the level

of signal accuracies is low, the departure from Bayesian updating is small, even for sizable differences in signal accuracies. In contrast, when the level of signal accuracies is high, even small differences in signal accuracies lead to large gaps from the predicted Bayesian behavior.

B Interface

Figure 11 shows the interface for the Baseline treatment. A detailed description can be found in the Online Appendix.

Figure 11: Baseline Interface



C Additional Task Difficulty Robustness Checks

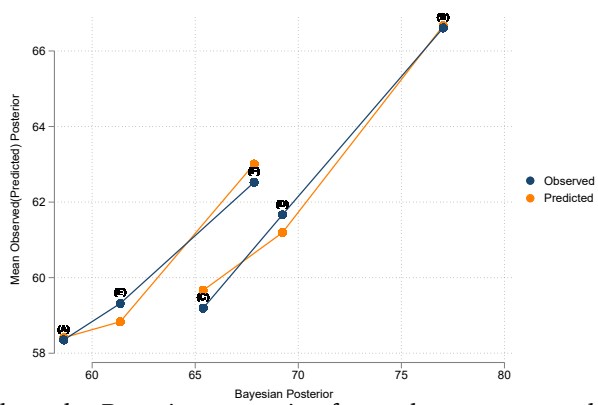
Task Difficulty and Accuracy Ratios. Previous research in psychology suggests that individuals may struggle to properly process information when its informativeness depends on ratios (Miller et al., 1989). To investigate whether this issue can explain our data, we shift focus from the absolute precision of high and low signals to their ratios and assess whether these ratios predict participant errors. By dividing the precision of the high signal by the precision of the low signal for treatments A, E, F, C, D, and B, respectively, we obtain {1.063, 1.059, 1.056, 1.133, 1.125, 1.118}. Alternatively, dividing the low precision by the high precision yields

{0.941, 0.944, 0.947, 0.882, 0.889, 0.895}. Comparing the first (last) three ratios, in either case, we see that the ratios barely change, whereas Figure 3 reveals that the mistakes significantly increase. Furthermore, in either calculation, the ratios are non-monotonic from F to C, whereas the mistakes are monotonically increasing. In summary, changes in signal precision ratios, regardless of the calculation method, fail to align with the documented errors.

Task Difficulty and Preferences for Simplicity. There is a growing body of research on preferences for simplicity. For example, Puri (2025) presents a model where agents incur a cost based on the number of states a lottery covers. As shown in Figure 3, we demonstrate that the rate of mistakes can vary significantly even when holding the support of the lottery fixed—in all six treatments, participants face two possible outcomes and receive two signals. One might suggest that, due to a preference for simplicity, participants may interpret signals as fully revealing when the signals approach near-certainty accuracy. However, this explanation does not align with our data. First, parametrizations C and D, with less precise signals than F, lead to higher error rates. Second, and more critically, participants tend to underreact to the high-precision signals rather than overreact.

Task Difficulty and Bayesian Posterior.

Figure 12: Bayesian and Observed (Predicted) Posteriors



Notes: The x-axis displays the Bayesian posterior for each treatment, while the y-axis shows the observed mean posterior in blue and the predicted posterior in orange. The predictions are based on the model estimates in Table 2. The posterior is flipped to make it in line with the observations from Benjamin (2019), that is, compared to the rest of the paper $\tilde{posterior} = 100 - posterior$.

In a comprehensive review of the literature Benjamin (2019) highlights that underinference tends to be more pronounced when the Bayesian posterior is greater.

We now examine whether this stylized fact alone can explain the observed patterns in our data. Figure 12 plots the Bayesian posterior on the x-axis, with the observed posteriors in blue and the predicted posteriors in orange on the y-axis. The predictions are based on the model estimates in Table 2. As shown, the observed posteriors in our data are not monotonic with respect to the Bayesian posteriors. These non-monotonicities are apparent in several cases—compare C with E, F with C, and D with F.[41] Notably, the figure underscores that varying the level—moving from A to E to F, or from C to D to B—and changing the gap between signal precisions are not equivalent. In other words, while both affect the Bayesian posterior, the posterior alone is not a sufficient statistic to capture these distinct effects. While this was not the focus of our study, it is possible to select different signal precision levels and gaps that result in the same Bayesian posterior but, according to our predictions, would lead to differing error rates.

References

Agranov, M. and Reshidi, P. (2026). Supplement to "Disentangling Suboptimal Updating: Task Difficulty, Structure, and Sequencing". Quantitative Economics Supplemental Material.

Augenblick, N., Lazarus, E., and Thaler, M. (2025). Overinference from weak signals and underinference from strong signals. *The Quarterly Journal of Economics*, 140(1):335–401.

Ba, C., Bohren, J. A., and Imas, A. (2024). Over-and underreaction to information. *Available at SSRN 4274617*.

Bar-Hillel, M. (1980). The base-rate fallacy in probability judgments. *Acta Psychologica*, 44:211–233.

Barbey, A. and Sloman, S. (2007). Base-rate respect: From ecological rationality to dual processes. *Behavioral and Brain Sciences*, 30.

Becker, G., DeGroot, M., and Marschak, J. (1964). Measuring utility by a singleresponse sequential method. *Behavioral Science*, 9:226–232.

Benjamin, D., Bodoh-Creed, A., and Rabin, M. (2019). Base-rate neglect: Foundations and implications. *working paper*.

[41] Non-monotonicities are also evident in errors. See, for example, Figure 3 where the gap is larger under parametrization C compared to parametrization F, despite the Bayesian posterior being higher in the latter.

Benjamin, D. J. (2019). Errors in probabilistic reasoning and judgment biases. *Handbook of Behavioral Economics: Applications and Foundations 1*, 2:69–186.

Bernheim, B. D. and Sprenger, C. (2020). On the empirical validity of cumulative prospect theory: experimental evidence of rank-independent probability weighting. *Econometrica*, 88(4):1363–1409.

Bordalo, P., Conlon, J., Gennaioli, N., Kwon, S., and Shleifer, A. (2026). How people use statistics. *Review of Economic Studies*, 93(1):250–285.

Brandts, J. and Charness, G. (2011). The strategy versus the direct-response method: a first survey of experimental comparisons. *Experimental Economics*, 14:375–398.

Camerer, C. (1987). Do biases in probability judgment matter in markets? experimental evidence. *The American Economic Review*, 77(5):981–997.

Charness, G. and Levine, D. (2005). When optimal choices feel wrong: A laboratory study of bayesian updating, complexity, and affect. *American Economic Review*, 95(4).

Chen, D. L., Schonger, M., and Wickens, C. (2016). otree—an open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance*, 9:88–97.

Cheng, P. and Holyoak, K. (1985). Pragmatic reasoning schemas. *Cognitive Psychology*, 17(4):391–416.

Danz, D., Vesterlund, L., and Wilson, A. J. (2022). Belief elicitation and behavioral incentive compatibility. *American Economic Review*, 112(9):2851–2883.

Eddy, D. M. (1982). Probabilistic reasoning in clinical medicine: problems and opportunities. *in Judgement Under Uncertainty: Heuristics and Biases by Kahneman, Slovic, and Tversky*, pages 249–267.

Edenborough, R. (1975). Order effects and display persistence in probabilistic opinion revision. *Bulletin of the Psychonomic Society*, 5 (1):39–40.

Enke, B., Gneezy, U., Hall, B., Martin, D., Nelidov, V., Offerman, T., and Van De Ven, J. (2023). Cognitive biases: Mistakes or missing stakes? *Review of Economics and Statistics*, 105(4):818–832.

Enke, B. and Graeber, T. (2023a). Cognitive uncertainty. *Quarterly Journal of Economics*.

Enke, B. and Graeber, T. (2023b). Replication Data for: "Cognitive Uncertainty". Harvard Dataverse, V2. https://doi.org/10.7910/DVN/IQ47ZB (CC0 1.0).

Enke, B., Graeber, T., and Oprea, R. (2025). Complexity and time. *Journal of the European Economic Association*, 23(5):1838–1867.

Enke, B. and Shubatt, C. (2023). Quantifying lottery choice complexity. Technical report, National Bureau of Economic Research.

Esponda, I., Vespa, E., and Yuksel, S. (2023). Mental models and learning: The case of base-rate neglect. *American Economic Review*.

Fan, T., Liang, Y., and Peng, C. (2022). The inference-forecast gap in belief updating. *manuscript*.

Fudenberg, D. and Puri, I. (2022). Evaluating and extending theories of choice under risk. *manuscript*.

Ganguly, A., Kagel, J., and Moser, D. (2000). Do asset market prices reflect traders' judgment biases? *Journal of Risk and Uncertainty*, 20(3):219–245.

Gigerenzer, G. and Hoffrage, U. (1995). How to improve bayesian reasoning without instruction: frequency formats. *Psychological review*, 102 (4).

Gneezy, U., Rockenbach, B., and Serra-Garcia, M. (2013). Measuring lying aversion. *Journal of Economic Behavior and Organization*, 93:293–300.

Gold, J. I. and Shadlen, M. (2007). The neural basis of decision making. *Annual Review of Neuroscience*, 30:535–574.

Goodman, A. and Puri, I. (2025). Overvaluing simple bets: Evidence from the options market. *Journal of Financial Economics*, 172:104140.

Grether, D. (1992). Testing bayes rule and the representativeness heuristic: Some experimental evidence. *Journal of Economic Behavior & Organization*, 17(1).

Grether, D. M. (1978). Recent psychological studies of behavior under uncertainty. *American Economic Review*.

Grether, D. M. (1980). Bayes rule as a descriptive model: The representativeness heuristic. *The Quarterly journal of economics*, 95(3):537–557.

Kahneman, D. and Tversky, A. (1972). On prediction and judgement. *ORI Research Monograph*, 12(4).

Kahneman, D. and Tversky, A. (1973). On the psychology of prediction. *Psychological review*, 80(4).

Knill, D. C. and Pouget, A. (2004). The bayesian brain: the role of uncertainty in neural coding and computation. *Trends in Neurosciences*, 27(12):712–719.

Koehler, J. J. (1996). The base rate fallacy reconsidered: Descriptive, normative, and methodological challenges. *Behavioral and brain sciences*, 19(1):1–17.

Larrick, R. and Soll, J. (2008). The mpg illusion. *Science*, 320:1593–1594.

Levy, M. and Tasoff, J. (2016). Exponential-growth bias and lifecycle consumption get access arrow. *Journal of the European Economic Association*, 14(3):545–583.

Levy, M. and Tasoff, J. (2017). Exponential growth bias and overconfidence. *Journal of Economic Psychology*, 58:1–14.

Madrian, B. C. and Shea, D. F. (2001). The power of suggestion: Inertia in 401(k) participation and savings behavior. *Quarterly Journal of Economics*, 116(4):1149–1187.

Miller, D. T., Turnbull, W., and McFarland, C. (1989). When a coincidence is suspicious: The role of mental simulation. *Journal of Personality and Social Psychology*, 57(4):581.

Oprea, R. (2024a). Complexity and its measurement. *Handbook of Experimental Methods in the Social Sciences*.

Oprea, R. (2024b). Decisions under risk are decisions under complexity. *American Economic Review*, 114(12):3789–3811.

Pitz, G. and Reinhold, H. (1968). Payoff effects in sequential decision-making. *Journal of Experimental Psychology*, 77 (2):249–257.

Pouget, A., Beck, J. M., Ma, W. J., and Latham, P. E. (2013). Probabilistic brains: knowns and unknowns. *Nature Neuroscience*, 16:1170–1178.

Puri, I. (2022). Preference for simplicity. *manuscript*.

Puri, I. (2025). Simplicity and risk. *The Journal of Finance*, 80(2):1029–1080.

Rabin, M. (2002). Inference by believers in the law of small numbers. *Quarterly Journal of Economics*, 117:775–816.

Rabin, M. and Vayanos, D. (2010). The gambler's and hot-hand fallacies: Theory and applications. *Review of Economic Studies*, 77:730–778.

Rees-Jones, A. and Taubinsky, D. (2020). Measuring schmeduling. *Review of Economic Studies*, 87:2399–2438.

Shaffer, B. (2020). Misunderstanding nonlinear prices: Evidence from a natural experiment on residential electricity demand. *American Economic Journal: Economic Policy*, 12(3):433–61.

Shubatt, C. and Yang, J. (2025). Tradeoffs and comparison complexity. *working paper*.

Stango, V. and Zinman, J. (2009). Exponential growth bias and household finance. *Journal of Finance*, 64(6):2807–2849.

Thaler, R. H. and Benartzi, S. (2004). Save more tomorrow: Using behavioral economics to increase employee saving. *Journal of Political Economy*, 112(51):5164–5187.

Thaler, R. H. and Sunstein, C. R. (2008). *Nudge: Improving Decisions About Health, Wealth, and Happiness.* Yale University Press: New Haven & London.

Valone, T. J. (2006). Are animals capable of bayesian updating? an empirical review. *Oikos*, 112:252–259.

Wagenaar, W. and Sagaria, S. (1975). Misperception of exponential growth. *Perception and Psychophysics*, 18(6):416–422.

Woodford, M. (2020). Modeling imprecision in perception, valuation, and choice. *Annual Review of Economics*, 12:579–601.