# Over- and Underreaction to Information*

Cuimin Ba[†] J. Aislinn Bohren[‡] Alex Imas[§]

July 20, 2023

This paper explores how properties of the learning environment determine how people react to information. We develop a two-stage model of belief formation where people first reduce complexity by channeling attention to a subset of states that are representative of the observed information, then evaluate this information using Bayes' rule subject to cognitive imprecision. The model predicts *overreaction* when environments are complex, signals are noisy, or priors are concentrated on intermediate states; it predicts *underreaction* when environments are simple, signals are precise, or priors concentrated on more extreme states. Results from a series of pre-registered experiments provide direct support for these predictions, as well as the proposed attentional mechanism. We show that the two-stage model is highly *complete* in capturing explainable variation in belief-updating; in particular, the interaction between the two psychological mechanisms is critical to explaining belief-formation in more complex settings. These results connect disparate findings in prior work: underreaction is typically found in laboratory studies, which feature simple learning settings, while overreaction is prevalent in financial markets, which feature more complex environments.

**Keywords**: overreaction, underreaction, beliefs, noisy cognition, representativeness, bounded rationality, attention, completeness, restrictiveness, behavioral economics, learning

*We thank Nageeb Ali, Nick Barberis, Francesca Bastianello, Cary Frydman, Thomas Graeber, Ingar Haaland, David Laibson, Ryan Oprea, Yuval Salant, Andrei Shleifer, Marciano Siniscalchi, Florian Zimmermann and seminar participants and conference participants at Drexel, Harvard, NBER Asset Pricing meeting, Northwestern, NYU Abu Dhabi, Stanford, Texas A&M, UC Berkeley, UC Santa Barbara, University of Alicante, and USC for helpful comments. Josh Hascher, Christy Kang, Enar Leferink and Marcus Tomaino provided expert research assistance. Bohren gratefully acknowledges financial support from NSF grant SES-1851629.

†Department of Economics, University of Pittsburgh: bacuimin@gmail.com

‡Department of Economics, University of Pennslyvania: abohren@sas.upenn.edu

§Booth School of Business, University of Chicago: alex.imas@chicagobooth.edu

## 1 Introduction

How do people interpret and react to new information? This question is fundamental to economic decision-making: investors adjust their beliefs about the quality of a stock based on past performance, managers learn from candidate interviews before making hiring decisions, and professional forecasters make economic predictions based on releases of new data. Standard models assume that people have an accurate representation of the signal-generating process and use Bayes' rule to draw inference. However, a large literature in economics, finance, and psychology documents systematic departures from these assumptions.

Laboratory experiments have been used extensively to study how people learn and update beliefs. In these studies, participants are told the information environment, observe a signal, and then report their belief about the realization of an unobserved state. The components of the information environment—the state space, prior, and signal distribution—are transparent and easy to exogenously manipulate. It is also straightforward to calculate the Bayesian posterior as a benchmark. Such experiments generally find that people deviate from this benchmark in a manner that suggests *underreaction* to information (Benjamin 2019).[1] Surveys and forecasts by households or industry professionals have also been used to study how people react to information, primarily in financial markets. Although it is not possible to calculate a Bayesian benchmark using such field data, the predictability of forecast *errors* can be used to identify deviations from rational expectations and determine the direction of such deviations. This literature mostly finds evidence for *overreaction* to information (Bordalo, Gennaioli, and Shleifer 2022).[2]

In this paper, we explore how properties of the information environment impact whether under- or overreaction emerges. We start by modeling belief-updating as a two-stage process where people first simplify the information structure and then form beliefs subject to cognitive imprecision. In the first stage, attention and memory constraints lead the agent to simplify complex information by focusing on a subset of states that are *representative* given the observed signal. In the second stage, cognitive imprecision leads the agent to form beliefs using a noisy representation of the simplified information structure. This model generates comparative static predictions about how the level of over- or underreaction varies with the information environment. Specifically, we show analytically that it predicts more overreaction as the state space becomes more complex, the signal is noisier, and the prior becomes concentrated on less representative states; it predicts more underreaction when the state space is simpler, the signal becomes more precise, and the prior becomes con-

[1] Benjamin (2019) writes: "The experimental evidence on inference taken as a whole suggests that even in small samples, people generally underinfer rather than overinfer."

[2] Bordalo et al. (2022) write: "The expectations of professional forecasters, corporate managers, consumers, and investors appear to be systematically biased in the direction of overreaction to news."

centrated on more representative states. Pre-registered experiments with $N = 2,210$ participants provide direct support for these predictions.

To test the proposed cognitive mechanism directly, we also measure and experimentally manipulate attention in our setting. We find that people overwhelmingly channel their attention towards representative states and, consistent with the framework, taxing cognitive resources exacerbates overreaction. We then use the experimental data to show that while cognitive imprecision alone can generate the observed belief-updating in simple environments, the mechanism quickly loses explanatory power as the environment becomes more complex. On the other hand, our two-stage model is highly *complete* in capturing the observed variation in belief-updating across all information environments (Fudenberg, Kleinberg, Liang, and Mullainathan 2022). Moreover, combining both stages of the model does not make the framework too flexible in its capacity to be falsified—it remains highly restrictive (Fudenberg, Gao, and Liang 2023). Taken together, our model and empirical results help rationalize the discrepancy between the predominant observation of underreaction in laboratory studies—which typically use a simple state space, a relatively precise signal, and a uniform prior—and overreaction in financial market studies—which feature a more complex environment and noisier signals.

To illustrate the two-stage model, consider the following example. An agent is deciding whether to invest in an asset that is either "good" or "bad" (the state) with equal probability (the prior). A good (bad) asset has a 70 (30) percent chance of increasing in price and a 30 (70) percent chance of decreasing in price. The agent observes a price increase (the signal). How should she update her belief? According to Bayes' rule, she should increase her belief that the asset is good from 50 to 70 percent. However, research from the lab suggests that the agent will underreact and increase her belief to less than 70 percent. Now suppose that there are five potential states—good and bad, as before, and three intermediate states with a 40, 50 or 60 percent chance of a price increase, respectively. Does the increased complexity of the information environment impact whether the agent over- or underreacts? What other properties of the information environment impact whether the agent exhibits one bias versus the other?

To answer these questions, we first turn to the literature on how people respond to complexity. Cognitive research shows that attention and working memory have a fixed capacity, limiting the number of objects in consideration at any given time (Oberauer, Farrell, Jarrold, and Lewandowsky 2016; Luck and Vogel 1997; Loewenstein and Wojtowicz 2023).[3] In the case of belief formation, this implies that people making likelihood judgments will be limited in the number of potential states that they can simultaneously consider. In our model, the first stage—which we term

[3]For example, in the case of visual stimuli, participants can attend to only three to four items at any given time (Bays, Gorgoraptis, Wee, Marshall, and Husain 2011)

the *editing* stage—determines which elements of the information structure are attended to and to what extent. We propose that, when confronted with a complex learning environment, an agent's attention will be channeled toward 'representative' states, i.e., states that generate the observed signal realization with the highest probability (Kahneman and Tversky 1972; Bordalo, Gennaioli, Porta, and Shleifer 2019; Bordalo, Coffman, Gennaioli, and Shleifer 2016). Upon observing a signal, this representativeness heuristic distorts the agent's posterior belief by overweighting states whose Bayesian posterior increases relative to the prior and underweighting states whose Bayesian posterior decreases relative to the prior. When the signal has a good news/bad news structure (Milgrom 1981) commonly used in the literature, this results in an overweighting of extreme states.

We then turn to how the 'edited' information structure is processed when forming subjective beliefs. A large body of work in cognitive psychology shows that an agent's response to information can be modeled using a noisy representation of the decision parameters (Green, Swets et al. 1966; Thurstone 1927; Woodford 2020; Gabaix 2019).[4] In the second stage of our model, termed the *evaluation* stage, the agent applies Bayes' rule to the (potentially) simplified information structure, but with cognitive imprecision. Specifically, the agent processes the information structure as if she is facing a signal extraction problem, treating the parameters as unbiased noisy signals of the true underlying values. Greater cognitive noise decreases the agent's sensitivity to changes in the parameters and biases her posterior belief toward a *cognitive default*. In simple binary state environments with a uniform prior, this leads to underreaction.

Returning to the example above, our model predicts that the agent underreacts to the price increase when the asset is simple (i.e., either good or bad), but overreacts when the asset is more complex. In both cases, the good state is representative of a price increase. When the asset is simple, there is little need to further simplify the information environment and the representativeness heuristic has limited bite. In turn, the impact of cognitive noise dominates: the agent does not fully internalize the informativeness of the price increase and underreacts to it as a result. On the other hand, when the asset is complex, channeling attention to the representative good state diverts attention from the four other states, which leads her to overweight the former at the expense of underweighting the latter. This results in overreaction.[5]

Beyond complexity, our model also predicts how other properties of the information environment impact belief-updating. Logic similar to the above yields a

[4] A more recent literature in economics applies the principles of such noisy cognition to explain anomalies in choice under uncertainty (Khaw, Li, and Woodford 2022; Frydman and Jin 2022; Enke and Graeber 2023) and forecasting (Azeredo da Silveira and Woodford 2019; Enke and Graeber 2023; Augenblick, Lazarus, and Thaler 2022).

[5] This prediction is stark: simulations show that even going from two to three states leads an agent to switch from underreaction to overreaction for a broad range of parameter values.

prediction that overreaction will decrease with signal informativeness. Specifically, the agent will overreact most (underreact least) to a noisy signal and overreact least (underreact most) to a precise signal. For example, in the complex asset example, our model predicts more overreaction when the good (bad) state generates a price increase with 60 (40) percent chance, compared to a more precise signal that generates a price increase with an 80 (20) percent chance.[6]

The shape of the prior also influences the magnitude of reaction. Our model predicts that overreaction increases as the prior becomes more concentrated on less representative states. In the case of an asymmetric prior, our model predicts underreaction to confirmatory signal realizations—those that increase the likelihood of states that are 'expected' ex-ante—and overreaction to disconfirmatory, 'surprising' signal realizations. For example, in the simple asset example, suppose there is an 80% chance of the good state and a 20% chance of the bad state. Then a price increase is 'expected' and a price decrease is 'surprising'; our model predicts underreaction to the former and overreaction to the latter. Notably, in the case of a confirmatory signal realization, our model also predicts that cognitive imprecision can lead agents to update in the *opposite* direction of the Bayesian posterior (e.g., placing less than an 80% chance on the good state after a price increase). Such wrong direction updating is not predicted for disconfirmatory realizations or under a symmetric prior.

Each of these predictions stem from how the interaction between representativeness and cognitive imprecision impact the weight an agent places on extreme versus moderate states relative to the Bayesian posterior. In environments where the Bayesian posterior places relatively less weight on the extreme states—either because there are more moderate states under consideration, the prior is more concentrated on these moderate states, a less informative signal leads to less movement from the prior, or the signal realization is surprising—representativeness dominates cognitive noise and overreaction is the predominant phenomenon. In contrast, in environments where the Bayesian posterior places relatively more weight on the extreme states—either because there are fewer moderate states under consideration, the prior places more weight on extreme states, a more informative signal leads to more movement from the prior, or the signal is consistent with the asymmetric prior—cognitive noise dominates and agents either overreact less or even underreact.

We test each of these predictions in a series of pre-registered experiments. We adopt the classic 'bookbag-and-poker-chip' design originally used in Edwards (1968) and employed extensively in the learning literature. In the standard paradigm, a

[6]Both Edwards (1968) and Benjamin (2019) show that underreaction in simple two-state settings decreases as the signal becomes noisier, even flipping to overreaction for very noisy signals. Augenblick et al. (2022) show that this relationship is consistent with a model of cognitive noise. Our model shows that the same pattern can be generated by representativeness, such that the degree of overreaction to noisy signals increases with the complexity of the state space—a prediction that is not captured by a model of cognitive noise alone.

set number of bags are filled with different color balls in known proportions. For example, Bag 1 contains 70 red balls and 30 blue balls while Bag 2 contains 30 red balls and 70 blue balls. One bag is chosen at random with known probability. A ball is drawn from it and shown to the participant. The participant then reports her belief about the likelihood that each bag was selected. Parameters in the design have a straightforward correspondence to our model: the bags represent the states, the probability that each bag is selected corresponds to the prior, and the proportion of balls in each bag represents the signal distribution. Our experiments have three main sources of treatment variation. First, we manipulate complexity by varying the state space from the standard 2-state setting up to 11 states. Second, we vary the signal informativeness. Finally, we vary the concentration and symmetry of the prior.

Increasing complexity has a striking effect on belief-updating. Focusing on uniform prior environments, we first replicate the standard finding that people generally underreact to information in simple 2-state environments. This result *qualitatively* flips when we add even a single additional state: the majority of participants *overreact* in 3-state environments, and do so across all signal distributions we consider. The level of overreaction increases monotonically with the complexity of the state space, such that the largest fraction of participants overreact in the 11-state environment. Note that these results are not consistent with people simply being insensitive to changes in the information environment; 'representativeness' is key because it highlights exactly *which* states receive more weight.[7] We also document the predicted relationship between signal informativeness and belief-updating. The decreasing relationship between level of overreaction and signal diagnosticity is observed across all complexities. For example, in the 3-state case, participants overreact the most to relatively uninformative signals and the least to relatively informative signals.

Consistent with the model's prediction, we document that overreaction increases as the prior becomes more concentrated on a moderate state; the most (least) overreaction is observed when the moderate state has greater (smaller) prior likelihood relative to the extreme states. Turning to an asymmetric prior, we document the hypothesized underreaction to confirmatory 'expected' signal realizations and overreaction to disconfirmatory surprising signal realizations. Even in simple two-state environments where people generally underreact when the prior is symmetric, we find evidence for significant overreaction when the prior is asymmetric—driven by overreaction to disconfirmatory realizations. Moreover, consistent with our model's prediction, we observe nearly three times as many wrong direction updates to confirmatory realizations compared to disconfirmatory realizations.

Our model has two key parameters which capture the level of representativeness

[7]Comparing the reported posteriors to the Bayesian benchmarks allows us to rule out the possibility that people are insensitive to changes in the information environment and that the shift from under- to overreaction is driven by a shifting Bayesian benchmark.

and cognitive imprecision. We use the experimental data to structurally estimate these parameters. At the aggregate level, both parameter estimates differ from the Bayesian benchmark and are in line with values found in prior work. At the individual level, we find that the vast majority of participants exhibit both significant representativeness and cognitive imprecision; small minorities exhibit either a single mechanism or neither one. Interestingly, the two parameter estimates are significantly correlated, suggesting that individual-level limits in cognitive resources lead to both representativeness and cognitive imprecision.

Next, we explore the relationship between bottom-up attention and belief-updating to directly test the proposed attentional mechanism in the editing stage. Specifically, we incorporate a paradigm commonly used in cognitive science to measure and manipulate attention (Payne, Bettman, and Johnson 1988) into our baseline design. Results in this treatment show that, upon observing the signal, participants' attention is overwhelmingly drawn to the most 'representative' state. We use the same paradigm to study how limiting cognitive resources impacts belief-updating. Our proposed mechanism predicts that such limits will exacerbate representativeness in complex environments and lead to greater overreaction. The results support this hypothesis: fixing the information environment, we observe a higher level of overreaction when cognitive resources are more taxed. Indeed, structural estimation shows that representativeness becomes more pronounced while cognitive imprecision remains the same.

Finally, we use the experimental data to evaluate the model's *completeness* in capturing predictable variation in belief-updating across information environments. Using the method of Fudenberg et al. (2022), we show that cognitive imprecision on its own is sufficient to explain belief-updating in simple 2-state environments, but precipitously loses explanatory power when complexity increases to 3 or more states (dropping from capturing 100% to 36% of the explainable variation in the data). Representativeness on its own also has low explanatory power. In contrast, our two-stage model has high explanatory power across simple and complex environments, notably capturing 92% of the explainable variation in the latter. Critically, this shows that the two processes are *cognitive complements*: their interaction plays a critical role in predicting belief-updating. Moreover, the two-stage model is nearly as *restrictive* as both the Bayesian benchmark and the cognitive-noise-only model in that it does not explain randomly-generated 'synthetic' belief data (Fudenberg et al. 2023). This shows that the high explanatory power for real belief data does not come at the expense of being flexible enough to explain *any* data.

Our findings contribute to a large literature on under- and overreaction in belief-updating. Section 5 provides an in-depth review of this work, and discusses how our results can help rationalize some of the disparate findings. Notably, our model predicts underreaction in simple settings such as the two-state experiments reviewed in Benjamin (2019), and overreaction in more complex environments such as the financial markets reviewed in Bordalo et al. (2022), where investors and forecasters need to consider a multitude of potential states (e.g., forming expectations about future returns of a stock or forecasting macroeconomic variables). We also discuss how our findings relate to the evidence on how investor behavior (prices) responds to news in financial markets (Daniel, Hirshleifer, and Subrahmanyam 1998; Barberis, Shleifer, and Vishny 1998; Klibanoff, Lamont, and Wizman 1998).

The paper contributes to the literature exploring the cognitive foundations of economic decision-making. Research on the role of complexity in judgment and decision-making argues that people are averse to complexity (Oprea 2020), and as a result, adopt simpler mental models (Kendall and Oprea 2021; Molavi 2022), form simpler hypotheses (Bordalo, Conlon, Gennaioli, Kwon, and Shleifer 2023), and use heuristics to reduce the mental cost of judgments and decisions (Banovetz and Oprea 2020; Oprea 2022). Another strand of research models an agent as optimally responding to a stimulus given a noisy representation of the decision problem. These models of noisy cognition have been used to explain phenomena such as small-stakes risk aversion (Khaw, Li, and Woodford 2021), state-dependent risk attitudes (Khaw et al. 2022), and myopia in time preferences (Gabaix and Laibson 2017). Awareness of this noise is correlated with the extent of people's insensitivity to the parameters of the decision problem (Enke and Graeber 2023). Our theoretical framework is linked to both lines of work: the proposed two-stage model of belief-updating incorporates a heuristic response to complexity in the editing stage and cognitive imprecision in the evaluation stage.

The rest of the paper proceeds as follows. Section 2 outlines the theoretical framework and its predictions. Section 3 presents the experimental paradigm, empirical findings, and structural estimation. Section 4 quantifies model completeness and restrictiveness. Section 5 reviews the prior literature and discusses our findings. Section 6 concludes.

## 2 Theoretical Framework

In this section, we formalize a two-stage model of belief formation. The first 'editing' stage guides what aspects of the information structure the agent attends to and to what extent. The second 'evaluation' stage determines how this edited input is processed to form a subjective posterior belief after observing the signal. We then define a general measure of over- and underreaction and use our model to derive a series of comparative static predictions on how the extent of under- or overreaction varies with properties of the information environment. We focus on four key properties: the complexity of the state space, the signal diagnosticity, and the concentration and symmetry of the prior.

### 2.1 Model

### 2.1.1 Information Environment

A state $\omega$ is drawn from finite state space $\Omega \equiv \{\omega_1, ..., \omega_N\} \subset [0,1]$ with $N$ distinct states in ascending order, $\omega_1 < ... < \omega_N$, and generic element $\omega_i$. The state is distributed according to prior $p_0 \in \Delta(\Omega)$, which we assume to have full support on $\Omega$. A binary signal $s$ with support $\mathcal{S} \equiv \{s_1, s_2\}$ and generic realization $s_i$ provides information about the state. In state $\omega_i$, the signal is distributed according to $Pr(s_2|\omega_i) = \omega_i$ and $Pr(s_1|\omega_i) = 1 - \omega_i$. For example, when $\Omega = \{0.3, 0.5, 0.7\}$, signal realization $s_2$ occurs with probability 0.3 in state $\omega_1$, probability 0.5 in state $\omega_2$, and probability 0.7 in state $\omega_3$. This signal has a good news/bad news structure (Milgrom 1981): the probability of realization $s_2$ is increasing in the state, so observing $s_2$ is good news about the state and observing $s_1$ is bad news. Since the signal distribution is pinned down by the state space, we refer to $\Omega$ as the *information structure* and $(\Omega, p_0)$ as the *information environment*.

We define several properties of information environments. An information structure $\Omega'$ is more *dispersed* than $\Omega$ if the minimum and maximum states in $\Omega'$ are weakly smaller and larger, respectively; i.e., $\omega'_1 \leq \omega_1$ and $\omega'_N \geq \omega_N$. An information structure $\Omega'$ is more *complex* than $\Omega$ if $\Omega'$ contains weakly more states than $\Omega$; i.e., $|\Omega'| \geq |\Omega|$. An information structure is *symmetric* if whenever $\omega_i \in \Omega$, then its reflection $1 - \omega_i \in \Omega$. A prior $p_0$ is *symmetric* if for any $\omega_i \in \Omega$, $\omega_i$ and $1 - \omega_i$ have the same mass; i.e., $p_0(\omega_i) = p_0(1 - \omega_i)$. Note that prior symmetry implies information structure symmetry but not vice versa. Related to individual states, state $\omega_j$ is more *interior* than state $\omega_i$ if it is closer to 1/2, $|\omega_j - \frac{1}{2}| \leq |\omega_i - \frac{1}{2}|$. The *diagnosticity* of the signal in state $\omega_i$ is the probability of the more likely signal realization, $d_i \equiv \max\{\omega_i, 1 - \omega_i\}$. Within the class of symmetric information structures, the set of diagnosticities is sufficient for the information structure. For example, in $\Omega = \{0.3, 0.5, 0.7\}$, the set of diagnosticities $\{0.5, 0.7\}$ pin down $\Omega$.[8]

Given a prior and an information structure, by Bayes' rule, the objective posterior probability of state $\omega_i$ following signal realization $s_2$ is

$$p(\omega_i|s_2) \equiv \frac{\omega_i p_0(\omega_i)}{\sum_{\omega_j \in \Omega} \omega_j p_0(\omega_j)}, \tag{1}$$

and analogously following $s_1$, $p(\omega_i|s_1) \equiv (1 - \omega_i)p_0(\omega_i) / \sum_{\omega_j \in \Omega}(1 - \omega_j)p_0(\omega_j)$. Let $p(s_i) = (p(\omega_1|s_i), ..., p(\omega_N|s_i))$ denote this objective posterior. The objective posterior expected state following signal realization $s_i$ is

$$E(\omega|s_i) \equiv \sum_{\omega_j \in \Omega} \omega_j p(\omega_j|s_i). \tag{2}$$

[8] Our model can be extended to asymmetric information structures with more involved notation and definitions. Our model's predictions introduced in Section 2.2 remain valid in these asymmetric environments.

This information structure mirrors the experimental environment in Section 3. We focus on a binary signal space with a good news/bad news structure, as this is the set-up typically used in both laboratory and financial studies (e.g., price increase versus decrease). Our belief-updating model is straightforward to extend to more general signal spaces and information structures.

### 2.1.2 Two-Stage Updating Process

We next model how an agent forms her subjective posterior belief, denoted by $\hat{p}(s_i) = (\hat{p}(\omega_1|s_i), ..., \hat{p}(\omega_N|s_i))$ following signal realization $s_i$. First, attention and working memory constraints lead her to *edit* (simplify) the information structure using the representativeness heuristic. Second, cognitive imprecision leads her to use a noisy representation of the information structure to *evaluate* the information (update her belief).

**Editing.** According to Tversky and Kahneman (1983), limits on cognitive resources lead people to overweight 'representative' objects when making judgments. For example, when predicting the hair color of someone from Ireland, people overweight the state of the world in which the person has red hair, as red hair is representative of someone from Ireland. Similarly, people overweight the state in which someone from Florida is a retiree, as retirees are representative of Floridians (Bordalo et al. 2016). In the information environment outlined above, Tversky and Kahneman (1983)'s conjecture would classify a state as representative of a signal realization if the relative frequency of the state following this signal realization is much higher than the frequency of the state in a reference distribution. We argue that agents use the representativeness heuristic in response to complexity as a means of simplifying the information structure, which leads to an overweighting of representative states.[9]

Following Bordalo et al. (2019), we define the representativeness of state $\omega_i$ following signal realization $s_j$ as the posterior frequency of $\omega_i$ relative to the prior frequency, $R(\omega_i, s_j) \equiv \frac{p(\omega_i|s_j)}{p_0(\omega_i)}$. A state is *more representative* if its objective likelihood increases more after observing $s_j$ relative to the prior belief before the signal arrived (i.e., the reference distribution is the prior). In the class of information structures we consider, $\omega_1$ is the most representative state following signal realization $s_1$ and $\omega_N$ is the most representative state following realization $s_2$.

When updating beliefs, the agent first simplifies the information structure by channeling more attention bottom-up to more representative states, which inflates their weights.[10] We use the representativeness-based discounting weighting func-

[9]Payne, Bettman, and Johnson (1993) reviews the evidence on heuristics as a simplification tool in the face of complex environments.

[10]This is related to the argument of Gennaioli and Shleifer (2010); Bordalo, Coffman, Gennaioli, Schwerter, and Shleifer (2021) that the most representative states are easier to recall and are therefore overweighted in judgment. See also Kahneman (2003) for discussion on the interaction between selective attention and recall, and how this relates to heuristics in judgement.

tion defined in Bordalo et al. (2016) to parameterize how representativeness distorts belief-updating. Specifically, the agent's edited posterior belief about state $\omega_i$ after observing signal realization $s_j$ is distorted proportionally to the representativeness $R(\omega_i, s_j)$ of $\omega_i$,

$$p_R(\omega_i|s_j) \equiv p(\omega_i|s_j)\frac{R(\omega_i, s_j)^\theta}{Z(s_j)}, \tag{3}$$

where $p(\omega_i|s_j)$ is the objective posterior. Parameter $\theta \geq 0$ captures the severity of the representativeness distortion: a higher $\theta$ corresponds to more distortion. The term

$$Z(s_j) \equiv \frac{\sum_{\omega_i \in \Omega} Pr(s_j|\omega_i)^{\theta+1} p_0(\omega_i)}{(\sum_{\omega_i \in \Omega} Pr(s_j|\omega_i) p_0(\omega_i))^{\theta+1}} \tag{4}$$

is a normalization factor that ensures the edited posterior sums to one across states. Let $p_R(s_i) \equiv (p_R(\omega_1|s_i), ..., p_R(\omega_N|s_i))$ denote the edited posterior belief following signal realization $s_i$. When $\theta = 0$, the edited posterior corresponds to the objective posterior. When $\theta > 0$, the edited posterior overweights more representative states and underweights less representative states.

The distortion captured in Eq. (3) is equivalent to an agent forming her posterior using Bayes' rule with respect to misspecified information structure

$$\hat{P}r(s_j|\omega_i) \equiv \frac{Pr(s_j|\omega_i)^{\theta+1}}{Pr(s_1|\omega_i)^{\theta+1} + Pr(s_2|\omega_i)^{\theta+1}}. \tag{5}$$

In the class of information environments we consider, Eq. (5) simplifies to $\hat{P}r(s_2|\omega_i) = \omega_i^{\theta+1}/((1-\omega_i)^{\theta+1} + \omega_i^{\theta+1})$ and analogously for $s_1$. This misspecified information structure overweights the probability of the more likely signal realization in each state and underweights the probability of the less likely signal realization, relative to the correctly-specified information structure. Intuitively, it corresponds to "counting" the signal realization $\theta + 1$ times, and has often been used in the theoretical literature to model overreaction (Bohren and Hauser 2021; Angrisani, Guarino, Jehiel, and Kitagawa 2020). For example, consider $\Omega = \{0.3, 0.7\}$ so that the probability of $s_2$ is 0.3 in state $\omega_1$ and 0.7 in state $\omega_2$. If the agent has representativeness parameter $\theta = 1$, then she updates beliefs as if the probability of $s_2$ is approximately 0.16 in state $\omega_1$ and 0.84 in state $\omega_2$. Following realization $s_j$, her posterior belief is as-if she had observed two realizations $s_j$.

*Discussion.* We refer to the first stage as *editing* because the agent responds to complexity in the information structure by focusing more on a subset of states while increasingly neglecting the other states. To see the intuition, consider an investor who forms beliefs about a tech company that recently entered the public market. The state space includes the possibility that the firm is a zombie (non-viable and set to crash), a unicorn (e.g., Google, Facebook), as well as a slew of intermediate possibilities. Upon observing a price increase (the signal realization), a boundedly rational investor does not have the cognitive capacity to consider all of the states when forming beliefs. Because unicorns are 'representative' of a price increase, the investor overweights the possibility of a unicorn, and the less extreme potential firm states are 'edited' out. Note that this does not imply that the investor is completely unaware of the intermediate firm states; these states just receive less weight compared to the Bayesian benchmark.

**Evaluation.** A large literature in cognitive psychology has shown that the variation in people's judgments and decisions can be explained by cognitive imprecision (Green et al. 1966; Thurstone 1927). Specifically, rather than using the parameters of the information environment directly, an agent treats them as if they are unbiased signals of the true underlying values. This generates a noisy representation of the information environment and leads to variation in actions. Notably, noisy cognition leads to reduced sensitivity to the parameters of any particular decision problem. Motivated by these insights, we argue that agents use a noisy representation of the edited posterior when evaluating information and updating beliefs.

Following the noisy cognition literature (Woodford 2020; Khaw et al. 2022), the agent does not know the edited posterior $p_R(s_i)$. Instead, for each signal realization $s_i$, she has a noisy internal representation $\tilde{y}(s_i) \equiv (\tilde{y}(\omega_1|s_i), ..., \tilde{y}(\omega_N|s_i)) \in \Delta(\Omega)$, which is an unbiased cognitive signal of $p_R(s_i)$. We assume that this representation is drawn from a multinomial distribution with $N$ categories (i.e., states), $\eta > 0$ trials, and event probabilities $p_R(s_i)$:

$$\tilde{y}(s_i) \sim \frac{1}{\eta} Multi(\eta, N, p_R(s_i)).$$

The parameter $\eta$ captures the precision of cognition: it is as-if the agent observed $\eta$ draws from distribution $p_R(s_i)$. Therefore, a higher $\eta$ corresponds to a more precise representation. The representation is unbiased, in that it has a mean equal to the edited posterior $p_R(s_i)$. The multinomial distribution is a natural choice for the distribution of a representation of a probability distribution, as any realization $y(s_i) = (y(\omega_1|s_i), ..., y(\omega_N|s_i))$ of the cognitive signal is indeed a probability distribution: each component is between 0 and 1 and the components sum to one, $y(\omega_j|s_i) \in [0, 1]$ and $\sum_{j=1}^{N} y(\omega_j|s_i) = 1$. It is the multi-state analogue of the binomial distribution used in Enke and Graeber (2023).

The agent has a prior about the edited posterior. Letting $\tilde{p}_R(s_i)$ denote the random variable describing the edited posterior, we assume the agent's prior about $\tilde{p}_R(s_i)$ is a Dirichlet distribution with $N$ categories (i.e., states) and concentration parameters $\nu\overline{p}_0$, where $\overline{p}_0 \in \Delta(\Omega)$ is the prior mean and the inverse of $\nu \geq 0$ scales the variance of the prior. As in Enke and Graeber (2023), $\overline{p}_0$ has the interpretation of a *cognitive default* prior that corresponds to an agent's belief about parameters of the information environment before internalizing a particular set of parameters. We assume that this default is the 'ignorance prior' that does not place greater weight on any given state; i.e., $\overline{p}_0(\omega_i) = 1/N$ is the uniform distribution.[11] Parameter $\nu$ determines how concentrated the agent's prior is around the default. The Dirichlet distribution is a natural choice for the prior over the edited posterior, since each draw from the Dirichlet distribution is a probability distribution over $N$ objects. It is the multi-state analogue of the Beta prior distribution used in Enke and Graeber (2023).

Given realized representation $y(s_i) \in \Delta(\Omega)$, the agent uses Bayes' rule to form her posterior belief about the edited posterior $\tilde{p}_R(s_i)$. Since the Dirichlet distribution is the conjugate prior of the multinomial distribution, this posterior also follows a Dirichlet distribution with concentration parameters $\nu\overline{p}_0 + \eta y(s_i)$ and mean

$$\mu(y(s_i)) \equiv E[\tilde{p}_R(s_i) \mid y(s_i)] = \lambda y(s_i) + (1-\lambda)\overline{p}_0, \tag{6}$$

where $\lambda \equiv \eta/(\eta+\nu) \in (0,1)$. For our predictions, we focus on the mean observed posterior, which corresponds to the expectation of $\mu(\tilde{y}(s_i))$ conditional on the realized edited posterior $p_R(s_i)$, i.e., $E[\mu(\tilde{y}(s_i))|p_R(s_i)]$. From Eq. (6), this is equal to

$$\hat{p}(s_i) \equiv \lambda p_R(s_i) + (1-\lambda)\overline{p}_0. \tag{7}$$

We refer to this as the agent's *subjective posterior*. When $\lambda < 1$, the agent biases her subjective posterior towards the cognitive default. As cognition becomes noisier (lower $\eta$) or the prior becomes more precise (higher $\nu$), the agent places more weight on her cognitive default (lower $\lambda$), and as cognition becomes more precise (higher $\eta$) or the prior becomes more diffuse (lower $\nu$), the agent places more weight on her edited belief (higher $\lambda$). When $\lambda = 1$ and $\theta = 0$, the subjective posterior is equal to the objective posterior.

In our analysis, we often focus on the expected state. Let $\hat{E}(\omega|s_i) = \lambda E_R(\omega|s_i) + (1-\lambda)\overline{E}(\omega_i)$ denote the subjective posterior expected state following signal realization $s_i$, where $\overline{E}(\omega) = \sum_{\omega_i \in \Omega} \overline{p}_0(\omega_i)\omega_i = 1/2$ is the expected state under the cognitive default and $E_R(\omega|s_i) = \sum_{\omega_j \in \Omega} p_R(\omega_j|s_i)\omega_j$ is the expected state with respect to the edited posterior following signal realization $s_i$.

*Discussion.* This cognitive noise model is related to the anchoring-and-adjustment heuristic in the judgment and decision-making literature (Tversky and Kahneman 1974), where the agent enters a decision environment with an 'anchor' belief $\overline{p}_0$ and insufficiently adjusts to new information (see Enke and Graeber (2023) for similar discussion). We are not the first to consider the relationship between the repre-

[11] We provide direct evidence for this assumption in Section 3 by eliciting beliefs *before* participants are provided with the parameters of the information environment. The 'ignorance prior' is the modal answer and the average is not significantly different from the uniform assumption.

sentativeness and anchoring-and-adjustment heuristics (see discussion in Griffin and Tversky (1992)), but our model is unique in formally outlining the predictions for belief-updating.

Our framework is part of a broader literature on how people use simplification strategies when making decisions or forming beliefs (e.g., Banovetz and Oprea (2023)). For example, Bordalo et al. (2023) present a model where an agent simplifies hypotheses using bottom-up attention, focusing on features that are salient to him at the time. The evaluation of these features generate different biases depending on which features are salient at the time, despite the same underlying information structure. In the context of our belief-updating framework, the salient features correspond to states that are most representative of the signal.

### 2.1.3 Defining Over- and Underreaction

We next define a measure of overreaction based on a comparison of the expected state under the subjective and objective posteriors. We refer to the objective (subjective) *expected movement* in the state following signal realization $s_i$ as the absolute value of the difference between the expected state under the objective (subjective) posterior belief following $s_i$ and the expected state under the prior belief, $|\hat{E}(\omega|s_i) - E(\omega)|$ and $|E(\omega|s_i) - E(\omega)|$, respectively. We say an agent *overreacts* to signal realization $s_i$ if her subjective expected movement is greater than the objective expected movement, and *underreacts* if it is less than the objective expected movement.

**Definition 1** (Over- and Underreaction)**.** *The agent* overreacts *to* $s_i$ *if* $|\hat{E}(\omega|s_i) - E(\omega)| > |E(\omega|s_i) - E(\omega)|$ *and* underreacts *if* $|\hat{E}(\omega|s_i) - E(\omega)| < |E(\omega|s_i) - E(\omega)|$.

When the objective and subjective expected movement coincide, $|\hat{E}(\omega|s_i) - E(\omega)| = |E(\omega|s_i) - E(\omega)|$, the agent neither overreacts nor underreacts to $s_i$. This is the case for all signal realizations when $\theta = 0$ and $\lambda = 1$ (the Bayesian benchmark). When $\theta > 0$ or $\lambda < 1$, the objective and subjective expected movement can differ and over- or underreaction can emerge. Whether an agent overreacts or underreacts potentially varies across signal realizations: an agent can overreact following one signal realization and underreact following the other.

In order to compare the magnitude of over- or underreaction across signal realizations and information environments, we need to define a measure of overreaction that accounts for the fact that the magnitude of the expected movement varies with the signal realization and information environment. For example, any signal realization results in less expected movement when the prior is tighter. Taking the difference between the objective and subjective expected movement does not account for this, so we standardize this difference by dividing it by the objective expected movement.[12]

[12]To see why this is necessary, note that when all possible states are close to each other, the numerator of $r(s_i)$ is naturally small, and the opposite is true if the states are very far apart. In

This brings us to the following definition of the *overreaction ratio* following signal $s_i$:

$$r(s_i) \equiv \frac{|\hat{E}(\omega|s_i) - E(\omega)| - |E(\omega|s_i) - E(\omega)|}{|E(\omega|s_i) - E(\omega)|}. \tag{8}$$

Given that the denominator of Eq. (8) is positive, by Definition 1 the agent overreacts to $s_i$ if $r(s_i) > 0$ and underreacts if $r(s_i) < 0$, and neither overreacts nor underreacts if $r(s_i) = 0$.

Definition 1 defines over- and underreaction with respect to the posterior expected state. This is consistent with both the finance and experimental literatures. The former typically studies asset prices and average forecasts instead of the entire belief distribution, which are similar in spirit to the expected state.[13] The latter typically compares the movement of the subjective and objective posterior beliefs in a binary state space. Given that a single number summarizes the posterior belief for a binary state space, this definition is equivalent to our comparison of posterior expected states.[14]

Given that Definition 1 is based on the absolute value of the difference between the posterior and prior expected states, it does not distinguish between whether an agent's subjective expected state moves in the same or opposite direction as the objective expected state. We define an additional property of updating to distinguish between subjective updates that move in the same versus opposite direction as the objective update.

**Definition 2** (Same and Wrong Direction Updates)**.** *An agent has a* same direction update *at $s_i$ if $\hat{E}(\omega|s_i) - E(\omega) \leq 0$ when $E(\omega|s_i) - E(\omega) \leq 0$ and $\hat{E}(\omega|s_i) - E(\omega) \geq$ when $E(\omega|s_i) - E(\omega) \geq 0$. Otherwise the agent has a* wrong direction update*. The agent has* same direction updates *if this holds for all signal realizations $s_i \in \mathcal{S}$.*

When the subjective and objective posterior expected states move in the same direction, $\hat{E}(\omega|s_i) - E(\omega)$ and $E(\omega|s_i) - E(\omega)$ have the same sign. In this case, the absolute values can be omitted from the expression for $r(s_i)$ and it reduces to $(\hat{E}(\omega|s_i) - E(\omega|s_i))/(E(\omega|s_i) - E(\omega))$.

addition, if we double the value of all states, the numerator is automatically doubled. Therefore, the numerator is not a sensible measure of magnitude.

[13] Our measure is closely linked to a common empirical test in the finance literature developed by Coibion and Gorodnichenko (2015). They examine the correlation between forecast errors and forecast revisions over time, where positive (negative) correlation corresponds to underreaction (overreaction). In our static belief-updating model, the counterparts of forecast errors and forecast revisions are $E(\omega|s_i) - \hat{E}(\omega|s_i)$ and $\hat{E}(\omega|s_i) - E(\omega)$, respectively. It is straightforward to verify that if the subjective posterior expected state moves in the same direction as the objective posterior expected state following $s_i$, then $r(s_i) < 0$ if and only if $(E(\omega|s_i) - \hat{E}(\omega|s_i))(\hat{E}(\omega|s_i) - E(\omega)) > 0$.

[14] When $\Omega$ is binary, we show that $r(s_i) > 0$ if and only if the subjective posterior for state $\omega_1$ moves further away from the prior than the objective posterior, i.e., $|\hat{p}(\omega_1|s_i) - p_0(\omega_1)| > |p(\omega_1|s_i) - p_0(\omega_1)|$, and similarly for $\omega_2$. See Appendix A for a proof.

## 2.2 Predictions

We next explore the interaction between representativeness and noisy cognition, and show that it gives rise to a predictable pattern of over- and underreaction. We derive comparative static predictions for how the overreaction ratio varies with respect to the complexity of the state space, the informativeness of the signal, the shape of the prior, and how the agent reacts to a *confirmatory* versus *disconfirmatory* signal realization. All proofs for this section are in Appendix A.

We first focus on the case of a symmetric prior. In this case, the expected state under the prior and the expected state under the cognitive default are both equal to 1/2; i.e., $E(\omega) = \overline{E}(\omega) = 1/2$. It is possible to simplify $r(s_i)$ to

$$r(s_i) = \lambda r_R(s_i) - (1 - \lambda), \tag{9}$$

where

$$r_R(s_i) \equiv \frac{|E_R(\omega|s_i) - E(\omega)| - |E(\omega|s_i) - E(\omega)|}{|E(\omega|s_i) - E(\omega)|}. \tag{10}$$

Note that $r(s_i)$ is linear and increasing in both $\lambda$ and $r_R(s_i)$. This highlights the opposing influences of representativeness and cognitive noise. In environments where the impact of representativeness dominates, indicated by $r_R(s_i) > (1-\lambda)/\lambda$, the agent exhibits overreaction. Conversely, when the impact of cognitive noise dominates, indicated by $r_R(s_i) < (1-\lambda)/\lambda$, the agent exhibits underreaction. While $(1-\lambda)/\lambda$ is a positive constant, $r_R(s_i)$ ranges from 0 to a potentially large number depending on $\theta$ and the information environment (as $\theta$ approaches $\infty$ and the objective expected movement approaches zero, $r_R(s_i)$ approaches $\infty$).

**Complexity of the State Space.** Consider two distinct information environments $(\Omega, p_0)$ and $(\Omega', p'_0)$ with symmetric state spaces and uniform priors. Let $r(s_i)$ denote the overreaction ratio for $(\Omega, p_0)$ given signal realization $s_i$ and $r'(s_i)$ analogously for $(\Omega', p'_0)$. We fix the dispersion of the state space (i.e., the minimum and maximum states) and vary the complexity of the state space by adding more interior states. Prediction 1 shows that when the representativeness parameter is sufficiently large, overreaction increases as the state space becomes more complex.

**Prediction 1** (Complexity)**.** *Suppose $\theta > 0$ and $\lambda \leq 1$. Consider two distinct information environments $(\Omega, p_0)$ and $(\Omega', p'_0)$ with symmetric state spaces, the same dispersion (i.e., $\omega_1 = \omega'_1$ and $\omega_N = \omega'_N$), and uniform priors. If $\Omega'$ is more complex than $\Omega$, and every state in $\Omega' \setminus \Omega$ is more interior than every state in $\Omega$, then for sufficiently large $\theta$, the agent overreacts more in $(\Omega', p'_0)$ than $(\Omega, p_0)$ following both signal realizations, $r'(s_i) > r(s_i)$ for $s_i \in \mathcal{S}$.*

For example, for any $x \in (0, 0.5)$ and $y \in (x, 0.5)$, the agent overreacts more in the 3-state environment $\Omega' = \{x, 0.5, 1 - x\}$ with a uniform prior or the four-state environment $\Omega'' = \{x, y, 1 - y, 1 - x\}$ with a uniform prior than in the binary state environment $\Omega = \{x, 1-x\}$ with a uniform prior, provided that $\theta$ is sufficiently large. Note that the 3-state environment is not directly comparable to the 4-state environment because the additional states $\{y, 1-y\}$ in $\Omega''$ are not more interior than the state 0.5 in $\Omega'$.

The intuition behind this result is as follows. Suppose the agent observes $s_2$. Under a uniform prior, when the state space becomes more complex, mass is shifted from extreme states to interior states under the prior. Since the highest state becomes less likely under the prior, the objective expected movement in the state following the signal is smaller. However, the representativeness-driven agent does not fully internalize this change because she overweights the highest state, as it remains the most representative state following $s_2$. When $\theta$ is sufficiently high, the agent's subjective expected state is primarily driven by this most representative state, which is the same in both the original and more complex state spaces. This leads to an increase in the relative difference between the movement of the subjective expected state and the objective expected state.

Note that the impact of increasing complexity on overreaction critically hinges on how the addition of the states changes the relative levels of representativeness. Such changes are substantial when the states are distinct from each other, but not when the states are extremely close to or even equal to each other. For example, the overreaction ratio moves continuously in $\varepsilon > 0$ as we move from state space $\Omega = \{0.3, 0.7\}$ with a uniform prior to state space $\Omega' = \{0.3, 0.3+\varepsilon, 0.7, 0.7+\varepsilon\}$ with a uniform prior. At $\varepsilon = 0$, $\Omega' = \{0.3, 0.3, 0.7, 0.7\}$ is an exact duplication of $\Omega$, and therefore, the two information environments yield equivalent overreaction ratios. In other words, the impact of increasing complexity is not determined by the number of new states per se, but by the number of new *distinct* states that affect how much the agent shifts her attention towards the more representative states.[15]

Prediction 1 also holds in a representativeness-only model ($\theta > 0$ and $\lambda = 1$), but such a model would predict overreaction in both $(\Omega, p_0)$ and $(\Omega', p'_0)$. This contrasts with our two-stage model, where it is possible to have underreaction in the simpler environment $(\Omega, p_0)$ and overreaction in the more complex environment $(\Omega', p'_0)$, or underreaction in both but less underreaction in the more complex environment $(\Omega', p'_0)$.

**Dispersion of the State Space.** We next explore changes in the extreme states while keeping complexity and the set of interior states constant. We maintain a uni-

[15]Indeed, Phillips and Edwards (1966) find significant underreaction in an experiment where there are ten states but each of them takes one of two unique values. In environments with duplicate states—or states so close that they are essentially duplicates—we conjecture that people will first simplify the environment by grouping these redundant states and then further reduce complexity via the representativeness heuristic. See, for example, Evers, Imas, and Kang (2022) for evidence on how agents simplify the evaluation of similar outcomes. Testing this prediction is outside the scope of the current paper, as our experiments focus on information environments with easily distinguishable states.

form prior throughout. The impact of changing the extreme states on the agent's reaction is more nuanced because moving the extreme states leads to non-trivial changes in both the objective expected state and the subjective expected state, even for an agent with a high level of representativeness. For example, consider an information structure with five states, $\Omega = \{x, 0.3, 0.5, 0.7, 1-x\}$, where $x < 0.3$. As $x$ increases, the objective expected state moves less following any signal realization, since $x$ and $1-x$ are closer to the prior expected state $1/2$. Meanwhile, the subjective expected state also moves less since representativeness causes the agent to overweight $x$ following $s_1$ or $1-x$ following $s_2$, and as in the objective case, $x$ and $1-x$ are both closer to the prior expected state. Reducing the dispersion of the state space results in a higher overreaction ratio if the impact on the objective expected state dominates. We show that under a uniform prior, this occurs when $W(\Omega) > 0$ and $W(\Omega') > 0$, where

$$W(\Omega) \equiv \sum_{i \in \{1,N\}} (\omega_i - 1/2)^2 - \sum_{i \notin \{1,N\}} (\omega_i - 1/2)^2 \,, \tag{11}$$

with $W(\Omega')$ defined analogously. When $W(\Omega) > 0$ and $W(\Omega') > 0$, the extreme states tend to be close to 0 and 1 and the interior states tend to be close to $1/2$. This results in a signal that is more informative about the extreme states and less informative about the interior states—and therefore, an objective posterior belief that attaches a high probability to one of the extreme states. Therefore, the objective expected state is relatively more sensitive to the values of the extreme states. To the contrary, if $W(\Omega) < 0$ and $W(\Omega') < 0$, the objective expected state is less sensitive to changes in the extreme states and the impact of the change in state space dispersion on the subjective expected state dominates. Reducing dispersion reduces the movement in the subjective expected state relatively more than the objective expected state, leading to a lower overreaction ratio. Taken together, this yields the following prediction.

**Prediction 2** (State Dispersion)**.** *Suppose $\theta > 0$ and $\lambda \leq 1$. Consider two distinct information environments $(\Omega, p_0)$ and $(\Omega', p_0')$ with the same complexity, the same set of interior states, symmetric state spaces, and uniform priors. If $\Omega'$ is less dispersed than $\Omega$, then for sufficiently large $\theta$:*

*(i) If $W(\Omega) > 0$ and $W(\Omega') > 0$, then the agent overreacts* more *in $(\Omega', p_0')$ than $(\Omega, p_0)$ following both signal realizations, $r'(s_i) > r(s_i)$ for $s_i \in \mathcal{S}$.*

*(ii) If $W(\Omega) < 0$ and $W(\Omega') < 0$, then the agent overreacts* less *in $(\Omega', p_0')$ than $(\Omega, p_0)$ following both signal realizations, $r'(s_i) < r(s_i)$ for $s_i \in \mathcal{S}$.*

As an example, consider $\Omega = \{x, 0.5, 1-x\}$ and $\Omega' = \{x', 0.5, 1-x'\}$, where $x < x' < 0.5$ so $\Omega$ is more dispersed than $\Omega'$. Since $W(\Omega) = 2(x-0.5)^2 > 0$ and

$W(\Omega') = 2(x' - 0.5)^2 > 0$, the agent overreacts more under the less dispersed state space $\Omega'$ than under $\Omega$ for sufficiently large $\theta$.

**Signal Diagnosticity.** In a symmetric binary state space, the diagnosticity of the signal is equal across states, $d_1 = d_2 \equiv d$. In this case, increasing the dispersion of the state space is equivalent to increasing the diagnosticity $d$. It follows as a corollary of Prediction 2 that a higher diagnosticity leads to less overreaction. In addition, the agent underreacts to precise signals and, for sufficiently high $\theta$, overreacts to noisy signals.

**Corollary 1** (Diagnosticity—Binary State Space)**.** *Suppose $\theta > 0$ and $\lambda \le 1$. Consider a set of information environments with symmetric binary state spaces parameterized by signal diagnosticity $d$ and a uniform prior. Then the overreaction ratio $r^d(s_i)$ is decreasing in $d$ for $s_i \in \mathcal{S}$. When $\lambda < 1$, there exists a cutoff $c \in [1/2, 1)$ such that the agent overreacts to both signal realizations if $d \in (1/2, c)$ and underreacts to both signal realizations if $d \in (c, 1)$. When $\theta$ is sufficiently large, $c > 1/2$.*

When the state space is symmetric but has more than two states, the vector of signal diagnosticities $(d_1, ..., d_N)$ can no longer be conveniently summarized by a single number. We combine Prediction 1 and Prediction 2 to establish that if $W(\Omega) > 0$, then the agent reacts more as the signal becomes less informative in all states—or equivalently, as *all* states become more interior—for sufficiently large $\theta$.

**Corollary 2** (Diagnosticity)**.** *Suppose $\theta > 0$ and $\lambda \le 1$. Consider two distinct information environments $(\Omega, p_0)$ and $(\Omega', p_0')$ with the same complexity, symmetric state spaces, and uniform priors such that $W(\Omega') > 0$ and $W(\Omega) > 0$. If $d_i' \le d_i$ for all $i = 1, ..., N$ and at least one inequality is strict, then for sufficiently large $\theta$, the overreaction ratio is larger in $(\Omega', p_0')$ than $(\Omega, p_0)$ following both signal realizations, $r'(s_i) > r(s_i)$ for $s_i \in \mathcal{S}$.*

For example, consider state space $\Omega = \{x, y, 1-y, 1-x\}$ with $x \in (0, 0.5)$, $y \in (x, 0.5)$ and a uniform prior. We can readily verify $W(\Omega) > 0$. Corollary 2 implies that the agent reacts more as $x$ and $y$ both move closer to $1/2$ for sufficiently large $\theta$.

**Prior Concentration.** Next, consider two symmetric information environments $(\Omega, p_0)$ and $(\Omega, p_0')$ with the same state space and different priors. We say that $p_0'$ is more *concentrated* than $p_0$ if there exists a cutoff $c \in (1/2, 1)$ such that $p_0'(\omega) \ge p_0(\omega)$ for all $\omega \in [1-c, c]$ and $p_0'(\omega) \le p_0(\omega)$ for all $\omega \in [0, 1-c] \cup [c, 1]$, and is strictly more concentrated if at least one of the inequalities is strict. In words, a more concentrated prior assigns higher probability to interior states and lower probability to extreme states. Prediction 3 establishes that the overreaction ratio increases in the concentration of the prior.

**Prediction 3** (Prior concentration)**.** *Suppose $\theta > 0$ and $\lambda \le 1$. Consider two symmetric information environments $(\Omega, p_0)$ and $(\Omega, p_0')$. If $p_0'$ is strictly more concen-*

*trated than $p_0$, then for sufficiently large $\theta$, the agent overreacts more in $(\Omega, p_0')$ than in $(\Omega, p_0)$ following both signal realizations, $r'(s_i) > r(s_i)$ for $s_i \in \mathcal{S}$.*

The intuition behind Prediction 3 is similar to that of Prediction 1. With a more concentrated prior, the objective expected state moves less following each signal realization, but the representativeness heuristic continues to result in the agent over-weighting the extreme states. Therefore, the agent exhibits more overreaction. Without cognitive noise, the representativeness-only model predicts overreaction under both priors. In contrast, the two-stage model with representativeness and cognitive noise can also result in underreaction under the less concentrated prior and overreaction under the more concentrated prior, or underreaction under both priors.

**Asymmetric Prior.** Finally, we consider information environments with asymmetric priors. We restrict attention to symmetric binary state spaces where it is straightforward to manipulate the symmetry of the prior. In such environments, we can define whether a signal realization is *confirmatory* or *disconfirmatory* based on its alignment with the prior.

**Definition 3.** *In a symmetric binary state space, a signal realization $s_i$ is* confirmatory *if either (i) $p_0(\omega_1) > p_0(\omega_2)$ and $Pr(s_i|\omega_1) > Pr(s_i|\omega_2)$, or (ii) $p_0(\omega_1) < p_0(\omega_2)$ and $Pr(s_i|\omega_1) < Pr(s_i|\omega_2)$. A signal realization $s_i$ is* disconfirmatory *if either (iii) $p_0(\omega_1) > p_0(\omega_2)$ and $Pr(s_i|\omega_1) < Pr(s_i|\omega_2)$, or (iv) $p_0(\omega_1) < p_0(\omega_2)$ and $Pr(s_i|\omega_1) > Pr(s_i|\omega_2)$.*

For example, if the prior assigns higher probability to state $\omega_1$, then a signal realization is confirmatory if it is more likely under $\omega_1$ than $\omega_2$ (the signal is expected), and is disconfirmatory if it is more likely under $\omega_2$ than $\omega_1$ (the signal is surprising). Note that in the case of a symmetric prior $p_0(\omega_1) = p_0(\omega_2)$, a signal realization is neither confirmatory nor disconfirmatory.

The two-stage model generates a rich set of predictions about reactions to confirmatory versus disconfirmatory signal realizations. When $\lambda < 1$, the agent can overreact to a disconfirmatory realization and underreact to a confirmatory realization. When the signal is relatively uninformative, the agent updates in the wrong direction following a confirmatory realization. Moreover, when representativeness is sufficiently high, the agent also overreacts to a confirmatory realization of a signal with intermediate diagnosticity.

**Prediction 4** (Asymmetric Prior)**.** *Suppose $\theta \geq 0$ and $\lambda < 1$. Consider a binary state space $\Omega$ with signal diagnosticity $d$ and prior $p_0$.*

*(i) Suppose $s_i$ is confirmatory. There exist cutoffs $1/2 < c_1 \leq c_2 \leq c_3 < 1$ such that the agent has a* wrong direction update *if $d \in (1/2, c_1)$,* overreacts *if $d \in (c_2, c_3)$, and* underreacts *if $d \in (c_1, c_2) \cup (c_3, 1)$. Moreover, $(c_2, c_3)$ is nonempty for sufficiently large $\theta$.*

(ii) *Suppose $s_i$ is disconfirmatory. There exist cutoffs $1/2 < c_4 \leq c_5 < 1$ such that the agent* overreacts *if $d \in (1/2, c_4)$ and* underreacts *if $d \in (c_5, 1)$.*

For intuition, consider the case where the agent is not subject to representativeness ($\theta = 0$). Suppose the prior places more weight on $\omega_2$, resulting in a prior mean $E(\omega) > 1/2$. A relatively uninformative confirmatory realization $s_2$ increases the objective expected state, $E(\omega|s_2) > E(\omega)$, but this increase is small. However, noisy cognition results in the agent compressing her subjective expected state towards the cognitive default 1/2 with a non-trivial weight of $1 - \lambda$, resulting in a wrong direction update, $\hat{E}(\omega|s_2) < E(\omega)$. In contrast, following a relatively uninformative disconfirmatory realization $s_1$, the objective expected state decreases slightly but remains above the cognitive default of 1/2, $1/2 < E(\omega|s_1) < E(\omega)$. Noisy cognition pulls the subjective expected state towards the cognitive default, decreasing it more than $E(\omega|s_1)$ and implying overreaction. As the diagnosticity increases, the signal outweighs the prior in determining the direction of the posterior, i.e., $E(\omega|s_1) < 1/2 < E(\omega)$. In this case, noisy cognition pulls the subjective expected state back up towards the cognitive default of 1/2, decreasing it less than $E(\omega|s_1)$ and resulting in underreaction. When the agent is also subject to representativeness ($\theta > 0$), she reacts more to both signal realizations, and more so when the signal is relatively uninformative. If this force is strong enough, the agent may even overreact to a confirmatory realization with an intermediate diagnosticity. As the diagnosticity approaches 1, similar to our observation in Corollary 1, the impact of representativeness is dominated by the impact of noisy cognition. Thus, the agent underreacts regardless of whether the signal realization is confirmatory or disconfirmatory.

## 3 Empirical Investigation

In this section, we test the predictions of our framework in a controlled experimental setting. In particular, we test how belief-updating depends on complexity, the informativeness of the signal, the shape of the prior, and the type of signal realization (i.e., confirmatory versus disconfirmatory). We also directly test the proposed attentional mechanism.

### 3.1 Method

Participants were recruited from the Prolific crowdsourcing platform. A total of 2,210 participants (49.3% female, 38.3 average age) took part in our experiment.[16] They first had to pass an attention check before reading any experimental instructions. Those who did not pass the first attention check did not proceed to the rest of the study; we did not collect data from these participants and they are not included in the participant totals. After passing the initial screen, participants were told that in addition to the base payment of $2, they could earn two additional bonus

[16] Preregistration materials can be found here: https://aspredicted.org/LTJ_CS7 and https://aspredicted.org/Q77_3LG.

payments. First, they earned \$1 for answering a comprehension check that followed the instructions. Second, they would earn an additional \$10 if their response to a randomly-chosen belief elicitation question was within 3% of the corresponding objective posterior.[17] We used this incentive procedure as opposed to more complex mechanisms (e.g., quadratic or binarized scoring rules) because recent evidence shows that these mechanisms can systematically bias truthful reporting.[18]

### 3.2 Design

Participants who passed the initial attention check were given the following description of the information environment:[19]

There is a deck of 100 cards, where each card has the number of a bag written on it, e.g., 'Bag 1' or 'Bag 2'. Each possible bag has 100 balls, which are either red or blue. The computer will randomly draw a card from the deck to select a bag, then randomly draw *one* ball from the selected bag and show it to you.

Participants completed multiple trials, each of which involved a new randomly selected bag and ball. In each trial, the participant was told the number of bags (the states), how many cards corresponded to each bag (the prior), and how many red versus blue balls each bag contained (the information structure). After observing the color of the randomly drawn ball (the signal realization, with $s_1 = b$ corresponding to blue and $s_2 = r$ corresponding to red), the participant's task was to report how likely she thought that each bag was selected (i.e., Bag 1, Bag 2, etc.) by reporting a percentage from 0 to 100. We required these percentages to add up to 100 across all possible bags. After reporting this probability assessment, the participant proceeded to the next trial.

This 'bookbag-and-poker-chip' design (Edwards 1968) cleanly maps onto the information environment in our model. The number of bags corresponds to the size of the state space, the number of cards for each bag corresponds to the objective prior, and the number of red versus blue balls in each bag corresponds to the information structure. As in Section 2, we set the value of the state corresponding to a given bag as the share of red balls in this bag, $\omega_i = Pr(r|\omega_i)$, and we define the diagnosticity of the signal in state $\omega_i$ as the probability of the more likely ball color, $d_i \equiv \max\{\omega_i, 1-\omega_i\}$. For example, in the 2-state case in which 'Bag 1' has 60 red balls and 'Bag 2' has 40 red balls, we have $d_1 = d_2 = 0.6$. Fig. 1 depicts an information environment with 3 states, a concentrated prior with more probability mass on the interior state (Bag 2) and less on the more extreme states (Bags 1 and

[17] See Enke, Graeber, and Oprea (2023) for similar use of objective posterior as the incentivized benchmark.

[18] Danz, Vesterlund, and Wilson (2022) show that the binarized scoring rule leads to conservatism in elicited beliefs and greater error rates compared to simpler mechanisms; they argue that incentives based on belief quantiles—such as the one we use here—will result in more truthful reporting and lower cognitive burden.

[19] The specific instructions can be found in Appendix C.

3), and information structure with signal diagnosticity 0.6 in Bags 1 and 3, and 0.5 in Bag 2. In this environment, Bag 1 corresponds to state $\omega_3 = 0.6$, Bag 2 corresponds to state $\omega_2 = 0.5$ and Bag 3 corresponds to state $\omega_3 = 0.4$.

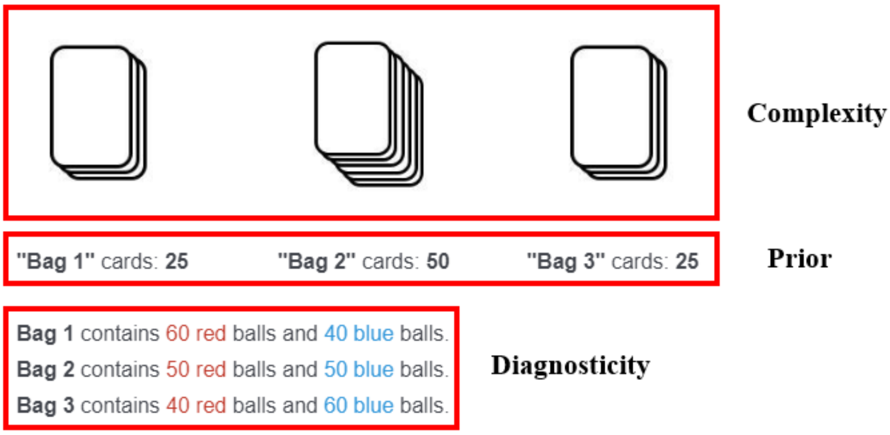



FIGURE 1. Experimental design for 3-state treatment

It is straightforward to manipulate the parameters of the information environment in this design. We manipulated four factors to test the predictions of our model:

- **Complexity of State Space:** The number of bags.
- **Information Structure:** The number of red versus blue balls in a given bag.
- **Prior Concentration:** In a symmetric state space with more than two bags, the number of cards corresponding to bags with a more extreme distribution of ball colors (i.e., more extreme states) versus a more moderate distribution (i.e., more interior states).
- **Prior Symmetry:** In a symmetric state space with two bags, the prior probability of one versus the other bag.

Table 9 in Appendix B.1 outlines the set of parameter combinations that we used in our experiments. As in the model, we focus on symmetric information structures (e.g., if there was a bag with 40 red balls, there was also a bag with 60 red balls). The most 'representative' bag always corresponded to Bag 1 or Bag $N$, depending on whether a red or blue ball, respectively, was drawn.[20]

Beyond the clean mapping to our model and the ease with which we can manipulate the parameters of the information environment, this 'bookbag-and-poker-chip' design has several advantages. First, it allows for a transparent calculation of the objective Bayesian benchmark. For example, suppose there are 50 'Bag 1' cards and

[20]For example, in the $N = 3$ state case with $Pr(r|\omega_1) = 0.4$, $Pr(r|\omega_2) = 0.5$ and $Pr(r|\omega_3) = 0.6$, if a red ball was drawn, the representative bag was the one with 60 red balls, and if a blue ball was drawn, the representative bag was the one with 40 red balls.

50 'Bag 2' cards, with Bag 1 containing 70 red balls and Bag 2 containing 30 red balls. Upon observing a red ball, the Bayesian posterior that Bag 1 was selected is 0.7. Second, the paradigm is used extensively in the literature, with the vast majority of papers using a simple state space with two bags. Consistent with our framework, these papers predominantly find underreaction. It would therefore be particularly noteworthy to show evidence for overreaction in the setting where underreaction is typically documented.

The timing of the experiment proceeded as follows. After reading the instructions, participants completed a set of comprehension questions. They were then randomized into one of the complexity conditions—2, 3, 4, 5 or 11 states—and completed a set number of trials in random order. Each complexity condition had at least 200 participants. The total number of possible unique trials for each complexity condition was equal to the product of the number of prior distributions, information structures, and signal realizations (which was always 2).[21] Each subject completed a maximum of 15 trials which were randomly drawn from the total set of possible trials. After completing all the trials, participants answered a set of basic demographic questions and exited the study.

To measure the cognitive default prior $\overline{p}_0$, we ran a version of the 3-state and 11-state uniform prior parameterizations where participants ($N = 149$) were presented with the basic structure of the experiment but not the actual underlying parameters.[22] Participants were then asked, based on the information provided, how many cards of each bag type were most likely to be in the deck. In addition to a \$1 completion fee, they received a \$1 bonus if one of their randomly-selected guesses was within 3% of the actual number of cards corresponding to that bag. A joint F-test cannot reject that participants were assigning the same probability to each bag in both the 3- and 11-state conditions. This is consistent with a uniform cognitive default, i.e., the 'ignorance prior.'

**Defining Over- and Underreaction.** Our main dependent variable compares participants' responses to the objective prior and posterior. Recall that our measure of reaction (Eq. (8) in Section 2) is based on the perceived versus objective expected state. Since in our experimental environment the numeric value of a given state is the fraction of red balls in the corresponding bag, the expected state is equal to the expected probability of drawing a red ball. In every trial, we calculate (a) the participant's expectation of drawing a red ball given their reported beliefs, (b) the objective prior expectation of drawing a red ball, and (c) the objective posterior expectation of drawing a red ball. Using these measures, we compute the reaction

[21]For example, the total number of unique trials in the 3-state condition was 3 (priors) $\times$ 8 (information structures) $\times$ 2 (signal realizations) = 48.

[22]Namely, participants were told that there were three or eleven potential bags but not the composition of bags in the deck nor the composition of balls in each bag.

ratio $r(s_i)$ defined in Eq. (8). A positive (negative) $r(s_i)$ corresponds to over- (under-) reaction in that trial. We use the average of the reaction ratios across participants in a given information environment as our primary measure of over- and underreaction in the analysis.

Experimental studies on belief-updating often measure over- and underreaction by running the so-called *Grether regression* (Grether 1980), which decomposes the logarithm of the posterior odds ratio into the logarithm of the prior ratio and the logarithm of the signal likelihood, $\log \frac{\hat{p}(\omega_2|s_i)}{\hat{p}(\omega_1|s_i)} = c_1 \log \frac{p_0(\omega_2)}{p_0(\omega_1)} + c_2 \log \frac{Pr(s_i|\omega_2)}{Pr(s_i|\omega_1)}$. These studies focus on binary state spaces in which the posterior belief can be summarized by a single likelihood ratio. This is no longer the case with more than two states, and therefore, the Grether regression is not applicable in our multi-state setting. Furthermore, the Grether regression imposes a log-linear structure on the form of over- and underreaction. This could mask important relationships between the parameters of the information structure and the extent of under- or overreaction. In contrast, our measure, which is based on expectations, is non-parametric and thus free from such restrictions.

## 3.3 Results

In this section, we test our predictions on how complexity, signal diagnosticity, prior concentration, and prior symmetry impact belief-updating. Per our pre-registration, unless otherwise noted, we exclude trials in which participants update in the wrong direction. Appendix B.2 replicates the analyses including wrong direction updates; the results do not qualitatively change.

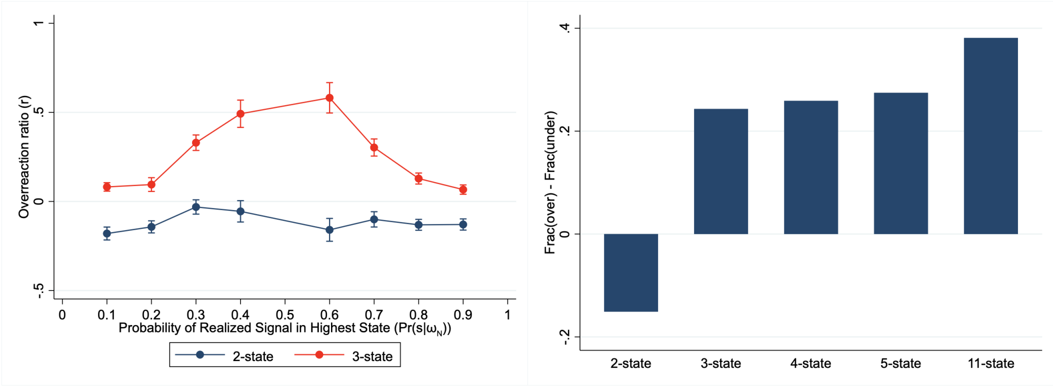
### 3.3.1 Complexity

To test Prediction 1, we compare belief-updating in information environments with a uniform prior that vary complexity while holding the dispersion of the state space (i.e., the highest and lowest states) constant. We begin by looking at how belief-updating changes when complexity is increased by just one state (comparing 2-state versus 3-state treatments). We then study the effect of adding progressively more interior states (comparing the 2-state treatment to 4-state and 5-state treatments).[23] Although not a direct test of Prediction 1, the 11-state treatment allows us to examine belief-updating with the addition of 'many' states. For convenience, we use $d$ to denote the signal diagnosticity associated with the extreme states, $d \equiv d_1 = d_N$.

In the simple 2-state treatment, we replicate the underreaction result from the experimental literature. Namely, on average, participants' reported posterior beliefs move significantly less than the Bayesian benchmark ($r < 0, p < .001$).[24] Fig. 2a

[23]We do not have a prediction for how the overreaction ratio in the 3-state treatment compares to the 4-state and 5-state treatments because the latter do not add interior states to the former.

[24]This $p$-value and others reported in the text come from a one-sample $t$-test against 0, unless otherwise noted.




(A) Overreaction Ratio (B) % Overreact - % Underreact

FIGURE 2. Complexity increases overreaction

plots the overreaction ratio for both signal realizations across different information structures. The x-axis corresponds to the probability of the realized signal in the highest state $\omega_N$. When a red ball is observed, this probability ranges from 0.6 to 0.9 depending on the signal diagnosticity $d$; when a blue ball is observed, this probability ranges from 0.1 to 0.4 depending on $d$ (see Table 9 for a list of the two-state information structures).[25] We observe significant underreaction to both signal realizations across nearly all information structures in the 2-state treatment. This is consistent with the evidence outlined in Benjamin (2019) that shows underreaction to signals in a host of experiments using a similar paradigm to our own.

Increasing the complexity of the state space reverses this result. Strikingly, increasing complexity by even a single state—going from 2 to 3 states—leads participants to report posterior beliefs that move significantly *more* than the Bayesian benchmark, $r > 0$ ($p < .001$). As illustrated in Fig. 2a, in the 3-state treatment we observe significant overreaction to both signal realizations across all information structures.

This pattern continues as we move to more complex settings. Table 1 compares the 2-state, 4-state, and 5-state treatments. At every diagnosticity $d$ for the 2-state case, the 4-state treatment adds two interior states and the 5-state treatment adds three interior states. Column 1 regresses the overreaction ratio on dummies corresponding to the 4-state and 5-state treatments, with the 2-state treatment as the control. As seen in the table, increasing the complexity of the state space leads to a significant increase in overreaction: the overreaction ratio is significantly higher in the 4-state and 5-state treatments compared to the simple 2-state treatment. Moreover, the overreaction ratio is significantly higher in the 5-state treatment compared to the 4-state treatment (the former adds one interior state to the latter): the coeffi-

[25]Due to the symmetry of the information structure, when a blue ball occurs with probability $x$ in state $\omega_N$, then a red ball occurs with probability $1 - x$. Therefore, on the x-axis of Fig. 2a, 0.1 and 0.9 correspond to blue and red signal realizations from the same information structure, and so on.

Table 1. Complexity increases overreaction

| | Overreaction Ratio | |
|---|---|---|
| | (1) | (2) |
| 4 States | 0.276*** | 0.371*** |
| | (0.0295) | (0.0315) |
| 5 States | 0.365*** | 0.455*** |
| | (0.0359) | (0.0383) |
| $d = 0.7$ | | -0.158*** |
| | | (0.0407) |
| $d = 0.8$ | | -0.355*** |
| | | (0.0422) |
| $d = 0.9$ | | -0.462*** |
| | | (0.0437) |
| Constant | -0.116*** | 0.127*** |
| | (0.0219) | (0.0409) |
| $N$ | 6253 | 6253 |
| adj. $R^2$ | 0.037 | 0.095 |

Notes: Baseline is 2 States and, in Column 2, diagnosticity $d = 0.6$. Includes information environments with a uniform prior and 2 states, 4 states or 5 states listed in Table 9; excludes wrong direction updates. Standard errors clustered at the individual level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

cient on the 4-state dummy is significantly smaller than the coefficient on the 5-state dummy ($p < .01$). Column 2 also controls for the information structure by including dummies for each diagnosticity, with $d = 0.6$ as the control. It finds a similar pattern of overreaction increasing with complexity. Finally, we also observe significant overreaction in the 11-state treatment, $r > 0$ ($p < .001$); this treatment has the highest average overreaction ratio.

Fig. 2b uses a discrete measure that captures the frequency of deviations from the Bayesian benchmark. For each complexity treatment, this measure computes the difference between the fraction of trials with overreaction and the fraction with underreaction. A positive value indicates a prevalence of overreaction and a negative value indicates a prevalence of underreaction. We again see that participants tend to underreact in the 2-state treatment, but overreact when complexity increases. Together, these results provide strong support for Prediction 1.

Table 2. Overreaction decreases in signal diagnosticity

| | Overreaction Ratio | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| | 2 States | 3 States | 4 States | 5 States |
| $d = 0.7$ | 0.0450 | -0.218*** | -0.370*** | -0.196** |
| | (0.0483) | (0.0502) | (0.0655) | (0.0863) |
| $d = 0.8$ | -0.0268 | -0.421*** | -0.597*** | -0.402*** |
| | (0.0498) | (0.0496) | (0.0692) | (0.0864) |
| $d = 0.9$ | -0.0432 | -0.461*** | -0.669*** | -0.558*** |
| | (0.0484) | (0.0505) | (0.0725) | (0.0878) |
| Constant | -0.110** | 0.535*** | 0.703*** | 0.644*** |
| | (0.0475) | (0.0554) | (0.0755) | (0.0942) |
| $N$ | 870 | 1347 | 2754 | 2629 |
| adj. $R^2$ | 0.002 | 0.070 | 0.117 | 0.059 |

Notes: Baseline is diagnosticity $d = 0.6$. Includes all information environments with a uniform prior listed in Table 9 except for the 11-state complexity; excludes wrong direction updates. Standard errors clustered at the individual level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

### 3.3.2 Signal Diagnosticity

To test Prediction 2 and Corollary 2, we compare belief-updating as a function of diagnosticity $d$ while holding the complexity constant. As above, we focus on information environments with a uniform prior and $d$ corresponds to the diagnosticity of the extreme states. Fig. 2a provides support for the predictions in the 3-state treatment: overreaction clearly decreases as signals become more precise, with the highest level of overreaction as $d$ approaches 0.5 (noisiest signal) and the lowest level of overreaction as $d$ approaches 0.9 (most precise signal), which corresponds to 0.1 and 0.9 on the x-axis. Table 2 provides further support in the other complexity treatments. Each column regresses the overreaction ratio in a given complexity treatment on dummy variables for the diagnosticity $d$, with $d = 0.6$ as the control.[26] Consistent with the prediction, there is progressively less overreaction as $d$ increases, i.e., as signals become more precise. For example, in the 5-state treatment (Column 4), the overreaction ratio decreases by 0.56 as $d$ increases from 0.6 to 0.9.[27]

[26] This analysis pools the information structures for a given diagnosticity of the extreme states. For example, in the 5-state treatment, when $d = 0.9$, the diagnosticity in the fourth states, $d_4$, ranges from 0.55 to 0.8. The results do not change qualitatively if we further split the analysis by diagnosticity of the interior states.

[27] Edwards (1968) and Augenblick et al. (2022) find overreaction to extremely noisy signals in a 2-state environment. We ran a version of the 2-state treatment with $d = 0.51$ and also find evidence for overreaction to this very noisy signal ($r = 0.08$, $p < .001$), although to a lesser extent than in the more complex environments.

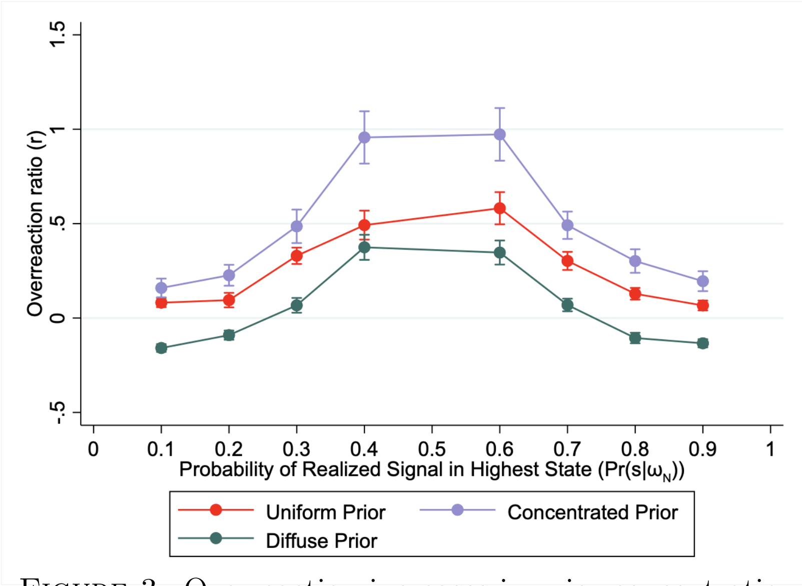



FIGURE 3. Overreaction increases in prior concentration

One potential concern is that changes in complexity and the information structure also change the Bayesian benchmark. Since our primary measure of overreaction is defined relative to the Bayesian benchmark, we may find the same pattern of under- versus overreaction even if participants do not understand the information environment and simply use a constant heuristic that reports the same posterior belief independently of changes in the information environment. To address this concern, Fig. 10 in Appendix B.3 presents the average reported posterior belief about each state for each information structure. There are several things to note. First, posterior beliefs shift from underreaction in the 2-state treatment to overreaction in the more complex treatments across all information structures. Second, it is readily apparent that this shift is due to participants actively changing their reported beliefs as a function of the information environment, in line with the theoretical framework. The changes in state-specific beliefs we observe are inconsistent with a constant heuristic response. Therefore, this is an unlikely alternative mechanism for our results.[28]

### 3.3.3 Prior Concentration

To test Prediction 3, we examine how the concentration of the prior affects belief-updating. Our experiment focused on 3-state environments, manipulating the prior from a diffuse prior of $(0.4, 0.2, 0.4)$ that placed twice as much mass on the extreme states relative to the interior state, to a uniform prior of $(0.33, 0.34, 0.33)$, to a concentrated prior of $(0.25, 0.50, 0.25)$ that placed twice as much mass on the interior state relative to the extreme states.[29] As illustrated in Fig. 3, consistent with Prediction 3,

[28] Section 3.5 provides further evidence against the constant heuristic mechanism in a setting that fixes the information environment and manipulates usable attention capacity.

[29] Note that our predictions hold for any number of states greater than two. The prior concentration cannot shift by definition in a two-state environment.

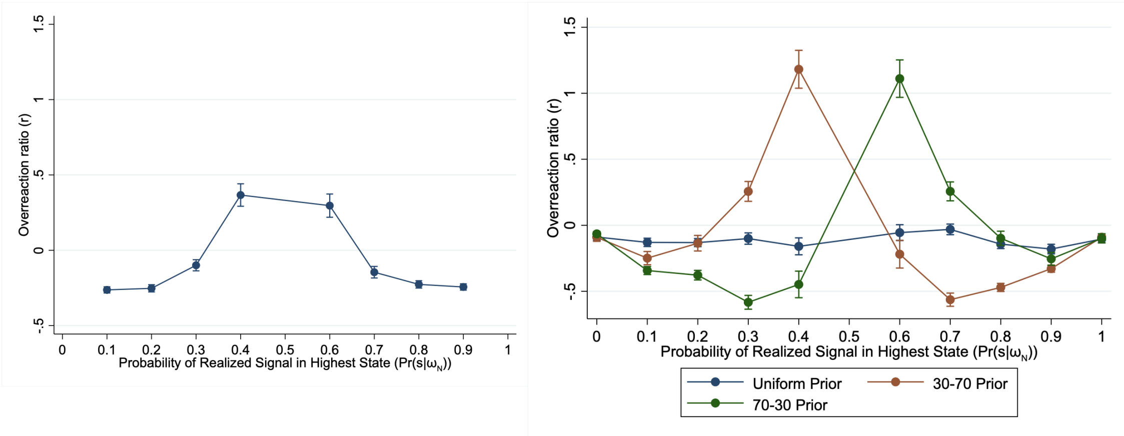
TABLE 3. Overreaction increases in prior concentration

| | Overreaction Ratio | |
|---|---|---|
| | (1) | (2) |
| Concentrated Prior | 0.213*** | 0.213*** |
| | (0.0547) | (0.0547) |
| Diffuse Prior | -0.215*** | -0.214*** |
| | (0.0321) | (0.0320) |
| $d = 0.7$ | | -0.311*** |
| | | (0.0321) |
| $d = 0.8$ | | -0.503*** |
| | | (0.0327) |
| $d = 0.9$ | | -0.557*** |
| | | (0.0332) |
| Constant | 0.260*** | 0.603*** |
| | (0.0253) | (0.0401) |
| $N$ | 4026 | 4026 |
| adj. $R^2$ | 0.048 | 0.127 |

Notes: Includes all information environments with three states listed in Table 9; excludes wrong direction updates. Baseline is uniform prior $(0.33, 0.34, 0.33)$ and, in Column 2, diagnosticity $d = 0.6$. Concentrated prior corresponds to $(0.25, 0.5, 0.25)$ and diffuse prior corresponds to $(0.4, 0.2, 0.4)$. Standard errors clustered at the individual level in parentheses. $^{*}$ $p < 0.10$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$.

we observe significantly more overreaction as the prior becomes more concentrated. This holds across both signal realizations for each possible information structure in the 3-state treatment. Table 3 Column 1 regresses the overreaction ratio on a dummy for a concentrated prior and a diffuse prior, with the uniform prior as the control. As shown in the table, participants overreact significantly more when the prior is concentrated and significantly less when the prior is diffuse. This comparative static continues to hold when controlling for the information structure (Column 2).

We observe significantly less overreaction for higher diagnosticities across all three priors, as shown in Fig. 3 and Column 2 of Table 3. Under a diffuse prior, we actually observe significant *underreaction* for high diagnosticities (e.g., the overreaction ratio is significantly less than zero for $d = 0.8$ and $d = 0.9$). Otherwise, we observe significant overreaction (i.e., the overreaction ratio is significantly greater than zero). Together, these findings provide strong support for Prediction 3.




(A) Underreaction to precise signals & overreaction to noisy signals

(B) More overreaction to disconfirmatory realizations

FIGURE 4. Overeaction with asymmetric priors

### 3.3.4 Prior Symmetry

To test Prediction 4, we examine how the type of signal realization (confirmatory versus disconfirmatory) affects belief-updating in information environments with an asymmetric prior. We focus on 2-state environments, manipulating the prior from an asymmetric prior of $(0.3, 0.7)$ or $(0.7, 0.3)$ to a symmetric prior of $(0.5, 0.5)$. Fig. 4a presents the overreaction ratio aggregated across all three priors. We continue to observe significant underreaction for more precise signals, but also observe overreaction to noisier signals.

However, aggregating across priors masks significant heterogeneity in belief-updating following a confirmatory (the more likely, or expected, realization under the prior) versus a disconfirmatory (the less likely, or surprising, realization under the prior) signal realization.[30] As illustrated in Fig. 4b and consistent with Prediction 4, we observe more underreaction to the 'expected' confirmatory realizations (left side of green curve, right side of red curve) and less underreaction or even *overreaction* to the 'surprising' disconfirmatory realizations (left side of red curve, right side of green curve). This overreaction occurs even in the simple 2-state case. Table 4 Column 1 regresses the overreaction ratio on dummies for whether a signal realization was confirmatory or disconfirmatory, with neutral realizations in the uniform prior environment as the control. As shown in the table, participants overreact significantly more to disconfirmatory realizations and significantly less to confirmatory realizations. This comparative static continues to hold when controlling for the information structure (Column 2). Participants indeed appear to overreact more to surprising news compared to news that is expected.

[30]Under prior $(0.3, 0.7)$, a red ball is confirmatory and a blue ball is disconfirmatory, with the opposite under prior $(0.7, 0.3)$. Under a uniform prior, both red and blue balls are neutral realizations (neither confirmatory nor disconfirmatory).

Table 4. More overreaction to disconfirmatory realizations

| | Overreaction Ratio | |
|---|---|---|
| | (1) | (2) |
| Confirmatory Realization | -0.302*** | -0.253*** |
| | (0.0255) | (0.0268) |
| Disconfirmatory Realization | 0.443*** | 0.422*** |
| | (0.0474) | (0.0443) |
| $d = 0.7$ | | -0.404*** |
| | | (0.0542) |
| $d = 0.8$ | | -0.484*** |
| | | (0.0532) |
| $d = 0.9$ | | -0.464*** |
| | | (0.0543) |
| Constant | -0.116*** | 0.223*** |
| | (0.0219) | (0.0505) |
| $N$ | 2432 | 2432 |
| adj. $R^2$ | 0.148 | 0.206 |

Notes: Includes all information environments with two states listed in Table 9; excludes wrong direction updates. Baseline is uniform prior $(0.5, 0.5)$ and, in Column 2, diagnosticity $d = 0.6$. Standard errors clustered at the individual level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Finally, we explore the prediction that people are more likely to update in the opposite direction from the Bayesian benchmark, i.e., a wrong direction update, for confirmatory realizations relative to neutral or disconfirmatory realizations (Prediction 4.i). Fig. 5 presents the share of wrong direction updates for confirmatory, disconfirmatory and neutral signal realizations. Consistent with Prediction 4, we observe a significant difference in these frequencies. While wrong direction updates occur relatively infrequently following neutral and disconfirmatory realizations, they occur significantly more often following confirmatory realizations. In the latter case, nearly 30% of updates are in the wrong direction—almost three times higher than in the former cases. Importantly, this incidence of wrong direction updates is not arbitrary noise (e.g., inattentive subjects), but is predicted by our model as a function of the information environment and type of signal realization.

### 3.4 Structural Estimation

We next use the experimental data to estimate the two parameters of the belief-updating model, $\theta$ and $\lambda$. We first estimate aggregate parameters across all participants and then explore individual-level heterogeneity.

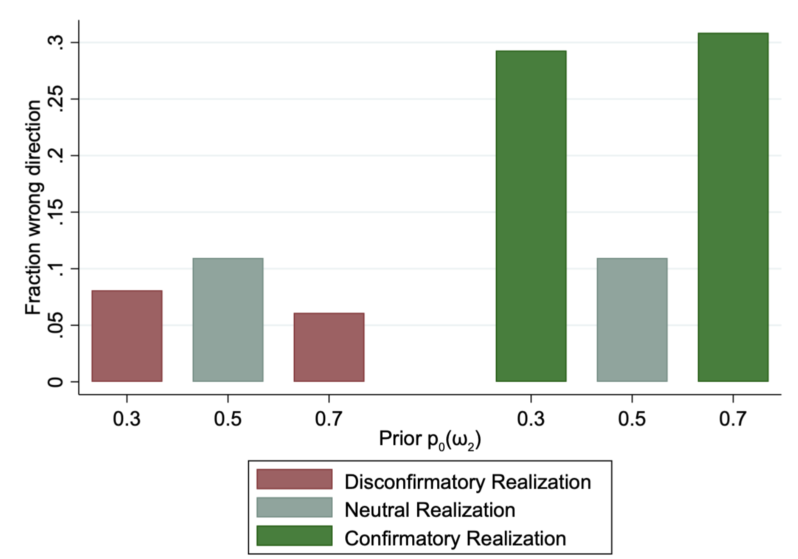



FIGURE 5. Higher share of wrong direction updates for confirmatory realizations

### 3.4.1 Aggregate-Level Estimation

We refer to the model-predicted posterior belief given parameter values $\theta$ and $\lambda$ as a *model prediction* and denote it by $\hat{p}_{\theta,\lambda}$ (see Eq. (7)). This prediction maps each information environment $(\Omega, p_0)$ and signal realization $s_i$ to a subjective posterior distribution $\hat{p}_{\theta,\lambda}(s_i; \Omega, p_0) \in \Delta(\Omega)$.

To estimate the parameters, we follow the literature on behavioral structural estimation (e.g., DellaVigna (2018) and Bordalo, Gennaioli, Ma, and Shleifer (2020)). We search a grid of parameters for the values that minimize the weighted sum of distances between the participants' reported posteriors and the model-predicted posteriors across all trials. We measure the distance between a reported posterior and a predicted posterior by the Kullback-Leibler (henceforth KL) divergence of the reported posterior from the predicted posterior.[31] This is a common measure of the statistical distance between two probability distributions. Since the KL divergence is undefined when $\hat{p}_{\theta,\lambda}(\omega_j|s_i; \Omega, p_0) = 0$, we restrict our analysis to information environments that generate predicted posteriors with full support on $\Omega$. Specifically, we include trials for all information environments listed in Table 9 except for the 11-state complexity.[32]

The estimated parameter values are $\theta = 0.85$ and $\lambda = 0.70$, as summarized in Table 5; both estimates are significantly different from the Bayesian benchmark of $\theta = 0$ and $\lambda = 1$. The estimates suggest that in the editing stage, directing attention

[31]The KL divergence of reported posterior $\hat{p}(s_i; \Omega, p_0)$ from predicted posterior $\hat{p}_{\theta,\lambda}(s_i; \Omega, p_0)$ is given by $\sum_{\omega_j \in \Omega} \hat{p}(\omega_j|s_i; \Omega, p_0) \log(\hat{p}(\omega_j|s_i; \Omega, p_0)/\hat{p}_{\theta,\lambda}(\omega_j|s_i; \Omega, p_0))$.

[32]An alternative distance measure is the quadratic difference between the expectation of the reported posterior and the predicted posterior. As a robustness check, in Appendix B.4 we estimate $\theta$ and $\lambda$ using this measure for the prediction loss function. We chose the KL divergence as our primary measure since it is independent of the values of the states, whereas the quadratic difference places a larger weight on higher states.

Table 5. Aggregate-level estimates of $\theta$ and $\lambda$

| | $\theta$ | 95% CI | $\lambda$ | 95% CI |
|---|---|---|---|---|
| Parameter Estimates | 0.85 | (0.82, 0.92) | 0.70 | (0.69, 0.70) |

Notes: Parameter estimates that minimize the average KL divergence at the aggregate level. Includes all information environments listed in Table 9, except for the 11-state complexity; excludes wrong direction updates. The 95% confidence intervals are obtained from 300 bootstrap samples.

to the representative states leads participants to update beliefs as-if they are counting the signal *nearly twice*. In the evaluation stage, participants' cognitive imprecision leads them to anchor on the cognitive default and adjust only 70% of the linear distance to the edited posterior.[33]

Our parameter estimates are qualitatively similar to others in the literature. Enke and Graeber (2023) estimate cognitive noise in a simple 2-state environment and obtain an estimate of $\lambda$ close to 0.5. Bordalo et al. (2019) examine forecasters' expectations about a series of economic indicators and find that $\theta$ ranges from 0.3 to 1.5, with an average of 0.6. It is noteworthy that we obtain a qualitatively similar value in a very different setting.

### 3.4.2 Individual-Level Estimation

Next, we estimate individual-level parameters for each participant. Although each participant was assigned to a single complexity treatment, we have sufficient data to estimate individual-level parameters for most participants ($N = 1546$) due to the variation in the prior and the information structure. We estimate the individual-level parameters in an analogous way to the aggregate estimates. For a given participant, we find the parameter values that minimize the average KL divergence of the participant's reported posteriors from the predicted posteriors across all her trials.

The results are presented in Fig. 6. Each point in the figure represents the parameter estimates for one participant. These estimates reveal significant heterogeneity across participants. Specifically, 70% of participants exhibit both the representativeness heuristic and cognitive imprecision, as characterized by estimates of $\theta > 0$ and $\lambda < 1$. Additionally, 9% of participants exhibit only cognitive imprecision ($\theta = 0$), 5% exhibit only representativeness ($\lambda = 1$), and the remaining 16% exhibit neither bias ($\theta = 0$ and $\lambda = 1$). The estimated values of $\theta$ and $\lambda$ exhibit a significant negative correlation, with a correlation coefficient of $-0.47$. The negative correlation implies that participants who are more prone to simplification through representativeness

[33] As a robustness check, in Appendix B.4 we estimate the parameters excluding information environments with an asymmetric prior. Our model predicts that cognitive noise may lead an agent to update in the wrong direction under an asymmetric prior (Prediction 4). Since we drop wrong direction updates in Table 5, this could potentially lead to an underestimation of $\lambda$. We find qualitatively similar parameter estimates in this robustness check.

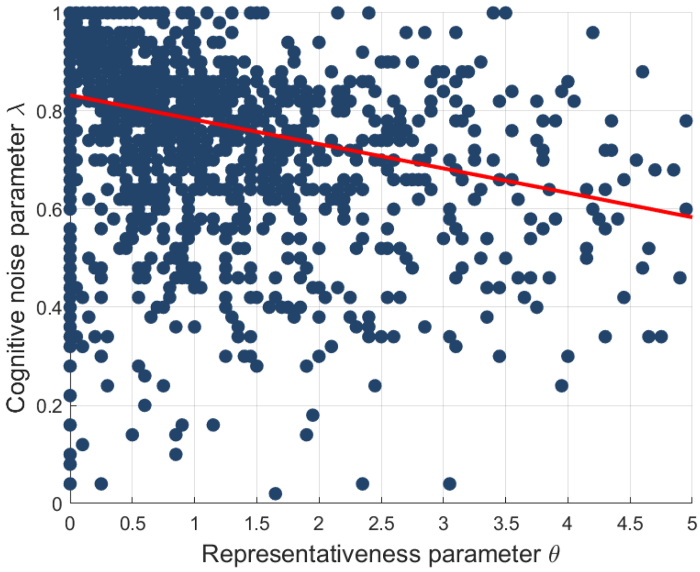



FIGURE 6. Individual level parameter estimates.

Notes: Parameter estimates that minimize the average KL divergence at the individual level. Includes all information environments listed in Table 9, except for the 11-state complexity; excludes wrong direction updates; excludes extreme estimates of $\theta$ larger than 5 (approx. 5.5% of sample).

(higher $\theta$) also tend to exhibit higher levels of cognitive imprecision (lower $\lambda$). This suggests that individual-level limits to cognitive capacity lead people to both engage in more simplification in the editing stage and noisier processing in the evaluation stage.

### 3.5 The Role of Attention

We next directly test the cognitive mechanism for our two-stage updating model. In the editing stage of our model, limits on attention and working memory lead agents to focus on representative states in complex learning environments. The framework thus predicts that (i) after observing the signal, agents' attention will be channeled bottom-up towards the state that is most representative of the signal realization, and (ii) any further limits on cognitive resources will exacerbate representativeness and lead to more overreaction.

To test these predictions, we employ the Mouselab paradigm of Payne et al. (1988), which is a commonly-used tool in cognitive psychology to study attention.[34] The Mouselab paradigm captures participants' attention to various features of the decision problem by the timing of the objects that they click on. For example, in a lottery choice task, participants are asked to click on the attributes of each gamble (e.g., the probability of winning each reward, the potential reward if a state is re-

[34]The Mouselab design, which has 2823 Google Scholar citations to date, has been used to study attention and information acquisition across a wide array of domains, from identifying decision strategies in consumer choice (Reisen, Hoffrage, and Mast 2008) to information search strategies in dynamic contexts (Callaway, Lieder, Krueger, and Griffiths 2017).

alized) before selecting a gamble. The first click is taken as a proxy for the feature that is attended to first, the second as a proxy for the feature that is attended to second, etc.[35] Research has also shown that the Mouselab paradigm, which requires participants to click on attributes, puts additional demands on cognitive attentional resources: while the ordering of clicks corresponds to ordering of attention, the process of clicking itself requires additional attention to implement (Meißner et al. 2010; Wolfe, Alvarez, and Horowitz 2000; Alvarez, Horowitz, Arsenio, DiMase, and Wolfe 2005).

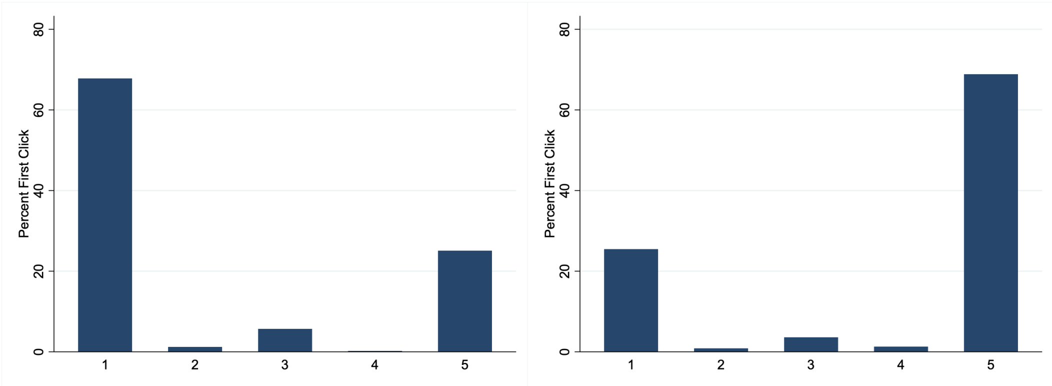
We used the Mouselab paradigm to measure the order in which participants clicked on the states as well as how the further attentional demands of the paradigm impacted belief-updating. This Limited Attention treatment required a participant to click on a state (e.g., Bag 5) before being able to enter her posterior belief about the state. Once a state was clicked, the participant could enter her belief for that state as before. As in the baseline treatment, the percentage assigned to each state had to sum to 100 and the order of states was randomized so that either the bag with the most red balls or the bag with the least red balls appeared first. We ran this Limited Attention treatment on all 5-state information environments listed in Table 9.

Two main predictions follow. First, participants will channel their attention and click on the representative state first. In other words, upon observing a blue (red) ball, the most likely first-click will be on the bag with the most blue (red) balls. Second, fixing the information environment, taxing attentional resources will increase overreaction in the Limited Attention treatment relative to the Baseline Attention treatment.

To examine the first prediction, Fig. 7 shows the distribution of first-clicks across all trials. Notably, even though the order of states was randomized, participants were much more likely to channel their attention—proxied by first-click—to the most representative state. The difference is stark: the representative state was three times more likely to be clicked first relative to the second-highest alternative ($p < .001$). The fact that the representative state varied with the realized signal and the random ordering rules out that this result is driven by an information-independent heuristic (e.g., always click on the left-most bag).

To examine the second prediction, Column 1 of Table 6 regresses the overreaction ratio on whether participants were in the Limited Attention or Baseline Attention version of the 5-state treatment. We find that overreaction was indeed significantly higher in the former than the latter. This is further illustrated in Fig. 8a, which shows that overreaction was higher in the Limited Attention treatment across nearly all signal diagnosticities.

[35]The use of click data as a proxy for channeled attention has been validated using eye-tracking tools (Meißner, Decker, and Pfeiffer 2010).




(A) State 1 Representative (B) State 5 Representative

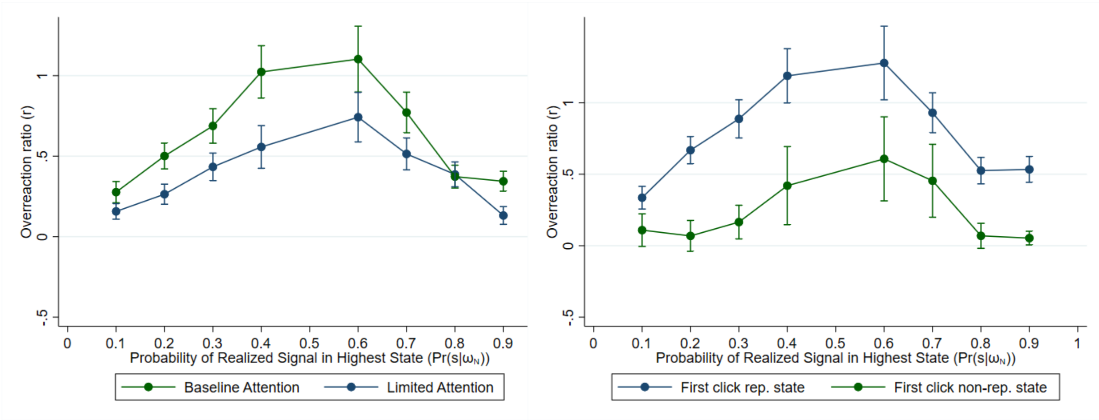
Figure 7. Most participants click on representative state first

We examine participant heterogeneity by looking at whether those who are more prone to the representativeness heuristic—proxied by their propensity to click the representative state first—also overreact more. Restricting attention to the Limited Attention treatment, Column 2 of Table 6 regresses the overreaction ratio on whether the participant clicked on the representative state versus a different state first. We find that the former group displayed significantly higher overreaction than the latter. This is illustrated in Fig. 8b, which shows that overreaction was substantially higher in the representative-state-first group across all signal diagnosticities. Taken together, these results support the two predictions outlined above, and provide further evidence against insensitivity and information-independent heuristics as alternative explanations for our results.

Table 6. Limited attention increases overreaction

| | Overreaction Ratio | |
|---|---|---|
| | (1) | (2) |
| Limited Attention | 0.179** | |
| | (0.0551) | |
| Click rep. state first | | 0.377*** |
| | | (0.0520) |
| Constant | 0.249*** | 0.156*** |
| | (0.0284) | (0.0458) |
| Observations | 4379 | 1740 |
| Adjusted $R^2$ | 0.012 | 0.036 |

Notes: Constant is the Baseline Attention treatment in Column 1 and first-click on a non-representative state in Column 2. Includes all information environments with five states listed in Table 9; excludes wrong direction updates. Standard errors clustered at the individual level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.




(A) Overreaction by attention treatment (B) Overreaction by first click

FIGURE 8. Limited attention increases overreaction

Finally, we provide further evidence for the proposed mechanism by structurally estimating the parameters in the Limited Attention treatment and comparing them to those obtained in the Baseline Attention treatment. Table 7 shows that the estimate of $\theta$ increases from 0.99 to 1.26, while the estimates of $\lambda$ are similar between the two treatments. This lends direct support to our prediction that decreasing attentional resources through the Mouselab paradigm increases reliance on the representativeness heuristic (as indicated by higher $\theta$), while leaving the level of cognitive imprecision unchanged.

TABLE 7. Limited attention increases representativeness $\theta$

| | $\theta$ | 95% CI | $\lambda$ | 95% CI |
|---|---|---|---|---|
| Limited Attention | 1.26 | (1.16, 1.38) | 0.74 | (0.72, 0.76) |
| Baseline Attention | 0.99 | (0.92, 1.08) | 0.73 | (0.72, 0.74) |

Notes: This table compares the parameter estimates that minimize the average KL divergence at the aggregate level for the limited attention treatment and the baseline attention treatment. Includes all information environments with 5 states listed in Table 9; excludes wrong direction updates. The 95% confidence intervals are obtained from 300 bootstrap samples.

## 4 Evaluating Model Performance

To evaluate the performance of our two-stage model of belief formation, we compute its completeness and restrictiveness following the methodology developed by Fudenberg et al. (2022, 2023). We then compare the performance of our model to a one-stage model of either only cognitive noise or only representativeness. We refer to these comparison models as the cognitive-noise-only model and representativeness-only model, respectively.

### 4.1 Completeness

We define model completeness as follows. Similar to the structural estimation in Section 3.4, we measure the prediction loss of a model by the KL divergence of the reported posterior from the predicted posterior. Let $e^B$ denote the expected prediction loss relative to the Bayesian prediction. Let $e^M$ denote the minimum expected loss relative to the prediction of model $M \in \{T, N, R\}$, where $M = T$ corresponds to our two-stage model, $M = N$ corresponds to the cognitive-noise-only model ($\theta = 0$), $M = R$ corresponds to the representativeness-only model ($\lambda = 1$), and the minimum is taken with respect to all feasible values of the model parameter(s). Finally, let $e^*$ denote the minimum expected loss relative to the best possible prediction. The completeness of model $M$ is given by

$$\kappa^M \equiv \frac{e^B - e^M}{e^B - e^*} \in [0, 1]. \tag{12}$$

That is, a model $M$ is 0% complete if it predicts no better than Bayesian updating and 100% complete if predicts as accurately as the best prediction.

Intuitively, completeness is a measure of how much of the explainable variation in data a model captures. It is distinct from the $R$-squared statistic typically reported for a regression analysis. As pointed out by Fudenberg et al. (2022), completeness measures whether a model captures regularities in the data, while $R$-squared captures the overall prediction error of the model, which could stem from either missing regularities or intrinsic, irreducible noise. A model could have high completeness but low $R$-squared—this would indicate that it successfully captures key regularities in the data but the environment is noisy.

Estimating completeness requires an estimate of $e^*$. As Fudenberg et al. (2022), we use ten-fold cross-validation to compute such an estimate. Estimates of $e^B$ and $e^M$ are straightforward to derive from the model and data. For this analysis, we do not exclude trials in which participants update in the wrong direction so as to capture the full extent of model fit to the data.

We first estimate completeness in the simple information environments with two states. As shown in Table 8, the cognitive-noise-only model ($M = N$) achieves essentially 100% completeness. In these simple environments, the addition of the editing stage does not yield any further improvement in model performance. This is consistent with our conjecture that the representativeness heuristic is used to respond to complexity, and therefore adds little explanatory power in simple environments.

However, increasing the complexity of the state space to three or more states decreases the completeness of the cognitive-noise-only model to a mere 36%. The representativeness-only model ($M = R$) also has little explanatory power in these more complex information environments. Yet taken together, the two-stage model with both psychological processes achieves a very high completeness—it captures

92% of the explainable variation in the data, relative to Bayes' rule. This shows that the two processes are critical *cognitive complements* in determining belief-updating in complex environments.

Taken together, while the cognitive-noise-only model effectively explains belief-updating in simple environments—potentially explaining its prominent role in organizing data from laboratory experiments that primarily use binary state spaces—the model's explanatory power declines rapidly in more complex settings. The interaction between representativeness and noisy cognition is key for our model's high explanatory power in complex environments.

Table 8. Completeness and Restrictiveness

| | Completeness | | Restrictiveness | |
|---|---|---|---|---|
| | 2 states | > 2 states | 2 states | > 2 states |
| Two-Stage Model | 1.00 | 0.92 | 0.73 | 0.91 |
| | (0.15) | (0.05) | (0.00) | (0.00) |
| Cognitive-noise-only Model | 1.00 | 0.36 | 0.76 | 0.97 |
| | (0.06) | (0.02) | (0.00) | (0.00) |
| Representativeness-only Model | 0.00 | 0.00 | 1.00 | 1.00 |
| | (0.15) | (0.04) | (0.00) | (0.00) |

Notes: Includes all information environments listed in Table 9 except for the 11-state complexity; includes wrong direction updates. Restrictiveness is estimated from 1000 simulations.

### 4.2 Restrictiveness

While our two-stage model has high completeness, the inclusion of an additional parameter could make the model so flexible that it could explain almost any dataset. To rule this out, we next estimate a measure of the two-stage model's restrictiveness using randomly generated synthetic belief data. We then compare the average prediction loss of the two-stage model on the synthetic dataset to the average prediction loss of Bayes' rule on this dataset. Intuitively, the model is too flexible if it has a good fit on the synthetic data relative to Bayes' rule.

Following Fudenberg et al. (2023), we randomly generate 1000 mappings, where each mapping assigns a posterior distribution over the state space to each information environment from our experimental set (see Table 9) and each signal realization $s_i \in \{b, r\}$. We draw mappings uniformly from an 'admissible' set of mappings that satisfy basic directional and monotonicity properties.[36] These properties hold for Bayes' rule and other common models of belief-updating. We impose such properties to ensure that our synthetic data is 'reasonable' belief data—without such restrictions

[36]For example, we require mappings to satisfy the property that the posterior probability of a state weakly increases in the signal diagnosticity of that state. At a more basic level, we require each posterior distribution in the mapping to in fact be a probability distribution, i.e., it assigns a number between 0 and 1 to each state and sums to one across states.

on the admissible set, any model that satisfies such basic properties could have high restrictiveness on a synthetic dataset, even if it is in fact quite flexible. Evaluating the restrictiveness of a model with respect to this 'admissible' synthetic data provides a sense of the additional restrictions on belief-updating imposed by the model.

Let $d^B$ denote the expected distance of the synthetic mapping from the Bayesian prediction, where distance is measured by the KL divergence and the expectation is taken with respect to the uniform distribution over the admissible set. Analogously, let $d^M$ denote the minimal expected distance of the synthetic mapping from the prediction of model $M$, where the minimum is taken with respect to the parameter(s) of model $M$. The restrictiveness of model $M$ is defined by the ratio of these two expected distances,

$$\rho^M \equiv \frac{d^M}{d^B} \in [0, 1]. \tag{13}$$

That is, a model is 0% restrictive if it fits synthetic data perfectly—the KL divergence of the synthetic mapping from the best fit of the model is zero—and 100% restrictive if it fits synthetic data no better than Bayes' rule—the KL divergence of the synthetic mapping from the best fit of the model is equal to the KL divergence of the synthetic mapping from Bayes' rule.

As indicated in Table 8, the two-stage model has high restrictiveness in simple information environments with two states (0.73), and very high restrictiveness in complex information environments with more than two states (0.91). Moreover, it has similar restrictiveness to the cognitive-noise-only model (0.73 versus 0.76 for simple environments and 0.91 versus 0.97 for more complex environments). This shows that the substantially higher explanatory power of the two-stage model relative to the cognitive-noise-only model does not come at the expense of a significant increase in flexibility.

While the representativeness-only model is more restrictive than the two-stage model in simple environments (1.00 versus 0.73), it is also very incomplete relative to the two-stage model. In complex environments, it has similar restrictiveness to the two-stage model (1.00 versus 0.91), but still features very low completeness. Therefore, although the representativeness-only model is as restrictive as Bayes' rule, it also adds little explanatory power relative to Bayes' rule. In contrast, the two-stage model has both high restrictiveness and high completeness—it is almost as restrictive as Bayes' rule while adding significant explanatory power relative to Bayes' rule.

To visualize the trade-off between explanatory power and flexibility, Fig. 9 plots the completeness and restrictiveness of the two-stage model, the cognitive-noise-only model and the representativeness-only model in complex information environments with more than two states. As the figure illustrates, once we go beyond a simple environment with two states, incorporating responses to complexity into a model of

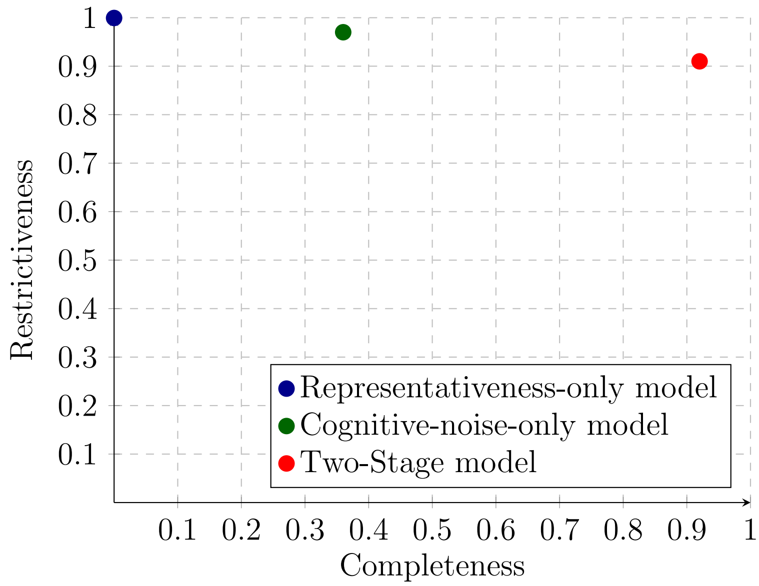



FIGURE 9. Completeness-restrictiveness trade-off ($> 2$ states)

belief-updating leads to a striking increase in explanatory power while only minimally increasing the model's flexibility.

## 5 Under- and Overreaction in Prior Work

In this section, we relate our findings to the theoretical and empirical literature on under- and overreaction. We primarily focus on settings where agents observe one signal draw, but also briefly discuss settings where agents observe multiple draws.

**Laboratory studies.** The key contribution of our paper is to explicitly consider how complexity of the information environment impacts belief-updating. As previously noted, the vast majority of laboratory experiments focus on a simple 2-state setting. Benjamin (2019) presents a meta-analysis of experiments with a binary state space, symmetric signal diagnosticity, and uniform prior, and finds that people generally underreact to information.

There are several noteworthy studies that do use more than two 'bags' in the design. Phillips and Edwards (1966) conduct 'bookbag-and-poker-chip' experiments in which the number of bags is increased to 10. However, there are only two unique states: each bag of $N$ chips has either $x$ red chips or $N - x$ red chips, with the remaining chips blue. Thus, this experiment is equivalent to varying the prior rather than expanding the state space. Consistent with our prediction, they predominantly find underreaction. Hartzmark, Hirshman, and Imas (2021) explore how people learn about owned versus non-owned goods. Their design features a uniform prior and 8 states, where each state is associated with a distinct signal distribution. Consistent with our framework, the authors document overreaction. But they do not explore how the size of the state space impacts the level of overreaction—their focus is on differences in belief-updating as a function of ownership. Prat-Carrabin and Woodford

(2022) find underreaction in an environment with a continuous state space $[0, 1]$ and uniform prior. Relating this result to our complexity predictions requires a model of how complexity is perceived for an uncountable state space. For example, participants may partition the state space into a finite set of intervals, with complexity corresponding to the cardinality of the partition. A partition into states that are greater or less than 0.5 would have the same complexity as a binary state space in our framework, predicting underreaction. A continuous state space may also prompt a different cognitive default. To test for this possibility, we ran a study that elicited the cognitive default in a 'continuous' version of our setting ($N = 100$).[37] Indeed, in contrast to a discrete state space, participants reported a cognitive default that placed substantially more weight on middle states relative to extreme states—similar to a (truncated) normal distribution. This difference in cognitive defaults could explain the underreaction they found in the continuous state space setting versus the overreaction we find in complex discrete state space settings.

In recent related work, Fan, Liang, and Peng (2023) show that people underinfer when making inferences after observing information and overinfer when forming forecasts about future information. Their main treatment featured a binary state space (a firm was either good or bad) and a discretized normal signal distribution (the firm's stock price growth this month). Over half of participants underreacted when asked to report their posterior about the firm's state and over half overreacted when asked to report their prediction of the next signal (the stock price growth next month). Our framework predicts such an inference-forecast gap in their environment: the state space is binary in the inference task but more complex in the forecasting task, since it involves considering a large space of potential signal realizations.[38] In addition, the current signal is more informative about the state than about the next-period signal, which also predicts more underreaction in the inference task relative to the forecasting task. Afrouzi, Kwon, Landier, Ma, and Thesmar (2023) also find overreaction in an experiment where the forecast variable has a complex state space.

Researchers have also studied how changes in signal diagnosticity affect belief-updating. Consistent with our predictions and empirical results, several studies have found that people exhibit greater underreaction to more precise signals. Edwards (1968) ran studies with a binary state space, uniform prior, and symmetric information structures with signal diagnosticities $d_i \in \{0.55, 0.7, 0.85\}$. When the signal was less precise ($d_i = 0.55$), subjects exhibited overreaction; as the diagnosticity increased, they exhibited more underreaction.[39] Kieren and Weber (2020) find un-

[37] The state space consisted of a set of bags ordered along the unit interval, where the state corresponded to the probability of drawing a red ball. We used the same method as in Section 3 to elicit the cognitive default.

[38] As discussed further below, the inherent complexity in forecasting can explain the prevalence of overreaction in that domain more generally.

[39] Similar patterns are documented in Phillips and Edwards (1966); Peterson, Schneider, and Miller (1965); Kahneman and Tversky (1972); Grether (1992); Holt and Smith (2009); Benjamin

derreaction to informative signals and overreaction to uninformative signals. Recent work by Augenblick et al. (2022) argues that this comparative static is consistent with a model of noisy cognition. Their paper complements our framework by extending the way in which cognitive noise can generate overreaction. They consider a simple two-state setting where the agent forms a noisy representation of the signal diagnosticity, and show that this predicts underreaction to precise signals and overreaction to sufficiently noisy signals. Our model generates the same comparative static on diagnosticity, but it stems from both representativeness and cognitive imprecision.

Our results also relate to findings on how the prior impacts belief-updating. A large body of work has shown that people are generally insensitive to base rates (e.g., Kahneman and Tversky (1973); Green, Halbert, and Robinson (1965); Grether (1992); Robalo and Sayag (2018)). However, as outlined in Prediction 4, whether base-rate neglect generates under- or overreaction depends on whether the signal realization is confirmatory or disconfirmatory. Holt and Smith (2009) vary the prior in a 2-state setting. In line with our findings, they show that when the prior is more asymmetric and a disconfirmatory realization is observed, people overreact; in contrast, following a confirmatory realization or under a more symmetric prior, people underreact. Kieren, Müller-Dethard, and Weber (2022) find that investors systematically overreact to disconfirmatory information in both experiments and financial market data.

A line of work explores belief-updating when agents observe multiple signals drawn from the same distribution. Griffin and Tversky (1992) find that people focus too much on the strength of evidence (e.g., sample proportions of each signal realization) and not enough on the weight (e.g., sample size). Massey and Wu (2005) find that people tend to neglect the possibility of a regime shift in a setting where the signal distribution probabilistically changes across time. This leads to under- or overreaction depending on the probability of a shift and the precision of the signal. Observing multiple signal draws introduces additional channels of bias that are outside of our framework. In future work, it would be interesting to explore how simplification heuristics and cognitive imprecision interact in such dynamic environments.

Our paper contributes to the theoretical literature that seeks to explain the prevalence of underreaction in laboratory studies. Phillips and Edwards (1966) propose that people suffer from *conservatism* bias: they underweight the likelihood ratio of the signal, which leads to underreaction. Benjamin, Rabin, and Raymond (2016) propose that people have *extreme-belief aversion*, i.e., an aversion to holding beliefs close to certainty. As pointed out by DuCharme (1970), both conservatism and extreme-

(2019). When the information structure is asymmetric, a similar pattern holds: agents tend to overreact when diagnosticities are close together (and thus close to 0.5) and underreact when they are further apart. See Peterson et al. (1965); Ambuehl and Li (2018).

belief aversion can lead to underreaction when the signal is precise. As discussed in Section 2, a model of noisy cognition also predicts underreaction (Woodford 2020).[40]

**Financial markets.** A growing literature in finance and macroeconomics uses surveys and forecasts by professionals and households to study departures from rational expectations (see Bordalo et al. (2022) for review). A common approach is to examine the predictability of forecast errors from forecast revisions (Coibion and Gorodnichenko 2015).[41] In contrast to the experimental findings, this research typically finds that people overreact to information. For example, Bordalo et al. (2020) analyze time series data on a large group of financial and macro variables and individual forecasts from professionals. They find that forecasts for the vast majority of these variables exhibit overreaction.[42] d'Arienzo (2020) and Wang (2021) find that individual analysts' forecasts of long-term interest rates exhibit overreaction. Bordalo et al. (2019) find overreaction in the expectations of long-term corporate earnings growth.

A workhorse theory in the financial literature is the diagnostic expectations model, where agents overreact to information due to a reliance on the representativeness heuristic (Bordalo et al. 2019, 2020). For example, Kwon and Tang (2021) show that such a model can explain overreaction to extreme corporate events and underreaction to non-extreme events. Our two-stage model incorporates the underlying psychology of the diagnostic expectations model into the 'evaluation' stage.

Our results can potentially reconcile the seemingly contradictory findings in the lab versus observational data. A prominent feature of real-world settings is that decision-makers tend to face much more complex information environments and noisier signals than in the lab. Consistent with the empirical results, our framework thus predicts that we should expect overreaction in these real-world settings. On the other hand, as noted above, laboratory studies tend to focus on simple binary state spaces and relatively informative signals. Again consistent with the findings in this literature, our framework predicts that we should see underreaction in these simple environments.

One important thing to note is that we focus on studies that collect belief data (either by eliciting them directly or through forecasts and surveys). A related literature starting with Ball and Brown (1968) and De Bondt and Thaler (1985) has

[40] A similar reduced form updating rule is found in Epstein, Noor, Sandroni et al. (2010), which considers the implication of underreaction on asymptotic learning.

[41] Augenblick and Rabin (2021) develop an alternative statistical test of under- and overreaction by exploiting the equivalence between the expected movement in beliefs and the expected uncertainty reduction for Bayesian learners. Greater (lesser) actual belief movement, relative to uncertainty reduction, is indicative over- (underreaction).

[42] In addition to identifying overreaction in individual forecasts, Bordalo et al. (2020) also document underreaction in consensus forecasts. They explain this underreaction with a model in which forecasters do not respond to other forecasters' information. The underreaction we identify differs in that it stems from cognitive noise at the individual level rather than a lack of information integration across forecasters.

examined under- and overreaction by looking at choice data—specifically, price movements. Prices have been found to adjust slowly to firm-specific (Ball and Brown 1968) and macro (Klibanoff et al. 1998) announcements, and to display short-term autocorrelation (i.e., momentum); these effects have been interpreted as underreaction (Hirshleifer, Lim, and Teoh 2009; Daniel et al. 1998). Prices also display long-term negative autocorrelation, which has been interpreted as overreaction. However, it is not clear whether price responses are driven by preferences or beliefs. For example, Frazzini (2006) shows that the slow price adjustment to earnings announcements—the famous post-earnings announcement drift (PEAD)—is consistent with the disposition effect, which has been explained through prospect theory preferences (Barberis 2012; Heimer, Iliewa, Imas, and Weber 2021). Charles, Frydman, and Kilic (2023) show that noisy cognition can weaken the link between beliefs and behavior, such that overreaction in the former can still generate underreaction in the latter. Since our paper focuses on belief-updating, we do not attempt to apply our framework to behavior.

## 6 Conclusion

This paper examines the incidence and underlying drivers of under- and overreaction. A key contribution of our framework is the two-stage model of belief-updating, which allows for the interaction between multiple psychological mechanisms. We empirically show that representativeness and cognitive noise are cognitive compliments and their interaction plays a crucial role in explaining how agents update beliefs across learning environments. While the majority of papers in psychology and behavioral economics have focused on identifying the implications of a single psychological mechanism, it is likely the case that observed judgments and choice are the product of multiple mechanisms. Our results show that heuristics do not just operate independently but also reinforce one another in important ways. This suggests that modeling and testing more 'unified' frameworks across economically-important domains is a fruitful area for further research.

Another contribution of our framework is explicitly consider the complexity of the learning environment as an important determinant of belief-updating. We empirically show that complexity leads agents to simplify the information structure, which impacts the form of bias that emerges. While we focus on the complexity of the state space, other aspects of the learning environment—such as the signal space or number of signal draws—can also vary in complexity. This suggests that modeling and testing how agents simplify other types of complexity when interpreting and using information is an important area for future research.

## References

Afrouzi, H., S. Y. Kwon, A. Landier, Y. Ma, and D. Thesmar (2023): "Overreaction in Expectations: Evidence and Theory," *The Quarterly Journal of*

*Economics*, 138, 1713–1764.

Alvarez, G. A., T. S. Horowitz, H. C. Arsenio, J. S. DiMase, and J. M. Wolfe (2005): "Do Multielement Visual Tracking and Visual Search Draw Continuously on the Same Visual Attention resources?" *Journal of Experimental Psychology: Human Perception and Performance*, 31, 643.

Ambuehl, S. and S. Li (2018): "Belief Updating and the Demand for Information," *Games and Economic Behavior*, 109, 21–39.

Angrisani, M., A. Guarino, P. Jehiel, and T. Kitagawa (2020): "Information Redundancy Neglect versus Overconfidence: A Social Learning Experiment," *American Economic Journal: Microeconomics*.

Augenblick, N., E. Lazarus, and M. Thaler (2022): "Overinference from Weak Signals and Underinference from Strong Signals," *Available at SSRN 4315007*.

Augenblick, N. and M. Rabin (2021): "Belief Movement, Uncertainty Reduction, and Rational Updating," *The Quarterly Journal of Economics*, 136, 933–985.

Azeredo da Silveira, R. and M. Woodford (2019): "Noisy Memory and Over-reaction to news," in *AEA Papers and Proceedings*, vol. 109, 557–61.

Ball, R. and P. Brown (1968): "An Empirical Evaluation of Accounting Income Numbers," *Journal of Accounting Research*, 159–178.

Banovetz, J. and R. Oprea (2020): "Complexity and Procedural Choice," *American Economic Journal: Microeconomics*.

——— (2023): "Complexity and procedural choice," *American Economic Journal: Microeconomics*, 15, 384–413.

Barberis, N. (2012): "A Model of Casino Gambling," *Management Science*, 58, 35–51.

Barberis, N., A. Shleifer, and R. Vishny (1998): "A Model of Investor Sentiment," *Journal of Financial Economics*, 49, 307–343.

Bays, P. M., N. Gorgoraptis, N. Wee, L. Marshall, and M. Husain (2011): "Temporal Dynamics of Encoding, Storage, and Reallocation of Visual Working Memory," *Journal of Vision*, 11, 6–6.

Benjamin, D. J. (2019): "Errors in Probabilistic Reasoning and Judgment Biases," *Handbook of Behavioral Economics: Applications and Foundations 1*, 2, 69–186.

Benjamin, D. J., M. Rabin, and C. Raymond (2016): "A Model of Nonbelief in the Law of Large Numbers," *Journal of the European Economic Association*, 14, 515–544.

Bohren, J. A. and D. N. Hauser (2021): "Learning with Heterogeneous Misspecified Models: Characterization and Robustness," *Econometrica*, 89, 3025–3077.

Bordalo, P., K. Coffman, N. Gennaioli, F. Schwerter, and A. Shleifer (2021): "Memory and Representativeness," *Psychological Review*, 128, 71.

Bordalo, P., K. Coffman, N. Gennaioli, and A. Shleifer (2016): "Stereotypes," *The Quarterly Journal of Economics*, 131, 1753–1794.

Bordalo, P., J. Conlon, N. Gennaioli, S. Kwon, and A. Shleifer (2023): "How People Use Statistics," Working paper.

Bordalo, P., N. Gennaioli, Y. Ma, and A. Shleifer (2020): "Overreaction in Macroeconomic Expectations," *American Economic Review*, 110, 2748–82.

Bordalo, P., N. Gennaioli, R. L. Porta, and A. Shleifer (2019): "Diagnostic Expectations and Stock Returns," *The Journal of Finance*, 74, 2839–2874.

Bordalo, P., N. Gennaioli, and A. Shleifer (2022): "Overreaction and Diagnostic Expectations in Macroeconomics," *Journal of Economic Perspectives*, 36, 223–44.

Callaway, F., F. Lieder, P. Krueger, and T. L. Griffiths (2017): "Mouselab-MDP: A New Paradigm for Tracing How People Plan," *OSF Preprints*.

Charles, C., C. Frydman, and M. Kilic (2023): "Insensitive Investors," *Available at SSRN 3971082*.

Coibion, O. and Y. Gorodnichenko (2015): "Information Rigidity and the Expectations Formation Process: A Simple Framework and New Facts," *American Economic Review*, 105, 2644–78.

Daniel, K., D. Hirshleifer, and A. Subrahmanyam (1998): "Investor Psychology and Security Market Under-and Overreactions," *the Journal of Finance*, 53, 1839–1885.

Danz, D., L. Vesterlund, and A. J. Wilson (2022): "Belief Elicitation and Behavioral Incentive Compatibility," *American Economic Review*.

d'Arienzo, D. (2020): "Maturity Increasing Overreaction and Bond Market Puzzles," *Available at SSRN 3733056*.

De Bondt, W. F. and R. Thaler (1985): "Does the Stock Market Overreact?" *The Journal of Finance*, 40, 793–805.

DellaVigna, S. (2018): "Structural Behavioral Economics," in *Handbook of Behavioral Economics: Applications and Foundations 1*, Elsevier, vol. 1, 613–723.

DuCharme, W. M. (1970): "Response Bias Explanation of Conservative Human Inference," *Journal of Experimental Psychology*, 85, 66.

Edwards, W. (1968): "Conservatism in Human Information Processing," *Formal Representation of Human Judgment*.

Enke, B. and T. Graeber (2023): "Cognitive Uncertainty," Working paper, National Bureau of Economic Research.

Enke, B., T. Graeber, and R. Oprea (2023): "Confidence, Self-selection, and Bias in the Aggregate," *American Economic Review*, 113, 1933–1966.

Epstein, L. G., J. Noor, A. Sandroni, et al. (2010): "Non-Bayesian Learning," *The BE Journal of Theoretical Economics*, 10, 1–20.

Evers, E. R., A. Imas, and C. Kang (2022): "On the Role of Similarity in Mental Accounting and Hedonic Editing," *Psychological Review*, 129, 777.

Fan, T. Q., Y. Liang, and C. Peng (2023): "The Inference-Forecast Gap in Belief Updating," *Available at SSRN 3889069*.

Frazzini, A. (2006): "The Disposition Effect and Underreaction to News," *The Journal of Finance*, 61, 2017–2046.

Frydman, C. and L. J. Jin (2022): "Efficient Coding and Risky Choice," *The Quarterly Journal of Economics*, 137, 161–213.

Fudenberg, D., W. Gao, and A. Liang (2023): "How Flexible is that Functional Form? Quantifying the Restrictiveness of Theories," *Available at SSRN 3580408*.

Fudenberg, D., J. Kleinberg, A. Liang, and S. Mullainathan (2022): "Measuring the Completeness of Economic Models," *Journal of Political Economy*, 130, 956–990.

Gabaix, X. (2019): "Behavioral Inattention," in *Handbook of Behavioral Economics: Applications and Foundations 1*, Elsevier, vol. 2, 261–343.

Gabaix, X. and D. Laibson (2017): "Myopia and Discounting," Working Paper 23254, National Bureau of Economic Research.

Gennaioli, N. and A. Shleifer (2010): "What Comes to Mind," *The Quarterly Journal of Economics*, 125, 1399–1433.

Green, D. M., J. A. Swets, et al. (1966): *Signal Detection Theory and Psychophysics*, vol. 1, Wiley New York.

Green, P. E., M. H. Halbert, and P. J. Robinson (1965): "An Experiment in Probability Estimation," *Journal of Marketing Research*, 2, 266–273.

Grether, D. M. (1980): "Bayes Rule as a Descriptive Model: The Representativeness Heuristic," *The Quarterly Journal of Economics*, 95, 537–557.

——— (1992): "Testing Bayes Rule and the Representativeness Heuristic: Some Experimental Evidence," *Journal of Economic Behavior & Organization*, 17, 31–57.

Griffin, D. and A. Tversky (1992): "The Weighing of Evidence and the Determinants of Confidence," *Cognitive Psychology*, 24, 411–435.

Hartzmark, S. M., S. D. Hirshman, and A. Imas (2021): "Ownership, Learning, and Beliefs," *The Quarterly Journal of Economics*, 136, 1665–1717.

Heimer, R., Z. Iliewa, A. Imas, and M. Weber (2021): "Dynamic Inconsistency in Risky Choice: Evidence from the Lab and Field," *Available at SSRN 3600583*.

Hirshleifer, D., S. S. Lim, and S. H. Teoh (2009): "Driven to Distraction: Extraneous Events and Underreaction to Earnings News," *The Journal of Finance*, 64, 2289–2325.

Holt, C. A. and A. M. Smith (2009): "An Update on Bayesian Updating," *Journal of Economic Behavior & Organization*, 69, 125–134.

Kahneman, D. (2003): "A Perspective on Judgment and Choice: Mapping Bounded Rationality." *American Psychologist*, 58, 697.

Kahneman, D. and A. Tversky (1972): "Subjective Probability: A Judgment of Representativeness," *Cognitive Psychology*, 3, 430–454.

——— (1973): "On the Psychology of Prediction." *Psychological Review*, 80, 237.

Kendall, C. and R. Oprea (2021): "On the Complexity of Forming Mental Models," Working paper.

Khaw, M. W., Z. Li, and M. Woodford (2021): "Cognitive Imprecision and Small-Stakes Risk Aversion," *The Review of Economic Studies*, 88, 1979–2013.

——— (2022): "Cognitive Imprecision and Stake-Dependent Risk Attitudes," Working Paper 30417, National Bureau of Economic Research.

Kieren, P., J. Müller-Dethard, and M. Weber (2022): "Can Agents Add and Subtract when Forming Beliefs? Evidence from the Lab and Field," *Proceedings of Paris December 2021 Finance Meeting EUROFIDAI - ESSEC*.

Kieren, P. and M. Weber (2020): "Expectation Formation under Uninformative Signals," *Available at SSRN 3971733*.

Klibanoff, P., O. Lamont, and T. A. Wizman (1998): "Investor Reaction to Salient News in Closed-End Country Funds," *The Journal of Finance*, 53, 673–699.

Kwon, S. Y. and J. Tang (2021): "Extreme Events and Overreaction to News," *Available at SSRN 3724420*.

Loewenstein, G. and Z. Wojtowicz (2023): "The Economics of Attention," *Available at SSRN 4368304*.

Luck, S. J. and E. K. Vogel (1997): "The Capacity of Visual Working Memory for Features and Conjunctions," *Nature*, 390, 279–281.

Massey, C. and G. Wu (2005): "Detecting Regime Shifts: The Causes of Under- and Overreaction," *Management Science*, 51, 932–947.

Meissner, M., R. Decker, and J. Pfeiffer (2010): "Ein empirischer Vergleich der Prozessaufzeichnungsmethoden: Mouselab und Eyetracking bei Präferenzmessungen mittels Choice-based Conjoint Analyse," *Marketing ZFP-Journal of Research and Management*, 32, 135–144.

Milgrom, P. R. (1981): "Good News and Bad News: Representation Theorems and Applications," *The Bell Journal of Economics*, 380–391.

Molavi, P. (2022): "Simple Models and Biased Forecasts," *arXiv preprint arXiv:2202.06921*.

Oberauer, K., S. Farrell, C. Jarrold, and S. Lewandowsky (2016): "What Limits Working Memory Capacity?" *Psychological Bulletin*, 142, 758.

Oprea, R. (2020): "What Makes a Rule Complex?" *American Economic Review*, 110, 3913–51.

——— (2022): "Simplicity Equivalents," Working paper.

Payne, J. W., J. R. Bettman, and E. J. Johnson (1988): "Adaptive Strategy Selection in Decision Making," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14, 534.

——— (1993): *The Adaptive Decision Maker*, Cambridge University Press.

Peterson, C. R., R. J. Schneider, and A. J. Miller (1965): "Sample Size and the Revision of Subjective Probabilities," *Journal of Experimental Psychology*, 69, 522.

Phillips, L. D. and W. Edwards (1966): "Conservatism in a Simple Probability Inference Task," *Journal of Experimental Psychology*, 72, 346.

Prat-Carrabin, A. and M. Woodford (2022): "Imprecise Probabilistic Inference from Sequential Data," *PsyArXiv Preprint*.

Reisen, N., U. Hoffrage, and F. W. Mast (2008): "Identifying Decision Strategies in a Consumer Choice Situation," *Judgment and Decision Making*, 3, 641–658.

Robalo, P. and R. Sayag (2018): "Paying is Believing: The Effect of Costly Information on Bayesian Updating," *Journal of Economic Behavior & Organization*, 156, 114–125.

Thurstone, L. L. (1927): "A Law of Comparative Judgment." *Psychological Review*, 101, 266.

Tversky, A. and D. Kahneman (1974): "Judgment under Uncertainty: Heuristics and Biases: Biases in judgments Reveal Some Heuristics of Thinking under Uncertainty." *Science*, 185, 1124–1131.

——— (1983): "Extensional versus Intuitive Reasoning: The Conjunction Fallacy in Probability Judgment," *Psychological review*, 90, 293.

Wang, C. (2021): "Under-and Overreaction in Yield Curve Expectations," *Available at SSRN 3487602*.

Wolfe, J. M., G. A. Alvarez, and T. S. Horowitz (2000): "Attention is Fast but Volition is Slow," *Nature*, 406, 691–691.

Woodford, M. (2020): "Modeling Imprecision in Perception, Valuation, and Choice," *Annual Review of Economics*, 12, 579–601.

## A Proofs

**Proof of Claim in Footnote 14.** We show that our definition of overreaction based on Definition 1 is equivalent to the binary state definition stated in Footnote 14. Fix any signal realization $s_i$. Note that

$$\begin{aligned}\left|\hat{E}\left(\omega|s_i\right)-E\left(\omega\right)\right| &= \left|\omega_2\left(\hat{p}\left(\omega_2|s_i\right)-p\left(\omega_2\right)\right)+\omega_1\left(\hat{p}\left(\omega_1|s_i\right)-p\left(\omega_1\right)\right)\right| \\ &= \left|\omega_2\left(p\left(\omega_1\right)-\hat{p}\left(\omega_1|s_i\right)\right)+\omega_1\left(\hat{p}\left(\omega_1|s_i\right)-p\left(\omega_1\right)\right)\right| \\ &= \left|\omega_2-\omega_1\right|\cdot\left|\hat{p}\left(\omega_1|s_i\right)-p\left(\omega_1\right)\right|,\end{aligned}$$

where $\hat{p}$ is the subjective posterior following signal realization $s_i$, and similarly

$$\left|E\left(\omega|s_i\right)-E\left(\omega\right)\right|=\left|\omega_2-\omega_1\right|\cdot\left|p\left(\omega_1|s_i\right)-p\left(\omega_1\right)\right|,$$

where $p$ is the objective posterior following signal realization $s_i$. Hence,

$$\begin{aligned} r(s_i) =& \frac{|\hat{E}(\omega|s_i) - E(\omega)| - |E(\omega|s_i) - E(\omega)|}{|E(\omega|s_i) - E(\omega)|} \\ =& \frac{|\hat{p}(\omega_1|s_i) - p(\omega_1)| - |p(\omega_1|s) - p(\omega_1)|}{|p(\omega_1|s_i) - p(\omega_1)|}. \end{aligned}$$

That is, $r(s_i) > 0$ if and only if $|\hat{p}(\omega_1|s_i) - p(\omega_1)| > |p(\omega_1|s_i) - p(\omega_1)|$, and similarly for $\omega_2$.

**Proof of Prediction 1.** Suppose the signal realization is $s_2$. The objective posterior of any state $\omega_i \in \Omega$ is

$$p(\omega_i|s_2) = \frac{p_0(\omega_i)\omega_i}{\sum_{\omega_j \in \Omega} p_0(\omega_j)\omega_j} = \frac{2\omega_i}{N}$$

We can write the Bayesian expected state as

$$E(\omega|s_2) = \sum_{\omega_i \in \Omega} p(\omega_i|s_2)\omega_i = \frac{2}{N} \sum_{\omega_i \in \Omega} \omega_i^2$$

Suppose $\Omega$ contains an even number of states and $N = 2K$, then

$$\begin{aligned} E(\omega|s_2) - E(\omega) &= \frac{2}{N} \sum_{\omega_i \in \Omega} \omega_i^2 - \frac{1}{2} \\ &= \frac{2}{N} \left[ (1 - \omega_N)^2 + ... + (1 - \omega_{K+1})^2 + \omega_{K+1}^2 + ... + \omega_N^2 - \frac{K}{2} \right] \\ &= \frac{4}{N} \left[ \left( \omega_{K+1} - \frac{1}{2} \right)^2 + ... + \left( \omega_N - \frac{1}{2} \right)^2 \right]. \end{aligned}$$

When $\Omega$ contains an odd number of states and $N = 2K - 1$, symmetry implies that the $K$th state must be $\frac{1}{2}$. We therefore obtain the same expression for $E(\omega|s_2) - E(\omega)$. On the other hand,

$$E_R(\omega|s_2) = \sum_{\omega_i \in \Omega} p_R(\omega_i|s_2)\omega_i = \sum_{\omega_i \in \Omega} \frac{p_0(\omega_i)\omega_i^{\theta+2}}{\sum_{\omega_j \in \Omega} p_0(\omega_j)\omega_j^{\theta+1}} = \frac{\sum_{\omega_i \in \Omega} \omega_i^{\theta+2}}{\sum_{\omega_i \in \Omega} \omega_i^{\theta+1}}.$$

Note that $E_R(\omega|s_2)$ converges to the most representative state as $\theta$ goes to infinity. That is, $\lim_{\theta \to \infty} E_R(\omega|s_2) = \omega_N$. It follows that

$$\begin{aligned} \lim_{\theta \to \infty} r_R(s_2) + 1 &= \lim_{\theta \to \infty} \frac{|E_R(\omega|s_2) - E(\omega)|}{|E(\omega|s_2) - E(\omega)|} \\ &= \frac{\omega_N - \frac{1}{2}}{\frac{4}{N} \left[ \left( \omega_{K+1} - \frac{1}{2} \right)^2 + ... + \left( \omega_N - \frac{1}{2} \right)^2 \right]}. \end{aligned} \tag{14}$$

A similar expression to Eq. (14) with respect to $\Omega'$ holds for $r'(s_2)$. Since $\Omega'$ is equally dispersed as $\Omega$, $\omega'_N = \omega_N$. Since $\Omega'$ is more complex than $\Omega$ and every state in $\Omega' \setminus \Omega$ is more interior than every state in $\Omega$,

$$\frac{4}{N'}\left[\left(\omega'_{K+1}-\frac{1}{2}\right)^2+...+\left(\omega'_{N'}-\frac{1}{2}\right)^2\right] < \frac{4}{N}\left[\left(\omega_{K+1}-\frac{1}{2}\right)^2+...+\left(\omega_{N}-\frac{1}{2}\right)^2\right].$$

Therefore, when $\theta$ is sufficiently large, it follows from Eq. (9) that $r'(s_2) > r(s_2)$. The proof is analogous for signal realization $s_1$. □

**Proof of Prediction 2.** As in the proof of Prediction 1, we can show that

$$\begin{aligned}\lim_{\theta\to\infty} r_R(s_2)+1 &= \lim_{\theta\to\infty}\frac{|E_R(\omega|s_2)-E(\omega)|}{|E(\omega|s_2)-E(\omega)|}\\ &= \frac{\omega_N-\frac{1}{2}}{\frac{4}{N}\left[\left(\omega_{K+1}-\frac{1}{2}\right)^2+...+\left(\omega_N-\frac{1}{2}\right)^2\right]},\end{aligned}$$

and analogously for $r'(s_2)$. Fixing $\omega_{K+1},...,\omega_{N-1}$, the above expression is increasing in $\omega_N$ if $\left(\omega_N-\frac{1}{2}\right)^2 < \left(\omega_{K+1}-\frac{1}{2}\right)^2+...+\left(\omega_{N-1}-\frac{1}{2}\right)^2$, i.e., $W(\Omega)<0$, and decreasing in $\omega_N$ if $\left(\omega_N-\frac{1}{2}\right)^2 > \left(\omega_{K+1}-\frac{1}{2}\right)^2+...+\left(\omega_{N-1}-\frac{1}{2}\right)^2$, i.e., $W(\Omega)>0$. The proof is analogous for signal realization $s_1$. □

**Proof of Corollary 1.** It follows from Prediction 2 that $r(s_i)$ decreases as the signal diagnosticity $d$ increases when $\theta$ is sufficiently large. Below we show that this holds for any $\theta>0$. Given any $s_i$, note that

$$\begin{aligned}r_R(s_i)+1 &= \frac{|E_R(\omega|s_i)-E(\omega)|}{|E(\omega|s_i)-E(\omega)|}\\ &= \frac{\frac{(1-\omega_2)^{\theta+2}+\omega_2^{\theta+2}}{(1-\omega_2)^{\theta+1}+\omega_2^{\theta+1}}-1/2}{(1-\omega_2)^2+\omega_2^2-1/2}\\ &= \frac{\omega_2^{\theta+1}-(1-\omega_2)^{\theta+1}}{2(\omega_2-1/2)((1-\omega_2)^{\theta+1}+\omega_2^{\theta+1})}.\end{aligned}$$

Therefore, $r_R(s_i)$ is decreasing in $\omega_2$ if and only if $f(x) \equiv \frac{(x-1/2)((1-x)^{\theta+1}+x^{\theta+1})}{x^{\theta+1}-(1-x)^{\theta+1}}$ is increasing in $x$ when $x>1/2$. Differentiating $f(x)$, we have

$$f'(x) = \frac{x^{\theta+1}(x^{\theta+1}-(\theta+1)(1-x)^{\theta})-(1-x)^{\theta+1}((1-x)^{\theta+1}-(\theta+1)x^{\theta})}{(x^{\theta+1}-(1-x)^{\theta+1})^2}.$$

Note that the numerator can be written as $g(x)-g(1-x)$, where $g(x)\equiv x^{\theta+1}(x^{\theta+1}-(\theta+1)(1-x)^{\theta})$. Since $g(x)$ is increasing in $x$, it follows that $f'(x)>0$ for $x>1/2$. Since $d=\omega_2$, by Eq. (9), $r(s_i)$, we know that $r(s_i)$ is decreasing in $d$ for all $s_i \in \mathcal{S}$.

When $\omega_2$ approaches 1, we have $\lim_{\omega_2\to 1} r_R(s_i)+1=1$. By Eq. (9), $r(s_i)$ converges to $-(1-\lambda)$, which is negative when $\lambda<1$. Hence, there exists a cutoff $c\in[1/2,1)$ such that the agent underreacts to all signals if $d\in(c,1)$. When $\omega_2$ approaches 1/2, by the L'Hospital's Rule, we have

$$\lim_{\omega_2 \to 1/2} r_R(s_i) + 1 = \lim_{\omega_2 \to 1/2} \frac{(\theta+1)(\omega_2^\theta + (1-\omega_2)^\theta)}{2((1-\omega_2)^{\theta+1} + \omega_2^{\theta+1})} = \theta + 1.$$

Therefore, we have $\lim_{\omega_2 \to 1/2} r(s_i) = \lambda(\theta+1) - (1-\lambda) > 0$ when $\theta > (1-\lambda)/\lambda - 1$. It follows that $c > 1/2$ when $\theta$ is sufficiently large. □

**Proof of Corollary 2** When $\Omega$ and $\Omega'$ contain no more than three states, Corollary 2 immediately follows from Prediction 2. When $|\Omega| = |\Omega'| \geq 4$, we construct a state space $\Omega'' = \{\omega_1, \omega_2', ..., \omega_{N-1}', \omega_N\}$, where states $\omega_1$ and $\omega_N$ come from $\Omega$ and $\omega_2', ..., \omega_{N-1}'$ come from $\Omega'$. Let $p_0'' = p_0$. By Prediction 1, we know that the agent reacts more in $(\Omega'', p_0'')$ than under $(\Omega, p_0)$. Since $\Omega'$ is less dispersed than $\Omega''$ and $W(\Omega'') > W(\Omega') > 0$, by Prediction 2, we know that the agent reacts more in $(\Omega', p_0')$ than in $(\Omega'', p_0'')$. It then follows that the agent reacts more in $(\Omega', p_0')$ than in $(\Omega, p_0)$. □

**Proof of Prediction 3.** Suppose $p_0'$ is strictly more concentrated than $p_0$ and both are symmetric. Let $\omega'$ and $\omega$ denote the random variables that are distributed according to $p_0'$ and $p_0$, respectively. Since the priors have the same support, $E_R(\omega'|s_i)$ coincides with $E_R(\omega|s_i)$ when $\theta$ diverges to infinity. Thus, to show that $r'(s_i) > r(s_i)$ when $\theta$ is sufficiently large, it suffices to show that $|E(\omega'|s_i) - E(\omega')| < |E(\omega|s_i) - E(\omega)|$.

Suppose the signal realization is $s_2$. Since $E(\omega'|s_2) > 1/2$, $E(\omega|s_2) > 1/2$, and $E(\omega') = E(\omega) = 1/2$, we only need to show $E(\omega'|s_2) < E(\omega|s_2)$. Let $\Delta(\omega_i) = p_0'(\omega_i) - p_0(\omega_i)$. Then $\Delta(\omega_i) \geq 0$ for $\omega_i \in [1-c, c]$ and $\Delta(\omega_i) \leq 0$ for $\omega_i \in [0, 1-c] \cup [c, 1]$, and at least one inequality is strict. We have

$$E(\omega'|s_2) = 2 \sum_{\omega_i \in \Omega} p_0'(\omega_i)\omega_i^2 = E(\omega|s_2) + 2 \sum_{\omega_i \in \Omega} \Delta(\omega_i)\omega_i^2.$$

Since $\Delta(\omega_i)$ is symmetric around $1/2$,

$$\begin{aligned}\sum_{\omega_i \in \Omega} \Delta(\omega_i)\omega_i^2 = &\sum_{\omega_i < 1-c} \Delta(\omega_i)\omega_i^2 + \sum_{\omega_i \in (1-c, c)} \Delta(\omega_i)\omega_i^2 + \sum_{\omega_i > c} \Delta(\omega_i)\omega_i^2 \\ =&2 \sum_{\omega_i \in (1/2, c)} \Delta(\omega_i)(\omega_i - 1/2)^2 + 2 \sum_{\omega_i \in [c, 1)} \Delta(\omega_i)(\omega_i - 1/2)^2 < 0,\end{aligned}$$

where the inequality holds because $|\omega_i - 1/2| < |\omega_j - 1/2|$ for any $\omega_i \in (1/2, c)$ and $\omega_j \in (c, 1)$. Therefore, $E(\omega'|s_2) < E(\omega|s_2)$. The proof is analogous for signal realization $s_1$. □

**Proof of Prediction 4.** For convenience, we denote the binary state space as $\Omega = \{1-x, x\}$ where $x > 1/2$ and the prior as $(1-p_0, p_0)$.

Part (i). First assume $p_0 > 1/2$ and consider a confirmatory realization $s_i$. We have

$$\overline{E}(\omega) = 1/2, \tag{15}$$

$$E(\omega) = (1-p_0)(1-x) + p_0 x, \tag{16}$$

$$E(\omega|s_i) = \frac{(1-p_0)(1-x)^2 + p_0 x^2}{(1-p_0)(1-x) + p_0 x}, \tag{17}$$

$$E_R(\omega|s_i) = \frac{(1-p_0)(1-x)^{\theta+2} + p_0 x^{\theta+2}}{(1-p_0)(1-x)^{\theta+1} + p_0 x^{\theta+1}}. \tag{18}$$

The agent has a wrong direction update at $s_i$ if $\hat{E}(\omega|s_i) - E(\omega) < 0$, which occurs if and only if

$$\lambda E_R(\omega|s_i) + (1-\lambda)\overline{E}(\omega) < E(\omega).$$

By Eqs. (15) to (18), the above inequality simplifies to the following,

$$\frac{p_0 x^{\theta+1} - (1-p_0)(1-x)^{\theta+1}}{p_0 x^{\theta+1} + (1-p_0)(1-x)^{\theta+1}} < \frac{2p_0 - 1}{\lambda}. \tag{19}$$

The agent overreacts to $s_i$ if $\hat{E}(\omega|s_i) > E(\omega|s_i)$, which occurs if and only if

$$\lambda E_R(\omega|s_i) + (1-\lambda)\overline{E}(\omega) > E(\omega|s_i).$$

This inequality simplifies to the following,

$$\frac{p_0 x^{\theta+1} - (1-p_0)(1-x)^{\theta+1}}{p_0 x^{\theta+1} + (1-p_0)(1-x)^{\theta+1}} > \frac{1}{\lambda}\frac{p_0 x - (1-p_0)(1-x)}{p_0 x + (1-p_0)(1-x)}. \tag{20}$$

The agent underreacts to $s_i$ if $E(\omega) < \hat{E}(\omega|s_i) < E(\omega|s_i)$, which occurs if and only if

$$\frac{2p_0 - 1}{\lambda} < \frac{p_0 x^{\theta+1} - (1-p_0)(1-x)^{\theta+1}}{p_0 x^{\theta+1} + (1-p_0)(1-x)^{\theta+1}} < \frac{1}{\lambda}\frac{p_0 x - (1-p_0)(1-x)}{p_0 x + (1-p_0)(1-x)}. \tag{21}$$

Let $t \equiv x/(1-x) > 1$ and $\ell(t) \equiv \frac{p_0 t - (1-p_0)}{p_0 t + (1-p_0)}$. Then $\ell(t)$ is increasing in $t$. By Eqs. (19) to (21), a wrong direction update occurs if $\ell(t^{\theta+1}) < \frac{2p_0-1}{\lambda}$, underreaction occurs if $\frac{2p_0-1}{\lambda} < \ell(t^{\theta+1}) < \frac{\ell(t)}{\lambda}$, and overreaction occurs if $\ell(t^{\theta+1}) > \frac{\ell(t)}{\lambda}$.

First note that $\lim_{t\to 1} \ell(t^{\theta+1}) = 2p_0 - 1$ and $\lim_{t\to\infty} \ell(t^{\theta+1}) = 1$. Since $\ell(s)$ is increasing, if $\lambda \le 2p_0 - 1$, then the agent updates in the wrong direction for all values of $x$. If $\lambda > 2p_0 - 1$, then there exists a cutoff $c_1 \in (1/2, 1)$ such that $\ell((c_1/(1-c_1))^{\theta+1}) = \frac{2p_0-1}{\lambda}$ and the agent updates in the wrong direction for all $x \in (1/2, c_1)$.

Second, note that

$$\frac{\ell(t^{\theta+1})}{\ell(t)} = \frac{(p_0 t + (1-p_0))(p_0 t^{\theta+1} - (1-p_0))}{(p_0 t - (1-p_0))(p_0 t^{\theta+1} + (1-p_0))}$$

$$=1+\frac{2}{\frac{p_0^2 t^{\theta+2}-(1-p_0)^2}{p_0(1-p_0)(t^{\theta+1}-t)}-1}.$$

It is then easy to show that when $p_0 > 1/2$, the ratio $\ell(t^{\theta+1})/\ell(t)$ is first increasing and then decreasing in $t$. Since $\frac{2p_0-1}{\lambda} = \ell((c_1/(1-c_1))^{\theta+1}) < \frac{\ell(c_1/(1-c_1))}{\lambda}$, by continuity we have $\frac{2p_0-1}{\lambda} < \ell(t^{\theta+1}) < \frac{\ell(t)}{\lambda}$ for $t$ strictly larger than but sufficiently close to $c_1/(1-c_1)$. Furthermore, for $t$ sufficiently large, both $\ell(t^{\theta+1})$ and $\ell(t)$ are close to 1, so we must have $\frac{2p_0-1}{\lambda} < \ell(t^{\theta+1}) < \frac{\ell(t)}{\lambda}$ for $\lambda < 1$. Lastly, notice that for any $\lambda > 2p_0 - 1$, we have $\frac{1}{\lambda}\lim_{t\to 1} l(t) = \frac{2p_0-1}{\lambda} < \lim_{t\to 1}\lim_{\theta\to\infty} l(t^{\theta+1}) = 1$. Therefore, if $\theta$ sufficiently large, there exists an $x$ close to $1/2$ such that the agent overreacts. Combining these observations, we know that there exist $c_1 \leq c_2 \leq c_3 \leq 1$ such that the agent underreacts when $x \in (c_1, c_2) \cup (c_3, 1)$, overreacts when $x \in (c_2, c_3)$, and $(c_2, c_3)$ is non-empty if $\theta$ is sufficiently large.

Part (ii). Next assume $p_0 < 1/2$ and consider a disconfirmatory realization $s_i$. Then $E(\omega|s_i) > E(\omega)$. As in Part (i), a wrong direction update occurs if $\ell(t^{\theta+1}) < \frac{2p_0-1}{\lambda}$, underreaction occurs if $\frac{2p_0-1}{\lambda} < \ell(t^{\theta+1}) < \frac{\ell(t)}{\lambda}$, and overreaction occurs if $\ell(t^{\theta+1}) > \frac{\ell(t)}{\lambda}$. Since $l(t)$ is increasing, $\ell(t^{\theta+1}) > \ell(1) = 2p_0 - 1 > \frac{2p_0-1}{\lambda}$, so a wrong direction update is impossible. It remains to determine whether the agent overreacts or underreacts by comparing $\ell(t^{\theta+1})$ and $\frac{\ell(t)}{\lambda}$. Let $c_4 = 1 - p_0$. Then $\ell(c_4/(1-c_4)) = 0$. Note that when $t < c_4/(1-c_4)$, we have $\ell(t) < 0$ and thus $\frac{\ell(t)}{\lambda} < \ell(t) < \ell(t^{\theta+1})$. That is, the agent overreacts when $x \in (1/2, c_4)$. When $t > c_4/(1-c_4)$, we have $\ell(t) > 0$ and $\ell(t^{\theta+1}) > 0$. Moreover, when $t$ is sufficiently large, both $\ell(t^{\theta+1})$ and $\ell(t)$ are close to 1, which implies that $\ell(t^{\theta+1}) < \ell(t)/\lambda$ and so the agent underreacts. Therefore, there exists a cutoff $c_5 \in (c_4, 1)$ such that the agent underreacts if $x \in (c_5, 1)$. □

## B Additional Details and Analyses

### B.1 Experimental Details

Table 9. Information environments used in experiments

| Complexity $\lvert\Omega\rvert$ | Prior $p_0$ | Information Structure $\Omega$ |
|---|---|---|
| 2 states | $p_0(\omega_1) \in \{0.3, 0.5, 0.7\}$<br>$p_0(\omega_2) = 1 - p_0(\omega_1)$ | $Pr(r\|\omega_2) \in \{0.6, 0.7, 0.8, 0.9\}$<br>$Pr(r\|\omega_1) = 1 - Pr(r\|\omega_2)$ |
| 3 states | $p_0(\omega_1) \in \{0.25, 0.33, 0.4\}$<br>$p_0(\omega_2) = 1 - 2p_0(\omega_1)$<br>$p_0(\omega_3) = p_0(\omega_1)$ | $Pr(r\|\omega_3) \in \{0.6, 0.7, 0.8, 0.9\}$<br>$Pr(r\|\omega_2) = 0.5$<br>$Pr(r\|\omega_1) = 1 - Pr(r\|\omega_3)$ |
| 4 states | $p_0(\omega_i) = 0.25$<br>$\forall \omega_i \in \Omega$ | $(Pr(r\|\omega_3), Pr(r\|\omega_4)) \in \{(0.55, 0.6),$<br>$(0.6, 0.7), (0.55, 0.7), (0.7, 0.8),$<br>$(0.6, 0.8), (0.55, 0.8), (0.8, 0.9),$<br>$(0.7, 0.9), (0.6, 0.9), (0.55, 0.9)\}$<br>$Pr(r\|\omega_2) = 1 - Pr(r\|\omega_3)$ |

Table 9. Information environments used in experiments

| Complexity $\|\Omega\|$ | Prior $p_0$ | Information Structure $\Omega$ |
|---|---|---|
| | | $Pr(r\|\omega_1) = 1 - Pr(r\|\omega_4)$ |
| 5 states | $p_0(\omega_i) = 0.2$<br>$\forall \omega_i \in \Omega$ | $(Pr(r\|\omega_4), Pr(r\|\omega_5)) \in \{(0.55, 0.6),$<br>$(0.6, 0.7), (0.55, 0.7), (0.7, 0.8),$<br>$(0.6, 0.8), (0.55, 0.8), (0.8, 0.9),$<br>$(0.7, 0.9), (0.6, 0.9), (0.55, 0.9)\}$<br>$Pr(r\|\omega_3) = 0.5$<br>$Pr(r\|\omega_2) = 1 - Pr(r\|\omega_4)$<br>$Pr(r\|\omega_1) = 1 - Pr(r\|\omega_5)$ |
| 11 states | $p(\omega_i) = 1/11$<br>$\forall \omega_i \in \Omega$ | $Pr(r\|\omega_i) = (i-1)/10$<br>$\forall i \in \{1, ..., 11\}$ |

Notes: States are ordered by number of red balls, with $\omega_1$ corresponding to the bag with the fewest red balls, and so on up through $\omega_N$ corresponding to the bag with the most red balls.

## B.2 Section 3.3 Analyses including Wrong Direction Updates

TABLE 10. Impact of complexity on belief updating

| | Overreaction Ratio | |
|---|---|---|
| | (1) | (2) |
| 4 States | 0.274*** | 0.372*** |
| | (0.0277) | (0.0298) |
| 5 States | 0.364*** | 0.457*** |
| | (0.0344) | (0.0368) |
| $d = 0.7$ | | -0.159*** |
| | | (0.0383) |
| $d = 0.8$ | | -0.364*** |
| | | (0.0399) |
| $d = 0.9$ | | -0.470*** |
| | | (0.0415) |
| Constant | -0.109*** | 0.139*** |
| | (0.0199) | (0.0379) |
| $N$ | 6714 | 6714 |
| adj. $R^2$ | 0.038 | 0.099 |

Notes: Baseline is 2 States and, in Column 2, diagnosticity $d = 0.6$. Includes information environments with a uniform prior and 2 states, 4 states or 5 states listed in Table 9; includes wrong direction updates. Standard errors clustered at the individual level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 11. Impact of signal diagnosticity on belief updating

| | Overreaction Ratio | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| | 2 States | 3 States | 4 States | 5 States |
| $d = 0.7$ | 0.0458 | -0.208*** | -0.381*** | -0.212*** |
| | (0.0435) | (0.0494) | (0.0644) | (0.0819) |
| $d = 0.8$ | -0.0382 | -0.414*** | -0.607*** | -0.436*** |
| | (0.0449) | (0.0470) | (0.0680) | (0.0826) |
| $d = 0.9$ | -0.0546 | -0.449*** | -0.683*** | -0.586*** |
| | (0.0432) | (0.0488) | (0.0711) | (0.0843) |
| Constant | -0.0972** | 0.528*** | 0.720*** | 0.677*** |
| | (0.0421) | (0.0540) | (0.0738) | (0.0905) |
| $N$ | 986 | 1404 | 2928 | 2800 |
| adj. $R^2$ | 0.005 | 0.068 | 0.117 | 0.065 |

Notes: Baseline is diagnosticity $d = 0.6$. Includes all information environments with a uniform prior listed in Table 9 except for the 11-state complexity; includes wrong direction updates. Standard errors clustered at the individual level in parentheses. $^{*}$ $p < 0.10$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$.

TABLE 12. Impact of prior concentration on belief updating

| | Overreaction Ratio | |
|---|---|---|
| | (1) | (2) |
| Concentrated Prior | 0.209*** | 0.209*** |
| | (0.0529) | (0.0529) |
| Diffuse Prior | -0.219*** | -0.219*** |
| | (0.0313) | (0.0313) |
| $d = 0.7$ | | -0.303*** |
| | | (0.0312) |
| $d = 0.8$ | | -0.497*** |
| | | (0.0314) |
| $d = 0.9$ | | -0.549*** |
| | | (0.0317) |
| Constant | 0.260*** | 0.597*** |
| | (0.0250) | (0.0389) |
| $N$ | 4220 | 4220 |
| adj. $R^2$ | 0.049 | 0.127 |

Notes: Includes all information environments with three states listed in Table 9; includes wrong direction updates. Baseline is uniform prior $(0.33, 0.34, 0.33)$ and, in Column 2, diagnosticity $d = 0.6$. Concentrated prior corresponds to $(0.25, 0.5, 0.25)$ and diffuse prior corresponds to $(0.4, 0.2, 0.4)$. Standard errors clustered at the individual level in parentheses. $^{*}$ $p < 0.10$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$.

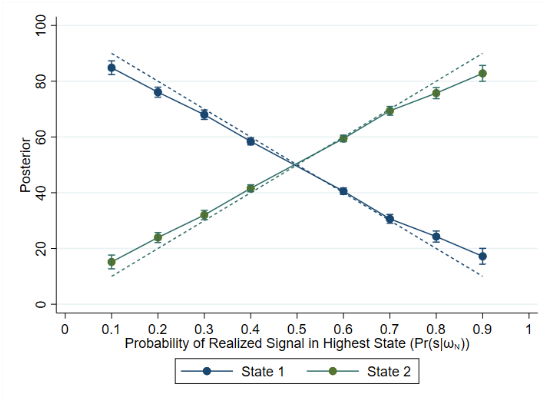
Table 13. Impact of signal type on belief updating

| | Overreaction Ratio | |
|---|---|---|
| | (1) | (2) |
| Confirmatory Realization | 0.121*** | 0.136*** |
| | (0.0384) | (0.0392) |
| Disconfirmatory Realization | 0.371*** | 0.348*** |
| | (0.0442) | (0.0412) |
| $d = 0.7$ | | -0.493*** |
| | | (0.0569) |
| $d = 0.8$ | | -0.590*** |
| | | (0.0565) |
| $d = 0.9$ | | -0.637*** |
| | | (0.0560) |
| Constant | -0.109*** | 0.321*** |
| | (0.0199) | (0.0488) |
| $N$ | 2961 | 2961 |
| adj. $R^2$ | 0.022 | 0.093 |

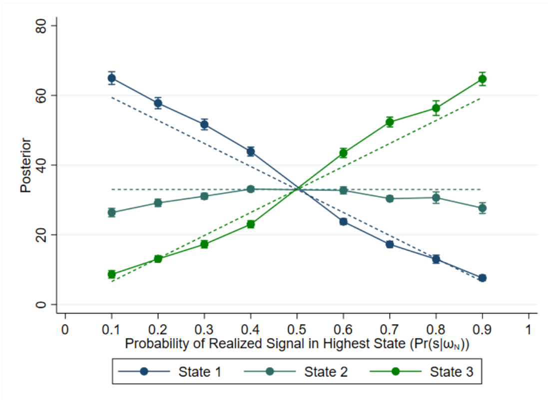
Notes: Includes all information environments with two states listed in Table 9; includes wrong direction updates. Baseline is uniform prior $(0.5, 0.5)$ and, in Column 2, diagnosticity $d = 0.6$. Standard errors clustered at the individual level in parentheses. $^{*}$ $p < 0.10$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$.

## B.3 Reported Posterior Beliefs

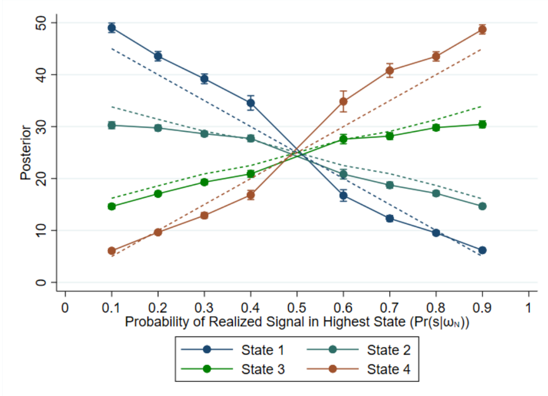
Fig. 10 presents the average reported posterior belief about each state and the Bayesian posterior belief for information environments with a uniform prior listed in Table 9. For a given realized signal and information structure, the dots with 95% error bars show the average reported posteriors and the dashed lines show the objective Bayesian posteriors.



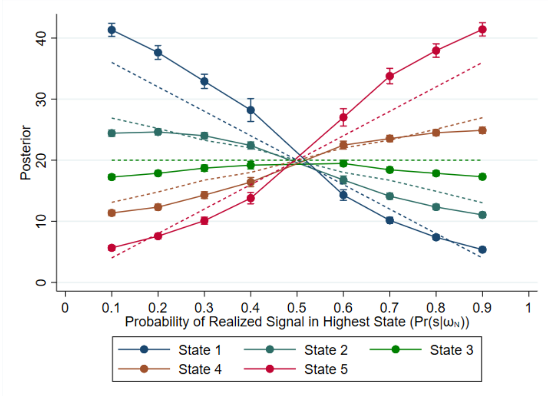

(A) 2 States




(B) 3 States




(C) 4 States




(D) 5 States

FIGURE 10. Reported Posterior Belief

## B.4 Structural Estimation Robustness Checks

We present two robustness checks for our structural estimation. First, we estimate the parameters $\theta$ and $\lambda$ for a prediction loss function that minimizes the average quadratic mean difference between the expected state under the reported posterior and predicted posterior.[43]

[43]The quadratic mean difference between reported posterior $\hat{p}(s_i; \Omega, p_0)$ and predicted posterior $\hat{p}_{\theta,\lambda}(s_i; \Omega, p_0)$ is given by $\left(\sum_{\omega_j \in \Omega} \omega_j \left(\hat{p}(\omega_j | s_i; \Omega, p_0) - \hat{p}_{\theta,\lambda}(\omega_j | s_i; \Omega, p_0)\right)\right)^2$.

Table 14. Structural Estimation with Quadratic Mean Loss Function

| | $\theta$ | 95% CI | $\lambda$ | 95% CI |
|---|---|---|---|---|
| Parameter Estimates | 0.39 | (0.18, 0.92) | 0.79 | (0.68, 0.86) |

Notes: Parameter estimates that minimize the average quadratic mean difference at the aggregate level. Includes all information environments listed in Table 9, except for the 11-state complexity; excludes wrong direction updates. The 95% confidence intervals are obtained from 300 bootstrap samples.

Second, we estimate the parameters for information environments with a symmetric prior. Specifically, we exclude information environments with two states and either a 30/70 or a 70/30 prior. The motivation behind this exercise stems from the model prediction that the agent may update in the wrong direction under an asymmetric prior (Prediction 4). In our main analysis, we drop wrong direction updates. This could potentially lead to an underestimation of cognitive noise. By excluding these information environments, we can drop wrong direction updates without introducing such a bias. The following table demonstrates that this exclusion does not meaningfully affect the parameter estimates.

Table 15. Structural Estimation for Symmetric Priors

| | $\theta$ | 95% CI | $\lambda$ | 95% CI |
|---|---|---|---|---|
| Parameter Estimates | 0.96 | (0.88, 0.99) | 0.69 | (0.68, 0.71) |

Notes: Parameter estimates that minimize average KL divergence at the aggregate level. Includes all information environments with a symmetric prior listed in Table 9, except for the 11-state complexity; excludes wrong direction updates. The 95% confidence intervals are obtained from 300 bootstrap samples.

# C Experimental Instructions

The following shows the experimental instructions for the 3-state treatment. The other complexity treatments are analogous.

**Page 1:**

**The Experiment**

In each guessing task, there are three bags, "Bag 1," "Bag 2," and "Bag 3." Each bag contains 100 balls, some of which are **red** and some of which are **blue**. One of the bags will be selected at random by the computer as described below. You will not observe which bag was selected. Instead, the computer will then randomly draw a ball from the secretly selected bag, and will show this ball to you.

Your task is to **guess the probability that each bag was selected** based on the available information. The exact procedure is described below.

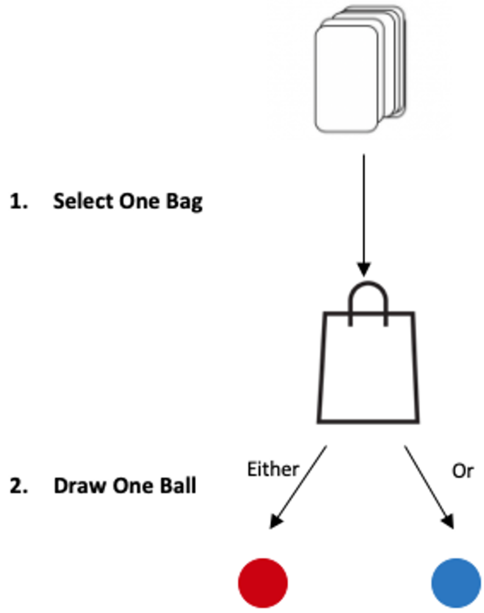
**Task Setup**

- There is a deck of cards that consists of 100 cards. Each card in the deck either has "Bag 1," "Bag 2," or "Bag 3" written on it. You will be informed about **how many** of these 100 cards have "Bag 1," "Bag 2," and "Bag 3" written on them.

- You will be informed about **how many red and blue balls** each bag contains.

**These numbers are very important for making accurate guesses.**

Page 2:




**Sequence of Events**

1. The computer **selects one** of the 100 cards.
    - If a "Bag 1" card was drawn, Bag 1 is selected.
    - If a "Bag 2" card was drawn, Bag 2 is selected.
    - If a "Bag 3" card was drawn, Bag 3 is selected.
2. Next, the computer randomly draws **one of the 100 balls** from the secretly selected bag. Each of the 100 balls is equally likely to be selected.
3. The computer will then **inform you about the color** of the randomly drawn ball.

After seeing the color of the ball, you will make your guess by **stating a probability between 0% and 100%** that each of Bag 1, Bag 2, and Bag 3 was drawn. Note that the probabilities have to sum to 100.

**One ball** will be drawn from a bag and you will make one guess after the ball is drawn.

**Please Note**

- The number of "Bag 1," "Bag 2," and "Bag 3" cards **can vary across tasks**.
- The number of red and blue balls in each bag **varies across tasks**.
- The computer **draws a new card for each task,** so you should **think about which bag was selected in a task independently of all other tasks.**

## Page 3:

**Comprehension Questions**

The following questions test your understanding of the instructions.

Click here to review the instructions.

Which statement about the number of cards corresponding to each bag is correct?

- ○ The number of "Bag 1" cards is always the same in all tasks.
- ○ The exact number of cards corresponding to each bag may vary across tasks.

---

Which statement about the allocation of red and blue balls in the bags is correct?

- ○ The exact fraction of red and blue balls in each bag may vary across tasks.
- ○ The fraction of red balls in each bag is the same in all tasks.

---

Which statement about the probabilities of each bag is correct?

- ○ In a given task, the probabilities that each bag was drawn must add up to 100.
- ○ In a given task, the probability that each bag was drawn is 100, summing up to 300 in total.

---

If Bag 1 has more red balls than blue balls and Bag 2 has more blue balls than red balls, and a red ball is drawn in the first round, which bag is more likely to have been chosen for this task? Write **Bag 1** or **Bag 2**.

---

If Bag 3 has more blue balls than red balls and Bag 1 has more red balls than blue balls, and a red ball is drawn in the first round, which bag is more likely to have been chosen for this task? Write **Bag 1** or **Bag 3.**