Over- and Underreaction to Information*

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

May 27, 2024

This paper explores the impact of the learning environment on how people react to information. We develop a model of belief-updating where people respond to complexity by forming a simplified representation of the environment via salience-driven channeled attention, then process information using Bayes' rule subject to cognitive imprecision. The model predicts *overreaction* when environments are complex, signals are noisy, information is surprising, or priors are concentrated on less salient states; it predicts *underreaction* when environments are simple, signals are precise, information is expected, or priors are concentrated on salient states. Results from a series of pre-registered experiments provide support for these predictions and evidence for the proposed cognitive mechanisms. Our model is highly *complete* in capturing explainable variation in belief-updating; the interaction between the two psychological mechanisms is critical to explaining belief data. These results connect disparate findings in prior work: underreaction is typically found in laboratory studies, which feature simple learning settings, while overreaction is more prevalent in financial markets with more complex environments.

Keywords: overreaction, underreaction, beliefs, noisy cognition, representativeness, bounded rationality, attention, mental representation, completeness, restrictiveness, behavioral economics, learning, forecasting, inference

*We thank Nageeb Ali, Nick Barberis, Dan Bartels, Francesca Bastianello, Cary Frydman, Xavier Gabaix, Thomas Graeber, Ingar Haaland, David Laibson, Annie Liang, Ryan Oprea, Pietro Ortoleva, Yuval Salant, Andrei Shleifer, Marciano Siniscalchi, Leeat Yariv, Sevgi Yuksel, Florian Zimmermann and seminar participants and conference participants at Drexel, Harvard, NBER Asset Pricing meeting, Northwestern, NYU Abu Dhabi, Princeton, Stanford, Texas A&M, UC Berkeley, UC Santa Barbara, University of Alicante, University of Connecticut, USC, and University of Zurich for helpful comments. Josh Hascher, Christy Kang, Enar Leferink, Matthew Murphy 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 its past performance, managers learn from candidate interviews before making hiring decisions, and professional forecasters make economic predictions based on data releases. Standard models assume that people have an accurate mental representation of the learning environment and use Bayes' rule to draw inference with respect to this representation. However, a large literature in economics, finance, and psychology documents systematic departures from these assumptions.

The literature uses a variety of methods to study how people deviate from the standard model of belief-updating, and the findings are mixed. In laboratory experiments, participants are told the information environment, observe a signal, and then report their belief about an unobserved state. Such experiments generally find that people *underreact* to information relative to the Bayesian benchmark (Benjamin 2019).[1] Another line of work studies belief-updating using surveys and forecasts of households and financial industry professionals. In contrast, these studies often find that people *overreact* to information (Bordalo, Gennaioli, and Shleifer 2022).[2]

This paper explores how properties of the learning environment—such as its complexity or the informativeness of signals—impact whether under- or overreaction emerges. We propose a two-stage framework of belief-updating where cognitive constraints on attention and processing capacity interact with the learning environment to systematically distort beliefs. We start with the premise that an agent's mental representation of her learning environment may differ from the environment she actually faces. In the *representational* stage, limits on attention prompt the agent to simplify her learning environment by focusing on states that are most salient given the observed signal. This channeled attention causes salient states to be overweighed in her representation. In the *processing* stage, limited processing capacity impacts how the agent evaluates information, generating cognitive imprecision when she forms her subjective belief. The interaction between attentional and processing constraints generates a rich set of theoretical predictions about how belief-updating varies with the learning environment. In particular, the framework predicts more overreaction when the state space is more complex, the signal is noisier, the information is surprising, or the prior is less concentrated on salient states; it predicts more underreaction when the state space is simpler, the signal is more precise, the information is expected, or the prior is more concentrated on salient states.

[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] In a review of how people update their beliefs in financial markets, 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." There are notable exceptions where underreaction is observed, however, such as the case of forecasting short-term interest rates.

A series of experiments provide direct support for these predictions and the proposed cognitive mechanisms. While cognitive imprecision alone can generate the observed pattern of belief-updating in simple binary-state environments, the mechanism quickly loses explanatory power as the environment becomes more complex—even when just moving to three states. In contrast, the model which incorporates both attentional and processing constraints is highly *complete* in capturing the observed patterns of belief-updating across a wide array of learning environments (Fudenberg, Kleinberg, Liang, and Mullainathan 2022).[3] We show that the two mechanisms act as *cognitive complements*: accounting for both generates belief predictions that are substantially closer to the data than either mechanism on its own. Importantly, the large increase in model completeness from simultaneously incorporating both mechanisms does not come at the expense of model flexibility—the proposed framework is also highly *restrictive* (Fudenberg, Gao, and Liang 2023), and only slightly less so than either mechanism on its own.[4] Taken together, our results help rationalize the discrepancy between the predominant observation of underreaction in laboratory studies—which typically use simple binary state spaces, relatively precise signals, and uniform priors—and the larger prevalence of overreaction in financial market studies—which feature more complex environments and noisier signals.

We illustrate how salience-channeled attention and cognitive imprecision interact with the learning environment with the following example. Suppose 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 signal distribution). The agent observes a price increase. 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, results from laboratory studies suggest that the agent will underreact and increase her belief from 50 to less than 70 percent. Now suppose that there are five potential states of equal prior likelihood: good and bad, with the same chances of a price increase as before, as well as three intermediate states with a 40, 50, or 60 percent chance of generating a price increase. Does the increased complexity of the state space impact how the agent updates her belief?

To answer this question, we turn to the literature on how people respond to complexity. We distinguish between two broad categories of complexity: *representational* and *computational*. Representational complexity corresponds to the number of objects one needs to consider to form an accurate mental representation of the environment (e.g., internalizing the information structure), while computational complexity

[3]Completeness is a measure of the extent to which a model captures the predictable variation in the data relative to Bayes' rule.

[4]Restrictiveness is a measure of the extent to which a model is rejected when tested on "synthetic" data (as opposed to actual data). High restrictiveness rules out that the model is so flexible that it can fit any data well.

corresponds to the cognitive costs of carrying out the task at hand given the mental representation (e.g., processing information to form beliefs).[5] In the case of belief-updating, representational complexity increases with the size of the state space, as a larger state space requires simultaneous consideration of more objects when internalizing the information structure. Attention and working memory constraints imply that an agent can fully attend to a limited number of objects at any given time.[6] As a result, the agent simplifies her learning environment by channelling attention to a limited number of states. These states are overweighed relative to other states, resulting in a distorted mental representation of the learning environment.

Importantly, attention is not channeled randomly: the agent focuses on states that are most salient, measured by their "representativeness."[7] A state is more representative of a given signal realization if it is more likely to generate it relative to other states. When the signal has a good news/bad news structure, as in our model and experiments, extreme states are the most salient and hence most overweighed in the mental representation.[8] Absent other factors, this generates overreaction in terms of excess movement of the agent's subjective expected state relative to the movement of the Bayesian expected state.[9]

Returning to the example, the "good" asset is most representative of a price increase in both the two- and five-state cases. The agent focuses her attention on this "good" state following a price increase, and as a result, her belief overweighs its likelihood relative to the other potential states. This mechanism has more bite as representational complexity increases: in the five-state case, the prior places more mass on less representative middle states, which implies that channeling attention to the "good" state will generate a larger distortion in the agent's representation.

[5]Research in cognitive psychology distinguishes between these two forms of complexity by differentiating between representational and computational capacity (see Shenhav, Musslick, Lieder, Kool, Griffiths, Cohen, and Botvinick (2017) for an overview). Research in computer science makes a similar distinction between the complexity of representing multiple objects, which is typically referred to as state complexity (Gao, Moreira, Reis, and Yu 2015; Oprea 2020), and the cost of implementing an algorithm conditional on the representation (Papadimitriou 2003).

[6]See Oberauer, Farrell, Jarrold, and Lewandowsky (2016); Luck and Vogel (1997); Loewenstein and Wojtowicz (2023). 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).

[7]A large theoretical and empirical literature shows that attention is channelled to objects as a function of their salience (see Bruce and Tsotsos (2009) for review). Representativeness is a salience cue that operates through bottom-up attention. It was initially identified in Kahneman and Tversky (1972) and its economic implications were explored in Bordalo, Coffman, Gennaioli, and Shleifer (2016); Bordalo, Gennaioli, Porta, and Shleifer (2019). We empirically study other salience cues—visual (bottom-up) and goal-directed (top-down)—in Section 4.3 and show that representativeness is the dominant driver of attention in our setting.

[8]In the good news/bad news structure (Milgrom 1981), signals are ordered so that "better" signals (good news) increase the likelihood of "better" states. This structure naturally maps onto many financial environments (e.g., equities, where a price increase increases the likelihood that an asset is "good"), and is the canonical structure used in laboratory experiments.

[9]Under alternative signal structures, our model still predicts that the representative state will be overweighed. Section 6.1 shows that this prediction is supported empirically. But whether this overweighing results in overreaction now depends on the parameters of the learning environment.

The agent also faces computational complexity when using her mental representation to process the signal and form a posterior belief. Prior work shows that such complexity can be modeled as optimal processing subject to noise—broadly termed *noisy cognition* (Green, Swets et al. 1966; Thurstone 1927; Woodford 2020).[10] This is captured in the *processing stage* of our framework via the agent updating with a noisy version of Bayes' rule with respect to her (potentially distorted) mental representation. This leads to insensitivity to the signal and thus underreaction. Combining both stages, channeled attention and cognitive imprecision interact with the properties of the learning environment to determine whether underreaction or overreaction emerges overall.

Returning again to the example, our model predicts that the agent underreacts to the price increase when the asset is simple (the two-state case), but overreacts when the asset is more complex (the five-state case). In the simple case, limited attention generates a relatively small distortion and cognitive imprecision dominates: the agent does not fully internalize the informativeness of the price increase and underreacts to it. On the other hand, when the asset is complex, limited attention generates a more distorted mental representation and this dominates the impact of cognitive imprecision, which leads to overreaction.[11]

Beyond state-space complexity, our framework predicts how the other properties of the learning environment impact belief-updating. With respect to signal informativeness, our model predicts that the extent of overreaction will decrease as signals become more diagnostic of the state. In the example above, this means that the agent will overreact more when the good and bad assets have a 60 and 40 percent chance of a price increase, respectively, compared to a more precise signal where these chances are 80 and 20 percent.[12] Regarding the prior, our framework predicts less overreaction when the prior concentrates more mass on the extreme states and more overreaction when it concentrates less mass on these states. In the case of an asymmetric prior, "surprising" disconfirmatory signals—those that increase the likelihood of states that the prior assigns lower probability—generate overreaction, and "expected" confirmatory signals generate underreaction or even wrong-direction updating.[13] As we show in Section 2, these predictions follow from the interaction

[10]A more recent literature in economics applies the principles of noisy cognition to explain anomalies in choice under uncertainty (Khaw, Li, and Woodford 2022; Frydman and Jin 2022; Enke and Graeber 2023; Woodford 2020) and forecasting (Azeredo da Silveira and Woodford 2019; Augenblick, Lazarus, and Thaler 2022; Gabaix 2019).

[11]By focusing on a limited number of states in the representational stage, the agent diminishes the impact of increasing states on computational complexity, and therefore, cognitive imprecision. As shown in Section 3, this is born out in the data.

[12]In a simple two-state setting, Edwards (1968) and Benjamin (2019) show that underreaction 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 salience-channeled attention as well.

[13]In the simple asset example, if there is an 80% chance of the good asset and a 20% chance of the bad asset, then a price increase is a confirmatory signal and a price decrease is disconfirma-

between salience-channeled attention and cognitive imprecision.

We test these predictions in a series of experiments. We adopt the classic "bookbag-and-poker-chip" design originally used in Edwards (1968) and employed extensively in the learning literature. A set number of bags have different colored 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 a 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: bags represent states, the probability that each bag is selected corresponds to the prior, and the proportion of balls in each bag represents the signal distribution. We employ three main sources of treatment variation: representational complexity via the number of states (varying from 2 as is standard up to 11), the signal distribution, and the symmetry and concentration of the prior.

Increasing complexity has a striking effect on belief-updating. We first replicate the standard finding that people generally underreact in simple 2-state uniform-prior environments. But this result flips when we add even a single additional state: the majority of participants overreact in 3-state uniform-prior environments across all signal distributions we consider. The share of participants overreacting and the level of overreaction both increase monotonically with the complexity of the state space up through 11 states. Importantly, our model not only makes predictions about average belief movement, but also on *which* states will be overweighed versus underweighed. We provide direct support for these predictions in a state-by-state analysis of subjective beliefs. As shown in Section 3.3, the interaction between the two mechanisms is crucial for this prediction; they act as cognitive complements and neither alone can explain our observed patterns of belief-updating.[14]

We next test the predictions on signal informativeness and the prior. Consistent with the predictions, overreaction decreases with signal diagnosticity and increases as the prior becomes more concentrated on intermediate states. Turning to asymmetric priors, we observe underreaction to confirmatory, expected signals and overreaction to surprising, disconfirmatory signals. Documenting the latter in a simple two-state setting contrasts with the observed underreaction under a symmetric prior. Moreover, consistent with our prediction, we observe nearly three times as many wrong-direction updates to confirmatory realizations compared to disconfirmatory realizations.

We use the experimental data to structurally estimate the two key parameters

tory. Wrong-direction updates occur when the agent's belief that the asset is good decreases after observing a price increase or vice verse.

[14]Specifically, cognitive imprecision alone predicts that the most representative state will be underweighed and the least representative state will be overweighed, while representativeness-driven salience alone predicts the opposite pattern. Neither of these patterns is borne out in the data, but the prediction from the interaction of the mechanisms is.

of our model—capturing the severity of the attentional and processing distortions. In aggregate, both estimates differ from the Bayesian benchmark and are in line with values found in prior work. At the individual level, the vast majority of participants exhibit significant distortions from both salience-channeled attention and cognitive imprecision. Moreover, these individual estimates are significantly positively correlated, suggesting underlying differences in cognitive capacity driving both the representational and processing stages of belief-updating.

We then directly test the proposed attentional mechanism in the representational stage. Employing a common paradigm from cognitive science to measure and manipulate attention (Payne, Bettman, and Johnson 1988), we find that upon observing the signal, participants' attention was indeed overwhelmingly drawn to the most "representative" state. Moreover, fixing the information structure, exogenously limiting attentional resources exacerbated overreaction. Structural estimates show that this is driven by a greater distortion in the representational stage without affecting the processing stage. We then proceed to study the causal effect of attention on belief-updating using a variation of our 5-state paradigm that suppresses representativeness as a salience cue. In this variation, attention is channeled to states as-if randomly. Consistent with the predicted distortion, this random state is overweighed and *underreaction* emerges on average. These results imply that underreaction is not a unique feature of the simple 2-state environment; it can also emerge in complex environments where the representativeness salience cue is suppressed or there is uncertainty over it. Finally, we compare representativeness to other salience cues considered in the literature (visual and goal-directed salience); we show that the former has a substantially stronger influence in channeling attention.

To evaluate model fit, we measure its *completeness* in capturing predictable variation in belief-updating (Fudenberg et al. 2022) relative to Bayes' rule. The model has high explanatory power across both simple and complex environments, capturing nearly all of the explainable variation in both cases (completeness 1.00 and 0.92, respectively, on a scale of 0 to 1). Each stage alone does not fare nearly as well. While cognitive imprecision can explain belief-updating in simple environments (completeness 1.00), it precipitously loses explanatory power in complex environments (dropping to 0.36). Salience-channeled attention alone has low explanatory power across both simple and complex environments. This further demonstrates that the two mechanisms act as cognitive complements: their interaction plays a critical role in predicting belief-updating. Notably, the two-stage model's completeness does not come at the expense of being too flexible: it is nearly as *restrictive* (Fudenberg et al. 2023) as each of the stages on its own.

Finally, we explore several variations of our model in other information environments and beyond inference. Under an alternative signal structure in which moderate states are representative (i.e., not a good news/bad news structure), the overweighing of states continues to be driven by their representativeness. However, whether this overweighing translates to over or underreaction now depends on the parameters of the information environment. Taken together with the findings from the suppressed-representativeness treatment described above, these results show that our framework does predict underreaction in some complex information environments. Next, we show that differences in representational complexity across forecasting versus inference environments can rationalize the "inference-forecast" gap in belief-updating (Fan, Liang, and Peng 2023). Manipulating such complexity can close or even reverse this gap. Lastly, applying our framework to financial markets, we experimentally show how the structure of an asset—binary option versus equivalent bull spread—determines whether people under- or overreact to news about its performance.

A large literature explores under- and overreaction in belief-updating. We provide an in-depth review of this work in Appendix A and discuss how our results can help rationalize some of the disparate findings. For instance, our model predicts underreaction in simple settings such as the binary-state experiments reviewed in Benjamin (2019), and overreaction in more complex environments with a good news/bad news signal structure, such as the studies in financial markets reviewed in Bordalo et al. (2022). Our framework also predicts the observed underreaction in settings where the representative state is not clear (e.g., US treasury rates) or the signal structure is not good news/bad news. 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 also contributes to the literature exploring the cognitive foundations of economic decision-making. Our two-stage model is similar in spirit to Schwartzstein (2014), where the agent selectively channels her attention to a subset of the available information and then uses this subset to update her beliefs using Bayes rule. Our findings on the role of complexity relate to research showing that people are averse to complexity (Oprea 2020), and as a result, adopt simpler mental models (Kendall and Oprea 2021; Molavi 2022; Molavi, Tahbaz-Salehi, and Vedolin 2023), form simpler hypotheses (Bordalo, Conlon, Gennaioli, Kwon, and Shleifer 2023), and use heuristics to reduce the mental cost of judgments and decisions (Salant and Spenkuch 2022; 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 environment (Gabaix and Laibson 2017; Khaw, Li, and Woodford 2021; Khaw et al. 2022). Such cognitive noise has been shown to generate insensitivity to the parameters of the environment (Enke and Graeber 2023). Our theoretical framework is linked to both areas of research: our proposed model incorporates a heuristic response to complexity and cognitive imprecision as two stages of the belief-updating process.

The rest of the paper proceeds as follows. Section 2 outlines the theoretical framework. Section 3 outlines the experimental paradigm, empirical findings, and structural estimation. Section 4 presents evidence for the proposed mechanism. Section 5 quantifies model completeness and restrictiveness. Section 6 tests the implications of our model in other settings. Section 7 concludes.

2 Theoretical Framework

In this section we formalize a two-stage model of belief formation, derive predictions on how the subjective belief distribution varies with the degree of bias in each stage, define a measure of over- and underreaction, and finally derive comparative static predictions on how this measure varies with properties of the information environment. All proofs are in Appendix B.

2.1 Information Environment

A state ω is drawn from state space $\Omega \equiv \{\omega_1, ..., \omega_N\} \subset (0,1)$ with $N > 1$ distinct states in ascending order, $\omega_1 < ... < \omega_N$, and generic element ω_i. The state is distributed according to full-support prior $p_0 \in \Delta(\Omega)$. A signal s provides information about the state. We focus on a binary signal with a good news/bad news structure (Milgrom 1981), as is typically used in laboratory and financial studies (e.g., price increase versus decrease).[15] Let $\mathcal{S} \equiv \{s_1, s_2\}$ denote the support of the signal, with generic realization s_j. In state ω_i, the signal is distributed according to $\pi(s_2|\omega_i) = \omega_i$ and $\pi(s_1|\omega_i) = 1 - \omega_i$. For example, when $\Omega = \{0.3, 0.5, 0.7\}$, signal realization s_2 occurs with probability 0.3 in state ω_1, 0.5 in state ω_2, and 0.7 in state ω_3. Since the probability of s_2 is increasing in the state and s_1 is decreasing, s_2 is indicative of higher states (good news) and s_1 is indicative of lower states (bad news). We refer to Ω as the *information structure*, since the signal distribution is pinned down by the state space, and (Ω, p_0) as the *information environment*. This information environment mirrors the experimental paradigm in Section 3.

Given an information environment (Ω, p_0), by Bayes' rule, the objective posterior probability of state ω_i following signal realization s_2 is

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

and analogously following s_1, $p_B(\omega_i|s_1) \equiv (1-\omega_i)p_0(\omega_i)/\sum_{\omega_j \in \Omega}(1-\omega_j)p_0(\omega_j)$. Let $p_B(s_j) = (p_B(\omega_1|s_j), ..., p_B(\omega_N|s_j))$ denote this objective posterior.

We next define several properties of information environments, which we will manipulate for our comparative static predictions. An information structure Ω' is more *complex* than Ω if Ω' contains weakly more states, $|\Omega'| \geq |\Omega|$, and more *dispersed* than Ω if the minimum and maximum states in Ω' are weakly smaller and larger, respectively, $\omega_1' \leq \omega_1$ and $\omega_N' \geq \omega_N$. An information structure Ω is

[15] The majority of our experiments focus on this case; in Section 6.1 we explore an alternative signal structure.

symmetric if $\omega_i \in \Omega$ implies $1 - \omega_i \in \Omega$. A prior p_0 is *symmetric* on Ω if for any $\omega_i \in \Omega$, ω_i and $1 - \omega_i$ have the same mass, $p_0(\omega_i) = p_0(1 - \omega_i)$. Note that prior symmetry implies information structure symmetry (but not vice versa), and therefore, if p_0 is symmetric we also refer to (Ω, p_0) as a symmetric information environment. Related to individual states, state ω_j is more *interior* than ω_i if it is closer to $1/2$, $|\omega_j - \frac{1}{2}| \leq |\omega_i - \frac{1}{2}|$. The *diagnosticity* in state ω_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.[16]

2.2 Two-Stage Model of Belief-Formation

We next model how an agent forms her subjective posterior belief. In the first stage, attention and working memory constraints lead her to form a distorted mental representation of the information structure. In the second stage, processing capacity constraints introduce cognitive imprecision in processing information (updating beliefs) with respect to this mental representation.

Stage 1: Mental Representation. Before processing the signal, an agent first forms a mental representation of the information environment. As discussed in the introduction, research on how people respond to complexity highlights two key categories: *representational* and *computational* (Shenhav et al. 2017). Representational complexity is the relevant category for this stage, as it captures the number of objects the agent needs to consider to form an accurate mental representation. In our setting, representational complexity is proportional to the size of the state space, since in larger state spaces the agent needs to simultaneously consider more objects when forming a belief.[17] With fixed constraints on attention and working memory, the agent simplifies the information environment in her mental representation by channeling attention to a limited number of states and neglecting others.[18] Therefore, higher representational complexity leads to a more distorted mental representation, as it results in the agent neglecting a larger number of states.

In the face of representational complexity, a key question is which states the agent channels attention to and which she neglects. Prior work finds that attention is channelled towards salient objects (Bruce and Tsotsos 2009).[19] Based on this

[16]For example, in $\Omega = \{0.3, 0.5, 0.7\}$, the set of diagnosticities $\{0.5, 0.7\}$ pin down Ω.

[17]The focus on state complexity as a key driver of representational complexity is mirrored in both theoretical and empirical work in finance (Molavi et al. 2023; Puri 2022) and computer science (Gao et al. 2015), and has been shown to have a large impact on choice (Oprea 2020).

[18]Research has shown that an agent can attend to a limited number of objects at a time—typically 3 or 4 (Oberauer et al. 2016; Luck and Vogel 1997; Loewenstein and Wojtowicz 2023).

[19]Salience can channel attention towards objects through both top-down and bottom-up processes (Talsma, Senkowski, Soto-Faraco, and Woldorff 2010; Yantis 2008; Tanner and Itti 2019). Bottom-up attention, also known as *stimulus-driven* attention (Li and Camerer 2022), corresponds to attention channeled through a subconscious response to a stimulus; attention is channeled based on the stimulus's inherent properties relative to the rest of the information environment (i.e., visual salience). Top-down attention corresponds to intentionally allocating attention through a conscious process, typically in response to incentives in the information environment (i.e., goal-directed

research, we propose that the agent channels attention proportional to the salience of each state given the observed stimulus cue. This results in a distorted mental representation $\hat{\pi}$ of the information structure that scales the likelihood of s_j in state ω_i proportional to the salience of ω_i when observing s_j,

$$\hat{\pi}(s_j|\omega_i) \equiv \pi(s_j|\omega_i)R(\omega_i, s_j)^{\theta} \tag{2}$$

where $R(\omega_i, s_j) \geq 0$ measures the salience of state ω_i given s_j and $\theta \geq 0$ captures the severity of the attentional distortion (higher θ corresponds to more distortion).[20] When $\theta = 0$, the mental representation is accurate, and when $\theta > 0$, the mental representation overweighs the probability of the signal realization in more salient states and underweighs it in less salient states.

An important driver of salience is the extent to which an object is "representative" of the stimulus cue—in our setting, the observed signal (Kahneman and Tversky 1972; Tversky and Kahneman 1983). For example, when predicting the hair color of someone from Ireland, people overweigh the likelihood that the person has red hair, as someone from Ireland is more likely to have red hair than the general population—red hair is representative of someone from Ireland (Bordalo et al. 2016). Motivated by these findings, we use a state's "representativeness" of the observed signal (Gennaioli and Shleifer 2010) as our main measure of its salience, where the representativeness of state ω_i for signal realization s_j is equal to the conditional probability of s_j in ω_i relative to the total probability of s_j,[21]

$$R(\omega_i, s_j) \equiv \frac{\pi(s_j|\omega_i)}{Pr(s_j)}. \tag{3}$$

A state is *more representative* if it is more likely to generate s_j relative to other states. In good news/bad news information environments, ω_1 is the most representative state for s_1 and ω_N is the most representative state for s_2.[22] Substituting this expression for $R(\omega_i, s_j)$ into Eq. (2) yields mental representation

$$\hat{\pi}(s_j|\omega_i) = \frac{\pi(s_j|\omega_i)^{\theta+1}}{Pr(s_j)^{\theta}}. \tag{4}$$

salience). The literature on rational inattention develops models of top-down attention (Maćkowiak, Matějka, and Wiederholt 2023).

[20] Representation $\hat{\pi}$ is a pseudo-information structure, in the sense that substituting it for the true information structure in Bayes' rule results in a well-defined posterior belief over the state space, but $\hat{\pi}$ is not necessarily a probability distribution ($\hat{\pi}(s_j|\omega_i)$ may not sum to one across signals). When there are two states, a well-defined probability distribution over signals—a misspecified model—that "represents" $\hat{\pi}$, in that it prescribes the same Bayesian updates as $\hat{\pi}$, exists. When there are more states than signals, it will generally be necessary to augment the signal space in order to find such a misspecified model representation. See Bohren and Hauser (2024).

[21] This is equivalent to the definition of representativeness in Gennaioli and Shleifer (2010) taking the prior as the comparison group, and is also the measure used in Bordalo et al. (2016, 2019).

[22] We empirically study other salience cues—visual (bottom-up) and goal-directed (top-down)—in Section 4.3 and show that representativeness is the dominant driver of attention in our setting.

This representation overweighs the probability of a signal realization in states that are more likely to generate it.[23]

Discussion. In the first stage, the agent responds to complexity in the information environment by honing in on a subset of states while neglecting the other states. To see the intuition, consider an investor who forms beliefs about a new tech company. The state space includes the possibility that the firm is a zombie (non-viable and set to crash), a unicorn (e.g., Google, Facebook), or a slew of intermediate possibilities. Upon observing a price increase (the signal), 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 overweighs the possibility of a unicorn, at the expense of other states. This does not imply that the investor is completely unaware of other states; these states just receive less weight compared to the Bayesian benchmark.

Our framework is part of a broader literature on how people use simplification strategies when making decisions or forming beliefs. For example, in Bordalo et al. (2023), an agent simplifies hypotheses using bottom-up attention, focusing on features that are salient. The evaluation of these features generate different biases depending on which features are salient, despite the same underlying information structure. Similarly, Banovetz and Oprea (2023) show that agents simplify decision rules by 'economizing' on the number of states that need to be tracked to execute the rule. See Payne, Bettman, and Johnson (1993) for a review of evidence on heuristics as a simplification tool in complex environments.

While we focus on a setting where representativeness channels attention in the context of "online" stimuli, Gennaioli and Shleifer (2010) and Bordalo, Coffman, Gennaioli, Schwerter, and Shleifer (2021) argue that the most representative states are also overweighed in judgment because they are easier to recall. See also Kahneman (2003) for a discussion on the interaction between selective attention and recall, and how this relates to heuristics in judgement. Exploring how attention interacts with memory in belief updating is an exciting avenue for future research.

Finally, we focus on representativeness as a driver of salience because we believe it to be particularly relevant for the environments our framework aims to capture. Section 4.3 provides empirical support for this claim. The model can also capture salience-based distortions stemming from other features of the information environment, including low-level bottom-up channels (e.g., visual salience) and top-down drivers of attention (e.g., goal-directed salience). We leave the exploration of other attention-based distortions to future work.

[23] As we show in Appendix B, applying Bayes rule to this representation results in an updating rule that "counts" a signal $\theta + 1$ times; it is equivalent to forming a posterior belief based on the representativeness-based discounting weighing function used in Bordalo et al. (2016, 2019).

Stage 2: Processing. After forming a mental representation, the agent processes the observed signal optimally but with cognitive noise. As discussed in the introduction, a large literature in cognitive psychology shows that people display cognitive imprecision when making judgments and decisions (Green et al. 1966; Thurstone 1927). In our context, this corresponds to an agent perceiving the parameters of the information environment with noise and treating the perceived parameters as signals of their underlying values, rather than using them directly. Notably, noisy cognition leads to reduced sensitivity to the parameters of the information environment when processing information and updating beliefs.

Following the literature (Woodford 2020; Khaw et al. 2022), we model cognitive imprecision in inference as the agent updating by applying a noisy version of Bayes' rule to her mental representation. Specifically, fixing signal realization s_j, she observes a noisy cognitive signal $Y(s_j) \equiv (Y(\omega_1|s_j), ..., Y(\omega_N|s_j))$ of the posterior belief $p_R(s_j) \equiv (p_R(\omega_1|s_j), ..., p_R(\omega_N|s_j))$ derived from applying Bayes' rule to her mental representation, where

$$p_R(\omega_i|s_j) \equiv \frac{\hat{\pi}(s_j|\omega_i)p_0(\omega_i)}{\sum_{\omega_k \in \Omega} \hat{\pi}(s_j|\omega_k)p_0(\omega_k)} \tag{5}$$

given mental representation $\hat{\pi}$. We assume that this cognitive signal is drawn from a multinomial distribution with $\eta > 0$ trials, $N = |\Omega|$ categories (i.e., states), and event probabilities $p_R(s_j)$:

$$Y(s_j) \sim \frac{1}{\eta} Multi(\eta, N, p_R(s_j)).$$

The cognitive signal is unbiased, in that its mean is equal to the non-noisy posterior $p_R(s_j)$. The multinomial distribution is a natural choice for the distribution of a signal of a probability distribution, as any realization $y = (y_1, ..., y_N)$ is indeed a probability distribution: each component y_i is between 0 and 1 and the components sum to one. It is the multi-state generalization of the binomial distribution used in Enke and Graeber (2023). The parameter η captures the precision of cognition: it is as-if the agent observed η draws from distribution $p_R(s_j)$. Therefore, a higher η corresponds to a more precise cognitive signal.

To form a belief about $p_R(s_j)$ from the cognitive signal, the agent needs a prior over $p_R(s_j)$. We assume this cognitive prior 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 mean and $1/\nu \geq 0$ scales the variance. The Dirichlet distribution is a natural choice for the cognitive prior distribution, since its support is the set of probability distributions over N objects. It is the multi-state generalization of the Beta prior distribution used in Enke and Graeber (2023). As in Enke and Graeber (2023), $\overline{p}_0$ has the interpretation of a *cognitive default*—that is, an agent's average prior belief about the posterior before processing the parameters of a particular learning environment. We assume that this default is the "ignorance prior", $\overline{p}_0(\omega_i) = 1/N$ for all $\omega_i \in \Omega$, such that, on average, the set of possible parameters do not result in a posterior belief over the state space that places greater weight on any one state. The parameter ν determines how concentrated the cognitive prior is around the default.

Given realized cognitive signal $y(s_j)$ after observing signal realization s_j, the agent uses Bayes' rule to form a posterior belief about $p_R(s_j)$. Since the Dirichlet distribution is the conjugate prior of the multinomial distribution, this posterior also follows a Dirichlet distribution with concentration parameters $\nu\overline{p}_0 + \eta y(s_j)$ and mean

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

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 $\mu(Y(s_j))$ conditional on $p_R(s_j)$, i.e., $E[\mu(Y(s_j))|p_R(s_j)]$. From Eq. (6), this is equal to

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

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 η) or the cognitive prior becomes more precise (higher ν), the subjective belief places more weight on the cognitive default (lower λ), and as cognition becomes more precise (higher η) or the cognitive prior becomes more diffuse (lower ν), the agent places more weight on the posterior belief derived from her mental representation without noise (higher λ). In a good news/bad news environment, cognitive imprecision leads to an underweighing of extreme states and generates underreaction.

This subjective posterior belief $\hat{p}(s) = (\hat{p}(\omega_1|s), ..., \hat{p}(\omega_N|s))$ forms the basis of our theoretical analysis. It incorporates how channeled attention and cognitive imprecision interact with the properties of the learning environment to impact belief formation. Note that when $\lambda = 1$ and $\theta = 0$, it is equal to the objective Bayesian posterior p_B.

Discussion. In the second stage, an agent with limited processing capacity does not fully internalize the parameters of the information environment. To see the intuition, return to the example of an investor who forms beliefs about a new tech company. In different markets, there are different prior probabilities over unicorns, zombies, and intermediate types, as well as probabilities of a price increase for each of these types. The boundedly rational investor faces cognitive imprecision when adjusting to the parameters of each particular market, which dampens her response to the signal relative to a cognitively precise investor. This does not imply that the investor completely ignores differences across markets; she just does not fully adjust.

Prior work models cognitive imprecision as noisy processing with respect to the objective information environment (Augenblick et al. 2022; Enke and Graeber 2023).

A contribution of this paper is to consider processing noise with respect to the agent's mental representation of the environment, which, as we argue above, may differ from the actual environment due to distortions from other cognitive constraints (e.g., limited attention, memory).

An important assumption is that the cognitive default is the "ignorance prior." We provide empirical evidence for this assumption in Section 3.1. A direct implication is that the agent exhibits insensitivity to both the prior—base-rate neglect—and the new information conveyed by the signal—signal-diagnosticity neglect. Both components play a role in our prior asymmetry prediction; hence, the empirical support for this prediction in Section 3 provides evidence of both forms of neglect. We compare our model with other models of cognitive imprecision in Appendix C, including Augenblick et al. (2022) which generates a more flexible form of signal-diagnosticity neglect by allowing the cognitive default to vary.

Noisy cognition is related to the anchoring-and-adjustment heuristic in the judgment and decision-making literature (Tversky and Kahneman 1974), where an agent enters a decision environment with an "anchor" belief $\overline{p}_0$ and insufficiently adjusts to new information (see Enke and Graeber (2023) for a similar discussion). We are not the first to consider the relationship between the representativeness and anchoring-and-adjustment heuristics (see discussion in Griffin and Tversky (1992)), but our model is unique in formally developing its predictions for belief-updating.

2.3 State-by-State Predictions

We first investigate which states are overweighed versus underweighed, and show that the two-stage model generates distinct predictions from either stage on its own. An agent *overweighs* state ω_i if her subjective posterior assigns a higher probability to it than the objective posterior, and similarly for any set of states, with the opposite for underweighing. The definition with respect to sets of states is analogous.

Definition 1. *An agent* overweighs ω_i *if* $\hat{p}(\omega_i|s_j) > p_B(\omega_i|s_j)$ *and* underweighs *it if* $\hat{p}(\omega_i|s_j) < p_B(\omega_i|s_j)$. *An agent* overweighs *set of states* $\Omega_i \subset \Omega$ *if* $\hat{p}(\Omega_i|s_j) > p_B(\Omega_i|s_j)$ *and* underweighs *it if* $\hat{p}(\Omega_i|s_j) < p_B(\Omega_i|s_j)$.

Note that if a state or set of states is overweighed, then at least one other state must be underweighed, and vice versa. This implies that in binary state environments, it is not possible to have both states simultaneously under- or overweighed.

Let $\omega_R(s_j)$ denote the most representative state for signal realization s_j (e.g., ω_1 for s_1 and ω_N for s_2) and analogously, ω_{NR} denote the least representative state. The following result establishes that there is a set of parameters (θ, λ) for which the agent overweighs both the most and least representative state. This prediction uniquely stems from the interaction between channeled attention and cognitive imprecision: it cannot arise when only one of these mechanisms is present. In contrast, for sufficiently low cognitive imprecision and high representativeness, the agent overweighs the most representative state and underweighs the least, as in the representativeness-only model, and for sufficiently high cognitive imprecision and low representativeness, the agent underweighs the most representative state and overweights the least, as in the cognitive-imprecision-only model. Finally, representativeness also drives the agent to underweigh the set of interior states.

Prediction 1. *Consider a symmetric information environment (Ω, p_0) with a uniform prior. For each $\theta > 0$, there exist cutoffs $0 < \overline{\lambda}_1(\theta) \leq \overline{\lambda}_2(\theta) < 1$, where $\overline{\lambda}_1(\theta) < \overline{\lambda}_2(\theta)$ when $|\Omega| > 2$ and $\overline{\lambda}_1(\theta) = \overline{\lambda}_2(\theta)$ when $|\Omega| = 2$, such that there are three regions: (i) Cognitive-imprecision-dominant: for $\lambda \in [0, \overline{\lambda}_1(\theta))$, the agent* underweighs *ω_R and* overweighs *ω_{NR}; (ii) Cognitive complementarity: for $\lambda \in (\overline{\lambda}_1(\theta), \overline{\lambda}_2(\theta))$, the agent* overweighs *ω_R and ω_{NR}; (iii) Representativeness-dominant: for $\lambda \in (\overline{\lambda}_2(\theta), 1]$, the agent* overweighs *ω_R and* underweighs *ω_{NR}. Moreover, for each $\theta > 0$, the agent underweighs the set of interior states $\Omega_I = \Omega \setminus \{\omega_R, \omega_{NR}\}$ when $\lambda > 0$ and neither under- nor overweighs Ω_I when $\lambda = 0$.*

The intuition is as follows. In the first stage, representativeness prompts the agent to overweigh the most representative state and underweigh the least. In the second stage, cognitive imprecision acts as a counteracting force by pulling beliefs towards the uniform cognitive default. When cognitive imprecision is low (the representativeness-dominant region), the agent continues to overweigh ω_R and underweigh ω_{NR}. In contrast, when cognitive imprecision dominates (the cognitive-imprecision-dominant region), the pattern is reversed.[24] Most importantly, in environments with at least three states, there is an intermediate range of representativeness and cognitive imprecision (the cognitive complementarity region) where the agent overweighs *both* ω_R and ω_{NR}.[25] Representativeness pulls mass towards the most representative state from multiple other states, and as a result, directs more probability mass to the most representative state than away from the least representative state. Hence, moderate levels of cognitive imprecision reverse the representativeness-driven underweighing for the least representative state but not the most representative state. Notably, as discussed above, the interaction between representativeness and cognitive imprecision generates this pattern.

Interior states are underweighed overall, as representativeness moves more mass to ω_R than away from ω_{NR}. Therefore, while individual interior states may be overweighed or underweighed, depending on their representativeness, as a whole this excess mass must be moved from interior states. This contrasts with the cognitive-imprecision-only model ($\lambda = 0$), where excess weight on interior states averages to

[24]This is also the case in the more flexible model of cognitive imprecision in Augenblick et al. (2022), as we show in Appendix C.

[25]Such a region cannot exist in binary state environments, as overweighing one state implies underweighing the other. This is another reason why restricting attention to binary state environments does not provide a complete picture of belief formation.

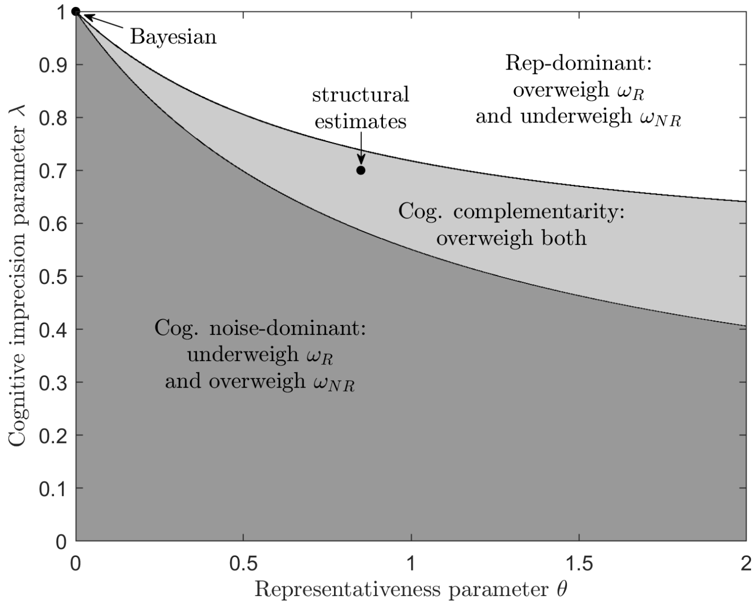


FIGURE 1. Illustration of Prediction 1 for $\Omega = (0.1, 0.3, 0.5, 0.7, 0.9)$

zero—as a set, they are neither under- nor overweighed.[26]

Fig. 1 illustrates the three regions for an information environment with five states. As shown in the figure, our parameter estimates from Section 3.2 fall in the cognitive complementarity region for this information environment.

Comparative static predictions on how properties of the information environment impact over- versus underweighing of individual states are difficult to derive, as varying these properties impacts both the expected belief movement and the impact of each mechanism. For example, as complexity increases, the prior probability of each state decreases, leading to less objective movement, but representativeness has a larger impact, leading to more subjective movement. Therefore, the overall impact is ambiguous. This motivates our next section, where we define a measure that is normalized to parse out differences in objective belief movement across environments.

2.4 Predictions on Under/overreaction

We next derive predictions on how under/overreaaction—measured across all states—varies with the information environment. Again, the two-stage model generates distinct predictions from either stage in isolation.

[26]The predictions for individual interior states are dependent on the details of the information environment. When representativeness is sufficiently strong, the agent channels most attention to the most representative state and underweighs all interior states, even relatively representative states that are more likely than average to generate a given signal realization. In this case, cognitive imprecision can actually counteract this underweighing: if the subjective posterior assigns less mass to an interior state than the cognitive default, then cognitive imprecision pulls the belief back towards the cognitive default, mitigating the underweighing of this state. This demonstrates that cognitive imprecision does not always contribute to the underweighing of more representative states.

2.4.1 Measuring Over- and Underreaction

We define overreaction based on a comparison of the expected state under the subjective and objective posteriors. Let $\hat{E}(\omega|s_j) = \sum_{\omega_i \in \Omega} \hat{p}(\omega_i|s_j)\omega_i$ denote the subjective posterior expected state following signal realization s_j, with an analogous definition for the objective posterior expected state $E_B(\omega|s_j)$. The *subjective expected movement* following signal realization s_j is the absolute value of the difference between the subjective posterior and prior expected state, $|\hat{E}(\omega|s_j) - E_0(\omega)|$, and analogously for the *objective expected movement* $|E_B(\omega|s_j) - E_0(\omega)|$. An agent *overreacts* to s_j if her subjective expected movement is greater than the objective expected movement and *underreacts* if it is less.

Definition 2 (Over- and Underreaction)**.** *The agent* overreacts *to s_j if $|\hat{E}(\omega|s_j) - E_0(\omega)| > |E_B(\omega|s_j) - E_0(\omega)|$ and* underreacts *if $|\hat{E}(\omega|s_j) - E_0(\omega)| < |E_B(\omega|s_j) - E_0(\omega)|$.*

For a given level of representativeness θ and cognitive imprecision λ, whether over- or underreaction emerges varies across signal realizations and information environments.

Our goal is to compare the extent of over/underreaction across these dimensions. In order to do so, we need a measure that accounts for variation in the objective expected movement. We measure the magnitude of over/underreaction by the difference between the objective and subjective expected movement, divided by the objective expected movement, which we refer to as the *overreaction ratio*:[27]

$$r(s_j) \equiv \frac{|\hat{E}(\omega|s_j) - E_0(\omega)| - |E_B(\omega|s_j) - E_0(\omega)|}{|E_B(\omega|s_j) - E_0(\omega)|}. \tag{8}$$

Given that the denominator of Eq. (8) is positive, by Definition 2 the agent overreacts to s_j if $r(s_j) > 0$ and underreacts if $r(s_j) < 0$.

To glean intuition for how representativeness and cognitive imprecision impact the overreaction ratio, consider a symmetric information environment (i.e., symmetric Ω and p_0). In this case, the overreaction ratio simplifies to

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

where

$$r_R(s_j) \equiv \frac{|E_R(\omega|s_j) - E_0(\omega)| - |E_B(\omega|s_j) - E_0(\omega)|}{|E_B(\omega|s_j) - E_0(\omega)|} \tag{10}$$

is the overreaction ratio and $E_R(\omega|s_j)$ is the posterior expected state under no cognitive imprecision (i.e., with respect to p_R).[28] Representativeness overweighs extreme

[27]This measure satisfies scaling irrelevance. When all possible states are close to each other, the numerator of $r(s_j)$ is naturally small, and the opposite is true if the states are very far apart. For example, if we double the value of all states, the numerator is automatically doubled. However, $r(s_j)$ stays the same when all states are scaled.

[28]In a symmetric environment, the prior expected state and the expected state under the cognitive

states, so $r_R(s_j) > 0$ for all $\theta > 0$. Therefore, when there is no cognitive imprecision, overreaction emerges: $r(s_j) > 0$ when $\theta > 0$ and $\lambda = 1$. This stems from the good news/bad news signal structure; as shown in Section 6.1, representativeness can generate underreaction for other types of signal structures. In contrast, when there is no representativeness, underreaction emerges: $r(s_j) = -(1-\lambda) < 0$ when $\theta = 0$ and $\lambda < 1$.[29] This highlights the opposing influences of representativeness and cognitive imprecision: when representativeness dominates, $r_R(s_j) > (1-\lambda)/\lambda$ and overreaction emerges, while when $r_R(s_j) < (1-\lambda)/\lambda$, cognitive imprecision dominates and underreaction emerges. While $(1-\lambda)/\lambda$ is a positive constant, $r_R(s_j)$ ranges from 0 to a potentially large number depending on θ and the information environment.[30] Therefore, the direction and magnitude of reaction can vary across environments.

Discussion. Our definition and measure of overreaction are based on the expected state. This is consistent with both the finance and experimental literatures. The former typically studies asset prices and average forecasts, which are summary statistics of the belief distribution similar in spirit to the expected state.[31] The latter typically compares the movement of subjective and objective beliefs in binary state environments—which is equivalent to our comparison of expected states.[32]

2.4.2 Comparative Statics

Our model gives rise to a rich set of predictions for belief-updating across information environments, as captured by the overreaction ratio. We derive comparative static predictions for how the overreaction ratio varies with respect to state space complexity, signal diagnosticity, and prior concentration and symmetry. Throughout this section, when we compare two environments (Ω, p_0) and (Ω', p_0'), we let $r(s_j)$ denote the overreaction ratio for (Ω, p_0) and $r'(s_j)$ analogously for (Ω', p_0').

Complexity of the State Space. To explore how complexity impacts belief-updating, we fix the dispersion of the state space—so that the most salient states are the same across environments—and vary complexity by adding more interior states. Prediction 2 shows that when the attentional distortion is sufficiently large,

default are both equal to 1/2. Eq. (9) follows from plugging $\hat{E}(\omega|s_j) = \lambda E_R(\omega|s_j) + (1-\lambda)\overline{E}(\omega)$ into Eq. (8), where $\overline{E}(\omega)$ is the expected state under the cognitive default $\overline{p}_0$.

[29] Section 4.2 provides direct empirical evidence for this prediction by suppressing representativeness as a salience cue.

[30] As θ approaches ∞ and $|E_B(\omega|s_j) - E_0(\omega)|$ approaches zero, $r_R(s_j)$ approaches ∞.

[31] 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 model, the counterparts of forecast errors and forecast revisions are $E_B(\omega|s_j) - \hat{E}(\omega|s_j)$ and $\hat{E}(\omega|s_j) - E_0(\omega)$, respectively. It is straightforward to verify that if $\hat{E}(\omega|s_j)$ moves in the same direction as $E_B(\omega|s_j)$, then $r(s_j) < 0$ if and only if forecast errors and revisions are negatively correlated, i.e. $(E_B(\omega|s_j) - \hat{E}(\omega|s_j))(\hat{E}(\omega|s_j) - E_0(\omega)) > 0$.

[32] When Ω is binary, we show that $r(s_j) > 0$ if and only if $|\hat{p}(\omega_1|s_j) - p_0(\omega_1)| > |p_B(\omega_1|s_j) - p_0(\omega_1)|$, and similarly for ω_2. See Appendix B for a proof.

overreaction increases as the state space becomes more complex.

Prediction 2 (Complexity)**.** *Consider two symmetric information environments* (Ω, p_0) *and* (Ω', p'_0) *with the same dispersion (i.e.,* $\omega_1 = \omega'_1$ *and* $\omega_N = \omega'_N$*) and uniform priors. If* Ω' *is more complex than* Ω*, and every state in* $\Omega' \setminus \Omega$ *is more interior than every state in* Ω*, then for sufficiently large* θ*, the agent overreacts more in* (Ω', p'_0) *than* (Ω, p_0)*,* $r'(s_j) > r(s_j)$ *for* $s_j \in \mathcal{S}$*.*

For example, for sufficiently large θ, the agent overreacts more in the 3-state environment $\Omega' = \{0.3, 0.5, 0.7\}$ or the four-state environment $\Omega'' = \{0.3, 0.4, 0.6, 0.7\}$ than in the binary state environment $\Omega = \{0.3, 0.7\}$, and overreacts even more in the five-state environment $\Omega''' = \{0.3, 0.4, 0.5, 0.6, 0.7\}$ than the four-state environment $\Omega'' = \{0.3, 0.4, 0.6, 0.7\}$.[33] The intuition is as follows. Under a uniform prior, as complexity increases, mass is shifted from extreme states to interior states. Since the extreme states become less likely under the prior, the objective expected movement is smaller. However, because the agent's mental representation neglects interior states, her posterior belief continues to concentrate on extreme representative states, resulting in less of a reduction in the subjective expected movement.

The impact of increasing complexity on overreaction critically hinges on how the addition of states changes the relative levels of representativeness. Such changes are substantial when the additional states are distinct, but not when they are very similar. For example, the overreaction ratio moves continuously in $\varepsilon > 0$ from $\Omega = \{0.3, 0.7\}$ to $\Omega' = \{0.3, 0.3 + \varepsilon, 0.7, 0.7 + \varepsilon\}$, maintaining a uniform prior. At $\varepsilon = 0$, $\Omega' = \{0.3, 0.3, 0.7, 0.7\}$ is equivalent to Ω, and therefore, the two state spaces have equal overreaction ratios. Therefore, the impact of increasing complexity is not determined by the number of new states per se, but by the number of new *distinct* states, as this is what alters how much attention the agent channels towards representative states.[34]

Prediction 2 also holds in a representativeness-only model ($\theta > 0$ and $\lambda = 1$), but such a model predicts overreaction across all levels of complexity. This contrasts with our two-stage model, where the interaction between the two cognitive mechanisms can generate underreaction in simple environments and overreaction in complex environments.

Prior Concentration. We next explore how the concentration of the prior impacts belief movement. Consider two symmetric information environments (Ω, p_0) and

[33]Note that the 3-state environment is not directly comparable to the 4-state or 5-state environment because the states $\{0.4, 0.6\}$ are not more interior than the state 0.5.

[34]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 simplify 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.

(Ω, p'_0) with the same state space. We say that prior p'_0 is more *concentrated* than p_0 if it assigns higher probability to interior states and lower probability to extreme states: for some $c \in (1/2, 1)$, $p'_0(\omega_i) \geq p_0(\omega_i)$ for all $\omega_i \in [1-c, c]$ and $p'_0(\omega_i) \leq p_0(\omega_i)$ for all $\omega_i \in [0, 1-c] \cup [c, 1]$, with at least one inequality strict. Prediction 3 establishes that overreaction increases in the concentration of the prior.

Prediction 3 (Prior concentration)**.** *Consider two symmetric information environments (Ω, p_0) and (Ω, p'_0) with the same state space. If p'_0 is more concentrated than p_0, then for sufficiently large θ, the agent overreacts more in (Ω, p'_0) than in (Ω, p_0), $r'(s_j) > r(s_j)$ for $s_j \in \mathcal{S}$.*

The intuition behind Prediction 3 is similar to that of Prediction 2. The objective expected movement decreases in the concentration of the prior, but representativeness continues to generate overweighing of extreme states. This leads to a smaller decrease in the subjective expected movement, and therefore, more overreaction.

While a representativeness-only model predicts overreaction for all priors and a cognitive-imprecision-only model predicts underreaction, our two-stage model allows for overreaction to some priors and underreaction to others—including those that are sufficiently diffuse. Specifically, the framework predicts a region where the two psychological mechanisms act as cognitive complements—their interaction plays a critical role in predicting whether over- or underreaction emerges in a given information environment. Our next prediction highlights this cognitive complementary.

Prediction 4. *Consider the set $\mathbf{\Omega}$ of symmetric information environments with state space Ω. For each $\theta > 0$, there exist cutoffs $0 < \overline{\lambda}_1(\theta) < \overline{\lambda}_2(\theta) < 1$ such that:*

(i) Cognitive-imprecision-dominant: for $\lambda \in [0, \overline{\lambda}_1(\theta))$, the agent underreacts to all information environments in $\mathbf{\Omega}$.

(ii) Cognitive complementarity: for each $\lambda \in (\overline{\lambda}_1(\theta), \overline{\lambda}_2(\theta))$, there exists a positive measure set of information environments in $\mathbf{\Omega}$ on which the agent overreacts and a positive measure set on which the agent underreacts. The latter set includes all environments with a sufficiently diffuse prior p_0 such that $p_0(\{\omega_1, \omega_N\}) > c_1$ for some $c_1 \in (0, 1)$.

(iii) Representativeness-dominant: for $\lambda \in (\overline{\lambda}_2(\theta), 1]$, the agent overreacts to all information environments in $\mathbf{\Omega}$.

Signal Diagnosticity. To examine how the signal diagnosticity impacts belief movement, we consider the differential reaction to information as the signal becomes (weakly) less diagnostic in *all* states—in other words, as all states move closer to 0.5. Moving interior states closer to 0.5 has a similar impact to adding interior states—it causes representativeness to generate a more distorted mental representation. The impact of moving extreme states is more nuanced, as this involves changing the value of the most representative state.

To illustrate this, consider information structure $\Omega_3 = \{1-x, 0.5, x\}$ and $p_0 = 1/3$, where $x \in (0.5, 1)$. As x decreases towards 0.5, the objective expected movement decreases, since the extreme states x and $1-x$ are closer to the prior expected state $E_0(\omega) = 0.5$. The subjective expected movement also decreases: representativeness causes the agent to overweigh x or $1-x$, which are moving closer to $E_0(\omega)$. Decreasing x results in a higher overreaction ratio if the objective expected movement decreases more. This turns out to hold for all values of x when the degree of representativeness is sufficiently high. More generally, our next result shows that under a uniform prior, decreasing the diagnosticity of the extreme states results in more overreaction for sufficiently large θ if

$$W(\Omega) \equiv \sum_{i \in \{1,N\}} (\omega_i - 0.5)^2 - \sum_{i \notin \{1,N\}} (\omega_i - 0.5)^2 > 0. \tag{11}$$

Note that $W(\Omega) > 0$ for all symmetric state spaces with 2, 3, 4 or 5 states, which includes all information structures we consider in the experiment (see Table D.1). When $W(\Omega) > 0$, the signal is more informative about extreme states and less informative about interior states; therefore, the objective posterior attaches higher probability to an extreme state relative to interior states. This makes the objective expected movement more sensitive to the values of the extreme states.[35]

Prediction 5 (Diagnosticity). *Consider two symmetric information environments (Ω, p_0) and (Ω', p_0') with the same complexity, uniform priors, and $W(\Omega) > 0$ and $W(\Omega') > 0$. If Ω' is less diagnostic than Ω, $d_i' \leq d_i$ for all $i = 1, ..., N$ with at least one inequality strict, then for sufficiently large θ, the agent overreacts more in (Ω', p_0') than (Ω, p_0), $r'(s_j) > r(s_j)$ for $s_j \in \mathcal{S}$.*

For example, consider $\Omega_4 = \{x, y, 1-y, 1-x\}$ with $x \in (0, 0.5)$ and $y \in (x, 0.5)$. Prediction 5 implies that the agent overreacts more or underreacts less as both x and y move closer to 0.5.

Analogous to Prediction 4, our two-stage model predicts three regions as we manipulate the signal diagnosticity, including a cognitive complementarity region in which an agent underreacts to sufficiently precise signals and overreacts to less precise signals.[36] In the interest of space, we present this result in Appendix B.

Asymmetric Prior. Finally, we consider information environments with asymmetric priors. We restrict attention to binary state spaces where it is straightforward to manipulate the symmetry of the prior; in this case, an asymmetric prior corresponds to $p_0(\omega_1) \neq p_0(\omega_2)$. In such environments, the two signal realizations are no

[35] When $W(\Omega) < 0$, the objective expected movement is less sensitive to changes in the extreme states. In this case, decreasing the diagnosticity of extreme states reduces the subjective expected movement more than the objective expected movement, leading to less overreaction.

[36] The more flexible model of cognitive imprecision in Augenblick et al. (2022) also predicts overreaction to noisy signals and underreaction to precise signals. See Appendix C for discussion.

longer equally likely ex-ante. Reaction may differ based on whether the agent observes the more likely expected—*confirmatory*—signal realization, or the less likely surprising—*disconfirmatory*—signal realization. For example, if the prior assigns higher probability to state ω_1, then a signal realization is confirmatory if it is more likely under ω_1 than ω_2, and is disconfirmatory if it is more likely under ω_2 than ω_1.

Definition 3. *Fix a symmetric information environment (Ω, p_0) with a binary state space. A signal realization s_j is* confirmatory *if (i) $p_0(\omega_1) > p_0(\omega_2)$ and $\pi(s_j|\omega_1) > \pi(s_j|\omega_2)$, or (ii) $p_0(\omega_1) < p_0(\omega_2)$ and $\pi(s_j|\omega_1) < \pi(s_j|\omega_2)$. A signal realization s_j is* disconfirmatory *if (iii) $p_0(\omega_1) > p_0(\omega_2)$ and $\pi(s_j|\omega_1) < \pi(s_j|\omega_2)$, or (iv) $p_0(\omega_1) < p_0(\omega_2)$ and $\pi(s_j|\omega_1) > \pi(s_j|\omega_2)$.*

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 reaction to confirmatory versus disconfirmatory information. If the information is surprising, then it predicts overreaction to imprecise signals and underreaction to precise ones. If the information is expected, the model predicts a non-monotonicity with respect to diagnosticity. When the signal is either very precise or relatively imprecise, it predicts underreaction; at the same time, there is a region of intermediate precision where overreaction can emerge. Importantly, the model also predicts that people may even react in the wrong direction for a sufficiently imprecise confirmatory signal (or for high enough cognitive imprecision, regardless of signal precision).[37]

Prediction 6 (Asymmetric Prior)**.** *Consider an information environment (Ω, p_0) with a binary symmetric state space, asymmetric prior, and diagnosticity d.*

(i) There exists cutoff $c_1 \in (0.5, 1)$ such that following a disconfirmatory signal realization, the agent overreacts *if $d \in (0.5, c_1)$ and* underreacts *if $d \in (c_1, 1)$.*

(ii) There exist cutoffs $0.5 < c_2 \leq c_3 \leq c_4 \leq 1$ such that following a confirmatory signal realization, the agent reacts in the wrong direction when $d \in (0.5, c_2)$, underreacts *when $d \in (c_2, c_3) \cup (c_4, 1)$ and* overreacts *when $d \in (c_3, c_4)$. If cognitive imprecision is sufficiently low (high λ), $c_2 < 1$ so the underreaction region exists, and if representativeness is sufficiently high (large θ), $c_3 < c_4$ so the overreaction region exists. If $c_3 < c_4$, then $c_2 < c_3$ and $c_4 < 1$.*

Cognitive imprecision drives wrong direction reaction. For intuition, consider the cognitive-noise-only model. In this case, the subjective expected state is equal to a weighted average of the objective expected state and the cognitive default, $\hat{E}(\omega|s_j) = \lambda E_B(\omega|s_j) + (1-\lambda)\overline{E}(\omega)$, where $\overline{E}(\omega) = 0.5$. Wrong direction reaction can arise when

[37] Wrong direction reaction occurs when the subjective and objective expected states move in opposite directions—$\hat{E}(\omega|s_j) - E_0(\omega)$ and $E_B(\omega|s_j) - E_0(\omega)$ have opposite signs. This is not possible in symmetric information environments, as $\overline{E}(\omega) = E_0(\omega)$.

E_B and $\overline{E}$ are on opposite sides of the prior expected state E_0—which can only occur for a confirmatory signal. Suppose the prior places more weight on ω_2, resulting in prior $E_0(\omega) > 0.5$. The objective expected state increases following confirmatory s_2, $E_B(\omega|s_2) > E_0(\omega)$. Cognitive imprecision compresses the subjective expected state towards the cognitive default, which is less than the prior, $\overline{E}(\omega) < E_0(\omega)$. For a sufficiently imprecise signal (low diagnosticity), $E_B(\omega|s_2)$ is close to $E_0(\omega)$ and this results in wrong direction reaction: $\hat{E}(\omega|s_2) \approx \lambda E_0(\omega) + (1-\lambda)\overline{E}(\omega) < E_0(\omega)$. For a more precise signal (higher diagnosticity), $\hat{E}(\omega|s_2)$ remains above $E_0(\omega)$ but below $E_B(\omega|s_2)$, resulting in underreaction.

In contrast, following disconfirmatory s_1, the objective expected state decreases, $E_B(\omega|s_1) < E_0(\omega)$. For a sufficiently imprecise signal, it remains above the cognitive default, $\overline{E}(\omega) < E_B(\omega|s_1)$; cognitive imprecision compresses $\hat{E}(\omega|s_1)$ towards $\overline{E}(\omega)$, decreasing it more than $E_B(\omega|s_1)$ and generating overreaction. As the signal becomes more precise, $E_B(\omega|s_1)$ decreases below the cognitive default, and cognitive imprecision instead increases $\hat{E}(\omega|s_1)$, resulting in underreaction.

When the agent is also subject to representativeness ($\theta > 0$), she reacts more to both signal realizations. If representativeness is strong enough, the agent may even overreact to a confirmatory signal of intermediate diagnosticity. Section 3.4.4 presents results consistent with these predictions, including the non-monotonicity with respect to signal diagnosticity.

Other Asymmetric Environments. Aside from the final result, we focus on symmetric information environments as this case yields tractable predictions. The insights generally continue to hold in asymmetric environments, albeit with more cumbersome notation. We do not explore these environments here for brevity.

3 Empirical Investigation

In this section, we test the predictions of our framework in a controlled experiment.

3.1 Experimental Set-up

Method. We recruited 3,797 participants from the Prolific crowdsourcing platform (49% female, 39 years average age).[38] They first had to pass an attention check before reading any experimental instructions. Those who did not pass did not proceed to the rest of the study, we did not collect data from them, and they are not included in the participant total. After passing the initial check, participants were told that in addition to the base payment of $2, they could earn two additional bonus payments. First, they earned $1 for correctly answering a comprehension check that followed the instructions. Second, they earned $10 if their response to a randomly-chosen belief

[38] Preregistration materials are available here: https://aspredicted.org/LTJ_CS7 and https://aspredicted.org/Q77_3LG.

elicitation question was within 3% of the objective Bayesian posterior.[39] 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.[40]

Design. After the initial attention check, participants read the experimental instructions that included the following description of the information environment:

> *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 next completed several comprehension questions, then proceeded to a series of inference tasks. Each trial involved a new information environment, a randomly selected bag, and a randomly drawn ball. 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 reported how likely she thought that each bag was selected (i.e., Bag 1, Bag 2, etc.) by reporting a percentage from 0 to 100.[41] 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. Each participant completed 8 to 15 inference tasks as described below, and then answered a set of basic demographic questions before exiting the study. See Appendix E for the full instructions.

This "bookbag-and-poker-chip" design (Edwards 1968) is used extensively in the literature—typically with a simple binary state space. It cleanly maps into 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

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

[40]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.

[41]To ensure that our results were not driven by the procedure of entering beliefs for every bag rather than the expectation, we ran a variation where participants reported their expectation $E(\omega|s_j)$ of the number of red balls directly. Importantly, they were aware of the complexity of the information environment and were instructed to consider each potential state (i.e., Bag) separately before reporting their expectation. This did not significantly change the results (see Appendix D.4). We did not use this elicitation method for the main analysis because it would prevent us from studying belief-updating state by state, which is critical for testing our model's predictions.

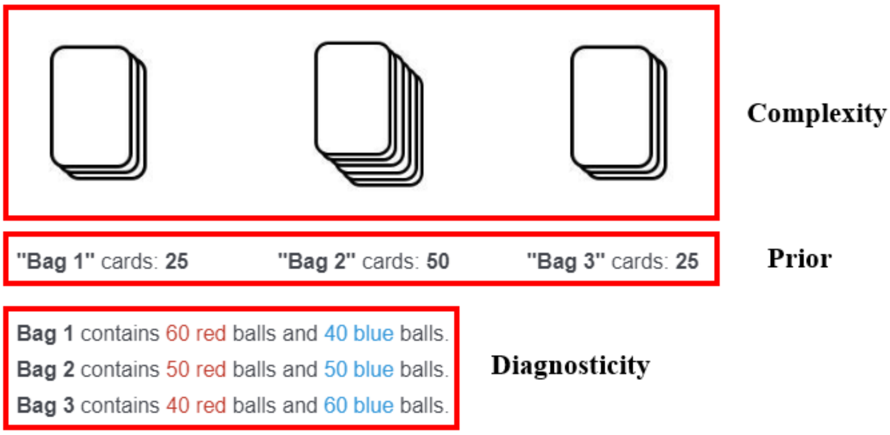


FIGURE 2. Experimental design for 3-state treatment

structure. As in Section 2, we set the share of red balls as the value of the state corresponding to a given bag, $\omega_i = Pr(r|\omega_i)$. Fig. 2 depicts an information environment with 3 states, with Bag 1 as state $\omega_3 = 0.6$, Bag 2 as $\omega_2 = 0.5$, and Bag 3 as $\omega_1 = 0.4$, a prior concentrated on the interior state (Bag 2), and a signal diagnosticity 0.6 in Bags 1 and 3 and 0.5 in Bag 2.

It is straightforward to manipulate the parameters of the information environment. 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 setting with three 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 setting with two bags, the prior probability of one versus the other bag.

Table D.1 in Appendix D.1 outlines the set of parameter combinations that we used. As in the model, we focused on symmetric information structures (e.g., if there is a bag with 40 red balls, there is also a bag with 60 red balls) and a good news/bad news structure. The most "representative" bag was Bag 1 or Bag N, depending on whether a red or blue ball, respectively, was drawn. Section 6.1 reports results from an informational environment that does not have a good news/bad news structure.

Participants were randomized into a complexity condition—2, 3, 4, 5 or 11 states. Those in the 3-state condition were also randomized into one of three prior concentration conditions—uniform, concentrated, or dispersed; those in the 2-state condition were randomized into one of two prior symmetry conditions. Participants in the 4-, 5-, and 11-state conditions all faced a uniform prior. Participants then completed a maximum of 15 trials randomly drawn from the set of possible trials for the respective complexity and prior condition.[42] Each complexity and prior condition had at least 200 participants.

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 specific parameters of the information environment.[43] 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 a randomly-selected guess was within 3% of the actual number of cards corresponding to that bag. Across both conditions, a joint F-test cannot reject that participants assigned the same probability to each bag. This is consistent with a uniform cognitive default, i.e., the "ignorance prior."

Analysis. We conduct three types of analyses on the experimental data. First, we structurally estimate the two model parameters, θ and λ. Next, we compare participants' reported posterior belief about each state to the objective Bayesian posterior. Finally, we calculate participants' subjective expected state from their reported posterior, and compare this to the objective posterior expected state using the overreaction ratio $r(s)$ defined in Section 2 (see Eq. (8)). These analyses are complementary: comparing the subjective and objective expected movement in beliefs allows us to examine over- versus underreaction, the state-by-state analysis demonstrates the unique predictions of our model, e.g., how the two psychological mechanisms interact, and addresses potential issues related to the overreaction ratio $r(s)$ (see Appendix D.1 for discussion), and the structural estimation lets us compare the observed patterns of belief updating in the data to the best-fit model prediction. For convenience, we use d to denote the signal diagnosticity associated with the extreme states, $d \equiv d_1 = d_N$.

It is worth noting that experimental studies on belief-updating often measure over- and underreaction by running a so-called *Grether regression* (Grether 1980). Due to our focus on multi-state settings, the rich set of model predictions cannot be tested using Grether regressions. See Appendix D.1 for further discussion.

3.2 Structural Estimation

We first use the experimental data to estimate the parameters of the belief-updating model, following the literature on behavioral structural estimation (e.g., DellaVigna (2018) and Bordalo, Gennaioli, Ma, and Shleifer (2020)) as outlined in Appendix D.6. The estimated parameter values of $\theta = 0.85$ and $\lambda = 0.70$ suggest that in the repre-

[42]For all conditions except the 2-state asymmetric prior, the total set of possible trials is equal to the product of the number of information structures and signal realizations (always 2). For the 2-state asymmetric prior condition, the total set of possible trials is equal to the product of the number of priors (2), information structures and signal realizations (2).

[43]Namely, participants were told that there were three or eleven potential bags but were not told the composition of bags in the deck or the composition of balls in each bag.

sentational stage, directing attention to the representative states leads participants to update beliefs as-if they are counting the signal *nearly twice*, while in the processing stage, participants' cognitive imprecision leads them to anchor on the cognitive default and adjust only 70% of the linear distance to the edited posterior. Both estimates are significantly different from the Bayesian benchmark of $\theta = 0$ and $\lambda = 1$.

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 λ close to 0.5. Bordalo et al. (2019) examine forecasters' expectations about a series of economic indicators and find that θ 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.

3.3 State-by-State Analysis

We next examine whether participants over- or underweigh a state by comparing the difference between the subjective and objective posterior belief at specific states. Recall that overweighing (underweighing) corresponds to a subjective posterior that places more (less) weight on a state or set of states than the objective posterior. We focus on states where our two-stage model makes a distinct theoretical prediction; from Prediction 1, these correspond to the most representative state, the least representative state, and the set of middle-representative states.

Fig. 3a plots the difference between participants' average subjective posterior and the objective posterior, aggregated across all 2, 3, 4 and 5-state uniform prior environments used in the experiment. As shown in the figure, participants overweigh both the most and least representative states—with the most representative state overweighed the most—and underweigh the set of middle-representative states.

This matches the prediction in the cognitive complementarity region of our two-stage model: as outlined in Prediction 1, the interaction between salience-channeled attention and cognitive imprecision leads the most salient state to be overweighed the most, the least salient state to be overweighted the second most, and the middle states to be underweighed. Fig. 3b plots this model prediction by computing the difference between the predicted subjective posterior at the structural estimates of θ and λ and the objective posterior, also aggregated across all 2, 3, 4 and 5-state uniform prior environments used in the experiment.

In contrast, the observed pattern of over- and underweighing is distinct from the prediction of each stage of our model in isolation.[44] As plotted in Figs. 3c and 3d, the representational stage alone predicts underweighing of the least representative state, while the processing stage alone predicts underweighing of the most representative state and accurate weighing of the set of middle states.

Aggregating across information environments masks interesting heterogeneity with

[44]It is also distinct from the alternative version of cognitive imprecision explored in Appendix C.

(A) Data

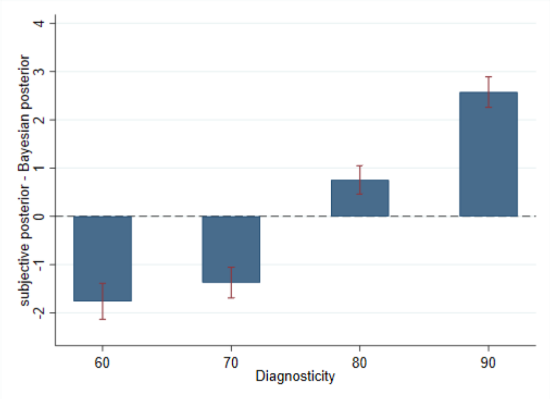
(B) Two-Stage Model

(C) Representativeness-Only Model

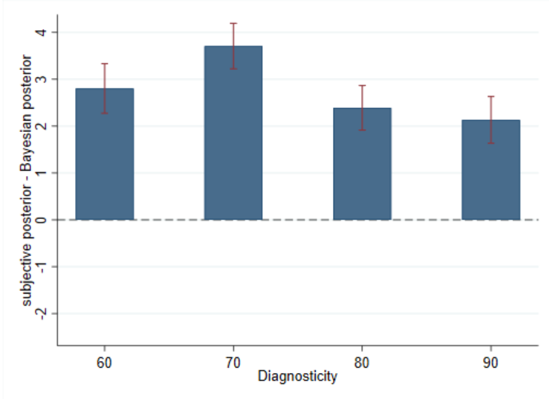
(D) Cognitive-Imprecision-Only Model

FIGURE 3. Over- and Underweighing by Representativeness of State. Each bar aggregates all 2, 3, 4 and 5-state uniform prior environments; (b)-(d) are weighted to match the share of experimental observations in each environment. Panels (b)-(d) are based on structural estimates of θ and λ: (b) $\theta = 0.85$, $\lambda = 0.7$; (c) $\theta = 0.85$, $\lambda = 1$; (d) $\theta = 0$, $\lambda = 0.7$.

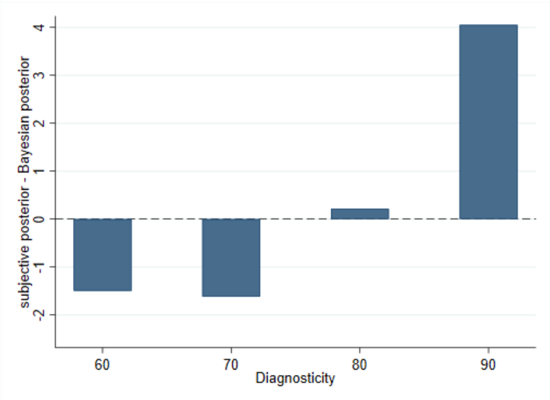
respect to signal informativeness. Consider the least representative state: at the estimated values of θ and λ, the two-stage model predicts underweighing in environments with an imprecise signal and overweighing in environments with a precise signal (Fig. 4c). When the signal is imprecise (low diagnosticity), the objective posterior is close to the cognitive default; cognitive imprecision has little room to generate overweighing and representativeness dominates, leading to underweighing. For precise signals (high diagnosticity), the effect reverses: the objective posterior is close to zero and representativeness has little room to generate underweighing; cognitive imprecision dominates, leading to overweighing. This pattern is borne out in the experimental data (Fig. 4a). On the other hand, the two-stage model predicts the overweighing of the most representative state across all diagnosticities, with the most overweighting for environments with intermediate precision (Fig. 4d). This is because representativeness leads to a larger distortion of the most representative state, as it pulls more weight to this state than from the least representative state. Again this



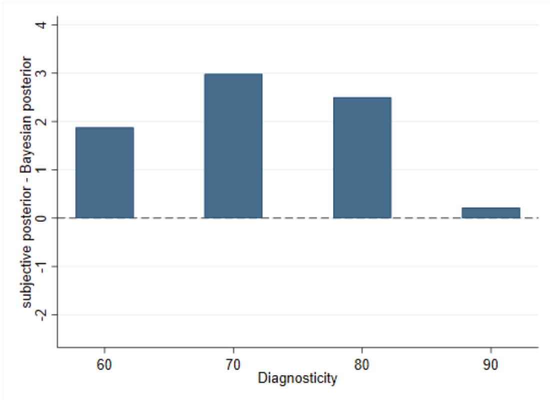
(A) Data: Least Rep. State



(B) Data: Most Rep. State



(C) Two-Stage Model: Least Rep. State



(D) Two-Stage Model: Most Rep. State

FIGURE 4. Over- and Underweighing by Diagnosticity. Each bar aggregates all uniform prior 2, 3, 4 and 5-state environments of diagnosticity d. Panels (c) and (d) are based on structural estimates $\theta = 0.85$ and $\lambda = 0.7$ and are weighted to match the share of experimental observations in each environment.

pattern matches the data (Fig. 4b).[45]

Overall, our parameter estimates generally fall in the cognitive complementarity region of Prediction 1 for individual environments as well: they are in this region for 15 of the 24 information environments with more than two states that we consider in the experiment. This suggests that the interaction between channeled attention and cognitive imprecision will play a key role in driving belief updating in many information environments. Our model also systematically predicts when one of the psychological mechanisms will be the predominant driver of belief updating in a given environment: the parameter estimates fall in the representativeness-dominant region for 8 of these environments—including those with lower diagnosticities, consistent with the experimental data—and in the cognitive-imprecision-dominant region for the remaining environment.

Finally, we examine belief heterogeneity among participants. The distribution of subjective beliefs for a given state and information environment is generally unimodal,

[45]As in the aggregate data, the observed pattern of over- and underweighing by diagnosticity is distinct from the prediction of each stage of our model in isolation (Fig. D.1 in Appendix D.2).

smooth, and centered around the objective posterior with an overweighing bias for the most representative state, an underweighing bias for the middle states, and a bias for the least representative state that varies in direction. Fig. D.2 plots the distribution of reported posteriors for 3-state environments.

3.4 Over- versus Underreaction

We next test our predictions on how properties of the information environment impact the extent of over- or underreaction, as measured by the overreaction ratio. In the experiment, the numeric value of a state corresponds to the fraction of red balls in the bag. The expected state is therefore equal to the expected probability of drawing a red ball. In each trial, we calculate (i) the participant's expected probability of drawing a red ball given their reported posterior belief, (ii) the objective prior expected probability of drawing a red ball, and (iii) the objective posterior expected probability of drawing a red ball. We use these statistics to compute the overreaction ratio $r(s)$ as defined in Eq. (8) for each participant, and use the average of the overreaction ratios across all participants in a given information environment to measure over- and underreaction.[46] Recall that a positive (negative) $r(s)$ corresponds to over- (under-) reaction.

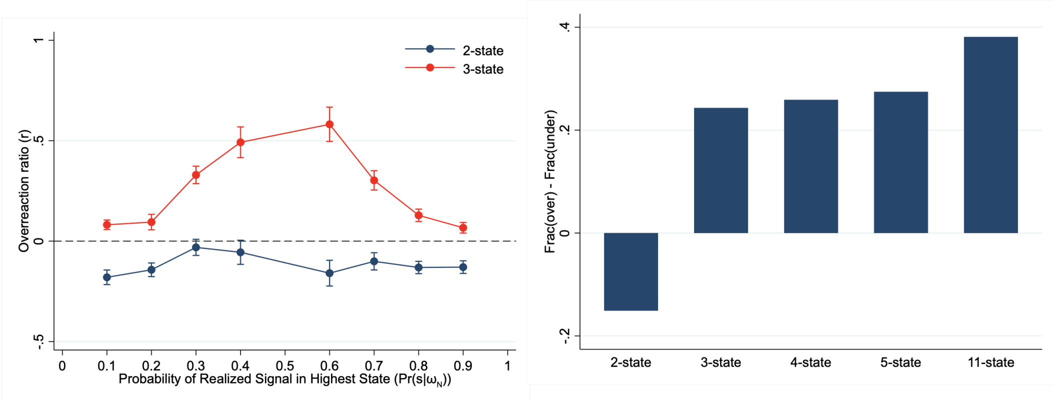
3.4.1 Complexity

To test Prediction 2, we compare the overreaction ratio across uniform prior environments that vary in complexity while holding fixed the dispersion of the state space (i.e., the highest and lowest states). In simple 2-state environments, we replicate the finding of underreaction from the experimental literature: on average, participants' overreaction ratio is negative across all the information environments in our experiment ($r < 0$, $p < .001$).[47] This finding also holds for individual information environments. Fig. 5a plots the overreaction ratio for each diagnosticity and signal realization. The x-axis corresponds to the probability of the realized signal in state ω_N, which ranges from 0.6 to 0.9 for a red ball and from 0.1 to 0.4 for a blue ball depending on the diagnosticity.[48] As can be seen in the figure, we observe significant underreaction to both signal realizations across nearly all environments in the 2-state treatment. This is consistent with the evidence summarized in Benjamin (2019), which documents systematic underreaction in lab experiments that

[46] Per our pre-registration, unless otherwise noted, we exclude trials in which participants react in the wrong direction. Appendix D.3.2 replicates the analyses including wrong direction observations; the results do not qualitatively change.

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

[48] Due to the symmetry of the information structure, when a blue ball occurs with probability x in state ω_N, then a red ball occurs with probability $1 - x$. Therefore, on the x-axis of Fig. 5a, 0.1 and 0.9 correspond to blue and red signal realizations, respectively, from the information structure with diagnosticity 0.9, and so on.



(A) Overreaction ratio

(B) Difference in share of overreaction versus underreaction

FIGURE 5. Complexity increases overreaction. Each data point aggregates all uniform prior environments of a given complexity—by diagnosticity and signal realization on the left and across all diagnosticities and signal realizations on the right.

overwhelmingly use binary state spaces.

Increasing the complexity of the state space reverses this result. Strikingly, adding even a single state—going from 2 to 3 states—leads participants to report posterior beliefs that move significantly *more* than the objective benchmark, resulting in a positive overreaction ratio ($r > 0$, $p < .001$). As illustrated in Fig. 5a, we observe significant overreaction to both signal realizations across all environments in the 3-state treatment.

This pattern continues in more complex settings. We compare each 2-state environment to 4-state environments with two additional interior states and 5-state environments with three additional interior states.[49] Regressing the overreaction ratio on dummies corresponding to the 4-state and 5-state treatments, we find that the overreaction ratio is significantly higher in the complex 4 and 5-state treatments compared to the simple 2-state treatment. Moreover, the overreaction ratio is significantly higher in the 5-state treatment than the 4-state treatment ($p < .01$), as predicted since the former adds an interior state to the latter. A similar pattern of overreaction increasing with complexity emerges when we control for the information structure by adding dummies for each diagnosticity. See Table D.2 for these results.

Finally, we study an 11-state treatment to examine belief-updating with "many" states. We find significant overreaction ($r > 0$ $p < .001$) and the highest overreaction ratio across all complexity treatments (although this is not a direct test of Prediction 2 since the state space dispersion differs).

As an alternative measure of overreaction, we compute the difference between the fraction of trials with overreaction versus underreaction; this measure is used in prior work e.g., Fan et al. (2023). A positive value indicates more trials with overreaction

[49]We do not have a prediction for how the 3-state treatment compares to the 4-state and 5-state treatments because the latter do not add interior states to the former as required for Prediction 2.

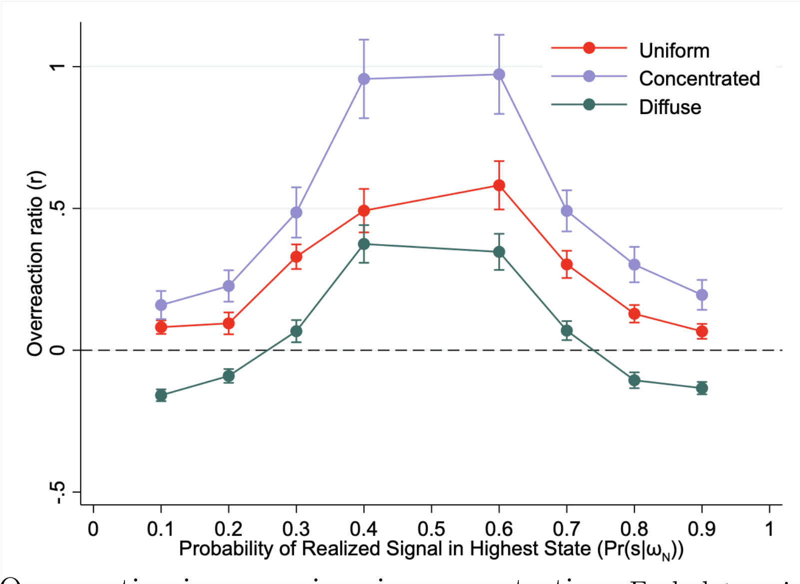


FIGURE 6. Overreaction increases in prior concentration. Each data point aggregates all 3-state environments of a given prior by diagnosticity and signal realization.

and a negative value indicates the opposite. As shown in Fig. 5b, we again find that participants tend to underreact in the 2-state treatment but overreact in treatments with 3 or more states. Together, these results provide strong support for Prediction 2.

3.4.2 Prior Concentration

To test Prediction 3, we examine how the overreaction ratio varies with the concentration of the prior. We focus on 3-state environments as this is the minimum number of states needed to manipulate the prior concentration. We varied the prior from a diffuse prior $(0.4, 0.2, 0.4)$ that placed twice as much mass on the extreme states as the interior state, to a uniform prior $(0.33, 0.34, 0.33)$, to a concentrated prior $(0.25, 0.50, 0.25)$ that placed twice as much mass on the interior state as the extreme states. Consistent with Prediction 3, we observe significantly more overreaction as the prior becomes more concentrated: regressing the overreaction ratio on dummies for each prior, we find that participants overreact significantly more when the prior is concentrated and significantly less when the prior is diffuse (Table D.3). As shown in Fig. 6, this holds for both signal realizations across all information environments—where again the x-axis corresponds to the probability of the realized signal in state ω_N, ranging from 0.6 to 0.9 for red and 0.1 to 0.4 for blue. Taken together, this provides strong support for Prediction 3.

At high diagnosticities, the data matches the cognitive complementarity region outlined in Prediction 4, where underreaction emerges for a sufficiently diffuse prior and overreaction emerges for more concentrated priors. Fig. 6 shows that this pattern emerges and is significant for the state spaces that correspond to diagnosticities $d = 0.8$ and 0.9 (0.2/0.8 and 0.1/0.9 on the x-axis). In contrast, at low diagnosticities, the data is consistent with the representativeness-dominant region outlined in

Prediction 4: for the state spaces that correspond to $d = 0.6$ and 0.7 ($0.3/0.7$ and $0.4/0.6$ on the x-axis), significant overreaction emerges for all three priors we consider. Indeed, consistent with this finding, our structural estimates of θ and λ lie in the cognitive complementarity region for $d = 0.8$ and 0.9 and the representativeness-dominant region for $d = 0.6$ and 0.7 (see Figure X). Taken together, this provides further evidence for how channeled attention and cognitive imprecision interact to generate distinct predictions on how the emergence of over- versus underreaction varies with the learning environment.

3.4.3 Signal Diagnosticity

To test Prediction 5, we examine how the overreaction ratio varies with signal diagnosticity. Fig. 6 provides support for the prediction in the 3-state treatment: there is significantly lower overreaction at higher diagnosticities, with the highest level of overreaction as d approaches 0.5 (noisiest signal; 0.5 on the x-axis) and the lowest as d approaches 0.9 (most precise signal; $0.1/0.9$ on the x-axis). While Prediction 5 is derived for a uniform prior, this pattern also holds for the diffuse and concentrated priors we consider.

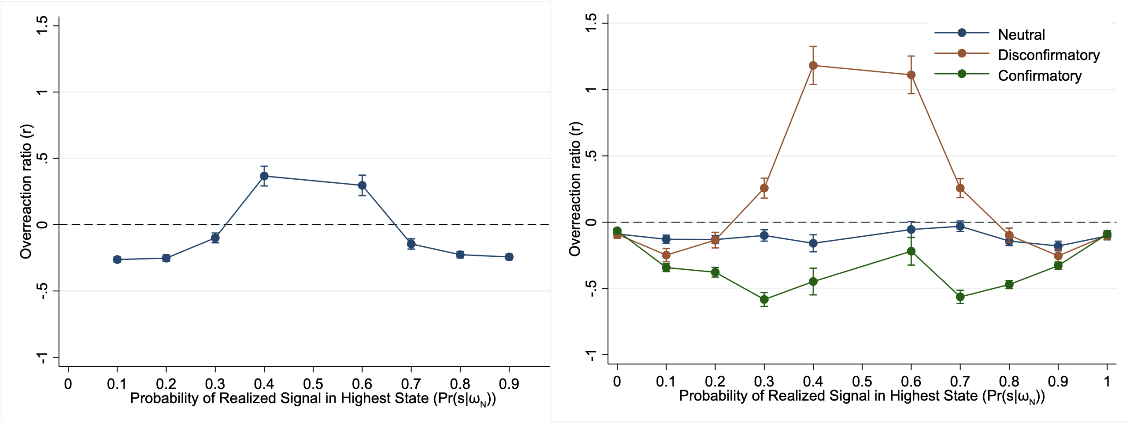
A similar pattern holds for the other complexity treatments. Table D.4 presents a regression analysis of the overreaction ratio on signal diagnosticity for each complexity treatment. Consistent with the prediction, there is progressively less overreaction as d increases and the signal is more precise. 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.[50]

3.4.4 Prior Symmetry

To test Prediction 6, we examine how the overreaction ratio varies with the type of signal realization—confirmatory or disconfirmatory—in asymmetric prior environments. We focused on 2-state environments and varied 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. 7a 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.

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)

[50] 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. We do not include this result in the figures because we did not run this signal diagnosticity in other complexity or prior information environments.



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

(B) Overreaction to imprecise & underreaction to precise disconfirmatory realizations; underreaction to confirmatory realizations

FIGURE 7. Reaction with an asymmetric prior. Each data point in the left figure aggregates across signal realizations in a 2-state environment with a given diagnosticity. Each data point in the right figure corresponds to one signal realization in a 2-state environment with a given prior and diagnosticity.

signal realization.[51] As illustrated in Fig. 7b and consistent with Prediction 6, 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 D.5 presents regression analyses. 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. 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.

Finally, we explore the prediction that people are more likely to update in the opposite direction from the objective posterior, i.e., wrong direction reaction, for confirmatory realizations relative to neutral or disconfirmatory realizations (Prediction 6.i). Fig. 10 presents the share of wrong direction reactions for confirmatory, disconfirmatory and neutral signal realizations. Consistent with Prediction 6, we observe a significant difference in these frequencies. While wrong direction reactions occur relatively infrequently following neutral and disconfirmatory realizations, they occur significantly more often following confirmatory realizations. In the latter case,

[51]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).

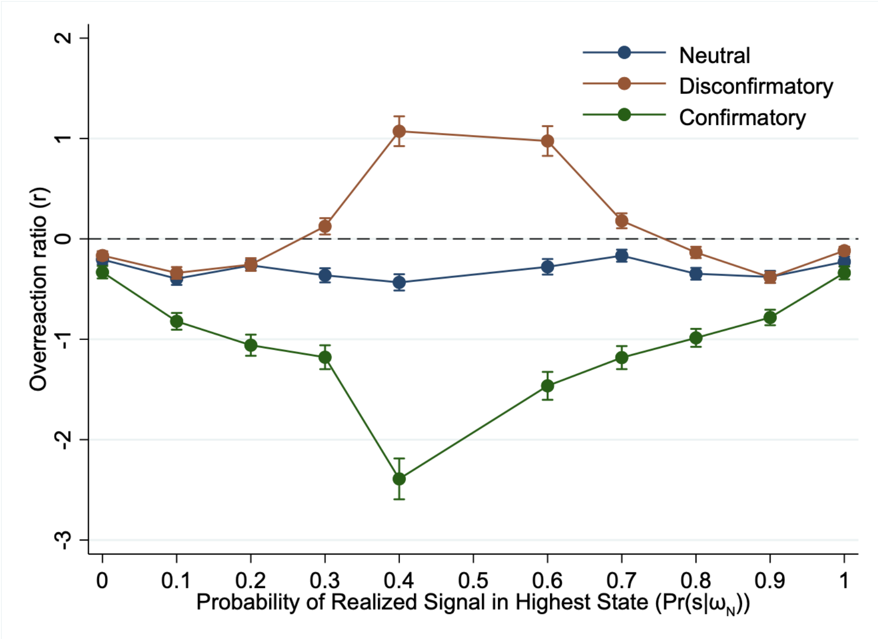


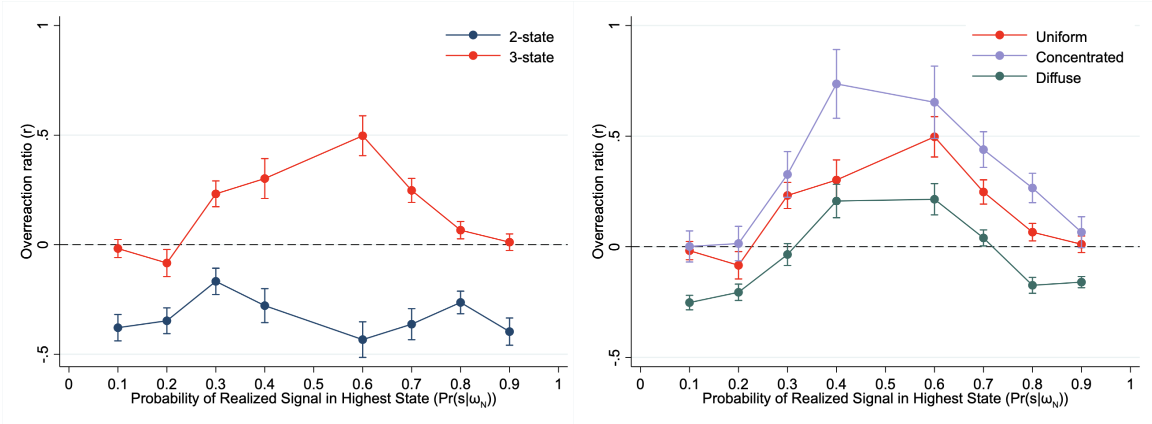
FIGURE 8. Overreaction ratio w/o absolute value and including wrong direction updates

nearly 30% of reactions are in the wrong direction—almost three times higher than in the former cases. Importantly, this incidence of wrong direction reaction 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.

3.5 Individual-Level Parameter Estimation

Finally, we explore individual-level heterogeneity in channeled attention and cognitive imprecision by estimating the parameters of the two-stage model 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. The results are presented in Fig. D.9.

These estimates reveal significant heterogeneity across participants. Specifically, 70% of participants exhibit distortions in both stages of the belief-updating process, 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 distortion ($\theta = 0$ and $\lambda = 1$). The estimated values of θ and λ exhibit a significant negative correlation, with a correlation coefficient of -0.47. The negative correlation implies that participants who are more prone to salience-driven attention through representativeness (higher θ) also tend to exhibit higher levels of cognitive imprecision (lower λ). This suggests that individual-level limits to cognitive capacity lead to both a greater distortion in the representational stage and noisier evaluation in the processing stage.



(A) CAPTION

(B) CAPTION

FIGURE 9. Overreaction ratio w/o absolute value and including wrong direction updates

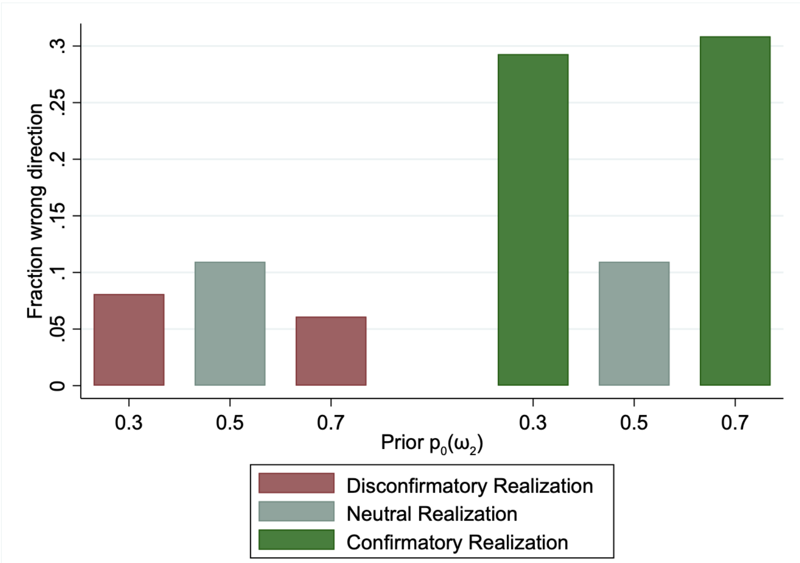


FIGURE 10. Higher share of wrong direction reactions for confirmatory realizations

4 Testing the Mechanism

In this section, we present direct evidence for the proposed cognitive mechanism. We first measure attention to see where it is channeled and then examine belief-updating when this resource is restricted. Next, we study the causal effect of attention on belief-updating and show how the results from our baseline paradigm change when salience cues are removed. Finally, we compare the attentional impact of representativeness to other salience cues considered in the literature.

4.1 Measuring and Restricting Attention

In the representational stage of our model, limits on attention lead agents to focus on representative states especially 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.[52] 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 realized) 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.[53] 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 the 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 D.1.

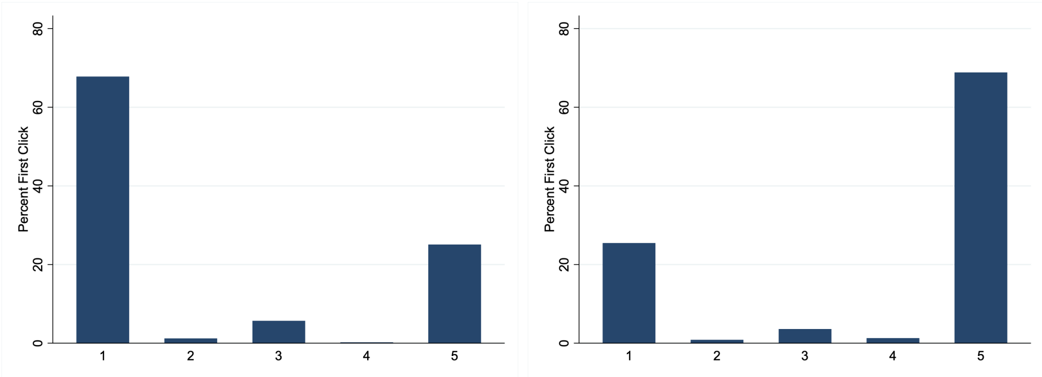
Two main predictions follow. First, participants will channel their attention and click on the representative state first. In other words, upon observing a blue (red) ball, the most likely first-click will be on the bag with the most blue (red) balls. Second, fixing the information environment, taxing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

To examine the first prediction, Fig. 11 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 their first click—to the most

[52]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).

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

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).



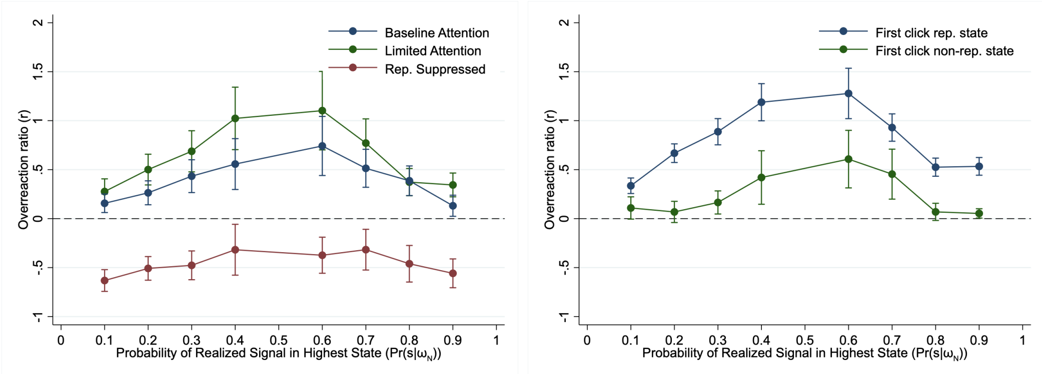
(A) State 1 Representative (B) State 5 Representative

FIGURE 11. Most participants click on representative state first

To examine the second prediction, Fig. 12a presents results on the overreaction ratio comparing the Baseline 5-state treatment to the Limited Attention treatment with the same exact information environment. Overreaction was indeed significantly higher in the latter than the former across nearly all signal diagnosticities. Table D.11, Column 1 presents regression analyses showing that the extent of overreaction increased in the Limited Attention treatment compared to the Baseline treatment.

We examine participant heterogeneity by looking at whether those who are more prone to salience-channeled attention—proxied by their propensity to click the representative state first—also overreact more. Restricting attention to the Limited Attention treatment, Fig. 12b shows that overreaction was substantially higher in the representative-state-first group across all signal diagnosticities. Column 2 of Table D.11 presents the same finding using regression analyses. Taken together, these results support the two predictions outlined above and provide further evidence against insensitivity and information-independent heuristics (e.g., partition dependence (Tversky and Koehler 1994)) as alternative explanations for our results.

Finally, we provide further evidence for the proposed predictions by structurally estimating the parameters in the Limited Attention treatment and comparing them to those obtained in the Baseline Attention treatment. Table D.12 shows that the estimate of θ increases from 0.99 to 1.26, while the estimates of λ are similar between the two treatments. This lends direct support to our prediction that decreasing attentional resources through the Mouselab paradigm increases the salience-driven distortion in the representational stage (as indicated by higher θ), while leaving the



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

FIGURE 12. Limited attention increases overreaction

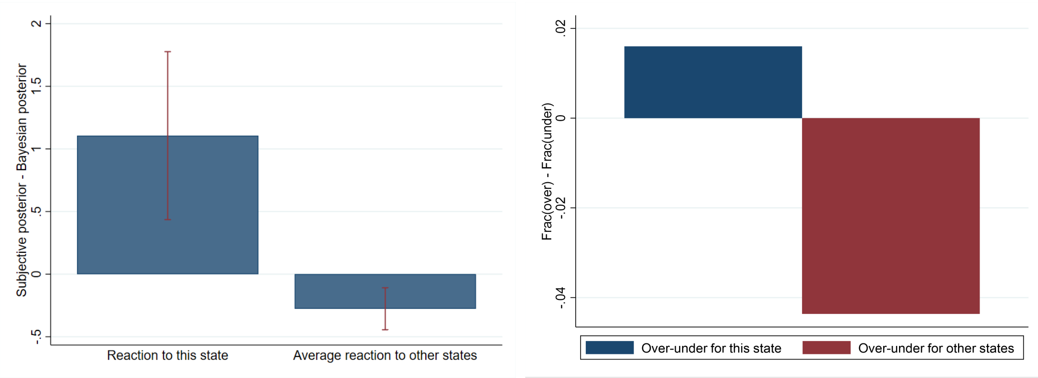
level of cognitive imprecision unchanged.

4.2 Causal Effect of Attention

To measure the causal effect of channeled attention on belief-updating, we adapted the Mouselab paradigm treatment described in the previous subsection to also remove representativeness-based salience cues. In this design, participants were presented with the same information about the learning environment as in our Baseline 5-state treatment, including the potential diagnosticities associated with the states. However, after seeing the signal, the participants needed to click on a state to see the value associated with it.[54] The order of states was completely randomized. Note that the learning environment was identical to the Baseline and Limited Attention treatments—by clicking on each state, the participant would have the same information as in the other treatments—but the initial salience cue of representativeness was gone. In this Representativeness Suppressed (Rep. Suppressed) treatment, the participant's initial attention could no longer be directed as a function of the state's propensity to generate a signal—it was channeled as-if randomly with respect to the underlying value. This suppresses representativeness as a salience cue and allows us to test the causal effect of attention on belief-updating.

We first test the conjecture that participants' attention was channeled to states as-if randomly by looking at the first-click data. Fig. D.5 shows that attention—as measured by first-click—was not associated with the salience features of the state, which is in stark contrast to Fig. 11. In Fig. 13 we present data on updated beliefs for the state that was attended to first compared to the average belief-updating on all other states. Both the measure of comparing the subjective and Bayesian posterior (Fig. 13a) and comparing the fraction of over- versus underweighing (Fig. 13a) show that, consistent with our proposed mechanism, attention leads to the overweighing of the target state in the agent's updated beliefs.

[54]For example, the participant had to click on the state to find out that this was $\omega_5 = 0.9$, i.e., that this was Bag 5 and had 90 (10) red (blue) balls in it.



(A) Belief Difference

(B) % overweigh - % underweigh

FIGURE 13. Participants overweighed the state they clicked on first and underweighed the rest of the states on average

Importantly, our framework predicts that suppressing representativeness as a salience cue will shift responses from overreacting to information to underreacting to it—while keeping the information environment constant. Fig. 12a shows that this is indeed the case. While participants overreacted to information across all signal diagnosticies when the representativeness-based salience cue was present—both in the Baseline and Limited Attention treatments—participants *underreacted* across all signal diagnosticities when representativeness was suppressed. **??** shows that this was driven by the as-if random channeling of attention: in the Representativeness Suppressed condition, the most representative state is underweighed and the least representative state is overweighed on average, consistent with the random allocation of attention. Together, the results in this subsection provide direct support for our proposed attentional mechanism and show that there is nothing special about the two-state case that systematically generates underreaction—people underreact in more complex environments as well when representativeness is suppressed (e.g., when there is uncertainty over which state is representative).

4.3 Other Drivers of Channeled Attention

Channeled attention plays a central role in our framework. While we focus on representativeness as a bottom-up salience cue that channels attention, there are other drivers of both bottom-up and top-down channeled attention. Here, we focus on two such alternatives that have been extensively explored in prior work—namely, visual salience and goal-directed attention—and compare their impact to the representativeness-based salience cue in our environment. Bottom-up visual salience is implemented by highlighting a particular color in a state that differs substantially from the background color (bright yellow), as typically done in the literature (Li and Camerer 2022), while top-down goal-directed salience is implemented by changing incentives such that participants are incentivized for their beliefs about a particular state (Maćkowiak et al. 2023). Appendix D.5 presents these results and discusses them in detail. First, when representativeness is suppressed as a salience cue, the bottom-up and top-down alternatives are effective in channeling attention to the respective state. Moreover, channelling attention to the representative state through these alternative salience cues brought back overreaction in the Representativeness Suppressed condition, providing further evidence for the critical role of attention in belief-updating.

However, when the representativeness-based salience cue is present it dominates the other alternatives: participants overwhelmingly channel their attention to the representative state and their beliefs move accordingly even in cases where the alternative drivers of attention are associated with a different state (e.g., the least representative state). These results suggest that while other drivers of attention may impact belief updating when representativeness is suppressed, the latter is likely to play a significant role in belief updating when the associated cue is available.

5 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 imprecision or only salience-channeled attention. We refer to these comparison models as the processing-only model and representational-only model, respectively.

5.1 Completeness

Completeness is a measure of how much of the explainable variation in data a model captures relative to an alternative, which we take to be Bayes' rule. 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. 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. Details of how we estimate completeness can be found in Appendix D.7.1.

We first estimate completeness in the simple information environments with two states. As shown in Table 1, the processing-only model ($M = P$) achieves essentially 100% completeness. In these simple environments, the addition of the representational stage does not yield any further improvement in model performance. This is consistent with our conjecture that limited attention only has bite in complex environments, 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 processing-only model to a mere 36%. The representational-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 processing-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. Together with the results in Section 3.3, these findings show that the interaction between limited attention and processing capacity is key for understanding belief-updating in complex environments.

Table 1. 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)
Processing-only Model	1.00	0.36	0.76	0.97
	(0.06)	(0.02)	(0.00)	(0.00)
Representational-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 D.1 except for the 11-state complexity; includes wrong direction updates. Restrictiveness is estimated from 1000 simulations.

5.2 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. A model is 0% restrictive if it fits synthetic data perfectly and 100% restrictive if it fits synthetic data no better than Bayes' rule. Details of the estimation procedure can be found in Appendix D.7.2.

As shown in Table 1, 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 processing-only model (0.73 versus 0.76 for simple environments

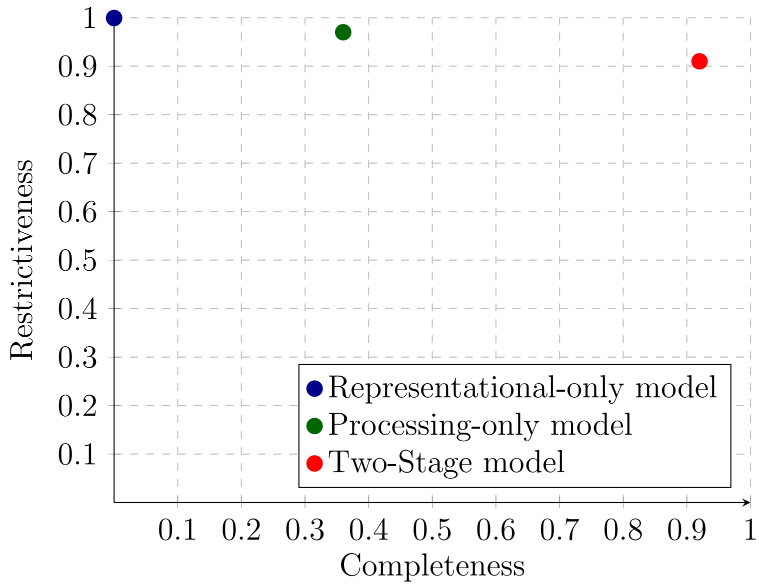


FIGURE 14. Completeness-restrictiveness trade-off (> 2 states)

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 processing-only model does not come at the expense of a significant increase in flexibility.[55]

While the representational-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 representational-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. 14 plots the completeness and restrictiveness of the two-stage model, the processing-only model and the representational-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 belief-updating leads to a striking increase in explanatory power while only minimally increasing the model's flexibility.

6 Extensions

This section presents several extensions that explore the predictions of our framework for different signal structures (beyond the good news/bad news setting), belief-

[55]In Appendix C we consider a variation of the processing-only model with more flexible types of cognitive imprecision. The results show that while the added flexibility achieves higher completeness, it is still much less complete than the two-stage model (0.65 vs 0.92 in complex environments) and also less restrictive (0.89 vs 0.91 in complex environments).

updating domains (inference versus forecasting), and learning objects (beliefs about financial assets). Experimental details are outlined in Appendix D.8.

6.1 Alternative Signal Structures

The preceding findings have all been in a setting with a clear good news/bad news signal structure. This signal structure is a natural one to adopt since it is used in the majority of prior experimental work and mirrors many real-world settings such as equity markets and some economic indicators (e.g., GDP), where signals have a clear ordering with respect to the underlying state.[56] Other settings, however, do not have such a structure.[57] The framework outlined in Section 2 outlines predictions for the relationship between representativeness and over versus underreaction in a good news/bad news setting; in alternative settings, the model still predicts that the most representative state will be overweighed in the agent's mental representation, but now the emergence of over versus underreaction depends on ordering of the states.

It is not possible to test this prediction with binary signals, so we designed a version of our 3-state paradigm with a uniform prior and three signals: balls could either be red, blue, or green. The bags contained a combination of these balls such that one bag was representative of each color.[58] The rest of the procedures were the same as in the baseline paradigm. Figure Fig. 15a shows belief-updating as the difference between the subjective and Bayesian posterior for each state, split by the signal; Table D.16 presents these results in regression form. In line with our prediction, the representative state is overweighed in each case, while the non-representative states are underweighed. Note, however, that unlike the good news/bad news setting, whether expectation corresponds to over or underreaction now depends on the ordering of the states.

6.2 Inference versus Forecasting

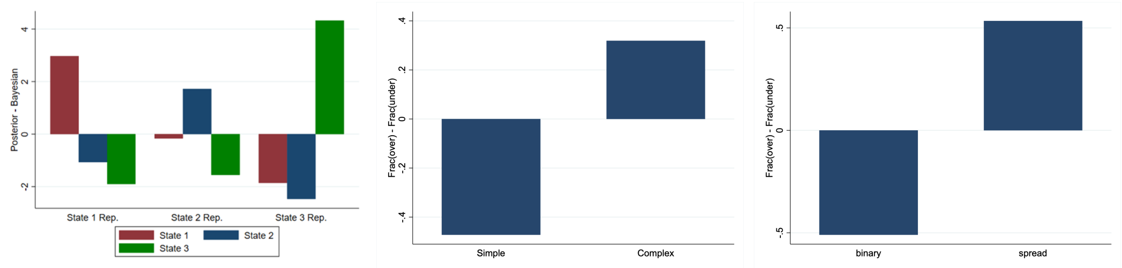
Recent work by Fan et al. (2023) shows that people underinfer when making inferences after observing information and overinfer when forming forecasts about future information. Their main treatment featured a binary state space (a firm was either good or bad) and a discretized normal signal distribution (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).

In our framework, attention is limited by the number of objects or outcomes in

[56]For example, a stock price increase (decrease) can be interpreted as a positive (negative) signal about the underlying value of the company. An increase (decrease) in the GDP can be interpreted as a positive (negative) signal about the economy.

[57]For example, it is not clear how changes in the US Treasury rate or inflation correspond to the underlying state.

[58]For example, Bag 1: (red, blue, green)=(45, 35, 20), Bag 2: (red, blue, green)=(20, 45, 35), and Bag 3: (red, blue, green)=(35, 20, 45). In this case, Bag 1, 2, and 3 are representative of the red, blue, and green signals, respectively.



(A) Three-Signal (B) Forecasting (C) Binary versus Spread

FIGURE 15. Fig. 15a shows that participants overweighed the most representative state even when the information environment does not feature a good news/bad news structure. For example, the middle state is the most representative state given a blue signal realization, and it was indeed overweighed when the blue signal was drawn. Fig. 15b displays the fraction of over versus undereaction in the forecasting task with two potential objects (Simple) versus 11 potential objects (Complex). Fig. 15c shows the fraction of over versus undereaction for news about a financial asset when the underlying asset is a binary option versus the equivalent spread. Results for both Fig. 15b and Fig. 15c do not qualitatively change when using the overreaction ratio, and are presented in Appendix D.8.

the information environment that one must form beliefs over. While the inference condition tasks participants with updating their beliefs about two objects (2 states, as in our Baseline paradigm), the forecasting condition tasks them with forming beliefs over many objects (the 11 potential price outcomes). Given the good news/bad news structure of both environments, our model predicts that the increased representational complexity in the forecasting condition will lead participants to overreact in that setting, and to underreact in the inference condition.

To test this conjecture, we replicate the setting from Fan et al. (2023) while manipulating the number of objects that need to be explicitly considered for belief-updating. In the forecasting task, participants were presented with a distribution of prices that differ depending on whether the underlying company is good or bad. They reported forecasts based on a price signal in one of two conditions, Complex and Simple. The two conditions were the same in every way except that in the Complex condition, participants made forecasts about the likelihood of each potential price, as in Fan et al. (2023), whereas in the Simple condition, the price space was partitioned into two bins and participants made forecasts on the likelihood of each. Importantly, the underlying information environment and price signals were the same across conditions—the only difference was the number of objects that participants made forecasts for.

As shown in Fig. 15b, which uses the same measure as Fan et al. (2023), we replicate the finding that people overreact when forecasting over a large number of objects (Complex condition). However, despite facing the *same* signal generating process, participants' forecasts displayed marked underreaction when tasked to updating over two objects. These results demonstrate the generality of our framework: people exhibit over versus underreaction in both forecasting as a function of representational complexity. We view these results as complementary to the mechanism proposed in Fan et al. (2023), where people use different simplifying heuristics across inference and forecasting environments.

6.3 Beliefs about Financial Assets

Recent work by Puri (2022) argues that people are averse to risk that is complex to evaluate, where her definition of complexity maps directly to the notion of representational complexity outlined in Section 2—as the number of objects one needs to consider when making a judgment. Goodman and Puri (2022) shows that attitudes towards complexity can explain the preference for binary options over bull-spreads over the same asset, even when the former is dominated by the latter. While a more detailed description of binary options and bull spreads is presented in Appendix D.8.3, the main substantive difference for the purposes of our investigation is that binary options have two potential outcomes—a pre-determined payoff if the price of an asset is above a certain threshold, and zero otherwise—while a bull spread has a larger number of payoff outcomes depending on the realized price of the same asset.[59]

Our framework predicts that the difference in complexity between the two financial instruments will generate a difference in belief-updating. Specifically, given the same signal, people will underreact about the payoff likelihoods in the case of binary options but overreact in the case of a bull spread, while keeping the signal generating process and underlying asset the same. We tested this prediction experimentally by endowing participants with and then giving them information about either a binary option or bull spread over the same underlying asset and asking them to forecast the likelihood of future payoffs. Results presented in Fig. 15c show that indeed, the majority of participants underreacted to information about the binary option while the majority of participants overreacted to the same information when considering the payoffs of a bull spread. Appendix D.8.3 shows that this pattern is robust across signal realizations.

7 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 salience-channelled attention and cognitive imprecision are cognitive complements and their interaction plays a crucial role in explaining how agents update

[59]For example, take a binary option that returns a pre-determined payoff if the price of an asset in a certain period is greater than S_2 and zero otherwise. A dominating bull spread would generate the same payoff if the price is greater than S_2, but also generate a series of smaller payoffs within the price range between S_1 and S_2, where $S_1 < S_2$.

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 choices 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 to 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.

——— (2024): "The Behavioral Foundations of Model Misspecification," *Working paper.*

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.

Bouchaud, J.-P., P. Krueger, A. Landier, and D. Thesmar (2019): "Sticky Expectations and the Profitability Anomaly," *The Journal of Finance*, 74, 639–674.

Bruce, N. D. and J. K. Tsotsos (2009): "Saliency, attention, and visual search: An information theoretic approach," *Journal of vision*, 9, 5–5.

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.

DellaVigna, S. and J. M. Pollet (2009): "Investor inattention and Friday earnings announcements," *The journal of finance*, 64, 709–749.

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*.

Fox, C. R., D. Bardolet, and D. Lieb (2005): "Partition dependence in decision analysis, resource allocation, and consumer choice," in *Experimental Business Research: Marketing, Accounting and Cognitive Perspectives Volume III*, Springer, 229–251.

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.

Gao, Y., N. Moreira, R. Reis, and S. Yu (2015): "A survey on operational state complexity," *arXiv preprint arXiv:1509.03254*.

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

Goodman, A. and I. Puri (2022): "Bulls and binaries: Price anomalies and behavioral biases," Tech. rep., Working paper.

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*.

Li, X. and C. F. Camerer (2022): "Predictable effects of visual salience in experimental decisions and games," *The Quarterly Journal of Economics*, 137, 1849–1900.

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.

Maćkowiak, B., F. Matějka, and M. Wiederholt (2023): "Rational inattention: A review," *Journal of Economic Literature*, 61, 226–273.

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.

Mohrschladt, H., M. Baars, and T. Langer (2020): "How General is the Strength-Weight Bias in Probability Updating?" *Working Paper*.

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

Molavi, P., A. Tahbaz-Salehi, and A. Vedolin (2023): "Model Complexity, Expectations, and Asset Prices," *The Review of Economic Studies*, rdad073.

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.

PAPADIMITRIOU, C. H. (2003): "Computational complexity," in *Encyclopedia of computer science*, 260–265.

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*.

PURI, I. (2022): "Simplicity and risk," *Available at SSRN*, 3253494.

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.

SALANT, Y. AND J. L. SPENKUCH (2022): "Complexity and choice," Tech. rep., National Bureau of Economic Research.

SCHWARTZSTEIN, J. (2014): "Selective attention and learning," *Journal of the European Economic Association*, 12, 1423–1452.

SHENHAV, A., S. MUSSLICK, F. LIEDER, W. KOOL, T. L. GRIFFITHS, J. D. COHEN, AND M. M. BOTVINICK (2017): "Toward a rational and mechanistic account of mental effort," *Annual review of neuroscience*, 40, 99–124.

TALSMA, D., D. SENKOWSKI, S. SOTO-FARACO, AND M. G. WOLDORFF (2010): "The multifaceted interplay between attention and multisensory integration," *Trends in cognitive sciences*, 14, 400–410.

TANNER, J. AND L. ITTI (2019): "A top-down saliency model with goal relevance," *Journal of vision*, 19, 11–11.

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.

TVERSKY, A. AND D. J. KOEHLER (1994): "Support theory: A nonextensional representation of subjective probability." *Psychological review*, 101, 547.

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.

Yantis, S. (2008): "The neural basis of selective attention: Cortical sources and targets of attentional modulation," *Current directions in psychological science*, 17, 86–90.

A 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 ($N = 100$).[60] 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.

Fan et al. (2023) show underreaction for inference in a simple two-state setting and overreaction when forecasting. As shown in 6.2, these results can be explained by the difference in complexity across the two settings. 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.[61] Kieren and Weber (2020) find underreaction 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 6, 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

[60]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.

[61]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 (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).

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). Mohrschladt, Baars, and Langer (2020) argue that the underweighting of signal weight does not generalize to a broader class of information environments. 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 underweigh 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-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).[62]

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).[63] 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.[64] d'Arienzo (2020) and Wang (2021) find that indi-

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

[63] 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).

[64] 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

vidual analysts' forecasts of long-term interest rates exhibit overreaction. Bordalo et al. (2019) find overreaction in the expectations of long-term corporate earnings growth.

Although overreaction has been found to be predominant for many financial variables, both in the case of macro news and news about individual stocks, there are notable exceptions. For example, Bordalo et al. (2020) find underreaction to news about the three-month US Treasury rate. As discussed in 6.1, this may be due to the absence of a good news/bad news signal structure in this setting. Bouchaud, Krueger, Landier, and Thesmar (2019) document underreaction of analyst forecasts of firms' short term earnings.[65] As also noted in Augenblick et al. (2022), earnings announcements tend to be fairly informative in the short-term, and this high diagnosticity would increase the likelihood of underreaction in our framework. Indeed, longer-term earnings forecasts, which are noisier, do exhibit overreaction (Bordalo et al. 2019). At the same time, we acknowledge that factors outside our model that can generate short-term unresponsiveness to information, such as inattention (DellaVigna and Pollet 2009), may also be driving the observed underreaction.

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 'representational' stage.

Our framework 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 real-world settings that feature noisy signals and a good news/bad news signal structure. 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 examined under- and overreaction by looking at choice data—specifically, price move-

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.

[65] Kwon and Tang (2021) similarly find short-term underreaction to earnings announcements in prices. As discussed further below, we focus on data on beliefs due to the potential for identification issues when interpreting price data.

ments. 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.

B Proofs

Proof of Claim in Footnote 23. In our context, the representativeness-based discounting weighing function generates a distorted belief

$$p_R(\omega_i|s_j) = \frac{p_B(\omega_i|s_j)R(\omega_i, s_j)^\theta}{\sum_{\omega_i\in\Omega} p_B(\omega_i|s_j)R(\omega_i, s_j)^\theta},$$

whereas applying Bayes' rule to $\hat{\pi}$ results in

$$\begin{aligned} p_R(\omega_i|s_j) = \frac{\hat{\pi}(s_j|\omega_i)p_0(\omega_i)}{\sum_{\omega_k\in\Omega}\hat{\pi}(s_j|\omega_k)p_0(\omega_k)} &= \frac{\pi(s_j|\omega_i)R(\omega_i, s_j)^\theta p_0(\omega_i)}{\sum_{\omega_k\in\Omega}\pi(s_j|\omega_k)R(\omega_k, s_j)^\theta p_0(\omega_k)} \\ &= \frac{p_B(\omega_i|s_j)R(\omega_i, s_j)^\theta}{\sum_{\omega_k\in\Omega} p_B(\omega_k|s_j)R(\omega_k, s_j)^\theta}, \end{aligned}$$

which are equal. To see the counting a signal $\theta + 1$ times property, note that $\hat{\pi}(s_k|\omega_i)/\hat{\pi}(s_k|\omega_j) \equiv (\pi(s_k|\omega_i)/\pi(s_k|\omega_j))^{\theta+1}$, so the mental representation is distorting the signal likelihood ratio by a factor of θ. This updating rule has often been used in the theoretical literature to capture overreaction (Bohren and Hauser 2021; Angrisani, Guarino, Jehiel, and Kitagawa 2020).

Proof of Claim in Footnote 32. We show that our definition of overreaction based on Definition 2 is equivalent to the binary state definition stated in Footnote 32. Fix any signal realization s_j. Note that

$$\begin{aligned} \left|\hat{E}\left(\omega|s_j\right) - E_0\left(\omega\right)\right| &= \left|\omega_2\left(\hat{p}\left(\omega_2|s_j\right) - p_0\left(\omega_2\right)\right) + \omega_1\left(\hat{p}\left(\omega_1|s_j\right) - p_0\left(\omega_1\right)\right)\right| \\ &= \left|\omega_2\left(p_0\left(\omega_1\right) - \hat{p}\left(\omega_1|s_j\right)\right) + \omega_1\left(\hat{p}\left(\omega_1|s_j\right) - p_0\left(\omega_1\right)\right)\right| \\ &= \left|\omega_2 - \omega_1\right| \cdot \left|\hat{p}\left(\omega_1|s_j\right) - p_0\left(\omega_1\right)\right|, \end{aligned}$$

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

$$|E(\omega|s_j) - E_0(\omega)| = |\omega_2 - \omega_1| \cdot |p_B(\omega_1|s_j) - p_0(\omega_1)|,$$

where p_B is the objective posterior following signal realization s_j. Hence,

$$\begin{aligned} r(s_j) =& \frac{|\hat{E}(\omega|s_j) - E_0(\omega)| - |E(\omega|s_j) - E_0(\omega)|}{|E(\omega|s_j) - E_0(\omega)|} \\ =& \frac{|\hat{p}(\omega_1|s_j) - p_0(\omega_1)| - |p_B(\omega_1|s) - p_0(\omega_1)|}{|p_B(\omega_1|s_j) - p(\omega_1)|}. \end{aligned}$$

That is, $r(s_j) > 0$ if and only if $|\hat{p}(\omega_1|s_j) - p_0(\omega_1)| > |p(\omega_1|s_j) - p_0(\omega_1)|$, and similarly for ω_2.

Proof of Prediction 1. We first prove that there exist the three regions as described in Prediction 1. Consider the distorted posterior derived from the first stage, $p_R(s_j)$. Note that for all $\omega_i \in \Omega$ such that $\omega_i \neq \omega_R$,

$$\frac{p_R(\omega_R|s_j)}{p_R(\omega_i|s_j)} = \left(\frac{p_B(\omega_R|s_j)}{p_B(\omega_i|s_j)}\right)^{\theta+1} = \left(\frac{\pi(s_j|\omega_R)}{\pi(s_j|\omega_i)}\right)^{\theta+1} > 1.$$

Since $\sum_{\omega_i \in \Omega} p_R(\omega_i|s_j) = \sum_{\omega_i \in \Omega} p_B(\omega_i|s_j) = 1$, it must be that $p_R(\omega_R|s_j) > p_B(\omega_R|s_j) > \frac{1}{N}$. Since $\hat{p}(\omega_R|s_j) = \lambda p_R(\omega_R|s_j) + (1-\lambda)\frac{1}{N}$, there exists threshold $\overline{\lambda}_1(\theta) \in (0,1)$ such that $\hat{p}(\omega_R|s_j) > p_B(\omega_R|s_j)$ if $\lambda > \overline{\lambda}_1(\theta)$ and $\hat{p}(\omega_R|s_j) < p_B(\omega_R|s_j)$ if $0 \leq \lambda < \overline{\lambda}_1(\theta)$.

Analogously, for all $\omega_i \in \Omega$ such that $\omega_i \neq \omega_{NR}$, $\frac{p_R(\omega_{NR}|s_j)}{p_R(\omega_i|s_j)} = \left(\frac{\pi(s_j|\omega_{NR})}{\pi(s_j|\omega_i)}\right)^{\theta+1} < 1$. It follows that $p_R(\omega_{NR}|s_j) < p_B(\omega_{NR}|s_j) < \frac{1}{N}$. Since $\hat{p}(\omega_{NR}|s_j) = \lambda p_R(\omega_{NR}|s_j) + (1-\lambda)\frac{1}{N}$, there exists threshold $\overline{\lambda}_2(\theta) \in (0,1)$ such that $\hat{p}(\omega_{NR}|s_j) < p_B(\omega_{NR}|s_j)$ if $\lambda > \overline{\lambda}_2(\theta)$ and $\hat{p}(\omega_{NR}|s_j) > p_B(\omega_{NR}|s_j)$ if $0 \leq \lambda < \overline{\lambda}_2(\theta)$.

When $|\Omega| = 2$, since it cannot be the case that both ω_R and ω_{NR} are overweighed, we must have $\overline{\lambda}_1(\theta) = \overline{\lambda}_2(\theta)$. We now show that $\overline{\lambda}_1(\theta) \leq \overline{\lambda}_2(\theta)$ if $|\Omega| > 2$. Note that

$$p_B(\omega_{NR}|s_j) + p_B(\omega_R|s_j) = \frac{\pi(s_j|\omega_R)p_0(\omega_R) + \pi(s_j|\omega_{NR})p_0(\omega_{NR})}{\sum_{\omega \in \Omega} \pi(s_j|\omega)p_0(\omega)} = \frac{2}{N}$$

where the second equality follows from the uniformity of the prior p_0 and the symmetry of the state space. Meanwhile, note that

$$p_R(\omega_R|s_j) + p_R(\omega_{NR}|s_j) = \frac{\pi(s_j|\omega_R)^{\theta+1}}{\sum_{\omega \in \Omega} \pi(s_j|\omega)^{\theta+1}} + \frac{\pi(s_j|\omega_{NR})^{\theta+1}}{\sum_{\omega \in \Omega} \pi(s_j|\omega)^{\theta+1}}.$$

Since $\{\omega_R, \omega_{NR}\} = \{\min \Omega, \max \Omega\}$ and $\theta > 0$, for all $\omega \in \Omega \setminus \{\omega_R, \omega_{NR}\}$ and its symmetric counterpart $\omega' = 1 - \omega$, we have

$$\pi(s_j|\omega_R)^{\theta+1} + \pi(s_j|\omega_{NR})^{\theta+1} > \pi(s_j|\omega)^{\theta+1} + \pi(s_j|\omega')^{\theta+1}.$$

Hence, $p_R(\omega_{NR}|s_j) + p_R(\omega_R|s_j) > 2/N$. It then follows from $p_R(\omega_R|s_j) > p_B(\omega_R|s_j) >$

$\frac{1}{N}$ and $p_R(\omega_{NR}|s_j) < p_B(\omega_{NR}|s_j) < \frac{1}{N}$ that

$$p_R(\omega_R|s_j) - \frac{1}{N} > \frac{1}{N} - p_R(\omega_{NR}|s_j) > 0.$$

By definition of $\overline{\lambda}_1(\theta)$,

$$\overline{\lambda}_1(\theta)p_R(\omega_R|s_j) + (1-\overline{\lambda}_1(\theta))\frac{1}{N} = p_B(\omega_R|s_j).$$

Using the previous inequality, we have

$$\begin{aligned}
&\overline{\lambda}_1(\theta)p_R(\omega_{NR}|s_j) + (1-\overline{\lambda}_1(\theta))\frac{1}{N} \\
>& \frac{1}{N} - \overline{\lambda}_1(\theta)\left(p_R(\omega_R|s_j) - \frac{1}{N}\right) \\
=& \frac{2}{N} - p_B(\omega_R|s_j) = p_B(\omega_{NR}|s_j).
\end{aligned}$$

Since

$$\overline{\lambda}_2(\theta)p_R(\omega_{NR}|s_j) + (1-\overline{\lambda}_2(\theta))\frac{1}{N} = p_B(\omega_{NR}|s_j),$$

it must be that $\overline{\lambda}_1(\theta) < \overline{\lambda}_2(\theta)$.

Moreover, since $p_R(\omega_{NR}|s_j) + p_R(\omega_R|s_j) > 2/N$ when $\theta > 0$,

$$\hat{p}(\omega_{NR}|s_j) + \hat{p}(\omega_R|s_j) = \lambda(p_R(\omega_{NR}|s_j) + p_R(\omega_R|s_j)) + (1-\lambda)\frac{2}{N} \geq \frac{2}{N},$$

where inequality holds if $\lambda < 1$ and equality holds otherwise. Therefore, for each $\theta > 0$, the agent underweighs interior states $\Omega_I = \Omega \setminus \{\omega_R, \omega_{NR}\}$ for $\lambda > 0$ and neither under- nor overweighs it for $\lambda = 0$. $\square$

Proof of Prediction 2. Suppose the signal realization is s_2. The objective posterior of any state $\omega_i \in \Omega$ is

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

We can write the Bayesian expected state as

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

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

$$\begin{aligned}
E_B(\omega|s_2) - E_0(\omega) &= \frac{2}{N}\sum_{\omega_i\in\Omega}\omega_i^2 - \frac{1}{2} \\
&= \frac{2}{N}\left[(1-\omega_N)^2 + ... + (1-\omega_{K+1})^2 + \omega_{K+1}^2 + ... + \omega_N^2 - \frac{K}{2}\right]
\end{aligned}$$

$$= \frac{4}{N}\left[\left(\omega_{K+1} - \frac{1}{2}\right)^2 + ... + \left(\omega_N - \frac{1}{2}\right)^2\right].$$

When Ω contains an odd number of states and $N = 2K - 1$, symmetry implies that the Kth state must be $\frac{1}{2}$. We therefore obtain the same expression for $E_B(\omega|s_2) - E(\omega)$. 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 θ goes to infinity. That is, $\lim_{\theta \to \infty} E_R(\omega|s_2) = \omega_N$. It follows that

$$\begin{aligned}
\lim_{\theta \to \infty} r_R(s_2) + 1 &= \lim_{\theta \to \infty} \frac{|E_R(\omega|s_2) - E_0(\omega)|}{|E_B(\omega|s_2) - E_0(\omega)|} \\
&= \frac{\omega_N - \frac{1}{2}}{\frac{4}{N}\left[\left(\omega_{K+1} - \frac{1}{2}\right)^2 + ... + \left(\omega_N - \frac{1}{2}\right)^2\right]}. \qquad (12)
\end{aligned}$$

A similar expression to Eq. (12) with respect to Ω' holds for $r'(s_2)$. Since Ω' is equally dispersed as Ω, $\omega'_N = \omega_N$. Since Ω' is more complex than Ω and every state in $\Omega' \setminus \Omega$ is more interior than every state in Ω,

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

Therefore, when θ 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 3. Suppose p'_0 is strictly more concentrated than p_0 and both are symmetric. Let ω' and ω denote the random variables that are distributed according to p'_0 and p_0, respectively. Since the priors have the same support, both $E_R(\omega'|s_j)$ and $E_R(\omega|s_j)$ converge to the highest state in the support, ω_N, when θ diverges to infinity. Thus, to show that $r'(s_j) > r(s_j)$ when θ is sufficiently large, it suffices to show that $|E_B(\omega'|s_j) - E_0(\omega')| < |E_B(\omega|s_j) - E_0(\omega)|$.

Suppose the signal realization is s_2. Since $E_B(\omega'|s_2) > 1/2$, $E_B(\omega|s_2) > 1/2$, and $E_0(\omega') = E_0(\omega) = 1/2$, we only need to show $E_B(\omega'|s_2) < E_B(\omega|s_2)$. 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

$$E_B(\omega'|s_2) = 2\sum_{\omega_i \in \Omega} p'_0(\omega_i)\omega_i^2 = E_B(\omega|s_2) + 2\sum_{\omega_i \in \Omega} \Delta(\omega_i)\omega_i^2.$$

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, $E_B(\omega'|s_2)<E_B(\omega|s_2)$. The proof is analogous for signal realization s_1. □

Proof of Prediction 4. Part (i). Given the symmetry in the information environment, we focus on signal realization s_2 without loss. Let $\overline{r}_R(\theta)$ denote the supremum of $r_R(s_2)$ over the set of all possible priors given state space Ω and parameter θ. Since $r(s_2)=\lambda r_R(s_2)-(1-\lambda)$, to show the existence of the cognitive-imprecision-dominant region, it suffices to show that $\overline{r}_R(\theta)<\infty$ for any $\theta>0$. Moreover, $\overline{\lambda}_1(\theta)$ is given by the solution to the following equation, $\lambda\overline{r}_R(\theta)-(1-\lambda)=0$.

For any state space Ω and prior p_0, the Bayesian posterior is

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

and the interim posterior is

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

Without loss of generality, assume $N=2K$ for a positive integer K (if N is odd, then we can duplicate the middle state to make the state space even). Note that

$$\begin{aligned}r_R(s_2)+1=&\frac{|E_R(\omega|s_2)-E_0(\omega)|}{|E_B(\omega|s_2)-E_0(\omega)|}\\ =&\frac{\sum_{\omega_i\in\Omega}p_0(\omega_i)\omega_i^{\theta+2}/(\sum_{\omega_i\in\Omega}p_0(\omega_i)\omega_i^{\theta+1})-\frac{1}{2}}{\sum_{\omega_i\in\Omega}2p_0(\omega_i)\omega_i^2-\frac{1}{2}}\\ =&\frac{\sum_{k=K+1}^{N}p_0(\omega_k)(\omega_k-1/2)(\omega_k^{\theta+1}-(1-\omega_k)^{\theta+1})}{2(\sum_{k=K+1}^{N}p_0(\omega_k)(\omega_k-1/2)^2)(\sum_{k=K+1}^{N}p_0(\omega_k)(\omega_k^{\theta+1}+(1-\omega_k)^{\theta+1}))}.\end{aligned}\tag{13}$$

Since $r_R(s_2)+1$ as a function of p_0 is continuous everywhere on $\Delta(\Omega)$, which is a compact set, $\overline{r}_R(\theta)<\infty$.

Part (ii) and (iii). Similarly define $\underline{r}_R(\theta)$ denote the infimum of $r_R(s_2)$ over the set of all possible priors given state space Ω and parameter θ and define $\overline{\lambda}_2(\theta)$ to be the solution to the following equation, $\lambda\underline{r}_R(\theta)-(1-\lambda)=0$. Since $r_R(s_2)$ as a function of p_0 is positive and continuous everywhere on $\Delta(\Omega)$, $\underline{r}_R(\theta)>0$. It follows that if $\overline{\lambda}_1(\theta)<\lambda<\overline{\lambda}_2(\theta)$, then there exists a positive measure of priors under which the agent overreacts and a positive measure under which the agent underreacts.

Moreover, if $\lambda > \overline{\lambda}_2(\theta)$, then the agent underreacts to any (Ω, p_0).

The remainder of this proof shows that given $\lambda \in (\overline{\lambda}_1(\theta), \overline{\lambda}_2(\theta))$, there exists c_1 such that the agent underreacts as long as $p_0(\{\omega_1, \omega_N\}) > c_1$.

Let $a(\omega_k) \equiv (\omega_k - 1/2)(\omega_k^{\theta+1} - (1-\omega_k)^{\theta+1})$, $b(\omega_k) \equiv (\omega_k - 1/2)^2$, $c(\omega_k) \equiv \omega_k^{\theta+1} + (1-\omega_k)^{\theta+1}$, and $f(\omega_k) \equiv \frac{a(\omega_k)}{b(\omega_k)c(\omega_k)}$. Then $f(\omega_k)$ is the hypothetical value of $r_R(s_2)+1$ if the state space Ω consists of only $1-\omega_k$ and ω_k. We now show that for any $p_0 \in \Delta\Omega$,

$$r_R(s_2) + 1 \geq \min_{k=K+1,...,N} f(\omega_k). \tag{14}$$

This obviously holds if $N = 2$. Suppose $N > 2$ and Eq. (14) does not hold, then for any $i = K+1, ..., N$, we have

$$r_R(s_2) + 1 = \frac{\sum_{k=K+1}^N p_0(\omega_k)a(\omega_k)}{2(\sum_{k=K+1}^N p_0(\omega_k)b(\omega_k))(\sum_{k=K+1}^N p_0(\omega_k)c(\omega_k))} < \frac{a(\omega_i)}{b(\omega_i)c(\omega_i)}. \tag{15}$$

We first show that Eq. (15) cannot hold for all $i = K+1, ..., N$ when p_0 is uniform, i.e. $p_0(\omega_i) = 1/N$ for any $\omega_i \in \Omega$. Rearrange and then summing up the inequalities, we obtain

$$K\left(\sum_{k=K+1}^N b(\omega_k)c(\omega_k)\right) - \left(\sum_{k=K+1}^N b(\omega_k)\right)\left(\sum_{k=K+1}^N c(\omega_k)\right) < 0.$$

This is further equivalent to

$$\sum_{k=K+1}^N \sum_{j=K+1}^N \left(b(\omega_k) - b(\omega_j)\right)\left(c(\omega_k) - c(\omega_j)\right) < 0.$$

However, this is impossible as both $b(\omega_k)$ and $c(\omega^k)$ increase in ω_k when $\omega_k > 1/2$. Therefore, *Eq.* (14) must hold for all N when p_0 is uniform. Suppose p_0 is not uniform but $p_0(\omega)$ is a rational number for each $\omega \in \Omega$. Then we can create an information environment $(\tilde{\Omega}, \tilde{p}_0)$ such that, for all $\omega_i \in \Omega$, $\tilde{\Omega}$ contains n_i copies of ω_i and $\tilde{p}_0$ assigns a total probability of $p_0(\omega_i)$ to this set. Since the overreaction ratio for $(\tilde{\Omega}, \tilde{p}_0)$ is equal to that for (Ω, p_0), we can use the same argument as above to show that Eq. (14) holds for the original environment (Ω, p_0). By continuity, Eq. (14) also holds when $p_0(\omega_i)$ is an irrational number for some ω_i.

It is easy to show that f is a strictly decreasing function of ω_k.[66] Therefore,

[66]Note that f is decreasing in ω_k if and only if $g(x) \equiv \frac{(x-1/2)((1-x)^{\theta+1}+x^{\theta+1})}{x^{\theta+1}-(1-x)^{\theta+1}}$ is increasing in x when $x > 1/2$. Differentiating $g(x)$, we have

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

Note that the numerator can be written as $h(x) - h(1-x)$, where $h(x) \equiv x^{\theta+1}(x^{\theta+1} - (\theta+1)(1-x)^{\theta})$. Since $h(x)$ is increasing in x, it follows that $g'(x) > 0$ for $x > 1/2$.

Eq. (14) implies that $r_R(s_2) + 1 \geq f(\omega_N)$. This minimum is attained when p_0 assigns probability 1 to $\{\omega_1, \omega_N\}$. Since $\lambda \in (\overline{\lambda}_1(\theta), \overline{\lambda}_2(\theta))$, it follows that the agent underreacts when $p_0(\{\omega_1, \omega_N\}) = 1$. By continuity, there exists $c_1 \in (0, 1)$ such that the agent underreacts when $p_0(\{\omega_1, \omega_N\}) > c_1$. □

Proof of Prediction 5. As in the proof of Prediction 2, we can show that

$$\begin{aligned}\lim_{\theta\to\infty} r_R(s_2) + 1 &= \lim_{\theta\to\infty} \frac{|E_R(\omega|s_2) - E_0(\omega)|}{|E_B(\omega|s_2) - E_0(\omega)|} \\ &= \frac{\omega_N - \frac{1}{2}}{\frac{4}{N}\left[\left(\omega_{K+1} - \frac{1}{2}\right)^2 + \ldots + \left(\omega_N - \frac{1}{2}\right)^2\right]}.\end{aligned}$$

The above expression is decreasing in $\omega_{K+1}, \ldots, \omega_{N-1}$. Moreover, fixing $\omega_{K+1}, \ldots \omega_{N-1}$, if $W(\Omega) > 0$, then $\left(\omega_N - \frac{1}{2}\right)^2 > \left(\omega_{K+1} - \frac{1}{2}\right)^2 + \ldots + \left(\omega_{N-1} - \frac{1}{2}\right)^2$, so the above expression is decreasing in ω_N. The proof is analogous for signal realization s_1. It then follows that the agent reacts more in (Ω', p_0') than in (Ω, p_0) for sufficiently large θ. □

Prediction 7. The following prediction is an analogue of Prediction 4, varying the signal diagnosticity instead of the prior.

Prediction 7. *Consider the set* $\mathbf{\Omega}_N$ *of symmetric information environments with complexity* N *and a uniform prior. For each* $\theta > 0$*, there exist cutoffs* $0 < \overline{\lambda}_1(\theta) < \overline{\lambda}_2(\theta) \leq 1$*, with* $\lambda_2(\theta) < 1$ *iff* N *is odd, such that:*

(i) Cognitive-imprecision-dominant: for $\lambda \in [0, \overline{\lambda}_1(\theta))$*, the agent underreacts to all information environments in* $\mathbf{\Omega}_N$*.*

(ii) Cognitive complementarity: for each $\lambda \in (\overline{\lambda}_1(\theta), \overline{\lambda}_2(\theta))$*, there exist a positive measure set of information environments in* $\mathbf{\Omega}_N$ *on which the agent overreacts and a positive measure set on which the agent underreacts. The latter set includes all precise environments with minimum diagnosticity* $\min_{\omega_i \in \Omega} d_i > c_1$ *for some* $c_1 \in (1/2, 1)$*.*

(iii) Representativeness-dominant: for $\lambda \in (\overline{\lambda}_2(\theta), 1]$*, the agent overreacts to all information environments in* $\mathbf{\Omega}_N$*.*

Fig. B.1 illustrates this result for a complexity of $N = 3$. For a given level of cognitive imprecision, it highlights the diagnosticity of the information environment where the agent is predicted to switch from overreaction to underreaction.

Proof. Part (i). Let $\overline{r}_R(\boldsymbol{\Omega_N}, \theta)$ denote the supremum of $r_R(s_j)$ over $\boldsymbol{\Omega_N}$ given parameter θ. Since $r(s_j) = \lambda r_R(s_j) - (1 - \lambda)$, to show the existence of the cognitive-imprecision-dominant region, it suffices to show that $\overline{r}_R(\boldsymbol{\Omega_N}, \theta) < \infty$ for any $\theta > 0$. Moreover, $\overline{\lambda}_1(\theta)$ is given by the solution to the following equation, $\lambda \overline{r}_R(\boldsymbol{\Omega_N}, \theta) - (1 -$

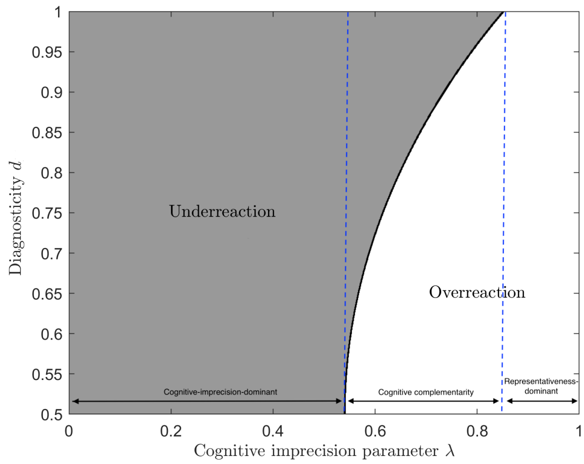


FIGURE B.1. Illustration of Prediction 7 ($\Omega_d = (1-d, 0.5, d)$ for $d \in (0.5, 1)$, uniform prior, $\theta = 0.85$)

$\lambda) = 0$. Note that

$$r_R(s_j) + 1 = \frac{|E_R(\omega|s_j) - E_0(\omega)|}{|E_B(\omega|s_j) - E_0(\omega)|}$$
$$= \frac{(\sum_{\omega_i \in \Omega} \omega_i^{\theta+2})/(\sum_{\omega_i \in \Omega} \omega_i^{\theta+1}) - \frac{1}{2}}{(\sum_{\omega_i \in \Omega} \omega_i^2)/(\sum_{\omega_i \in \Omega} \omega_i) - \frac{1}{2}},$$

where $r_R(s_j) > 0$ if $\theta > 0$. Letting $K = N/2$ if N is even and $K = (N-1)/2$ if N is odd,

$$r_R(s_j) + 1 = \frac{\sum_{k=K+1}^{N} 2\left(\omega_k - \frac{1}{2}\right)\left(w_k^{\theta+1} - (1-\omega_k)^{\theta+1}\right)}{\frac{4}{N}\left(\sum_{\omega_i \in \Omega} \omega_i^{\theta+1}\right)\left[\sum_{k=K+1}^{N}\left(\omega_k - \frac{1}{2}\right)^2\right]}$$
$$\leq \frac{1}{2\left(\frac{1}{2}\right)^{\theta+1}} \frac{\sum_{k=K+1}^{N}\left(\omega_k - \frac{1}{2}\right)\left(w_k^{\theta+1} - (1-\omega_k)^{\theta+1}\right)}{\sum_{k=K+1}^{N}\left(\omega_k - \frac{1}{2}\right)^2}.$$

It suffices to show that

$$h(\omega_k) \equiv \frac{\left(\omega_k - \frac{1}{2}\right)\left(w_k^{\theta+1} - (1-\omega_k)^{\theta+1}\right)}{\left(\omega_k - \frac{1}{2}\right)^2}$$

is bounded above for any $\omega_k \in (1/2, 1)$. This follows from the fact that $h(\omega_k)$ is a continuous positive function over $(1/2, 1]$ and, in addition,

$$\lim_{\omega_k \to 1/2} h(\omega_k) = \lim_{\omega_k \to 1/2} \frac{w_k^{\theta+1} - (1-\omega_k)^{\theta+1}}{\left(\omega_k - \frac{1}{2}\right)}$$
$$= \lim_{\omega_k \to 1/2} (\theta+1)(w_k^{\theta} + (1-\omega_k)^{\theta}) = (\theta+1)\left(\frac{1}{2}\right)^{\theta-1} < \infty.$$

Part (ii). From Part (i), we know that for any $\theta > 0$ and $\lambda > \overline{\lambda}_1(\theta)$, there exists a set of information environments where $r(s_j)$ is strictly positive and the agent overreacts. This set has a positive measure because $r(s_j)$ is a continuous function of the information structure.

Let $\underline{r}_R(\boldsymbol{\Omega_N}, \theta)$ denote the infimum of $r_R(s_j)$ over $\boldsymbol{\Omega_N}$ given parameter θ. We show below that $\underline{r}_R(\boldsymbol{\Omega_N}, \theta) = 0$ for any $\theta > 0$ and N even, and $\underline{r}_R(\boldsymbol{\Omega_N}, \theta) > 0$ for any $\theta > 0$ and N odd. Let $\overline{\lambda}_2(\theta)$ be the solution to the following equation, $\lambda \underline{r}_R(\boldsymbol{\Omega_N}, \theta) - (1 - \lambda) = 0$. Then $\overline{\lambda}_2(\theta) = 1$ if N is even, and $\overline{\lambda}_2(\theta) < 1$ if N is odd. By definition, $\underline{r}_R(\boldsymbol{\Omega_N}, \theta) < \overline{r}_R(\boldsymbol{\Omega_N}, \theta)$, so $\overline{\lambda}_1(\theta) < \overline{\lambda}_2(\theta)$.

Suppose N is even and $N = 2K$. Then letting $\omega_{K+1}, ..., \omega_N$ converge to 1 from below, we have $lim_{\omega_{K+1},...,\omega_N \to 1} r_R(s_j) + 1 = 1$. Hence in this case $\underline{r}_R(\boldsymbol{\Omega_N}, \theta) = 0$.

Suppose N is odd and $N = 2K + 1$. Note that $\omega_{K+1} = 1/2$ and $\omega_i > 1/2$ for $i > K + 1$. Therefore,

$$
\begin{aligned}
r_R(s_j) + 1 &= \frac{\sum_{k=K+1}^{N} 2\left(\omega_k - \frac{1}{2}\right)\left(w_k^{\theta+1} - (1 - \omega_k)^{\theta+1}\right)}{\frac{4}{N}\left(\sum_{\omega_i \in \Omega} \omega_i^{\theta+1}\right)\left[\sum_{k=K+1}^{N}\left(\omega_k - \frac{1}{2}\right)^2\right]} \\
&= \frac{\sum_{k=K+2}^{N}\left(\omega_k - \frac{1}{2}\right)\left(w_k^{\theta+1} - (1 - \omega_k)^{\theta+1}\right)}{\frac{2}{N}\left((1/2)^{\theta+1} + \sum_{i \neq K+1} \omega_i^{\theta+1}\right)\left[\sum_{k=K+2}^{N}\left(\omega_k - \frac{1}{2}\right)^2\right]} \\
&> \frac{\sum_{k=K+2}^{N}\left(\omega_k - \frac{1}{2}\right)\left(w_k^{\theta+1} - (1 - \omega_k)^{\theta+1}\right)}{\frac{2}{N-1}\left(\sum_{i \neq K+1} \omega_i^{\theta+1}\right)\left[\sum_{k=K+2}^{N}\left(\omega_k - \frac{1}{2}\right)^2\right]} > 1,
\end{aligned}
$$

where the first inequality follows from $(1/2)^{\theta+1} < \frac{1}{N-1}\sum_{i \neq K+1} \omega_i^{\theta+1}$ and the second inequality follows from the fact that the left-hand side is equal to the value of $\tilde{r}_R(s_j) + 1$ for an alternative state space $\Omega \setminus \{\omega_{K+1}\}$ and it is strictly larger than 1 when $\theta > 0$. Hence, $r_R(s_j)$ is bounded strictly away from 0 for any $\theta > 0$, and as a result, $\underline{r}_R(\boldsymbol{\Omega_N}, \theta) > 0$.

Part (iii). Immediately follows from Part (ii). □

Proof of Prediction 6. For convenience, we denote the binary state space as $\Omega = \{1 - x, x\}$ where $x > 1/2$ and the prior as $(1 - p_0, p_0)$. We first prove Part (ii) of Prediction 6 since it is more involved.

Part (ii). Without loss of generality, we assume $p_0 > 1/2$ and consider a confirmatory realization s_j. We have

$$
\overline{E}(\omega) = 1/2, \tag{16}
$$

$$
E_0(\omega) = (1 - p_0)(1 - x) + p_0 x, \tag{17}
$$

$$
E_B(\omega|s_j) = \frac{(1 - p_0)(1 - x)^2 + p_0 x^2}{(1 - p_0)(1 - x) + p_0 x}, \tag{18}
$$

$$
E_R(\omega|s_j) = \frac{(1 - p_0)(1 - x)^{\theta+2} + p_0 x^{\theta+2}}{(1 - p_0)(1 - x)^{\theta+1} + p_0 x^{\theta+1}}. \tag{19}
$$

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

$$\lambda E_R(\omega|s_j) + (1-\lambda)\overline{E}(\omega) < E_0(\omega).$$

By Eqs. (16) to (19), the above inequality simplifies to the following,

$$\frac{p_0 x^{\theta+1} - (1-p_0)(1-x)^{\theta+1}}{p_0 x^{\theta+1} + (1-p_0)(1-x)^{\theta+1}} < \frac{2p_0 - 1}{\lambda}. \tag{20}$$

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

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

This inequality simplifies to the following,

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

The agent underreacts to s_j if $E_0(\omega) < \hat{E}(\omega|s_j) < E_B(\omega|s_j)$, which occurs if and only if

$$\frac{2p_0 - 1}{\lambda} < \frac{p_0 x^{\theta+1} - (1-p_0)(1-x)^{\theta+1}}{p_0 x^{\theta+1} + (1-p_0)(1-x)^{\theta+1}} < \frac{1}{\lambda}\frac{p_0 x - (1-p_0)(1-x)}{p_0 x + (1-p_0)(1-x)}. \tag{22}$$

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. (20) to (22), a wrong direction reaction 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 reacts in the wrong direction for all values of $x \in (1/2, 1]$. If $\lambda > 2p_0 - 1$, then there exists a cutoff $c_2 \in (1/2, 1)$ such that $\ell((c_2/(1-c_2))^{\theta+1}) = \frac{2p_0-1}{\lambda}$ and the agent reacts in the wrong direction for all $x \in (1/2, c_2)$.

Second, note that

$$\begin{aligned}\frac{\ell(t^{\theta+1})}{\ell(t)} =& \frac{(p_0 t + (1-p_0))(p_0 t^{\theta+1} - (1-p_0))}{(p_0 t - (1-p_0))(p_0 t^{\theta+1} + (1-p_0))} \\ =& 1 + \frac{2}{\frac{p_0^2 t^{\theta+2} - (1-p_0)^2}{p_0(1-p_0)(t^{\theta+1}-t)} - 1}.\end{aligned}$$

Let $h(t, p_0) \equiv \frac{p_0^2 t^{\theta+2} - (1-p_0)^2}{p_0(1-p_0)(t^{\theta+1}-t)}$. We now show that h is first decreasing and then increasing in t. Note that

$$h_t(t, p_0) = \frac{p_0^2 t^{2\theta+2} + (1+\theta)t^\theta((1-p_0)^2 - p_0^2 t^2) - (1-p_0)^2}{(1-p_0)p_0 t^2 (t^\theta - 1)^2} \tag{23}$$

we can solve for each t a unique value of $p_0 \in (0, 1)$ such that $h_t(t, q_0) = 0$, and write this as $p_0^*(t)$. In particular,

$$1/p_0^*(t) = 1 + \frac{\sqrt{t^{\theta+2}(1-t^\theta+\theta)(t^\theta(1+\theta)-1)}}{t^\theta(1+\theta)-1}. \tag{24}$$

Let $g(t,p_0)$ denote the numerator of $h_t(t,p_0)$ in Eq. (23). We can show that for any $t>1$, $g_t(t,p_0)>0$ if $p_0=p_0^*(t)$.[67] Since $g(1,p_0)<0$ when $p_0>1/2$ and $g(t,p_0)>0$ for any $t>(\theta+1)^{1/\theta}$, it follows that $g(t,p_0)=0$ for at most one value of t at any p_0 (otherwise there exists a root $\hat{t}$ such that g crosses 0 from above and then $g_t<0$ at this point). This further implies that when $p_0>1/2$, h is first decreasing and then increasing in t, and hence the ratio $\ell(t^{\theta+1})/\ell(t)$ is first increasing and then decreasing in t.

Since $\frac{2p_0-1}{\lambda} = \ell((c_2/(1-c_2))^{\theta+1}) < \frac{\ell(c_2/(1-c_2))}{\lambda}$, 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 $c_2/(1-c_2)$. 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 $\frac{1}{\lambda}\lim_{t\to1}\ell(t) = \frac{2p_0-1}{\lambda} < \lim_{t\to1}\lim_{\theta\to\infty}\ell(t^{\theta+1}) = 1$. Therefore, if θ sufficiently large, there exists an x close to $1/2$ such that the agent overreacts. Combining these observations, we know that there exist $c_2 \le c_3 \le c_4 \le 1$ such that the agent underreacts when $x\in(c_2,c_3)\cup(c_4,1)$, overreacts when $x\in(c_3,c_4)$. In addition, $(c_2,c_3)\cup(c_4,1)$ is non-empty if $\lambda>2p_0-1$, and (c_3,c_4) is non-empty if θ is sufficiently large.

Part (i). Next assume $p_0<1/2$ and consider a disconfirmatory realization s_j. Then $E(\omega|s_j)>E(\omega)$. As in Part (i), a wrong direction reaction 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 reaction is impossible. It remains to determine whether the agent overreacts or underreacts by comparing $\ell(t^{\theta+1})$ and $\frac{\ell(t)}{\lambda}$. Note that when $t<(1-p_0)/p_0$, we have $\ell(t)<\ell((1-p_0)/p_0)=0$ and thus $\frac{\ell(t)}{\lambda}<\ell(t)<\ell(t^{\theta+1})$. That is, the agent overreacts when $x\in(1/2,(1-p_0)/p_0)$. When $t>(1-p_0)/p_0$, we have $\ell(t^{\theta+1})>\ell(t)>0$. Repeating the steps in the proof of Part (ii), we can obtain the same expressions as Eqs. (23) and (24). For any $t>(1-p_0)/p_0$, since $\ell(t^{\theta+1})>\ell(t)>0$, we know that $h(t,p_0)>1$ for all $p_0\in(0,1/2)$. However, Eq. (24) does not have a solution $p_0^*(t)\in(0,1/2)$ for any t. So for any $p_0\in(0,1/2)$, $h(t,p_0)$ must be either increasing or strictly decreasing in t for all $t>1$. This combined with the fact that $h_t((\theta+1)^{1/\theta},p_0)>0$ for any $p_0\in(0,1/2)$ implies that $h(t,p_0)$ is strictly increasing in t. Therefore, $\ell(t^{\theta+1})/\ell(t)$ is strictly decreasing in t for any $p_0\in(0,1/2)$

[67] Differentiating g yields $g_t(t,p_0)=t^{\theta-1}(1+\theta)(\theta-2p_0\theta+p_0^2(2t^{\theta+2}+\theta-t^2(\theta+2)))$. Plugging in $p_0=p_0^*(t)$ we obtain

$$g_t(t,p_0^*(t)) = t^{\theta-1}(1+\theta)\frac{t^2(t^\theta-1)^2(\theta+2)(t^\theta(1+\theta)-1)}{\left(-1+t^\theta(1+\theta)+\sqrt{t^{\theta+2}(1-t^\theta+\theta)(t^\theta(1+\theta)-1)}\right)^2} > 0.$$

and $t > (1-p_0)/p_0$. 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 $c_1 \in ((1-p_0)/p_0, 1)$ such that the agent overreacts if $x \in (0, c_1)$ and underreacts if $x \in (c_1, 1]$. $\square$

C Alternative Models of Cognitive Imprecision

C.1 Comparison with Augenblick et al. (2022)

In this section, we compare our two-stage model with the cognitive imprecision model proposed by Augenblick et al. (2022) (abbreviated as ALT below). In contrast to our multi-state setting, ALT restricts attention to a setting with binary states $\{\omega_0, \omega_1\}$. In their model, the agent perceives the *strength* of a signal s, denoted by $\mathbb{S} = \left|\ln\left(\frac{P(s|\omega=\omega_1)}{P(s|\omega=\omega_0)}\right)\right|$, with cognitive imprecision. Similar to our agent in the processing stage, their agent is endowed with a "cognitive prior" about the logarithm of the signal strength and updates their belief after observing a noisy representation of it denoted by r. This leads to perceived signal strength given by $\log \hat{\mathbb{S}}(r) = (1-\eta)\log \bar{\mathbb{S}} + \eta \cdot r$, where $\bar{\mathbb{S}}$ is the prior mean over signal strengths and η is a constant whose value depends on the amount of cognitive noise. Since the agent biases towards a moderate level of signal strength, it is clear that similar to our Prediction 5, this model also predicts overreaction to noisy signals (low signal strength) and underreaction to precise signals (high signal strength).

Apart from the multi-state versus binary-state settings, there are two major conceptual differences between ALT and our model. First, while our model imposes cognitive noise on the agent's posterior directly, the agent in ALT first perceives the signal strength with cognitive noise and then applies Bayes' rule using the correct prior. Hence, while our processing stage implies both base-rate neglect and signal-diagnosticity neglect, ALT only implies the latter. It follows that ALT does not predict our Prediction 6, namely, the agent may update in the wrong direction after observing noisy confirmatory signals.

Second, although both ALT and our model predict underreaction to precise signals and overreaction to noisier signals, the driving mechanisms are fundamentally different: in ALT this results from the assumption of a moderate cognitive default, while in our model this is generated by the interaction between channeled attention and cognitive imprecision. Distinguishing the two mechanisms is challenging in binary-state environments because of similar predictions in beliefs. This motivates the next section, where we extend an adapted version of ALT to multi-state settings and compare its predictions with our two-stage model.

C.2 Flexible Cognitive Imprecision Model

We now explore a more general version of the processing stage of our model, allowing the cognitive default to deviate from the ignorance prior and vary by the signal realization and state space complexity. When restricted to binary information environments, this flexible cognitive imprecision model captures the same spirit as Augenblick et al. (2022).[68] We derive the predictions of the flexible cognitive imprecision model and demonstrate that despite introducing more parameters, the flexible cognitive imprecision model does not fit the experimental data as well as our two-stage model, especially in complex information environments. We focus on information environments with a uniform prior for a clean comparison.

Consider an agent who perceives signal diagnosticities with a flexible form of cognitive imprecision: his prior belief about the objective posterior centers around cognitive default $\overline{p}_0(s_j, N) \in \Delta(\Omega)$, which may vary according to the signal realization s_j and the complexity of the state space, $N \equiv |\Omega|$ denotes. The agent combines this prior belief with a noisy representation extracted from the information environment, resulting in an average subjective posterior given by

$$\hat{p}(s_j) \equiv \lambda p_B(s_j) + (1-\lambda)\overline{p}_0(s_j, N). \tag{25}$$

When $\overline{p}_0(s_j, N) \equiv \overline{p}_0$ for all s_j and N, this reduces to the processing stage of our model. Allowing the cognitive default to deviate from the "ignorance prior" can capture the notion that the agent thinks that the signal should be somewhat informative by default. To maintain discipline, we make the following assumptions. We assume that $\overline{p}_0(s_j, N)$ takes the value of a Bayesian posterior derived from a *default information environment* with a symmetric *default state space* $\overline{\Omega}(N) \equiv \{\overline{\omega}_1, ..., \overline{\omega}_N\}$ and a uniform *default prior* $\overline{p}_0$. That is, $\overline{p}_0(s_j, N) = \mathcal{B}(s_j, \overline{\Omega}(N), \overline{p}_0)$, where $\mathcal{B}$ denotes the Bayesian operator.[69] We assume $0 < \overline{\omega}_1 \leq ... \leq \overline{\omega}_N < 1$, which rules out the case that the cognitive default assigns probability 0 to some states. Moreover, we assume that the cognitive default is symmetric across signal realizations and it aligns with the direction of the signal realization relative to a uniform prior, $(\overline{E}(\omega|s_j, N) - 1/2)(E(\omega|s_j) - 1/2) \geq 0$. For example, suppose the agent's default state space for binary information environments is $\overline{\Omega}(2) = \{0.3, 0.7\}$. Upon observing s_2, he compresses his posterior towards a cognitive default with $\overline{p}_0(\omega_1|s_2, N) = 0.3$; upon observing s_1, he biases towards $\overline{p}_0(\omega_1|s_1, N) = 0.7$. Compared to the cognitive-imprecision-only model, this model has seven additional parameters for the set of information environments we considered.[70]

Similar to ALT and our two-stage model, the flexible cognitive imprecision model predicts that the agent tends to overreact to precise signals and underreact to noisy

[68] ALT incorporates cognitive imprecision in signal strength exponentially while the flexible cognitive imprecision model considered here incorporates it linearly, but this does not affect the main qualitative predictions.

[69] For any information environment (Ω,p_0) and signal realization s_j, let $\mathcal{B}(s_j, \Omega, p_0)$ represents the implied Bayesian posterior.

[70] This includes one diagnosticity parameter for binary-state information environments, one for 3-three environments, two for 4-state environments, and another two for 5-state environments. Notably, the number of free parameters increases as one considers more information structures with higher complexities.

signals (Predictions 5 and 7). However, the flexible cognitive imprecision model does not predict our key result (Prediction 2) that higher complexity leads to more overreaction unless substantial assumptions are imposed on how the cognitive default varies across complexities. For illustration, suppose the agent's default state space for complexity $N = 2$ is given by $\overline{\Omega}(2) = \{0.3, 0.7\}$. Then he overreacts in a binary-state information environment with state space $\{1-d, d\}$ iff the signal diagnosticity $d > 0.5$ is below 0.7 and underreacts iff d is above 0.7. Now moving on to more complex information environments, the agent does not necessarily overreact more. For example, given a natural choice of the 3-state default state space, $\overline{\Omega}(3) = \{0.3, 0.5, 0.7\}$, the agent overreacts in the more complex environment with $\{1 - d, 0.5, d\}$ if and only if he also overreacts in the simpler environment with $\{1 - d, d\}$.

Analyzing the subjective belief state-by-state provides the simplest test to distinguish the flexible cognitive imprecision model and the two-stage model. Prediction 8 below shows that the agent always distorts his probabilistic assessments of the most and least representative states in different directions—underweighing one and overweighing the other. In addition, if the signal diagnosticity associated with the extreme states is sufficiently high, the agent *underweighs* the most representative state and *overweighs* the least representative state since cognitive imprecision pulls his posterior back to the moderate cognitive default. The proof of Prediction 8 is straightforward.[71]

Prediction 8 (Flexible Cognitive Imprecision Model)**.** *Fix any symmetric information environment* (Ω, p_0) *with* $|\Omega| = N \geq 2$ *and a uniform prior. Consider an agent who updates according to a flexible cognitive imprecision model with parameter* $\lambda \in (0, 1)$ *and default state space* $\overline{\Omega}(N)$*. Given a fixed set of interior states* $\Omega \setminus \{\omega_R, \omega_{NR}\}$*, there exists a cutoff* $d \in (1/2, 1)$ *such that:*

(i) If $\omega_R = 1 - \omega_{NR} > d$*, the agent* underweighs ω_R *and* overweighs ω_{NR}*.*

(ii) If $\omega_R = 1 - \omega_{NR} < d$*, the agent* overweighs ω_R *and* underweighs ω_{NR}*.*

Moreover, the agent neither under- nor overweighs the set of interior states $\Omega_I = \Omega \setminus \{\omega_R, \omega_{NR}\}$*.*

Fig. C.1 depicts the predictions of the flexible noise model, aggregating across uniform prior information environments used in experiments. In contrast, as shown in Prediction 1, the two-stage model allows the agent to *overweigh* both the most and the least representative state, as well as *overweigh* the most representative state and *underweigh* the least representative state even after observing signals with high diagnosticity at the extreme states. Comparing Fig. C.1 and Fig. 3, we observe that

[71] Note that $\hat{p}(\omega_i|s_j) = \lambda p_B(\omega_i|s_j) + (1-\lambda)\overline{p}_0(\omega_i|s_j, N) = \frac{2}{N}\left(\lambda\omega_i + (1-\lambda)\overline{\omega}_i\right)$. Letting $d = \overline{\omega}_N = 1 - \overline{\omega}_1$, then the agent overweighs ω_{NR} and underweight ω_R if and only if $\omega_R = 1 - \omega_{NR} > d$, and the opposite holds if and only if $\omega_R = 1 - \omega_{NR} < d$.

the data is consistent with the two-stage model and inconsistent with the flexible cognitive imprecision model.

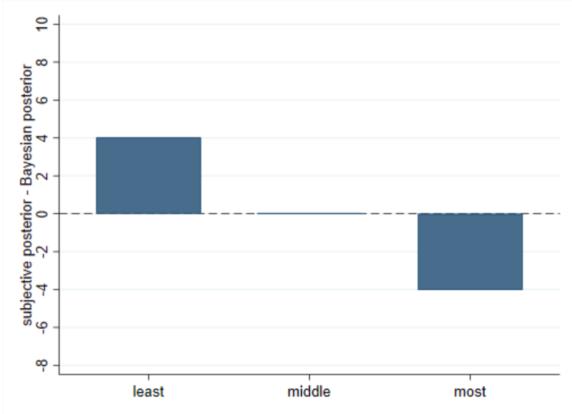


Figure C.1. The predictions of the flexible cognitive imprecision model on the difference between subjective posterior beliefs and the Bayesian posterior beliefs for the least and most representative states and states that are in between, aggregating across uniform prior informational environments used in the experiment. Structural estimates of the cognitive imprecision parameter λ and default state spaces are used to generate the plotted predictions, where $\lambda = 0.5$, $\overline{\Omega}(2) = \{0.49, 0.51\}, \overline{\Omega}(3) = \{0.3, 0.5, 0.7\}, \overline{\Omega}(4) = \{0.2, 0.4, 0.6, 0.8\}, \overline{\Omega}(5) = \{0.2, 0.4, 0.5, 0.6, 0.8\}$.

We also compute the completeness and the restrictiveness of the flexible cognitive imprecision model. As shown in Table C.1, the flexible cognitive model achieves 100% completeness in simple binary environments and 65% completeness in complex environments with more than two states. Note that the former is unsurprising since even the processing stage of our model alone achieves perfect completeness in simple environments, and the flexible cognitive imprecision model strictly nests it. However, it is noteworthy that the two-stage model achieves much higher completeness in complex environments (92% versus 65%). This is rather remarkable considering the fact that our two-stage model only adds one single representativeness parameter to the processing-only model whereas the flexible cognitive imprecision model adds a total of six more parameters. This is also reflected from the restrictiveness analysis—the flexible cognitive model is less restrictive than both the two-stage model and the processing-only model in all environments.

Table C.1. Completeness and Restrictiveness

	Completeness		Restrictiveness	
	2 states	>2 states	2 states	>2 states
Flexible Cognitive Noise Model	1.00	0.65	0.70	0.89
	(0.07)	(0.03)	(0.00)	(0.00)

Notes: Includes all information environments listed in Table D.1 except for the 11-state complexity; includes wrong direction updates. Restrictiveness is estimated from 1000 simulations.

D Additional Details and Analyses

D.1 Experimental Details

TABLE D.1. Information environments used in experiments

COMPLEXITY $\lvert\Omega\rvert$	PRIOR p_0	INFORMATION STRUCTURE Ω
2 states	$p_0(\omega_1) \in \{0.3, 0.5, 0.7\}$ $p_0(\omega_2) = 1 - p_0(\omega_1)$	$Pr(r\lvert\omega_2) \in \{0.6, 0.7, 0.8, 0.9\}$ $Pr(r\lvert\omega_1) = 1 - Pr(r\lvert\omega_2)$
3 states	$p_0(\omega_1) \in \{0.25, 0.33, 0.4\}$ $p_0(\omega_2) = 1 - 2p_0(\omega_1)$ $p_0(\omega_3) = p_0(\omega_1)$	$Pr(r\lvert\omega_3) \in \{0.6, 0.7, 0.8, 0.9\}$ $Pr(r\lvert\omega_2) = 0.5$ $Pr(r\lvert\omega_1) = 1 - Pr(r\lvert\omega_3)$
4 states	$p_0(\omega_i) = 0.25$ $\forall\omega_i \in \Omega$	$(Pr(r\lvert\omega_3), Pr(r\lvert\omega_4)) \in \{(0.55, 0.6),$ $(0.6, 0.7), (0.55, 0.7), (0.7, 0.8),$ $(0.6, 0.8), (0.55, 0.8), (0.8, 0.9),$ $(0.7, 0.9), (0.6, 0.9), (0.55, 0.9)\}$ $Pr(r\lvert\omega_2) = 1 - Pr(r\lvert\omega_3)$ $Pr(r\lvert\omega_1) = 1 - Pr(r\lvert\omega_4)$
5 states	$p_0(\omega_i) = 0.2$ $\forall\omega_i \in \Omega$	$(Pr(r\lvert\omega_4), Pr(r\lvert\omega_5)) \in \{(0.55, 0.6),$ $(0.6, 0.7), (0.55, 0.7), (0.7, 0.8),$ $(0.6, 0.8), (0.55, 0.8), (0.8, 0.9),$ $(0.7, 0.9), (0.6, 0.9), (0.55, 0.9)\}$ $Pr(r\lvert\omega_3) = 0.5$ $Pr(r\lvert\omega_2) = 1 - Pr(r\lvert\omega_4)$ $Pr(r\lvert\omega_1) = 1 - Pr(r\lvert\omega_5)$
11 states	$p(\omega_i) = 1/11$ $\forall\omega_i \in \Omega$	$Pr(r\lvert\omega_i) = (i-1)/10$ $\forall i \in \{1, ..., 11\}$

Notes: States are ordered by number of red balls, with ω_1 corresponding to the bag with the fewest red balls, and so on up through ω_N corresponding to the bag with the most red balls. All environments are symmetric, aside from the 2-state environments with $p_0(\omega_1) \in \{0.3, 0.7\}$.

Discussion of Measurement. 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,

$$\log \frac{\hat{p}(\omega_2|s_j)}{\hat{p}(\omega_1|s_j)} = c_1 \log \frac{p_0(\omega_2)}{p_0(\omega_1)} + c_2 \log \frac{\pi(s_j|\omega_2)}{\pi(s_j|\omega_1)}.$$

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 where multiple likelihood ratios are needed to capture all distinct pairs of states. Adapting the Grether regression to the multi-state setting yields that

$$\log \frac{\hat{p}(\omega_i|s_j)}{\hat{p}(\omega_k|s_j)} = \tilde{c}_1 \log \frac{p_0(\omega_i)}{p_0(\omega_k)} + \tilde{c}_2 \log \frac{\pi(s_j|\omega_i)}{\pi(s_j|\omega_k)},$$

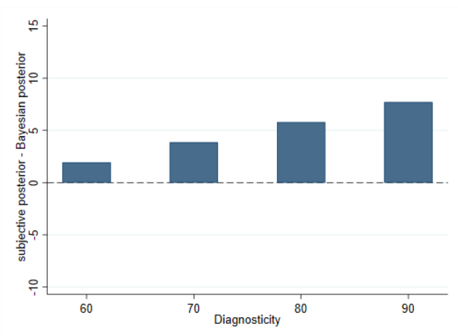
where $i, k \in \{1, ..., N\}$ and $i > k$. However, this imposes a strong assumption on the underlying distortionary force, namely that the agent distorts the prior odds and signal likelihoods of each pair of states in an identical way, which is clearly inconsistent with our experimental data as participants' reactions to different states are often non-monotone (see Section 3.3).[72] Furthermore, the Grether regression imposes a log-linear structure in the decomposition of over- and underreaction into prior distortion and signal likelihood distortion. While our measures based on expectations and state-by-state belief movement do not aim to distinguish between prior-based and signal-based distortions, they serve as better measures of over- and underreaction since they are non-parametric and thus free from restrictions mentioned above.

One potential concern with using the overreaction ratio $r(s)$ is that changes in complexity and the information structure also change the Bayesian benchmark. Since the measure of overreaction used in $r(s)$ is defined relative to the Bayesian benchmark, we may find a shift towards overreaction if participants use a constant heuristic that reports the same posterior belief independently of changes in the information environment or are subject to some version of partition dependence (Fox, Bardolet, and Lieb 2005; Tversky and Koehler 1994; Benjamin 2019).[73] We address this concern in several ways. First, the state-by-state analysis reported in Section 3.3 is not subject to this issue as it tests the predictions of our model for each state in the information environment; for example, Fig. D.2 shows that beliefs do not follow a simple information-independent heuristic. Second, Section 4 presents evidence for our framework in a setting that keeps the information environment constant, which rules out mechanisms such as partition dependence.

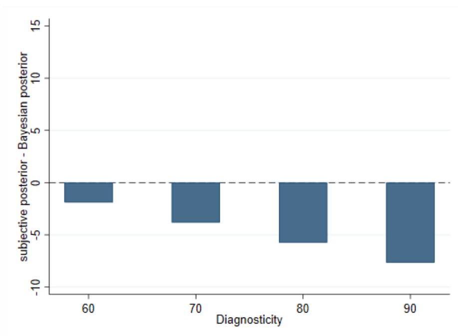
D.2 Additional Analysis: State-by-State

[72] For example, this assumption implies that under a uniform prior (so that $\tilde{c}_1$ does not matter), the agent either exaggerate all the posterior odds ($\tilde{c}_2 > 1$) or understate all posterior odds ($\tilde{c}_2 < 1$). This is strongly rejected by our data even when there are only three states. For instance, when $|\Omega| = 3$ and the prior is uniform, participants often overweigh both ω_1 and ω_3 and underweigh ω_2, suggesting that participants exaggerate $\log \frac{\pi(s_2|\omega_3)}{\pi(s_2|\omega_2)}$ and understate $\log \frac{\pi(s_2|\omega_2)}{\pi(s_2|\omega_1)}$.

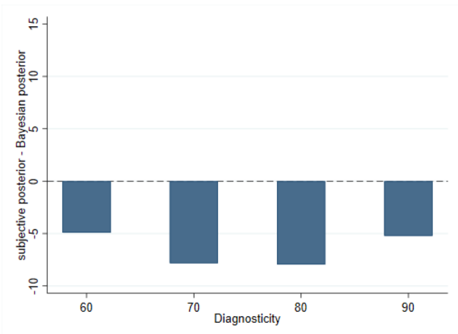
[73] Partition dependence leads to subadditivity of judgments, where people place a greater likelihood on an event when it is partitioned into mutually exclusive sub-events. Tversky and Koehler (1994) first demonstrated this phenomenon and offered Support Theory as the explanation, which posits that judgment likelihoods are a reflection of the evidence that 'comes to mind' when events are described. Partition dependence emerges from Support Theory because the description of the sub-events increases people's perceived likelihood of each event, thereby increasing their total perceived likelihood.



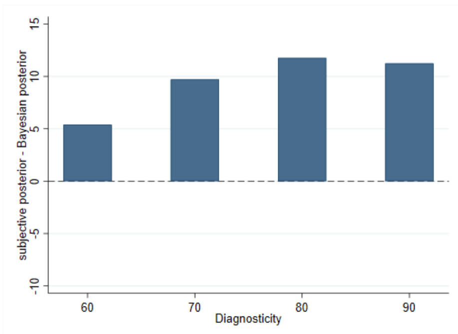
(A) CI-Only: Least Rep. State



(B) CI-Only: Most Rep. State

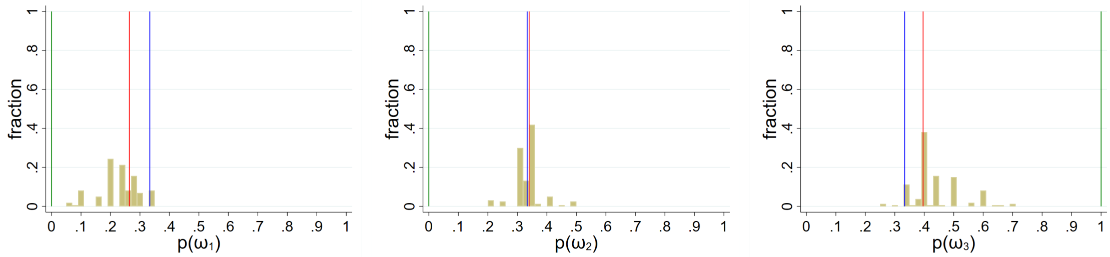


(C) Rep-Only: Least Rep. State

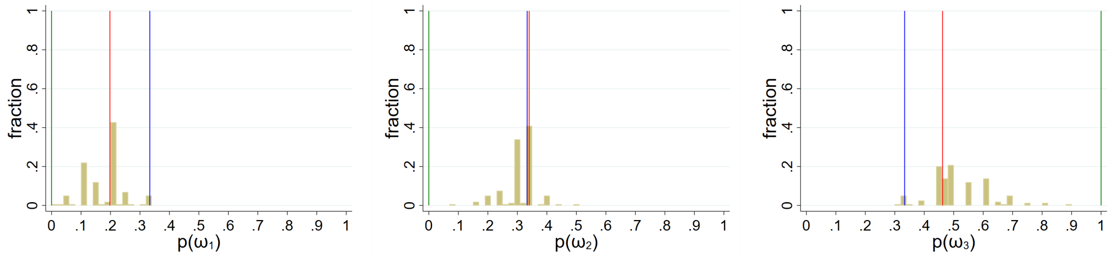


(D) Rep-Only: Most Rep. State

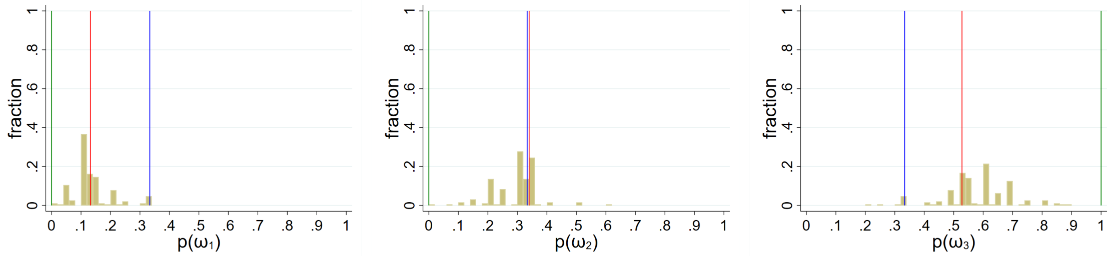
FIGURE D.1. Over- and Underweighing by Diagnosticity for One-Stage Models. Each figure aggregates all uniform prior environments of a given maximal diagnosticity d_N. Based on structural estimates of θ and λ ((a) and (b) $\theta = 0$, $\lambda = 0.7$, (c) and (d) $\theta = 0.85$, $\lambda = 1$).



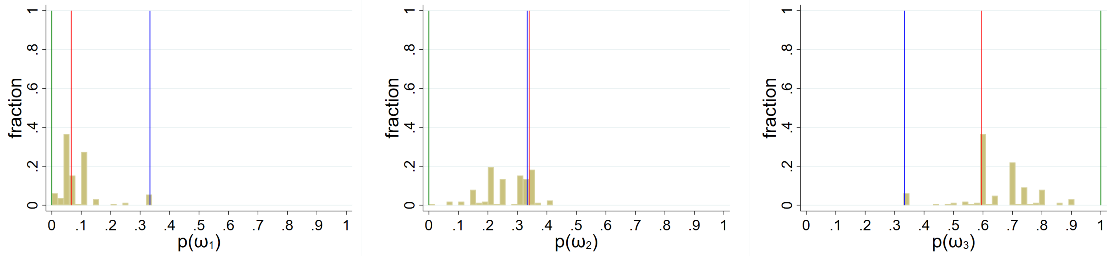
(A) $\Omega = \{0.4, 0.5, 0.6\}$, uniform prior



(B) $\Omega = \{0.3, 0.5, 0.7\}$, uniform prior



(C) $\Omega = \{0.2, 0.5, 0.8\}$, uniform prior



(D) $\Omega = \{0.1, 0.5, 0.9\}$, uniform prior

FIGURE D.2. Distribution of reported posteriors (by state) in information environments with three states. The red line indicates the corresponding Bayesian posterior, the blue line indicates the cognitive default as well as the uniform prior, and the green line corresponds to a value of 1 for the most representative state and 0 if otherwise (i.e. the subjective posterior if $\theta = \infty$ and $\lambda = 1$).

D.3 Additional Analysis: Overreaction Ratio

D.3.1 Regression Analyses of Overreaction Ratio

Table D.2. 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 uniform prior information environments with 2, 4 or 5 states listed in Table D.1; excludes wrong direction updates. Standard errors clustered at the individual level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

TABLE D.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: Baseline is uniform prior and, in Column 2, diagnosticity $d = 0.6$. Includes all 3-state information environments listed in Table D.1; excludes wrong direction updates. Standard errors clustered at the individual level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table D.4. 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 uniform prior information environments listed in Table D.1 except for the 11-state complexity; excludes wrong direction updates. The results do not change qualitatively if we further split the analysis by diagnosticity of the interior states. Standard errors clustered at the individual level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table D.5. 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: Baseline is uniform prior and, in Column 2, diagnosticity $d = 0.6$. Includes all 2-state information environments listed in Table D.1; excludes wrong direction updates. Standard errors clustered at the individual level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

D.3.2 Regression Analyses of Overreaction Ratio including Wrong Direction Reactions

TABLE D.6. 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 uniform prior information environments with 2, 4 or 5 states listed in Table D.1; includes wrong direction updates. Standard errors clustered at the individual level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

TABLE D.7. 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 uniform prior information environments listed in Table D.1 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 D.8. 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

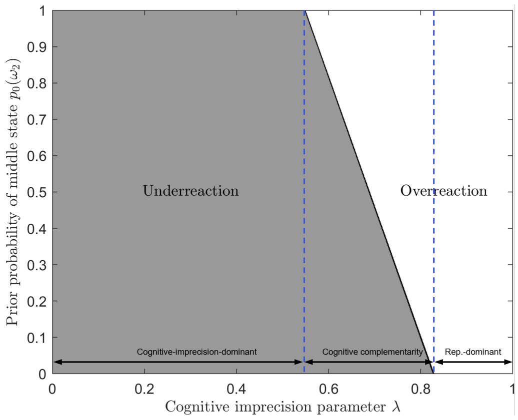
Notes: Baseline is uniform prior and, in Column 2, diagnosticity $d = 0.6$. Includes all 3-state information environments listed in Table D.1; includes wrong direction updates. Standard errors clustered at the individual level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

TABLE D.9. Impact of signal type on belief updating

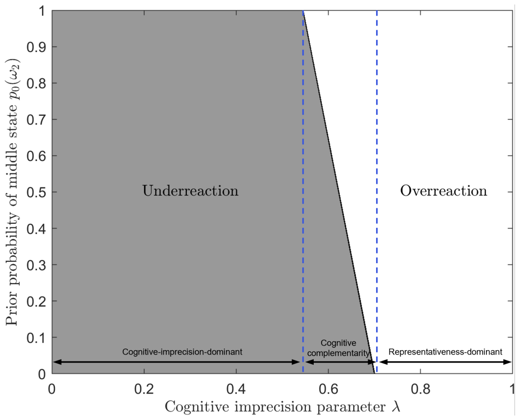
	Overreaction Ratio	
	(1)	(2)
Confirmatory Realization	0.121***	0.136***
	(0.0384)	(0.0392)
Disconfirmatory Realization	0.371***	0.348***
	(0.0442)	(0.0412)
$d = 0.7$		-0.493***
		(0.0569)
$d = 0.8$		-0.590***
		(0.0565)
$d = 0.9$		-0.637***
		(0.0560)
Constant	-0.109***	0.321***
	(0.0199)	(0.0488)
N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

Notes: Baseline is uniform prior and, in Column 2, diagnosticity $d = 0.6$. Includes all 2-state information environments listed in Table D.1; includes wrong direction updates. Standard errors clustered at the individual level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

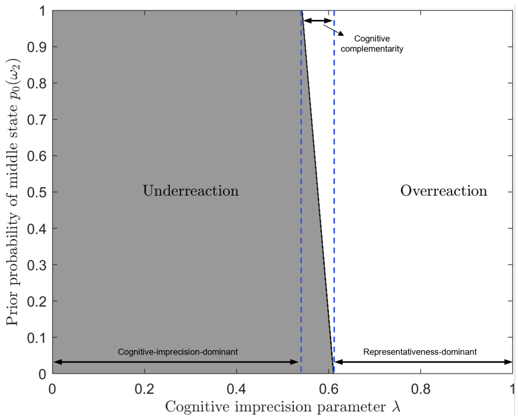
D.3.3 Illustration of Prediction 4 for 3-State Environments in Experiment



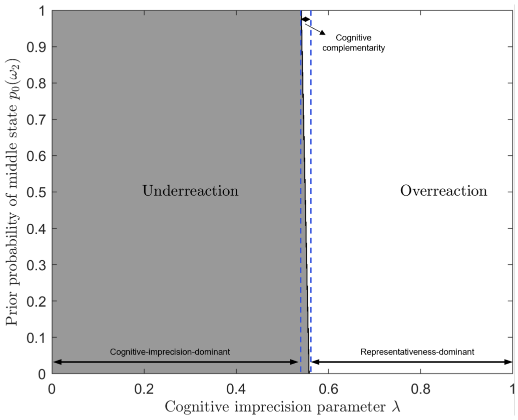
(A) $\Omega = (0.1, 0.5, 0.9)$, symmetric prior, $p_0(\omega_2) \in (0, 1)$, $\theta = 0.85$



(B) $\Omega = (0.2, 0.5, 0.8)$, symmetric prior, $p_0(\omega_2) \in (0, 1)$, $\theta = 0.85$



(C) $\Omega = (0.3, 0.5, 0.7)$, symmetric prior, $p_0(\omega_2) \in (0, 1)$, $\theta = 0.85$



(D) $\Omega = (0.4, 0.5, 0.6)$, symmetric prior, $p_0(\omega_2) \in (0, 1)$, $\theta = 0.85$

FIGURE D.3. Illustration of Prediction 4

D.4 Alternative Design: Reporting Expectations

This experiment mirrored the main study design with one modification. On each trial, after being presented with the information structure, participants reported their beliefs as the probability of drawing a Red ball; note that this belief is a function of their beliefs about the state-by-state probabilities. We used their expectation of a Red ball being drawn out of the chosen bag as the main variable of interest. This corresponds directly to $E(\omega|s_i)$ which is used to calculate the overreaction ratio $r(s_i)$.

We used their subjective expectations to replicate the 2-state, 3-state, and 5-state complexity treatments of our main study. These results are presented in Figure D.4 and Table D.10 below. As can be seen, all of the main results replicate. We see significant underreaction in the 2-state condition but significant overreaction as complexity increases. We also see the predicted relationship with signal diagnosticity, with more overreaction for noisier signals.

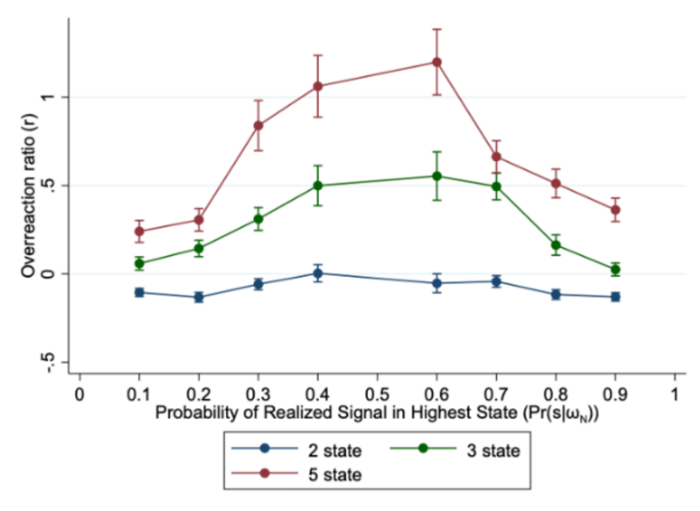


FIGURE D.4. Overreaction Ratio by State and Signal Strength

TABLE D.10. Complexity increases overreaction

	Overreaction Ratio	
	(1)	(2)
3 States	0.365***	0.368***
	(0.0458)	(0.0459)
5 States	0.466***	0.547***
	(0.0518)	(0.0556)
$d = 0.7$		-0.124***
		(0.0375)
$d = 0.8$		-0.349***
		(0.0430)
$d = 0.9$		-0.449***
		(0.0479)
Constant	-0.0788***	0.150***
	(0.0232)	(0.0412)
N	4063	4063
adj. R^2	0.072	0.117

Notes: Baseline is 2 states and, in Column 2, diagnosticity $d = 0.6$. Includes uniform prior information environments with 2, 4 and 5 states listed in Table D.1; excludes wrong direction updates. Standard errors clustered at the individual level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

D.5 Analyses of Mechanism

D.5.1 Measuring Attention

To pin down the attention mechanism, we first develop a method of measuring attention based on the Mouselab paradigm of Payne et al. (1993). We modified the

5 state conditions in our Baseline design by asking participants to click on a state before entering their beliefs. As outlined in Section 4.1, the paradigm itself restricts the stock of attention, while first-click is a validated measure of channelled attention. Importantly, this Limited Attention treatment does not change the informational environment relative to the standard Baseline condition.

The first column of Table D.11 shows that restricting attention increased overreaction significantly. The second column of the same table breaks down the Limited Attention treatment into trials in which the first click was on the representative state or not. This is meant to divide participants into those who employ representativeness as a salience cue or not. Those who appear to use representativeness as a salience cue display significantly more overreaction than those who do not. Finally, Table D.12 presents the sturctural estimates from the Limited Attention treatment in comparison to the Baseline condition with the same information structure. Consistent with our prediction, restricting attention exacerbates the distortion in the mental representation, captured by the higher θ in the Limited Attention treatment, while not affecting processing capacity, captured by the unchanged λ.

TABLE D.11. 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 D.1; excludes wrong direction updates. Standard errors clustered at the individual level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

D.5.2 Causal Effect of Attention

We proceeded to explore the proposed attentional mechanism further through several ways. Each of these methods build on the Limited Attention condition and use it the baseline.

We first developed a paradigm to mirror situations in which there is uncertainty over which state is representative or the representativeness cue is absent all-together.

Table D.12. Limited attention increases representativeness θ

	θ	95% CI	λ	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)

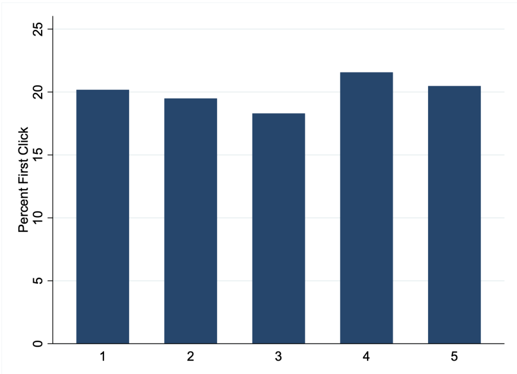
Notes: This table compares the parameter estimates that minimize the average KL divergence at the aggregate level for the limited attention treatment and the baseline attention treatment. Includes all information environments with 5 states listed in Table D.1; excludes wrong direction updates. The 95% confidence intervals are obtained from 300 bootstrap samples.

We used a version of the Limited Attention paradigm but suppressed the representativeness cue by hiding the number of red and blue balls associated with each state until participants clicked a "reveal" button for that state. Otherwise, the design was identical to the original Limited Attention condition. Because information on the representativeness of each state was initially not available, attention was predicted to be channeled as-if randomly in this Representativeness Suppressed. Our framework predicts that this will generate *underreaction* in the same information environment as the Limited Attention condition, where marked overreaction was observed.

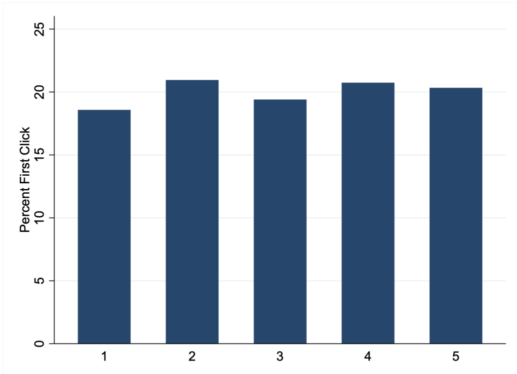
Fig. D.5 shows that, in contrast to the Limited Attention condition, participants' clicking behavior was not associated with the state's representativeness. This suggests that attention was channeled as-if randomly in the Representativeness Suppressed condition. As shown in Fig. D.7b, consistent with our framework, we find that this lead to underreaction across all signal diagnosticities. These results highlight that the emergence of over versus underreaction depend critically on the presence of representativeness as a salience cue.

We then sought to explore the impact of low-level (visual) and top-down (goal-directed) salience in channeling attention and driving belief-updating in our setting. We did this by first increasing the salience of the most representative state in the Representativeness Suppressed condition. The most representative state was visually highlighted in yellow against a neutral background, similar to the method of Li and Camerer (2022); we also instructed participants that they would be paid based on their reported beliefs for that state to manipualte top-down salience. Even though representativeness was suppressed as a salience cue, we predicted that the visual and top-down salience cues would channel participants' attention to the representative state. Fig. D.6 shows that this was indeed the case. Moreover, Fig. D.7b shows that the introduction of these salience cues to the representative state brought back overreaction, which provides additional evidence the critical role of attention in belief-updating.

To examine the impact of the representativeness salience cue relative to visual and top-down salience, we then added just visual salience (Visual Salience condition) and visual salience plus goal directed salience (Goal-Directed and Visual Salience) cues to the *least representative* state in the Limited Attention condition. Note that by placing these alternative salience cues on the least representative state, if visual and goal-directed salience dominate representativeness, then we should see undrreaction in these conditions. Instead, as shown inFigure D.7a, we still observe overreaction in both the Visual Salience and Goal-Directed and Visual Salience conditions. The extent of overreaction is similar to the Limited Attention condition. A state-by-state analysis, depicted in Figure D.8, shows a similar picture: when representativeness is present as a salience cue, it dominates both the visual and goal-directed salience cues in overweighting the beliefs about the associated state.[74] Together, these results suggest that the representativeness-based salience cue, when present, plays a significant role in channeling attention in belief-updating.

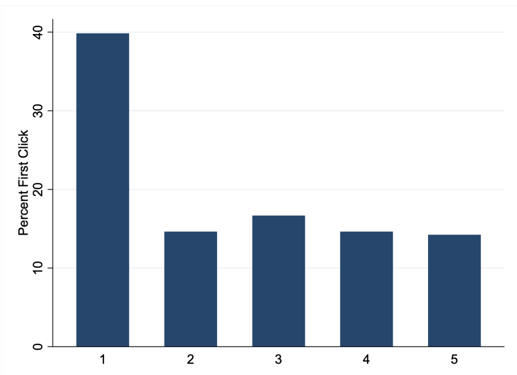


(A) State 1 Representative

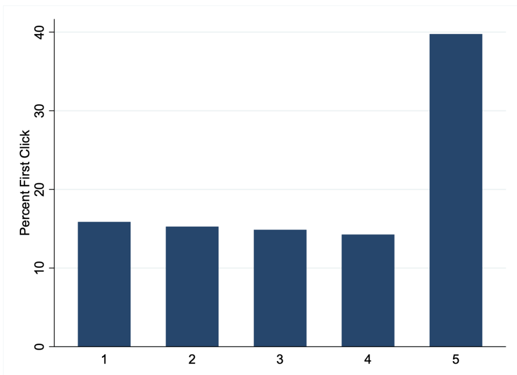


(B) State 5 Representative

FIGURE D.5. Participants' first clicks are as-if random when representativeness is suppressed as a salience cue



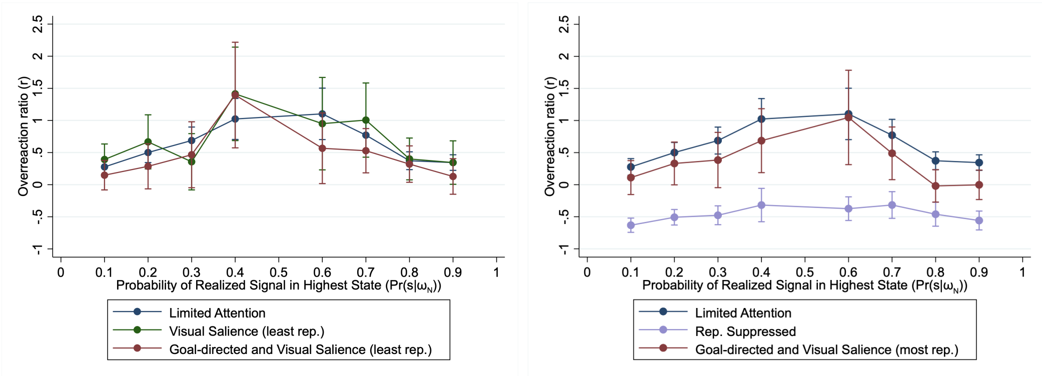
(A) State 1 Highlighted



(B) State 5 Highlighted

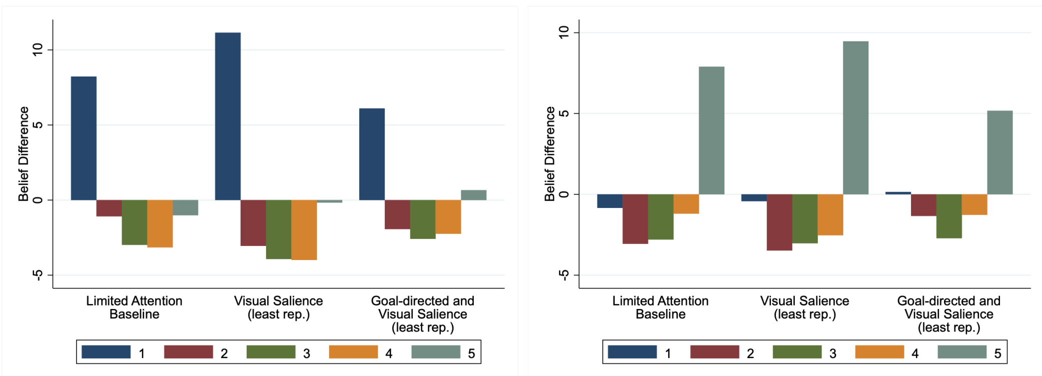
FIGURE D.6. Participants click on the state associated with the visual and goal-direct salience cue first when representativeness is suppressed as a salience cue

[74]Note that in the Visual Salience and Goal-Directed and Visual Salience conditions, the non-representative-based salience cues were associated with the least representative state: State 5 in Panel A and State 1 in Panel B.



(A) Representativeness is not suppressed (B) Representativeness is suppressed

FIGURE D.7. Visual and goal-directed salience cues are dominated by representativeness (left). However, when representativeness is suppressed, visual and goal-directed salience cues bring overreaction back (right).



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

FIGURE D.8. Representativeness-based salience dominates goal-directed and visual salience in influencing beliefs.

D.6 Structural Estimation

Aggregate-Level Estimation. We refer to the model-predicted posterior belief given parameter values θ and λ as a *model prediction* and denote it by $\hat{p}_{\theta,\lambda}$ (see Eq. (7)). This prediction maps each information environment (Ω, p_0) and signal realization s_i to a subjective posterior distribution $\hat{p}_{\theta,\lambda}(s_i; \Omega, p_0) \in \Delta(\Omega)$. 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.[75] 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 Ω. Specifically, we include trials for all information environments listed in Table D.1 except for the 11-state complexity. The results are summarized in Table D.13.

Table D.13. Aggregate-level estimates of θ and λ

	θ	95% CI	λ	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 D.1, except for the 11-state complexity; excludes wrong direction updates. The 95% confidence intervals are obtained from 300 bootstrap samples.

We present two robustness checks for our structural estimation. First, we estimate the parameters θ and λ for a prediction loss function that minimizes the average quadratic mean difference between the expected state under the reported posterior and predicted posterior. [76] 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 D.14. Structural Estimation with Quadratic Mean Loss Function

	θ	95% CI	λ	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 D.1, except for the 11-state complexity; excludes wrong direction updates. The 95% confidence intervals are obtained from 300 bootstrap samples.

[75] 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 $\sum_{\omega_j \in \Omega} \hat{p}(\omega_j|s_i; \Omega, p_0) \log(\hat{p}(\omega_j|s_i; \Omega, p_0)/\hat{p}_{\theta,\lambda}(\omega_j|s_i; \Omega, p_0))$.

[76] 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 $\left(\sum_{\omega_j \in \Omega} \omega_j \left(\hat{p}(\omega_j|s_i; \Omega, p_0) - \hat{p}_{\theta,\lambda}(\omega_j|s_i; \Omega, p_0)\right)\right)^2$.

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 react in the wrong direction under an asymmetric prior (Prediction 6). In our main analysis, we drop wrong direction reactions. This could potentially lead to an underestimation of cognitive noise. By excluding these information environments, we can drop wrong direction reactions without introducing such a bias. The following table demonstrates that this exclusion does not meaningfully affect the parameter estimates.

TABLE D.15. Structural Estimation for Symmetric Priors

	θ	95% CI	λ	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 D.1, except for the 11-state complexity; excludes wrong direction updates. The 95% confidence intervals are obtained from 300 bootstrap samples.

Individual-Level Estimation. 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. D.9. Each point in the figure represents the parameter estimates for one participant.

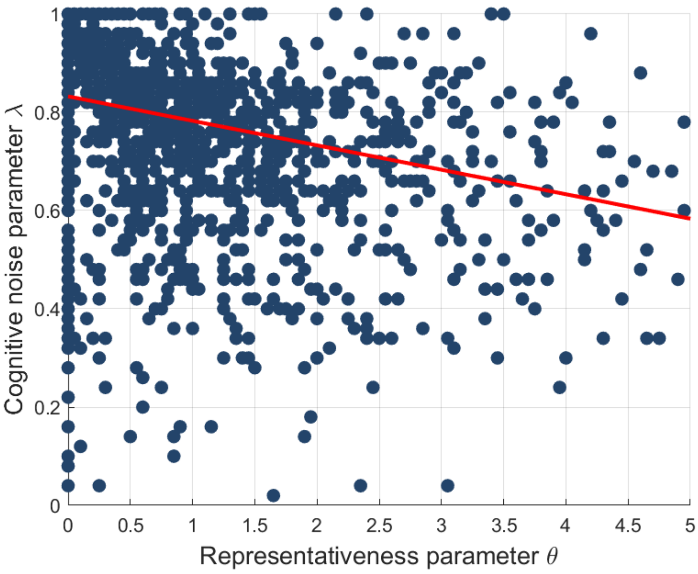


FIGURE D.9. Individual level parameter estimates.

Notes: Parameter estimates that minimize the average KL divergence at the individual level. Includes all information environments listed in Table D.1, except for the 11-state complexity; excludes wrong direction reactions; excludes extreme estimates of θ larger than 5 (approx. 5.5% of sample).

D.7 Evaluating Model Performance

D.7.1 Model Completeness

We define model completeness as follows. Similar to the structural estimation in Section 3.2, 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 $M \in \{T, P, R\}$, where $M = T$ corresponds to our two-stage model, $M = P$ corresponds to the processing-only model ($\theta = 0$), $M = R$ corresponds to the representational-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

$$\kappa^M \equiv \frac{e^B - e^M}{e^B - e^*} \in [0, 1]. \tag{26}$$

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.

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 react in the wrong direction so as to capture the full extent of model fit to the data.

D.7.2 Model Restrictiveness

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 D.1) 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.[77] 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 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

[77] 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.

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,

$$\rho^M \equiv \frac{d^M}{d^B} \in [0, 1]. \tag{27}$$

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.

D.8 Alternative Settings

D.8.1 Alternative Signal Structures

Table D.16 below presents regression results on over versus underweighting of specific states depending on whether they are representative or not. In line with our framework and Fig. 15a, representative states are overweighted in belief-updating even outside the good news/bad news structure studied in our main experiments.

TABLE D.16. Representative states are overweighted in a 3-signal environment

	(1)
	Post - Bayes
Representative	4.156***
	(0.830)
Not representative	-1.020***
	(0.210)
N	1389
adj. R^2	0.050

Notes: Outcome variable is the difference between the reported posterior and the Bayesian posterior for a given state. Excludes wrong direction updates. Standard errors clustered at the individual level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

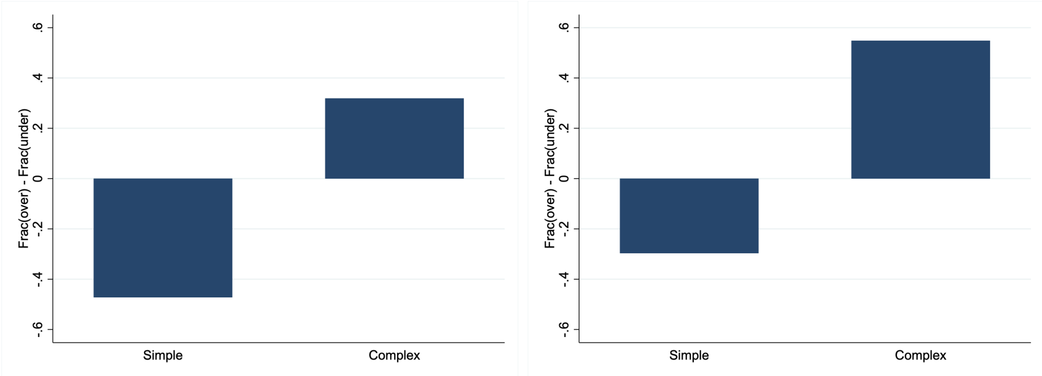
D.8.2 Inference versus Forecasting

In this section, we will describe the design of the Forecasting experiment reported in Section 6.2. The design largely follows Fan et al. (2023). Participants were first shown the stock price growth distribution for good and bad firms as in Fig. D.10. Note that the total number of potential stock price growths was 11, and participants were told that the average stock price growth of a Good firm was +100 and the average stock price growth of a Bad firm was -100. Across all conditions, participants were told that a firm would be selected at random, with Good and Bad firms equally likely to be selected. Each was told that they would see the selected firm's stock price growth for the current month and, in line with the graph, that a Good firm was more likely to generate a higher price growth signal than a Bad firm. They were then shown the selected firm's stock price growth for the current month.

As in Fan et al. (2023), each participant made forecasts by reporting their beliefs about the likelihood of future price realizations (next month) based on the current signal. They did so in one of two conditions that differed in representational complexity: Complex versus Simple. In the Complex condition, participants forecasted the likelihood that the selected firm would experience each of the possible eleven stock price growths in any given month. The Simple condition was the same as the Complex condition except for one change: the price growth space was partitioned into two bins, one with an average price growth of +100 and one with an average price growth of -100. Participants forecasted the likelihood that the future price growth fell into each of these two bins.

Despite the underlying structure of the forecasting problem being the same, we found that the change in representational complexity significantly affected belief-updating. Using the same measure of under versus overreaction as Fan et al. (2023), the Complex condition replicates their results that people tend to overreact to information. However, participants underreact to the same set of signals when making forecasts in the Simple condition. Figure D.11 and Fig. D.12 present these results.

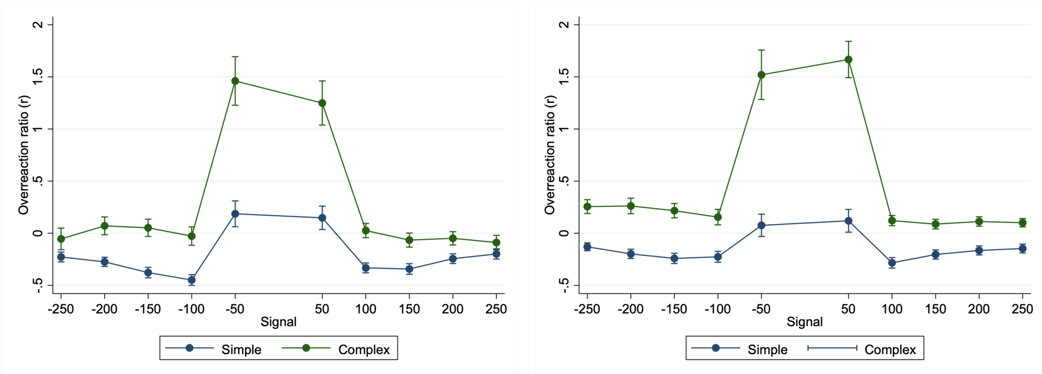
FIGURE D.10. The inference-forecast problem



(A) Forecast Condition

(B) Inference Condition

FIGURE D.11. In the Forecast (Inference) condition, participants overreact when asked to predict (infer) the full 11-bin distribution of price growth and underreact when asked how likely it is that the same stock's price growth is 100 cents versus -100 cents in the next (any given) month.



(A) Forecast Condition

(B) Inference Condition

FIGURE D.12. In the Forecast (Inference) condition, participants overreact when asked to predict (infer) the full 11-bin distribution of price growth and underreact when asked how likely it is that the same stock's price growth is 100 cents versus -100 cents in the next (any given) month.

D.8.3 Financial Assets

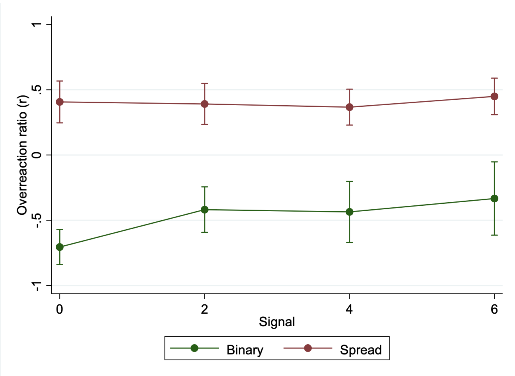
This section provides additional details of the financial assets extension described in Section 6.3. Similar to the Forecasting experiment reported in Section 6.2, participants were told that there was a pool of Good and Bad firms with respective stock price growth distributions shown in Fig. D.13. One firm would be selected at random, and each type of firm was equally likely to be selected.

The experiment was designed to mirror a setting where people form beliefs about the future performance of a financial product whose payoff space is either simple—binary option—or more complex—bull spread—while keeping the structure of the underlying asset fixed. Participants were endowed with a financial product based on the randomly selected firm drawn from the distribution in Fig. D.13. In the Bull

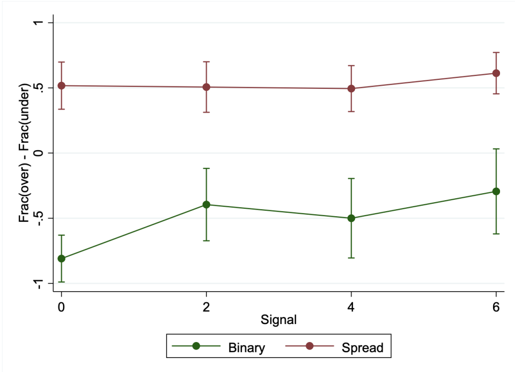
Spread condition, participants were told that they would receive a payoff corresponding to the price growth of the selected firm: \$0 if the price increase was \$0, \$2 if the price increase was \$2, etc. In the Binary Option condition, they were told that they would receive a payoff of \$0 if the price growth was less than \$3 (i.e., \$0 or \$2) and \$6 if the price growth was greater than \$3 (i.e., \$4 or \$6). Note that the average payoff, given the signal structure, and the underlying information environment was the same across the two conditions. Each participant was shown a price growth signal of the selected firm in the current month and asked to forecast the likelihood of the potential payoffs of their asset in the next month; this amounted to making forecasts over 2 objects in the Binary Option condition and 4 objects in the Bull Spread condition.

Consistent with our framework, the majority of participants underreacted to information about the binary option while the majority of participants overreacted to the same information when considering the payoffs of a Bull Spread (Fig. D.14b). The same pattern of results is obtained when considering the average overreaction ratio.

FIGURE D.13. The financial asset problem



(A) Overreaction ratio by signal



(B) % Overreact - % Underreact by signal

FIGURE D.14. Participants consistently underreact to all signal realizations about the payoff likelihoods in the case of a binary option but overreact in the case of a bull spread.

E 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.

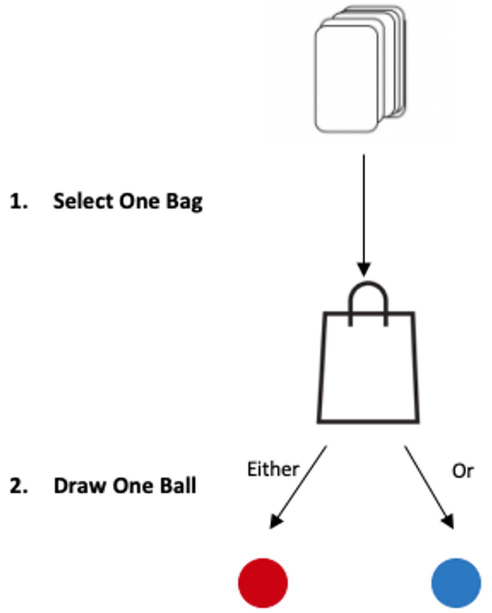
Task Setup

- 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.**