

Over- and Underreaction to Information: Belief Updating with Cognitive Constraints^{*}

Cuimin Ba[†]

J. Aislinn Bohren[‡]

Alex Imas[§]

August 15, 2025

This paper explores how cognitive constraints interact with the information environment to determine whether people overreact or underreact to information. In our model of belief updating, limited attention leads people to form a distorted mental model or *representation* of the information environment, and limited processing capacity generates cognitive imprecision when using this representation to update beliefs. The model predicts *overreaction* when facing complex environments, noisy or surprising signals, or priors concentrated on moderate states; it predicts *underreaction* when facing simple environments, precise or confirmatory signals, or priors concentrated on extreme states. A series of pre-registered experiments provide support for these predictions and direct evidence for the proposed cognitive mechanisms. Crucially, the interaction between the cognitive constraints generates the observed pattern of bias: neither constraint on its own can explain the data. These results connect prior disparate findings on whether underreaction versus overreaction arises.

Keywords: overreaction, underreaction, beliefs, attention, mental representation, behavioral economics, learning, forecasting, inference

*We thank Nageeb Ali, Nick Barberis, Dan Bartels, Francesca Bastianello, Cary Frydman, Xavier Gabaix, Thomas Graeber, Ingmar Haaland, Spencer Kwon, David Laibson, Chen Lian, 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 Pennsylvania: a.bohren@sas.upenn.edu

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

1 Introduction

How do people react to new information? This question is fundamental to economic decision-making: investors update their beliefs about the quality of a stock based on its performance, managers learn from interviews before making hiring decisions, and professional forecasters adjust their predictions based on economic indicators. Standard models assume that people have an accurate mental model of the learning environment and apply Bayes' rule to this model when learning from new information. However, a large literature in economics, finance, and psychology documents systematic departures from this benchmark. The findings are mixed: laboratory experiments generally find that people *underreact* to information relative to the objective posterior ([Benjamin 2019](#)), while results from financial markets—using both household surveys and professional forecasts—often show that people *overreact* ([Bordalo, Gennaioli, and Shleifer 2022a](#)).¹

This paper proposes and tests a model of belief updating with cognitive constraints—specifically, limited attention and processing capacity—and shows that the interaction between these constraints and the learning environment determines whether overreaction or underreaction emerges. In the model, an individual observes a signal, forms a mental model or *mental representation* of the information environment, and then uses this representation to update her belief about an unknown state (e.g., whether an asset is good or bad). When forming the mental representation, limited attention leads the individual to focus on salient values of the state—overweighing the probability of the observed signal in these states—and neglect others.² A primary determinant of salience is “representativeness:” a state is more representative of the observed signal if it is more likely to generate the signal than other states ([Gennaioli and Shleifer 2010](#)). The individual then applies Bayes' rule to her mental representation. However, limited processing capacity leads her to execute the update with noise, generating *cognitive imprecision*: the posterior belief is insensitive to the signal, underweighing (overweighing) more (less) likely states relative to a precise update.³

We use this model to analyze how reaction to information varies with features of

¹In a survey of belief updating in experiments, [Benjamin \(2019\)](#) writes: “The experimental evidence on inference taken as a whole suggests... people generally underinfer rather than overinfer.” In contrast, in a review of belief updating in financial markets, [Bordalo et al. \(2022a\)](#) 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 in financial markets, however, such as the case of forecasting short-term interest rates ([Bordalo, Gennaioli, Ma, and Shleifer 2020](#)) and inflation ([Kučinskas and Peters 2022](#)).

²See [Panichello and Buschman \(2021\)](#); [Oberauer \(2019\)](#) for evidence of the role attention plays in forming mental representations and [Bruce and Tsotsos \(2009\)](#) for a review of the literature on salience as a driver of attention.

³Prior work incorporates processing constraints as optimal processing subject to noise—broadly termed noisy cognition or cognitive imprecision ([Green, Swets et al. 1966](#); [Thurstone 1927](#); [Woodford 2020](#)). In the case of belief updating, this corresponds to a noisy version of Bayes' rule.

the learning environment—including how differences in the learning environment can explain the opposing biases documented in laboratory versus financial market settings. To illustrate, consider an example. Suppose an individual 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%) chance of increasing in price and a 30% (70%) chance of decreasing in price (the signal). The individual observes a price increase. How would she update her belief? According to Bayes’ rule, she should increase her belief that the asset is good from 50% to 70%. However, results from laboratory studies suggest that the individual will underreact and increase her belief to less than 70%. Now suppose the environment is more complex: instead of two states, there are five states of equal prior likelihood—good and bad (with the same chances of a price increase as before) plus three moderate states with a 60%, 50%, 40% chance of generating a price increase.⁴ The additional states necessitate attending to more objects when updating beliefs. How does this increase in complexity impact how she updates her belief?

Our model predicts how this complexity impacts belief updating. When the signal has a good news structure—as in the example and our main analysis—the good state is most representative following a price increase.⁵ Limited attention leads the individual to direct attention to the good state, overweighing its likelihood of generating a price increase in her mental representation. Absent processing constraints, applying Bayes’ rule to this mental representation results in a “cognitively precise” posterior that overweights the good state and underweights lower states, generating excess movement in beliefs—overreaction. But processing constraints lead to a noisy implementation of Bayes’ rule, reducing sensitivity to the signal: relative to the cognitively precise posterior, the good state is underweighted and the bad state is overweighted. Absent the distortion in the mental representation, this generates insufficient movement in beliefs—underreaction.⁶

Given the opposing impact of the two cognitive constraints in this example, whether over- or underreaction emerges depends critically on the learning environment. In the simple environment (the two-state case), focusing on the “good” state entails neglecting only one other state. Thus, limited attention generates a relatively small distortion in the representation, and the impact of cognitive imprecision dominates: the individual underreacts to the price increase. In contrast, in the complex environment (the five-state case), directing attention to the “good” state entails

⁴In our framework, we distinguish between two categories of complexity—*representational* and *computational*. The number of objects one must consider when forming a mental representation taxes attention resources, and therefore impacts representational complexity.

⁵In a good news signal structure (Milgrom 1981), signals are ordered so that “better” signals (good news) increase the likelihood of “better” states. This is the canonical signal structure used in many economic models and laboratory experiments.

⁶Under non-uniform priors, our model still predicts that the individual will be insensitive to the signal. But this insensitivity can sometimes result in overreaction.

neglecting *four* other states. Limited attention thus generates a larger distortion in the representation, which dominates the impact of cognitive imprecision and the individual overreacts to the price increase.

Beyond state-space complexity, our model yields a rich set of predictions on how other features of the learning environment impact belief updating. Holding the number of states constant, the level of overreaction will decrease as the signal becomes more precise or the prior concentrates more mass on more extreme states. For instance, in the five-state example, the individual will overreact less when the good and bad states have a 80% and 20% chance of a price increase, respectively, compared to the case where these chances are 70% and 30%.⁷ Likewise, the individual will overreact less when the prior places most of the mass on the good and bad state compared to when it places equal mass on all five states. Under an asymmetric prior, “surprising” disconfirmatory signals—those that increase the likelihood of states that are unlikely given the prior—generate overreaction, and “expected” confirmatory signals generate underreaction or even wrong direction reaction, i.e., updating in the opposite direction from the objective posterior.⁸

To test these predictions, we designed an experimental paradigm in which we can manipulate key features of the information environment, specifically the complexity, signal precision, and prior. The design builds on the classic “bookbag-and-poker-chip” paradigm originally used in [Edwards \(1968\)](#) and employed extensively in the learning literature. In our design, 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, and 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 direct mapping 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 structure. We employ four sources of treatment variation: complexity via the number of states, signal structure, concentration of the prior on extreme states, and symmetry of the prior.

We find that increasing complexity has a striking effect on belief updating. In the simple two-state uniform-prior environment typically studied in the literature, we replicate the standard finding that people underreact to information. But this result reverses when we add even a single additional state: the majority of participants

⁷In a 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, Lazarus, and Thaler \(2022\)](#) show that this relationship is consistent with a model of cognitive noise. Our model shows that the same pattern can be generated by salience-channelled attention.

⁸In 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 disconfirmatory. Wrong direction reaction occurs when the individual’s belief that the asset is good decreases after observing a price increase or vice versa.

overreact in three-state uniform-prior environments across all signal structures we consider. As we increase the number of states beyond three, the share of participants overreacting and the level of overreaction both increase monotonically. Our predictions on signal precision, prior concentration and prior symmetry are also borne out in the data: the level of overreaction decreases as the signal becomes more precise and as the prior becomes more concentrated on extreme 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 reactions to confirmatory signals compared to disconfirmatory signals.

We also use the experimental data to structurally estimate the two key cognitive parameters in our model that capture the attentional and processing constraints. In aggregate, both estimates substantially differ from the objective benchmark and are in line with values found in prior work. At the individual level, the vast majority of participants exhibit significant attention and processing distortions. Individual estimates of each parameter are significantly positively correlated, suggesting underlying differences in cognitive capacity that impact both stages of belief updating.

We next test the proposed attentional mechanism in the representational stage of our model. We incorporated a common instrument to measure attention ([Payne, Bettman, and Johnson 1988](#)) into our experimental paradigm and manipulated where attention was directed. The first variation was designed to measure and exogenously constrain attention in our baseline paradigm. We found that participants' attention was indeed overwhelmingly drawn to the state most representative of the observed signal, and beliefs overweighed this state the most. We also found that exogenously increasing attentional constraints while holding the information environment fixed led to more overreaction, consistent with our model. Structural estimates of the attentional and processing parameters show that more constrained attention increased the distortion in the representational stage but not the processing stage.

In the second variation, we suppressed representativeness as a salience cue to identify the causal effect of attention on beliefs. Absent representativeness, we found that individuals directed their attention to states as-if randomly and the state attended to first was overweighed in the belief data. This establishes the causal effect of attention on beliefs: directing attention to a state causes beliefs to place excess mass on the state. Moreover, consistent with the prediction from a random-attention variation of our model, *underreaction* emerged on average—despite holding complexity (and the information environment) constant.

In the third variation, we compared representativeness to other salience cues considered in the literature: specifically, visual and goal-directed salience. We empirically showed that representativeness had a stronger influence on attention than other

salience cues in our setting, highlighting its importance in learning environments.

In the final variation, we tested the predictions of our model for non-good-news signal structures. Representativeness continued to draw attention and drive the overweighing of states, as predicted by our model. However, this overweighing led to underreaction when the moderate state was representative. Taken together, the underreaction observed in the representativeness-suppressed and non-good-news signal variations shows that underreaction is not a unique feature of simple two-state environments, but rather emerges from the interaction between cognitive constraints and observable features of the information environment.

Finally, we evaluate model fit by studying predictions on the full subjective belief distribution and measuring model performance. In addition to predictions about underreaction and overreaction to information, our model yields predictions on *which* states will be overweighed versus underweighed in response to new information. We show that the interaction between the cognitive constraints is crucial for explaining observed patterns of belief updating across the full belief distribution: neither constraint on its own can explain the data. Next, to measure model performance, we compute our model’s *completeness* in capturing predictable variation in belief updating relative to Bayes’ rule ([Fudenberg, Kleinberg, Liang, and Mullainathan 2022](#)). 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 on a scale of 0 to 1, in binary versus more complex environments, respectively). In contrast, cognitive imprecision alone can only explain results in simple environments (completeness 1.00); its explanatory power precipitously drops in more complex environments (0.36). Salience-channeled attention also has low explanatory power on its own. This further demonstrates that the interaction between the two cognitive constraints plays a critical role in driving belief updating. Notably, our model’s completeness does not come at the expense of being too flexible: it remains highly *restrictive* in fitting arbitrary belief data ([Fudenberg, Gao, and Liang 2023](#)).

A large literature explores under- and overreaction in belief updating. 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, noisier signals, and a good news signal structure. For instance, simple settings such as the binary-state experiments reviewed in [Benjamin \(2019\)](#) generally document underreaction, while more complex environments such as the studies in financial markets reviewed in [Bordalo et al. \(2022a\)](#) find overreaction. Underreaction has also been observed in complex settings where the representative state is unclear or moderate, e.g., US treasury rates and inflation expectations ([Bordalo et al. 2020; Kučinskas and Peters 2022; DellaVigna and Pollet 2009](#)), in line with our model’s predictions. We provide a

more in-depth review of this work and its relation to our model in [Appendix A](#), along with a discussion of how our findings relate to price responses to financial market news ([Daniel, Hirshleifer, and Subrahmanyam 1998](#); [Barberis, Shleifer, and Vishny 1998](#); [Klibanoff, Lamont, and Wizman 1998](#)).

Our paper also contributes to the literature exploring the cognitive foundations of economic decision-making. Our model is similar in spirit to [Schwartzstein \(2014\)](#), where an individual selectively channels her attention to a subset of the available information and then uses Bayes' rule to update her beliefs on this subset. 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 2023](#); [Oprea 2022](#)). Our work also builds on the strand of research that models an individual as optimally responding to a stimulus given a noisy representation of the environment ([Gabaix and Laibson 2017](#); [Azeredo da Silveira and Woodford 2019](#); [Khaw, Li, and Woodford 2021](#); [Augenblick et al. 2022](#); [Frydman and Jin 2022](#); [Enke and Graeber 2023](#)).

The rest of the paper proceeds as follows. [Section 2](#) outlines the theoretical framework and [Section 3](#) outlines the experimental paradigm. [Section 4](#) presents the analytical results and empirical findings on reaction to information. [Section 5](#) provides evidence for the proposed attentional mechanism. [Section 6](#) presents predictions for the full belief distribution, evaluates model performance, and analyzes heterogeneity across participants. [Section 7](#) concludes. All proofs are in [Appendix B](#).

2 A Model of Belief Updating

In this section, we describe the information environment and introduce a model of belief updating.

2.1 Information Environment

Nature selects one of $N > 1$ distinct states $\omega \in \Omega := \{\omega_1, \dots, \omega_N\} \subset (0, 1)$ according to a full support prior $p_0 \in \Delta(\Omega)$. Without loss of generality, order the state space Ω in ascending order, $\omega_1 < \dots < \omega_N$. Let ω_i denote a generic element of Ω . We refer to states ω_1 and ω_N as extreme states and the remaining states as moderate states.

An agent learns about the state from a binary signal with a good news structure ([Milgrom 1981](#)).⁹ Specifically, the signal is drawn from signal space $\mathcal{S} := \{b, r\}$, with

⁹The majority of our analysis focuses on a good news structure, as this case is used in the majority of prior experimental work and mirrors many real-world settings where the expectation of the relevant economic variable monotonically increases in the signal (e.g., equity markets where a price increase increases the likelihood that an asset is “good,” economic indicators such as GDP, etc.). In [Section 5.4](#) we theoretically and empirically explore an information environment that does not have a good news structure.

a generic realization denoted by s . In state ω_i , the signal is distributed according to $\pi(r|\omega_i) = \omega_i$ and $\pi(b|\omega_i) = 1 - \omega_i$. For example, if $\Omega = \{0.3, 0.5, 0.7\}$, then r 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 r is increasing in the state, it is indicative of higher states (good news). Similarly, b is indicative of lower states (bad news). The diagnosticity of the signal in state ω_i is the probability of the more likely signal realization, $d_i := \max\{\omega_i, 1 - \omega_i\}$, and the diagnosticity of the signal is the maximum diagnosticity across all states, $d := \max_{\omega_i \in \Omega} d_i$. Under the assumed signal distribution $\pi(r|\omega_i) = \omega_i$, the state space Ω fully determines the signal distribution π . Therefore, we refer to Ω as both the state space and the signal structure, and to (Ω, p_0) as the *information environment*.

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

$$p_B(\omega_i|r) := \frac{\omega_i p_0(\omega_i)}{\sum_{\omega_k \in \Omega} \omega_k p_0(\omega_k)}. \quad (1)$$

Analogously following b , $p_B(\omega_i|b) := (1 - \omega_i)p_0(\omega_i)/\sum_{\omega_k \in \Omega}(1 - \omega_k)p_0(\omega_k)$. Let $p_B(s) := (p_B(\omega_1|s), \dots, p_B(\omega_N|s))$ denote this objective posterior.

We define two symmetry properties for our analysis. A signal structure Ω is symmetric if $\omega_i \in \Omega$ implies $1 - \omega_i \in \Omega$. In this case, the set of diagnosticities uniquely determines the signal structure, and the maximal diagnosticity is $d = \omega_N$. A prior p_0 is symmetric if for any $\omega_i \in \Omega$, states ω_i and $1 - \omega_i$ have the same mass, $p_0(\omega_i) = p_0(1 - \omega_i)$. Note that prior symmetry implies signal structure symmetry but not vice versa. Therefore, when p_0 is symmetric, we refer to (Ω, p_0) as a symmetric information environment.

2.2 The Two-Stage Model

We model the belief updating of an agent with limited cognitive resources as a two-stage process: (i) forming a mental representation of the information environment and (ii) processing the signal.¹⁰ In the first stage, attentional and working memory constraints lead to a distorted mental representation of the information environment. In the second stage, processing and computational capacity constraints introduce cognitive imprecision in processing the signal with respect to this mental representation. [Fig. 1](#) illustrates each stage of the model.

Stage 1: Mental representation of information environment. The first *representational* stage of belief updating draws on attention and working memory resources to simultaneously consider multiple objects (in our setting, we focus on considering multiple states). Constraints on these resources lead the agent to channel

¹⁰This two-stage process is motivated by work from the cognitive psychology literature showing that people first form a mental representation of an environment and then use that representation to process stimuli (Johnson-Laird 2010; Walsh, McGovern, Clark, and O'Connell 2020).

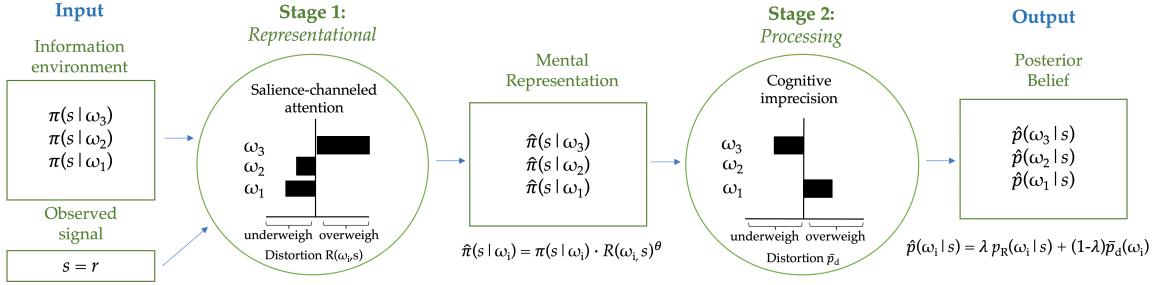


FIGURE 1. Illustration of model for 3-state environment.

attention to some states and neglect others.¹¹ Motivated by prior work documenting salience as a key driver of attention (Bruce and Tsotsos 2009; Awh, Belopolsky, and Theeuwes 2012), we propose that the agent channels attention to a state ω_i proportional to its salience $R(\omega_i, s) \geq 0$ given signal realization s , where R denotes a salience function (see Bordalo, Gennaioli, and Shleifer (2022b) for a definition).¹² Salience-channeled attention leads the agent to perceive the likelihood of signal realization s in state ω_i to be

$$\hat{\pi}(s|\omega_i) := \pi(s|\omega_i)R(\omega_i, s)^\theta \quad (2)$$

where $\theta \geq 0$ captures the severity of the attentional constraint (higher θ corresponds to more limited attention). We refer to $\hat{\pi}(s|\omega)$ as the mental representation of signal distribution $\pi(s|\omega)$.¹³ When $\theta = 0$, the mental representation is accurate, and when $\theta > 0$, the mental representation overweights the probability of the signal realization in more salient states and underweights it in less salient states.

¹¹ Attention and working memory—which are thought to share the same neural mechanism and therefore draw on the same limited resource—play a critical role in forming mental representations (Panichello and Buschman 2021; Oberauer 2019). An agent can attend to and keep in mind a limited number of objects at a time—typically 3 or 4 (Oberauer, Farrell, Jarrold, and Lewandowsky 2016; Luck and Vogel 1997; Loewenstein and Wojtowicz 2023; Bays, Gorgoraptis, Wee, Marshall, and Husain 2011).

¹² Salience is determined by a combination of top-down (incentive-driven), bottom-up (stimulus-driven), and selection history (memory-driven) factors, which jointly determine the allocation of attention across objects(Awh et al. 2012; Talsma, Senkowski, Soto-Faraco, and Woldorff 2010; Yantis 2008; Tanner and Itti 2019). Top-down attention is allocated through a conscious process, typically in response to incentives (e.g., goal-directed salience, rational inattention (Maćkowiak, Matějka, and Wiederholt 2023)). Bottom-up or stimulus-driven attention (Bruce and Tsotsos 2009; Awh et al. 2012) is a subconscious response to a stimulus based on the stimulus's inherent properties relative to the environment (e.g., visual salience as in Li and Camerer (2022), payoff-based salience as in Bordalo, Gennaioli, and Shleifer (2013))). History-dependent attention is driven by the objects and stimuli that the agent had attended to in the past.

¹³ Representation $\hat{\pi}$ is a pseudo-distribution: substituting it for true distribution π in Bayes' rule results in a well-defined posterior belief over the state space, but $\hat{\pi}$ is not necessarily a probability distribution, as $\hat{\pi}(s|\omega_i)$ may not sum to one across signals. When there are two states, a suitable normalization of the salience function R results in a well-defined probability distribution $\hat{\pi}$ (also commonly referred to as a misspecified model). 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).

An important determinant of salience in information settings is the extent to which the state is “representative” of the observed signal (Gennaioli and Shleifer 2010).¹⁴ For example, when predicting the hair color (the state) of an Irish person (the signal), people overweigh the likelihood that the person has red hair, as an Irish person is more likely to have red hair than the general population (Bordalo, Coffman, Gennaioli, and Shleifer 2016). Formally, the representativeness of state ω_i for signal realization s corresponds to the conditional probability of s in ω_i relative to the total probability of s ,¹⁵

$$R(\omega_i, s) := \frac{\pi(s|\omega_i)}{\Pr(s)}. \quad (3)$$

A state is *more representative* if it is more likely to generate s relative to other states. Under a good news signal structure, ω_1 is the most representative state for signal realization b and ω_N is the most representative for r . Substituting Eq. (3) into Eq. (2) yields mental representation

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

This representation overweighs (underweights) the probability of the observed signal in states that are more (less) representative.¹⁶ Fig. 1 illustrates this first stage, taking a three-state information environment and observed signal $s = r$ as inputs. Attention is channeled to most representative state ω_3 while the least representative state ω_1 is neglected. This generates a mental representation that overweighs the likelihood of the observed signal in ω_3 and underweights it in ω_1 , $\hat{\pi}(s|\omega_3) > \pi(s|\omega_3)$ and $\hat{\pi}(s|\omega_1) < \pi(s|\omega_1)$.

As an example, 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 attend to all of the states when forming beliefs. Because unicorns are “representative” of a price increase, the investor overweighs the probability a unicorn generates a price increase and underweights the probability that other states do. Note that the investor does not entirely neglect the possibility that other states generate a price increase; these states are just down-weighed relative to

¹⁴Representativeness is a salience cue that operates through bottom-up attention. It was initially identified by Kahneman and Tversky (1972); Tversky and Kahneman (1983).

¹⁵This 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); Bordalo, Gennaioli, Porta, and Shleifer (2019).

¹⁶Applying Bayes rule to this representation results in an updating rule that “counts” a signal $\theta + 1$ times (as shown in Appendix B.1). This updating rule is equivalent to forming a posterior belief based on the representativeness-based discounting weighing function used in Bordalo et al. (2016, 2019).

the true likelihood.

While our main analysis focuses on representativeness as the salience cue, our model can incorporate other salience cues through alternative salience functions $R(\omega_i, s)$. In Section 5.3, we theoretically and empirically study visual (bottom-up) and goal-directed (top-down) salience cues—and show that representativeness is an important driver of attention in our setting.

Stage 2: Processing signal. During the *processing* stage, the agent observes the signal and updates her beliefs given her mental representation. This process requires computation, which draws on controlled processing resources. A large literature in cognitive psychology models constraints on controlled processing as optimal processing subject to cognitive imprecision (Green et al. 1966; Thurstone 1927). Following Woodford (2020); Khaw et al. (2021), we model cognitive imprecision in belief updating as Bayesian updating with noise. We extend their framework by applying this noisy updating process to the agent’s mental representation $\hat{\pi}(s|\omega_i)$ rather than the true signal structure $\pi(s|\omega_i)$.

Let $p_R(s) := (p_R(\omega_1|s), \dots, p_R(\omega_N|s)) \in \Delta(\Omega)$ denote the posterior belief that arises from applying Bayes’ rule (without noise) to mental representation $\hat{\pi}(s|\omega)$ following signal realization s , where

$$p_R(\omega_i|s) := \frac{\hat{\pi}(s|\omega_i)p_0(\omega_i)}{\sum_{\omega_k \in \Omega} \hat{\pi}(s|\omega_k)p_0(\omega_k)}. \quad (5)$$

We refer to this as the *cognitively precise posterior*. We model noisy Bayesian updating as follows: (i) the agent observes a noisy cognitive signal of $p_R(s)$, and (ii) the agent applies Bayes’ rule to the cognitive signal to form a cognitive posterior about $p_R(s)$. Let $Y(s) \in \Delta(\Omega)$ denote the cognitive signal and assume it is drawn from a multinomial distribution $(1/\eta)\text{Multi}(\eta, N, p_R(s))$ with $\eta \geq 0$ trials (the cognitive precision), N categories (the number of states), and event probabilities $p_R(s)$.¹⁷ The cognitive signal is unbiased with mean equal to the cognitively precise posterior $p_R(s)$. Higher η corresponds to a more precise cognitive signal, as-if the agent observed η draws from distribution $p_R(s)$.

To use Bayes’ rule to form a cognitive posterior after observing $Y(s)$, the agent needs a prior over $p_R(s)$. We take this cognitive prior to be the Dirichlet distribution with N categories and vector of concentration parameters $\nu \bar{p}_d$, where $\bar{p}_d \in \Delta(\Omega)$ is the mean and $1/\nu \geq 0$ scales the variance.¹⁸ As in Enke and Graeber (2023), \bar{p}_d has the interpretation of the *cognitive default*—the agent’s average prior about the

¹⁷The multinomial distribution is a natural choice for the distribution of a signal of a probability distribution. Any realization $y = (y_1, \dots, y_N)$ 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).

¹⁸The 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).

posterior before internalizing a particular learning environment—and we take it to be uniform, $\bar{p}_d(\omega_i) = 1/N$ for all $\omega_i \in \Omega$. This implies that on average, the prior over the posterior places equal weight on all states. We provide empirical support for this assumption in [Appendix C.3](#). The parameter ν determines how concentrated the cognitive prior is around the default.

The noise in the cognitive signal induces randomness in the cognitive posterior. For our analysis, we focus on the expectation of the observed cognitive posterior—which we refer to as the *subjective posterior*—as it is a deterministic function of the information environment and the signal realization. The subjective posterior is derived by applying Bayes’ rule to signal $Y(s)$ and the cognitive prior and taking the expectation of the resulting cognitive posterior conditional on $p_R(s)$.

Lemma 1. *Given cognitive signal $Y(s)$ and cognitive prior \bar{p}_d , the subjective posterior $\hat{p}(s) = (\hat{p}(\omega_1|s), \dots, \hat{p}(\omega_N|s))$, is equal to*

$$\hat{p}(s) := \lambda p_R(s) + (1 - \lambda)\bar{p}_d. \quad (6)$$

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), λ is lower and the subjective belief places more weight on the cognitive default \bar{p}_d . As cognition becomes more precise (higher η) or the cognitive prior becomes noisier (lower ν), λ is higher and the agent places more weight on the cognitively precise posterior $p_R(s)$. [Fig. 1](#) illustrates this second stage, taking the mental representation as the input and executing Bayes’ rule with cognitive imprecision that compresses the cognitively precise posterior $p_R(s)$ towards the cognitive default \bar{p}_d . This compression attenuates the response to the signal: since $\bar{p}_d(\omega_3) < p_R(\omega_3|s)$ and $\bar{p}_d(\omega_1) > p_R(\omega_1|s)$, it reduces mass on the most likely state ω_3 and increases mass on the least likely state ω_1 .

As an illustration, return to the example of the investor who forms beliefs about a new tech company. In different markets, there are different prior beliefs about the possibility that a company is a unicorn, zombie, or intermediate type, as well as beliefs about the likelihood of price increases for each of these types. The boundedly rational investor faces cognitive imprecision when adjusting to the parameters of each particular market. This dampens her response to the signal relative to a cognitively precise investor. The investor does not completely ignore differences across markets, she just does not fully adjust to them.

The subjective posterior $\hat{p}(s)$ described by [Eq. \(6\)](#) forms the basis of our analysis. As shown in [Fig. 1](#), it incorporates how salience-channeled attention and cognitive imprecision interact to impact belief updating. Note that when $\lambda = 1$ and $\theta = 0$, $\hat{p}(s)$ is equal to the objective posterior $p_B(s)$. We also compare our two-stage model to a *limited-attention-only model* in which cognitive constraints are only present in the

first stage (i.e., $\theta > 0$ and $\lambda = 1$), and a *cognitive-imprecision-only model* in which cognitive constraints are only present in the second stage (i.e., $\theta = 0$ and $\lambda < 1$). The limited-attention-only model generates posterior $p_R(s)$ and the cognitive-imprecision-only model generates posterior $\lambda p_B(s) + (1 - \lambda)\bar{p}_d$.

Discussion of Model. We now briefly discuss several aspects of the model.

Information Environment. We restrict attention to information environments with binary signals and a good news signal distribution, as the majority of our theoretical and empirical analysis focuses on this case. This class of information environments is structured enough to yield sharp comparative static predictions about belief updating yet also general enough to mirror many real-world learning settings. Our belief-updating model immediately extends to environments with more than two signal realizations or alternative signal distributions. By imposing sufficient structure, one could also derive comparative static predictions in these alternative environments. We analyze such a setting in [Section 5.4](#) and [Appendix E](#).

Stage 1. Our framework for forming a mental representation is part of a broader literature on how people form mental models when executing tasks (see, e.g., [Gabaix \(2019\)](#) for a review). For example, in [Bordalo et al. \(2023\)](#), an agent focuses on features that are salient. This mimics the representational stage of our model, but there attention is channeled to features of the decision environment (e.g., signal diagnosticity, prior) rather than states. They show that different biases arise depending on which features are salient, despite holding the signal structure fixed. Similarly, [Banovetz and Oprea \(2023\)](#) show that agents “economize” on the number of states that need to be tracked to execute a decision rule. There is also a large body of evidence on how heuristics are used in complex environments, as reviewed in [Payne, Bettman, and Johnson \(1993\)](#).

Our framework posits that limited attention leads people to focus on some states while neglecting others. We operationalize neglect as a continuous attention function: people overweigh the former and underweigh the latter. This partial neglect formulation follows the literature on attention that models focus as a “zoom lens” on the object of interest (in our case, salient states) with a gradual drop-off of attention to objects outside the center of focus ([Loewenstein and Wojtowicz 2023](#); [Goodhew, Lawrence, and Edwards 2017](#)). It contrasts with a discontinuous “all-or-nothing” model of attention where all attention is directed to the object(s) of focus and other objects are fully neglected ([Posner 1980](#); [Bordalo, Gennaioli, Lanzani, and Shleifer 2025](#)).¹⁹ To understand when partial versus full neglect emerges, it is important to distinguish between whether attention is being allocated across objects within a feature (e.g., states in a state space) versus across features (e.g., line length versus

¹⁹For example, [Hanna, Mullainathan, and Schwartzstein \(2014\)](#) show that seaweed farmers attend to a limited number of features (e.g., line length) and completely neglect others (e.g., pod size) when making decisions.

pod size). Research in cognitive psychology suggests that people will display partial neglect across objects within a feature (Egly, Driver, and Rafal 1994; O’Craven, Downing, and Kanwisher 1999; Saenz, Buracas, and Boynton 2002) and full neglect across features (Found and Müller 1996; Huang and Pashler 2007). This distinction aligns models that focus on allocating attention across features (Hanna et al. 2014; Bordalo et al. 2023, 2025) with ours, which focuses on allocating attention within a feature.

In our model, representativeness channels attention in the context of online stimuli, i.e., stimuli that are available at the time of judgment. Relatedly, Gennaioli and Shleifer (2010); Bordalo, Coffman, Gennaioli, Schwerter, and Shleifer (2021) argue that representative states are also overweighed in judgment because they are easier to recall. See Kahneman (2003) for a discussion on the interaction between selective attention and bounded memory.

Finally, we focus on representativeness as the main driver of salience based on empirical evidence that it is a dominant salience cue in the environment we consider. In Section 5.3, we provide evidence for this assertion and also explore alternative salience cues, including visual and goal-directed salience.

Stage 2. Prior work models cognitive imprecision as noisy processing with respect to the objective information environment (Woodford 2020; Frydman and Jin 2022; Enke and Graeber 2023). A contribution of this paper is to consider noisy processing with respect to the agent’s mental representation of the information environment which, as we argue above, may be systematically distorted relative to the objective information environment.

A key assumption in our analysis is that the cognitive default is the uniform “ignorance prior.” This implies that the agent exhibits insensitivity to both the prior (base-rate neglect) and the signal precision (signal-precision neglect), both of which play a role in our theoretical predictions. Hence, empirical support for these predictions provides evidence of both forms of neglect. In Appendix D, we compare our model with other models of cognitive imprecision in belief updating, including Augenblick et al. (2022) where the agent perceives the strength of the signal with cognitive imprecision.

Noisy cognition is also 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 \bar{p}_d 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 predictions for belief updating.

3 Experiment

We ran a controlled experiment to estimate and test the predictions of our two-stage model. We next describe our experimental paradigm so that the theoretical predictions and empirical findings can be presented side-by-side.

Design. Our design mirrors the information environment described in Section 2.1. We recruited 3,856 participants from the Prolific crowdsourcing platform (48% female, 39 years average age).²⁰ 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 elicitation question was within 3% of the objective posterior.²¹

Participants then 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.

Following these instructions, participants completed several comprehension questions. They then proceeded to a series of inference tasks. Each task 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 signal structure). After observing the color of the randomly drawn ball (the signal realization, with $s = b$ corresponding to blue and $s = r$ 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. We required these percentages to add up to 100 across all possible bags. After reporting this probability assessment, the participant proceeded to the next task. Each participant completed 8 to

²⁰Preregistration materials are available here: https://aspredicted.org/LTJ_CS7 and https://aspredicted.org/Q77_3LG.

²¹We 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. For example, [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. See also [Enke, Graeber, and Oprea \(2023\)](#) for similar use of the objective posterior as the incentivized benchmark.

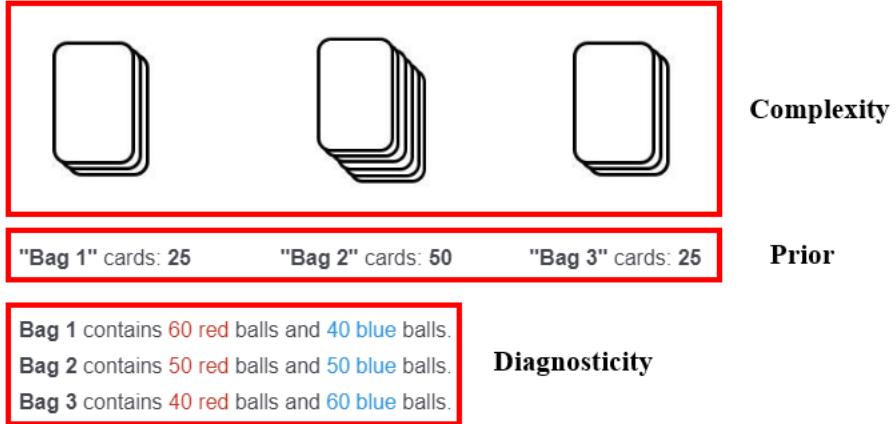


FIGURE 2. Experimental design for 3-state treatment

15 inference tasks as described below, and then answered a set of basic demographic questions before exiting the study. See [Appendix F](#) for the full instructions.

This “bookbag-and-poker-chip” design ([Edwards 1968](#)) is used extensively in the literature on belief updating—typically with a binary state space. It mirrors the information environment in our framework. 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 $\omega_1 = 0.4$, and a prior that puts more mass on the moderate state than the extreme states, $p_0(\omega_2) = 0.5$ and $p_0(\omega_1) = p_0(\omega_3) = 0.25$.

It is straightforward to manipulate the parameters of the information environment in this paradigm, which allows us to test how cognitive constraints interact with the information environment to affect belief updating. We manipulate the number of states via the number of bags, the signal structure via the number of red versus blue balls in a given bag, and the prior via the number of cards for each bag. As in the model, we focus on a good news signal structure. In the example, Bag 1 (state ω_N) is most “representative” following a red ball and Bag N (state ω_1) is most representative following a blue ball.

[Table C.1](#) in [Appendix C.1](#) outlines the set of information environments that we used. Participants were randomized into a complexity condition corresponding to the number of states—2, 3, 4, 5 or 11. Those in the 2 and 3-state conditions were also randomized into a prior condition. In the 2-state case, participants faced either a symmetric or asymmetric prior which corresponded to whether the prior assigned both states equal mass or not. In the 3-state case, participants faced either a uniform, concentrated, or dispersed prior which corresponded to how much mass was concentrated on state ω_2 versus states ω_1 and ω_3 . Participants in the 4, 5, and 11-state conditions all faced a uniform prior. Participants completed a maximum of 15 tasks randomly drawn from the set of possible information environments and signal

realizations for the respective complexity and prior condition.²² Each complexity and prior condition had at least 200 participants.

Structural Estimation. Our main analysis uses the experimental data to measure how bias—specifically, under- or overreaction to the signal—varies with the information environment. Our model makes predictions on how the direction of this bias depends on the interaction between cognitive constraints. Because these predictions depend on the model parameters, it is useful to first use the experimental data to estimate them. We follow the literature on behavioral structural estimation (e.g., DellaVigna (2018) and Bordalo et al. (2020)) to estimate parameters θ and λ . The details of this estimation are outlined in Appendix C.7.

Averaged across all participants, we estimate parameter values of $\theta = 0.85$ and $\lambda = 0.70$. Both estimates are significantly different from the objective benchmark of $\theta = 0$ and $\lambda = 1$, suggesting significant attentional and processing constraints.²³ We reference these estimates throughout the paper to help interpret the other empirical results through the lens of our model. In Section 6.3, we explore heterogeneity in attention and processing constraints across participants and the relationship between these two constraints within participant.

4 Over- versus Underreaction

We next examine how salience-channeled attention and cognitive imprecision interact with each other and the information environment to generate distortions in belief updating. We focus on whether this distortion takes the form of overreaction or underreaction to information. We first define a measure of overreaction, then derive and test comparative static predictions on how this measure varies with properties of the information environment. Per our pre-registration, in this and subsequent sections we exclude tasks in which participants react in the wrong direction (i.e., update in the opposite direction from the objective posterior) unless otherwise noted.

4.1 Measuring Overreaction.

Our definition of overreaction compares the subjective and objective movements in beliefs in terms of the expected state. Let $\hat{E}(\omega|s) := \sum_{\omega_i \in \Omega} \omega_i \hat{p}(\omega_i|s)$ denote the subjective posterior expected state following signal realization s . Analogously define the objective posterior expected state as $E_B(\omega|s) := \sum_{\omega_i \in \Omega} \omega_i p_B(\omega_i|s)$ and the prior expected state as $E_0(\omega) := \sum_{\omega_i \in \Omega} \omega_i p_0(\omega_i)$. The subjective movement in beliefs is the

²²For all conditions except the 2-state asymmetric prior, the set of possible tasks consists of the set of signal structures crossed with the set of signal realizations (always 2). For the 2-state asymmetric prior condition, the set of possible tasks consists of the set of asymmetric priors (2) crossed with the set of signal structures and the set of signal realizations.

²³Our 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.

difference between the subjective posterior and prior expected states, $\hat{E}(\omega|s) - E_0(\omega)$, and similarly the objective movement is $E_B(\omega|s) - E_0(\omega)$. This leads to the following definition.

Definition 1 (Overreaction and Underreaction). *Given signal realization s :*

- (i) *the agent overreacts if her subjective movement is greater than the objective movement and in the same direction, $|\hat{E}(\omega|s) - E_0(\omega)| > |E_B(\omega|s) - E_0(\omega)|$ and $(\hat{E}(\omega|s) - E_0(\omega))(E_B(\omega|s) - E_0(\omega)) > 0$,*
- (ii) *the agent underreacts if her subjective movement is less than the objective movement and in the same direction, $|\hat{E}(\omega|s) - E_0(\omega)| < |E_B(\omega|s) - E_0(\omega)|$ and $(\hat{E}(\omega|s) - E_0(\omega))(E_B(\omega|s) - E_0(\omega)) > 0$, and*
- (iii) *the agent wrong direction reacts if the subjective and objective movements are in the opposite direction, $(\hat{E}(\omega|s) - E_0(\omega))(E_B(\omega|s) - E_0(\omega)) < 0$.*

For example, if the objective expected state moves from a prior of $E_0(\omega) = 0.5$ to a posterior of $E_B(\omega|s) = 0.7$, the agent overreacts if her posterior expected state is $\hat{E}(\omega|s) > 0.7$, underreacts if $\hat{E}(\omega|s) \in [0.5, 0.7]$, and wrong direction reacts if $\hat{E}(\omega|s) < 0.5$. Note that if the agent does not pay attention to an informative signal and simply reports the prior, this corresponds to underreaction.

We next define a measure of overreaction in order to compare the magnitude of bias across information environments. A simple comparison of the difference between the subjective and objective movements fails to account for variation in the objective movement. For example, if the objective movement is $0.9 - 0.5 = 0.4$ versus $0.7 - 0.5 = 0.2$, the agent has more room to underreact and less room to overreact in the former environment. To account for such variation, our measure divides this difference by the objective movement.

Definition 2 (Level of overreaction). *The level of overreaction to signal realization s is:*

$$\beta(s) := \frac{(\hat{E}(\omega|s) - E_0(\omega)) - (E_B(\omega|s) - E_0(\omega))}{E_B(\omega|s) - E_0(\omega)} = \frac{\hat{E}(\omega|s) - E_B(\omega|s)}{E_B(\omega|s) - E_0(\omega)}. \quad (7)$$

From Definition 1, the agent overreacts if $\beta(s) > 0$, underreacts if $\beta(s) \in [-1, 0)$, and wrong direction reacts if $\beta(s) < -1$.²⁴

To glean intuition for how salience-channeled attention and cognitive imprecision each impact the level of overreaction, consider a symmetric information environment (i.e., symmetric Ω and p_0). The level of overreaction decomposes to

$$\beta(s) = \lambda\beta_R(s) - (1 - \lambda). \quad (8)$$

²⁴Our pre-registration expressed the level of overreaction in terms of the absolute value of the subjective and objective movements. This is equivalent to Eq. (7) when excluding wrong direction reaction observations, as we also specified in our pre-registration.

Here $\beta_R(s) := (E_R(\omega|s) - E_B(\omega|s))/(E_B(\omega|s) - E_0(\omega))$ captures the level of overreaction arising from salience-channeled attention distorting the agent's mental representation, where $E_R(\omega|s) := \sum_{\omega_i \in \Omega} \omega_i p_R(\omega_i|s)$ denotes the expected state under the cognitively precise posterior.²⁵ The parameter λ captures the impact of cognitive imprecision. In the first stage, salience-channeled attention leads to excess mass on the high state following $s = r$ and excess mass on the low state following $s = b$. This results in $\beta_R(s) > 0$ for $\theta > 0$. If there is no cognitive imprecision ($\lambda = 1$), Eq. (8) simplifies to $\beta(s) = \beta_R(s)$ and overreaction emerges for $\theta > 0$. In the second stage, cognitive imprecision pulls mass away from the representative state. If there is no attentional constraint ($\theta = 0$), $\beta_R(s) = 0$, Eq. (8) simplifies to $\beta(s) = -(1 - \lambda)$ and underreaction emerges for $\lambda < 1$.

This decomposition also highlights how the two stages of belief updating interact. When both cognitive constraints are present, salience-channeled attention dominates when $\beta_R(s) > (1 - \lambda)/\lambda$, leading to overreaction (Eq. (8) is positive), and otherwise cognitive imprecision dominates, leading to underreaction. For fixed levels of θ and λ , $(1 - \lambda)/\lambda$ is a positive constant and $\beta_R(s)$ varies with the information environment from zero to a positive number. Therefore, our model makes predictions for how the direction and level of overreaction will vary across information environments. Note that wrong direction reaction does not arise in symmetric information environments (as shown in Lemma 2 in Appendix B.2).

When comparing two information environments (Ω, p_0) and (Ω', p'_0) in the theoretical analysis, we let $\beta(s)$ and $\beta'(s)$ denote the respective levels of overreaction. We use our experimental data to estimate $\beta(s)$ across different information environments as follows. For each task a participant completed we use her reported posterior belief to calculate the subjective expected state, where the numeric value of a state is the fraction of red balls in the corresponding bag. We calculate the objective prior and posterior expected states from the parameters of the information environment. Combining these calculations yields an estimate of $\beta(s)$ for the task. Our analysis averages these estimates over participants within each information environment.²⁶

4.2 Complexity of the State Space.

Understanding how the interaction between cognitive constraints and complexity impacts belief updating is a key focus of this paper. We turn to the literature on how people respond to complexity to formalize a notion of complexity. This litera-

²⁵In a symmetric environment, $E_0(\omega) = \bar{E}(\omega) = 1/2$, where $\bar{E}(\omega)$ is the expected state under cognitive default \bar{p}_d . Eq. (8) follows from plugging $\hat{E}(\omega|s) = \lambda E_R(\omega|s) + (1 - \lambda)\bar{E}(\omega)$ into Eq. (7).

²⁶To ensure that our results are not driven by the procedure of entering beliefs for every bag rather than the expectation, we ran a variation of the experiment where participants reported their expectation $\hat{E}(\omega|s)$ of the number of red balls directly. Importantly, they were aware of the number of states 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 C.9). We did not use this elicitation method for the main analysis because it would prevent us from also analyzing the full belief distribution, which is critical for testing our model's predictions.

ture distinguishes between two broad categories of complexity—*representational* and *computational*—based on which cognitive resources are taxed as complexity increases (Shenhav, Musslick, Lieder, Kool, Griffiths, Cohen, and Botvinick 2017; Botvinick and Cohen 2014).²⁷ Representational complexity taxes attention and working memory resources. It increases with the number of objects a person needs to consider to form an accurate mental representation of the environment, as this requires attending to and keeping track of more objects.²⁸ Computational complexity taxes resources related to controlled processing. It increases with the cognitive costs of carrying out the task at hand given the mental representation. Representational complexity interacts with attention and memory bounds to distort the mental representation, while computational complexity interacts with processing bounds to inject noise in the implementation of the relevant algorithm—in our case, Bayes’ rule.

While other work has explored the impact of computational complexity in a variety of domains (e.g., Enke and Shubatt (2023) in risk), this paper focuses on how variation in representational complexity affects belief updating. In our setting, representational complexity is proportional to the size of the state space, since the larger the state space the more objects (i.e., states) the agent needs to consider simultaneously.²⁹ This leads to the following definition.

Definition 3 (Measure of Representational Complexity). *An information environment (Ω', p'_0) is more (representationally) complex than (Ω, p_0) if Ω' contains weakly more states, $|\Omega'| \geq |\Omega|$.*

Importantly, an increase in objects (in our case, states) does not necessarily increase computational complexity, since in the representational stage the individual focuses on a subset of states and the processing stage takes the mental representation as an input. If attention is sufficiently constrained, the number of focal states may not increase with the number of states.³⁰

²⁷The cognitive psychology literature distinguishes between representational and computational capacity constraints because they correspond to different cognitive resources. This distinction is mirrored in the two categories of complexity. The relationship between complexity and cognitive resources implies that perceptions of complexity are inherently subjective: different people will perceive an environment as more or less complex depending on their level of the relevant cognitive resource.

²⁸Limits on attention and working memory resources imply that a person can fully attend to a limited number of objects at any given time. See Oberauer et al. (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 et al. 2011).

²⁹The focus on state complexity as a key component of representational complexity is mirrored in both theoretical and empirical work in finance (Molavi et al. 2023; Puri 2022) and computer science (Gao, Moreira, Reis, and Yu 2015; Papadimitriou 2003). It has been shown to have a large impact on choice (Oprea 2020).

³⁰As discussed in Section 2, attention is directed to a subset of states at the expense of neglecting others. We find empirical support for computational complexity remaining similar as state space complexity increases: specifically, salience-channeled attention dominates cognitive imprecision in determining how the level of reaction changes with the number of states.

To explore how representational complexity (hereafter complexity) impacts the level of overreaction, we fix the values of the most salient states following each signal realization (i.e., ω_1 and ω_N) and vary complexity by adding moderate states. We say a state ω_k is more *moderate* than state ω_i if it is closer to $1/2$, $|\omega_k - \frac{1}{2}| \leq |\omega_i - \frac{1}{2}|$. [Prediction 1](#) shows that when attentional resources are sufficiently constrained, the level of overreaction increases in the complexity of the information environment.

Prediction 1 (Complexity). *Consider two symmetric information environments (Ω, p_0) and (Ω', p'_0) with uniform priors and the same most representative states, $\omega_1 = \omega'_1$ and $\omega_N = \omega'_N$. Suppose Ω' is more complex than Ω , $|\Omega'| > |\Omega|$, and every state in $\Omega' \setminus \Omega$ is more moderate than every state in Ω . There exists a $\bar{\theta} > 0$ such that for $\theta > \bar{\theta}$ and $\lambda > 0$, the agent overreacts more in (Ω', p'_0) than (Ω, p_0) , $\beta'(s) > \beta(s)$ for $s \in \mathcal{S}$.*

For example, for sufficiently large θ , the agent overreacts more in three-state environment $\Omega_3 = \{0.3, 0.5, 0.7\}$ and four-state environment $\Omega_4 = \{0.3, 0.4, 0.6, 0.7\}$ than in binary state environment $\Omega_2 = \{0.3, 0.7\}$, and more in five-state environment $\Omega_5 = \{0.3, 0.4, 0.5, 0.6, 0.7\}$ than Ω_4 .³¹

This result is driven by the interaction between limited attention and the signal structure. For good news signal, the most salient states are the most extreme states. Channeling attention to these states results in a mental representation that neglects the possibility of moderate states. As complexity increases, the uniform prior places less weight on the extreme states. Thus, the objective posterior also puts less weight on these states, leading to less objective movement in beliefs compared to simpler environments. However, an attention-constrained agent continues to channel attention towards the extreme states and overweigh them. This leads to relatively more excessive movement in beliefs, i.e., greater overreaction.

Notably, the interaction between cognitive constraints in each stage of belief updating plays a critical role in predicting the direction of reaction. This interaction can generate underreaction in simple environments—where the distortion in the mental representation is minimal—and overreaction in complex ones. In contrast, a limited-attention-only model that considers cognitive constraints only in the first representational stage (i.e., $\theta > 0$ and $\lambda = 1$) predicts overreaction regardless of complexity. Likewise, a cognitive-imprecision-only model that considers cognitive constraints only in the second processing stage (i.e., $\theta = 0$ and $\lambda < 1$) predicts underreaction. [Prediction 12](#) in [Appendix B.4](#) formally highlights these insights.

Empirical Findings. We test [Prediction 1](#) by comparing the level of overreaction while varying complexity, holding constant the prior (uniform) and the most representative states. In simple binary state environments, we replicate the finding of

³¹Note that Ω_3 is not directly comparable to Ω_4 or Ω_5 by [Prediction 1](#) because the states $\Omega_4 \setminus \Omega_3 = \{0.4, 0.6\}$ and $\Omega_5 \setminus \Omega_3 = \{0.4, 0.6\}$ are not more moderate than state $0.5 \in \Omega_3$.

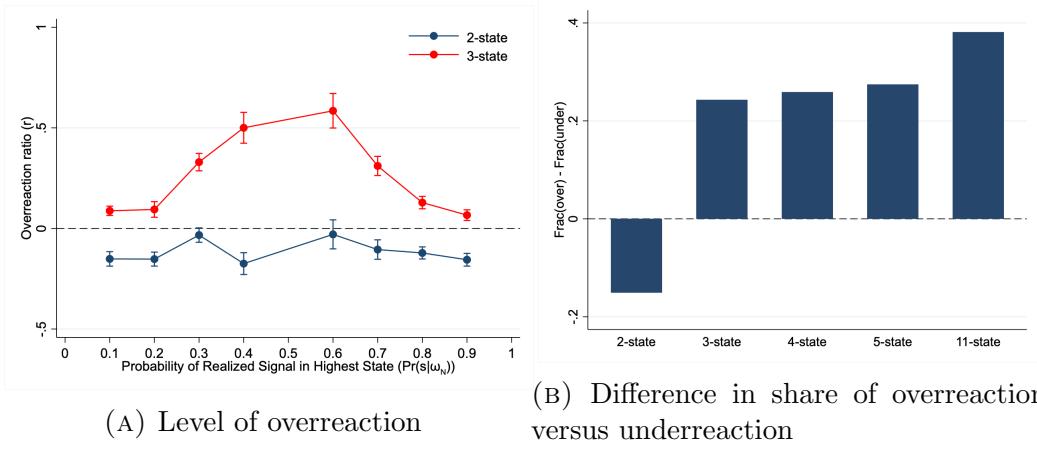


FIGURE 3. Complexity increases overreaction. Each data point aggregates all uniform prior environments of a given complexity—by signal structure and signal realization in Panel (A) and across all signal structures and signal realizations in Panel (B).

underreaction from the experimental literature: participants' level of overreaction is negative across all 2-state environments considered ($\beta < 0$, $p < .001$).³² This finding also holds separately for each signal structure. Fig. 3a plots the level of overreaction for each signal structure and signal realization, where the x-axis corresponds to the probability of the realized signal in state ω_N , ranging from 0.6 to 0.9 for a red ball and 0.1 to 0.4 for a blue ball.³³ As shown in the figure, we observe significant underreaction in nearly all 2-state environments. This is consistent with the experimental evidence summarized in Benjamin (2019), which documents systematic underreaction in lab experiments with binary state spaces.

Strikingly, increasing the complexity of the state space by adding even a single state changes the direction of the bias: participants switch from underreacting in the 2-state case to overreacting in the 3-state case ($\beta > 0$, $p < .001$ across all 3-state signal structures). This finding also holds separately for each signal structure: as shown in Fig. 3a, we observe significant overreaction in all 3-state environments.

The pattern of increasing overreaction continues in more complex settings. We compare 2-state environments to 4-state and 5-state environments with additional moderate states. Regressing the level of overreaction in these 2, 4 and 5-state environments on dummies for 4 and 5 states, we find that the level of overreaction increases with complexity: it is significantly higher in 4-state environments compared to 2-state environments ($p < .01$), and in 5-state environments compared to 4-state environments ($p < .01$) (see Table C.3 in Appendix C.4.1).³⁴ A similar pat-

³²This p -value and others reported in the text are from a one-sample t -test against 0, unless otherwise noted.

³³In symmetric signal structures, when a blue ball occurs with probability x in state ω_N , then a red ball occurs with probability $1 - x$. Therefore, 0.1 and 0.9 on the x-axis correspond to blue and red signal realizations for the same signal structure, and so on.

³⁴We do not include 3-state environments in this regression because Prediction 1 cannot be used to compare 3-state environments to more complex environments. This is because it is not possible

tern emerges when we control for the signal structure by adding dummies for each diagnosticity. Finally, we study an 11-state environment to examine belief updating with “many” states. We find significant overreaction ($\beta > 0$, $p < .001$) and a higher level of overreaction than in all other complexities (although this is not a direct test of [Prediction 1](#) since the representative states also change).

As an alternative measure of overreaction, we compute the difference between the fraction of tasks in which participants exhibit overreaction versus underreaction (this measure is also used in prior work, e.g., [Fan, Liang, and Peng \(2023\)](#)). A positive value indicates more tasks with overreaction and a negative value indicates the opposite. As shown in [Fig. 3b](#), underreaction is more prevalent in 2-state environments, and overreaction is more prevalent in environments with 3 or more states.

Together, these results provide strong evidence that overreaction increases with complexity ([Prediction 1](#)). Moreover, observing underreaction in simple environments and overreaction in complex environments provides evidence for the presence of both attentional and processing constraints ([Prediction 12](#)).

4.3 Signal Diagnosticity.

Limited cognitive resources also generate a relationship between the level of overreaction and other properties of the information environment. We next explore how the precision of information impacts overreaction by fixing the complexity of the state space and decreasing the diagnosticity of the signal in *all* states, i.e., moving all states closer to 0.5. [Prediction 2](#) shows that when attentional resources are sufficiently constrained, the level of overreaction increases as the signal becomes less informative, i.e., the diagnosticity decreases.

Prediction 2 (Diagnosticity). *Consider two symmetric information environments (Ω, p_0) and (Ω', p'_0) with the same complexity, five or fewer states, and uniform priors. Suppose Ω' is less diagnostic than Ω , $d'_i \leq d_i$ for all $i = 1, \dots, N$ with at least one inequality strict. There exists a $\bar{\theta} > 0$ such that for $\theta > \bar{\theta}$ and $\lambda > 0$, the agent overreacts more in (Ω', p'_0) than (Ω, p_0) , $\beta'(s) > \beta(s)$ for $s \in \mathcal{S}$.*

For example, consider $\Omega_4 = \{x, y, 1 - y, 1 - x\}$ with $x \in (0, 0.5)$ and $y \in (x, 0.5)$. [Prediction 2](#) establishes that overreaction increases as x and y move closer to 0.5.³⁵

Moving moderate states closer to 0.5 has a similar impact to adding moderate states in [Prediction 1](#): it results in a more distorted mental representation. The impact of moving extreme states is more nuanced, as this involves changing the value of the most salient states. As the extreme states move towards 0.5, the objective movement in beliefs decreases in magnitude, as the extreme states are closer to the prior expected state $E_0(\omega) = 0.5$. The subjective movement also decreases for

to add moderate states to a symmetric 3-state environment since it includes state 0.5.

³⁵In [Appendix B.2](#), we provide a sufficient condition on Ω for an analogous result to hold for more than five states ([Prediction 6](#)). We present the simpler result here since all information environments in our experiment have five or fewer states.

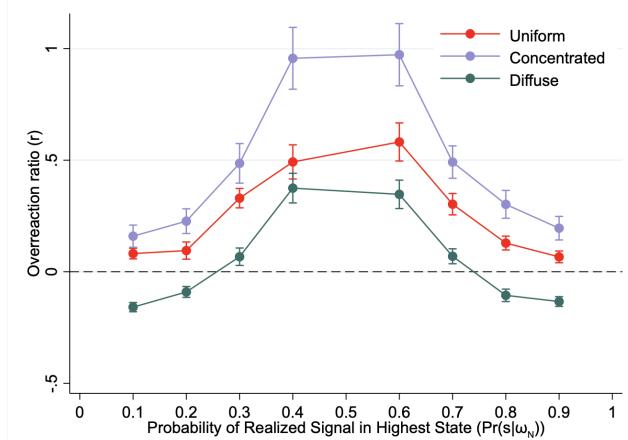


FIGURE 4. Overreaction decreases in signal diagnosticity and increases in prior concentration. Each data point aggregates all 3-state environments of a given prior by diagnosticity and signal realization.

the same reason. The level of overreaction increases when the objective movement decreases by more than the subjective movement. In information environments with five or fewer states, this turns out to always hold when the attentional constraint is sufficiently high.³⁶

Empirical Findings. We test Prediction 2 by examining how the level of overreaction varies with signal diagnosticity. Regressing the level of overreaction on signal diagnosticity, consistent with our prediction, we observe significantly more overreaction as the diagnosticity decreases. This holds for each level of complexity with the exception of the 2-state environment between diagnosticities $d = 0.6$ and $d = 0.7$. See Table C.5 in Appendix C.4.1 for these results. Fig. 4 illustrates this finding in 3-state uniform prior environments: overreaction is highest at $d = 0.6$ ($0.4/0.6$ on the x-axis) and lowest at $d = 0.9$ ($0.1/0.9$ on the x-axis). While Prediction 2 is derived for a uniform prior, similar patterns hold for the other priors shown in Fig. 4.³⁷

Similar to the complexity comparative static, the interaction between cognitive constraints determines whether and how the direction of reaction varies with signal diagnosticity (Prediction 13 in Appendix B.4). Given our structural estimates of the attentional and processing parameters, our model predicts that for all signal diagnosticities that we consider, cognitive imprecision dominates and underreaction emerges in simple 2-state environments—which do not strain attention—while salience-channeled attention dominates and overreaction emerges in more complex

³⁶Similarly, the model in Augenblick et al. (2022)—which imposes cognitive imprecision through signal strength—predicts that the level of overreaction varies with signal diagnosticity, with overreaction to noisy signals and underreaction to precise signals. See Appendix D for discussion.

³⁷Edwards (1968) and Augenblick et al. (2022) find overreaction to extremely noisy signals in a 2-state environment. As a comparison, we ran a version of the 2-state environment with $d = 0.51$. We also found overreaction to this very noisy signal ($r = 0.08$, $p < .001$). We do not include this result in the regression analyses or figures because we did not run this signal diagnosticity in any other complexity or prior information environment.

environments.³⁸ Indeed, in the experiment, we observe underreaction at all diagnosticities in the 2-state environments and overreaction at all diagnosticities in the higher complexity environments, holding the prior uniform. Thus complexity dominates signal diagnosticity in determining the direction of reaction.

4.4 Prior Concentration on Moderate States.

We next explore how the shape of the prior impacts overreaction by fixing the state space and varying the mass that the prior assigns to moderate versus extreme states. Given symmetric information environments (Ω, p_0) and (Ω, p'_0) , we say prior p'_0 is more *concentrated on moderate states* than p_0 if it assigns higher probability to moderate states and lower probability to extreme states.³⁹ [Prediction 3](#) shows that when attentional resources are sufficiently constrained, the level of overreaction increases as the prior becomes more concentrated.

Prediction 3 (Prior concentration). *Consider two symmetric information environments (Ω, p_0) and (Ω, p'_0) and suppose p'_0 is more concentrated on moderate states than p_0 . There exists a $\bar{\theta} > 0$ such that for $\theta > \bar{\theta}$ and $\lambda > 0$, the agent overreacts more in (Ω, p'_0) than (Ω, p_0) , $\beta'(s) > \beta(s)$ for $s \in \mathcal{S}$.*

The intuition is similar to [Prediction 1](#). The more weight that the prior places on moderate states, the less the objective posterior moves in response to new information. However, an attention-constrained agent continues to overweigh extreme states, leading to increasingly excessive movement in average beliefs relative to the objective benchmark, i.e., greater overreaction.

Empirical Findings. We use the 3-state environments to test [Prediction 3](#), as three is the minimum number of states needed to manipulate the prior concentration. We varied the prior from a diffuse distribution, $p_0 = (0.4, 0.2, 0.4)$, which places twice as much mass on the extreme states as the moderate state; to a uniform distribution, $p_0 = (0.33, 0.34, 0.33)$; to a concentrated distribution, $p_0 = (0.25, 0.50, 0.25)$, which places twice as much mass on the moderate state as the extreme states.

Consistent with [Prediction 3](#), we observe significantly more overreaction as the prior becomes more concentrated: regressing the level of overreaction 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 C.4](#) in [Appendix C.4.1](#)). As shown in [Fig. 4](#), this holds for both signal realizations across all information environments. This provides strong support for [Prediction 3](#).

As before, the interaction between cognitive constraints in our framework plays

³⁸For example, our structural estimates of $\theta = 0.85$ and $\lambda = 0.70$ lie in the cognitive-interaction region of the 3-state uniform-prior environments we consider, near the border with the limited-attention-dominant region. As shown in [Fig. B.1](#) in [Appendix B.4](#), these parameter estimates predict overreaction for all but the highest diagnosticities.

³⁹Formally, 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.

a critical role in determining how the direction of reaction varies with the prior. This interaction can generate overreaction for more concentrated priors and underreaction for less concentrated priors (see [Prediction 14](#) in [Appendix B.4](#)).⁴⁰ In the experiment, we find significant overreaction for the uniform and concentrated priors and significant underreaction for the diffuse prior at high diagnosticities ($d = 0.9$ and 0.8 ; see $0.1/0.9$ and $0.2/0.8$ on the x-axis of [Fig. 4](#)). In contrast, we find significant overreaction for all three priors at low diagnosticities ($d = 0.7$ and 0.6 ; see $0.3/0.7$ and $0.4/0.6$ on the x-axis).⁴¹ Taken together, these findings provide further evidence for both attentional and processing constraints and highlight how they interact with the learning environment to determine the direction of reaction.

4.5 Prior Symmetry.

Finally, we explore how prior symmetry impacts overreaction. We focus on binary state spaces where the degree of prior symmetry can be measured by the difference between $p_0(\omega_1)$ and $p_0(\omega_2)$, with $p_0(\omega_1) = p_0(\omega_2)$ corresponding to a symmetric prior. Under an asymmetric prior, the two signal realizations are no longer ex-ante equally likely. We refer to the more likely “expected” realization as *confirmatory* and the less likely “surprising” realization as *disconfirmatory*. Note that when $p_0(\omega_1) > p_0(\omega_2)$, b is confirmatory and r is disconfirmatory, with the reverse when $p_0(\omega_1) < p_0(\omega_2)$.

The interaction between cognitive constraints yields a rich set of predictions about how the level of overreaction differs for surprising versus expected information. For a disconfirmatory signal, the level of overreaction decreases in the signal diagnosticity, with overreaction to imprecise signals and underreaction to precise ones. For a confirmatory signal, the level of overreaction is non-monotonic in the signal diagnosticity: people update in the wrong direction when the signal is imprecise, but can potentially underreact and even overreact to more precise signals.

Prediction 4 (Asymmetric Prior). *Consider an information environment (Ω, p_0) with a symmetric binary state space and an asymmetric prior. Suppose $\lambda < 1$ and let d denote the diagnosticity of Ω .*

1. *The level of overreaction $\beta(s)$ to a disconfirmatory signal realization s is decreasing in d , with overreaction to imprecise signals (low d) and underreaction to precise signals (high d).*
2. *The level of overreaction $\beta(s)$ to a confirmatory signal realization s is single-peaked in d , first increasing and then possibly decreasing. The agent reacts*

⁴⁰In contrast, the limited-attention-only model predicts overreaction for all priors and the cognitive-imprecision-only model predicts underreaction for all priors

⁴¹Again, these findings are predicted by our model. Our structural estimates of $\theta = 0.85$ and $\lambda = 0.70$ ([Section 3](#)) lie in the cognitive-interaction region of [Prediction 14](#)—where the direction of reaction depends on the prior—for diagnosticities $d = 0.8$ and 0.9 , but in the limited-attention-dominant region—where overreaction emerges for all priors—for diagnosticities $d = 0.6$ and 0.7 . See [Fig. B.2](#) in [Appendix B.4](#) for an illustration.

in the wrong direction to imprecise signals (low d) and, for sufficiently low cognitive imprecision, underreacts to precise signals (high d).

Recall that in symmetric environments, overreaction was primarily driven by salience-channeled attention. However, in asymmetric environments, cognitive imprecision alone generates overreaction to disconfirmatory signals. For intuition, consider the cognitive-imprecision-only model with $\theta = 0$. In this case, the subjective expected state is equal to a weighted average of the objective expected state and the cognitive default, $\hat{E}(\omega|s) = \lambda E_B(\omega|s) + (1 - \lambda)\bar{E}(\omega)$, where $\bar{E}(\omega) = 0.5$ since the cognitive default is uniform. Suppose $p_0(\omega_1) < p_0(\omega_2)$, which implies b is disconfirmatory and $1/2 < E_0(\omega)$. Upon observing b , for a sufficiently imprecise signal the objective expected state decreases below the prior but remains above the cognitive default, $\bar{E}(\omega) < E_B(\omega|b) < E_0(\omega)$. Relative to $E_B(\omega|b)$, cognitive imprecision compresses $\hat{E}(\omega|b)$ towards $\bar{E}(\omega)$, which leads to overreaction. For a more precise signal, $E_B(\omega|b)$ decreases below the cognitive default, $E_B(\omega|b) < \bar{E}(\omega) < E_0(\omega)$, and cognitive imprecision instead increases $\hat{E}(\omega|b)$, which results in underreaction.

Another key contrast with symmetric environments is the prediction of wrong direction reaction to confirmatory signals—also driven by cognitive imprecision. Again consider the cognitive-imprecision-only model and $p_0(\omega_1) < p_0(\omega_2)$. Wrong direction reaction can arise when $E_B(\omega|s)$ and $\bar{E}(\omega)$ are on opposite sides of the prior expected state $E_0(\omega)$, which can only occur for a confirmatory signal. Suppose $p_0(\omega_1) < p_0(\omega_2)$, which implies r is confirmatory and $E_0(\omega) > 1/2$. The objective expected state increases following r , $E_B(\omega|r) > E_0(\omega)$. Cognitive imprecision compresses the subjective expected state towards the cognitive default, which is less than the prior, $\bar{E}(\omega) < E_0(\omega)$. For a sufficiently imprecise signal (low diagnosticity), $E_B(\omega|r)$ is close enough to $E_0(\omega)$ that $\hat{E}(\omega|s) = \lambda E_B(\omega|s) + (1 - \lambda)\bar{E}(\omega) < E_0(\omega)$ and wrong direction reaction emerges. Provided cognitive imprecision is not too high (λ is sufficiently large), at higher diagnosticities $\hat{E}(\omega|r)$ remains above $E_0(\omega)$ but below $E_B(\omega|r)$, resulting in underreaction. Incorporating salience-channeled attention ($\theta > 0$) increases the level of overreaction to both types of signals—as in symmetric environments.

Empirical Findings. We test Prediction 4 by comparing how the level of overreaction to confirmatory and disconfirmatory signals varies with signal diagnosticity. We focus on 2-state environments with asymmetric priors $p_0 = (0.3, 0.7)$ or $p_0 = (0.7, 0.3)$.⁴² Since Prediction 4 predicts wrong direction reaction at some diagnosticities, we include such observations in this analysis. We regress the level of overreaction on signal type (confirmatory or disconfirmatory) and diagnosticity, using 2-state environments with symmetric priors as the baseline neutral signal type.

⁴²Under prior $(0.3, 0.7)$, a red ball is confirmatory and a blue ball is disconfirmatory, with the opposite under prior $(0.7, 0.3)$.

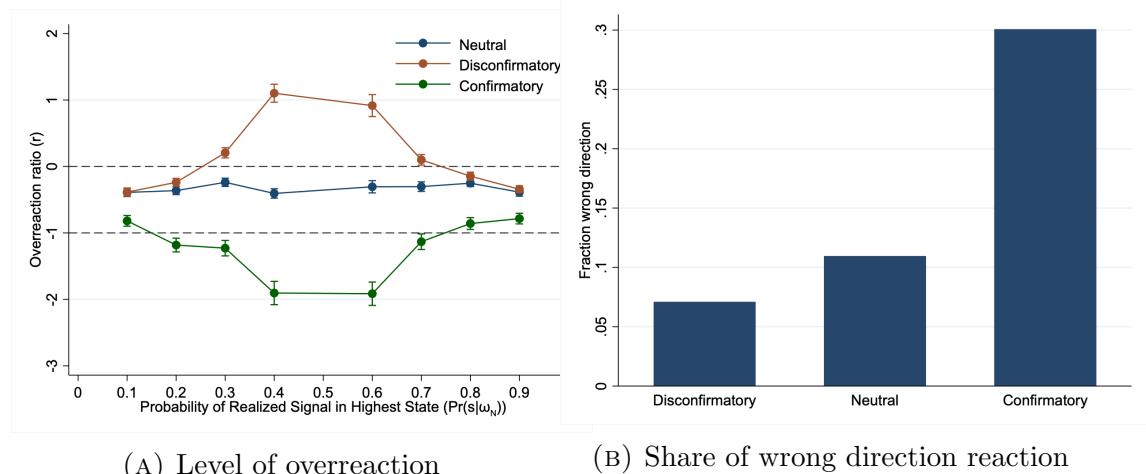


FIGURE 5. Reaction varies by signal type. Both figures include wrong direction reaction observations. In Panel (A), each data point aggregates all 2-state environments by diagnosticity and signal type. In Panel (B), each bar aggregates all 2-state environments by signal type.

Consistent with [Prediction 4](#), we observe significant overreaction to disconfirmatory signals at low diagnosticities and underreaction at high diagnosticities, with the level of overreaction monotonically decreasing in diagnosticity. For confirmatory signals, we observe significant wrong direction reaction at low diagnosticities ($\beta < -1$) and underreaction at high diagnosticities, with the level of overreaction monotonically increasing in diagnosticity. [Fig. 5a](#) illustrates these findings (again $d = 0.6$ corresponds to 0.4/0.6 on the x-axis and so on) and [Table C.10](#) in [Appendix C.4.2](#) presents analogous regression results.⁴³ It is notable that when the prior is asymmetric, we observe overreaction even in simple 2-state environments. This is consistent with cognitive imprecision being the dominant cognitive constraint in simple environments—as we observed for symmetric priors—since cognitive imprecision generates overreaction at low signal diagnosticities when the prior is asymmetric.

Finally, we explore whether participants are more likely to wrong direction react to confirmatory signals—where it is predicted to arise—relative to disconfirmatory and neutral signals, where it is not. As illustrated in [Fig. 5b](#), wrong direction reaction to confirmatory signals occurs roughly three times as frequently as in the other two cases: participants react in the wrong direction in nearly 30% of confirmatory signal tasks, but in only 8% and 11% of disconfirmatory and neutral signals, respectively. Importantly, this incidence of wrong direction reaction is not arbitrary noise (e.g., inattentive subjects), but a direct prediction of our model.

4.6 Discussion.

Measure of overreaction. Our definition and measure of overreaction are based on the expected state. This is consistent with the finance and experimental liter-

⁴³For the sake of comparability with the other analyses, [Table C.6](#) in [Appendix C.4.1](#) and [Fig. C.1](#) in [Appendix C.4.3](#) replicates this analysis excluding wrong direction observations.

atures. The former typically studies asset prices and average forecasts, which are summary statistics of the belief distribution similar to the expected state.⁴⁴ The latter typically compares subjective and objective belief movements in binary state environments, which is equivalent to our comparison of expected states (see [Lemma 3](#) in [Appendix B.1](#)). The experimental literature also typically uses a Grether regression ([Grether 1980](#)) to assess the degree of overreaction. As discussed further in [Appendix C.1](#), we developed the measure $\beta(s)$ in part because Grether regressions become difficult to implement and interpret when there are more than two states.

Our measure $\beta(s)$ satisfies several desirable properties. First, it is scale-invariant: $\beta(s)$ stays constant when all states are scaled proportionally (e.g., doubled). This is not true for the numerator of $\beta(s)$ alone, which is why the raw difference between subjective and objective movements is not a good measure. Second, in environments where our model does not predict wrong direction reaction (e.g., symmetric environments), $\beta(s)$ is equivalent to a measure that compares the absolute values of subjective and objective movements. However, in environments where wrong direction reaction can arise, $\beta(s)$ distinguishes between movement in the wrong direction and equal movement in the correct direction, whereas the absolute value measure does not. Finally, $\beta(s)$ generates a complete order over posterior beliefs as any two posterior beliefs are comparable. This contrasts with measures based on the entire posterior distribution (e.g., movement along the line through the prior and the objective posterior), which may be incomplete when there are more than two states.

State-space complexity. The result on increasing complexity ([Prediction 1](#)) critically hinges on how adding states changes the relative representativeness of existing states. Such changes are substantial when the additional states are distinct from existing states, but not when they are very close together. For example, consider an environment with a uniform prior and a simple state space $\Omega = \{0.3, 0.7\}$. The level of overreaction moves continuously in $\varepsilon > 0$ when moving to a more complex state space $\Omega' = \Omega \cup \{0.3 + \varepsilon, 0.7 - \varepsilon\}$ with arbitrarily similar additional states. At $\varepsilon = 0$, Ω' and Ω are equivalent and the two environments have equal levels of overreaction. Therefore, the impact of increasing complexity is not determined by the number of new states per se, but by the number of new *dissimilar* states, as this is what impacts the distortion due to salience-channeled attention.⁴⁵ Our experiments focus on

⁴⁴Our 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 $E_B(\omega|s) - \hat{E}(\omega|s)$ and $\hat{E}(\omega|s) - E_0(\omega)$, respectively. It is straightforward to verify that if $\hat{E}(\omega|s)$ moves in the same direction as $E_B(\omega|s)$, then $\beta(s) < 0$ if and only if forecast errors and revisions are negatively correlated, i.e. $(E_B(\omega|s) - \hat{E}(\omega|s))(\hat{E}(\omega|s) - E_0(\omega)) > 0$.

⁴⁵Indeed, [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. [Enke and Graeber \(2023\)](#) find the same when duplicating one state in a binary state environment. Because duplicate states—or states so similar that they are treated as the same in the mental representation—do not constitute

information environments with distinct states.

A continuous state space, at least conceptually, could be viewed as infinitely complex due to the infinite number of states one must consider. However, evidence from the category perception literature suggests that similar nearby states will be grouped together within a coarse category and treated as a single object in the mental representation. For example, when confronted with a continuous stimulus, people perceive it through the lens of coarse categories with little fuzziness in between (Harnad 2003; Nosofsky, Clark, and Shin 1989; Shanks and Gluck 1994); people also discretize continuous probabilities into coarse categories such as “might happen” and “will happen” (Schley, Ferecatu, Chan, and Gunadi 2023). Graeber, Roth, and Sammon (2025) argue that this coarse categorization of continuous stimuli is a response to representational complexity. Therefore, our model can be readily applied to settings with a continuous state space when categories can either be inferred (as in Graeber et al. (2025) and Schley et al. (2023)) or elicited.

Beyond symmetric good news environments. When the information environment is symmetric and has a good news structure, representativeness is the key driver of overreaction, and cognitive imprecision is the key driver of underreaction. However, there are also environments in which representativeness generates underreaction—for example, those with a non-good news signal structure (see Section 5.4)—or cognitive imprecision generates overreaction—for example, environments with an asymmetric prior (Prediction 4). We focus on symmetric information environments as these settings are sufficiently tractable to yield sharp predictions. The predictions generally continue to hold in asymmetric environments, albeit with more cumbersome notation.

Inference versus forecasting. Our analysis has thus far focused on the inference domain, where people are tasked with inferring the likelihood of states after observing information. But much of the evidence for overreaction in the finance literature comes from the forecasting domain, where the relevant representational objects are either the same as the signal space (e.g., predicting a future price based on today’s price) or a function of it (e.g., predicting the future payoff of an option—a function of the future price—based on today’s price). Since inference and forecasting both require forming a mental representation and using it to process information, our model can be readily applied to forecasting. In Appendix E we present experimental evidence showing that similar insights obtain: people overreact (underreact) in complex (simple) forecasting tasks for both abstract financial assets (Appendix E.1) and more realistic financial products (Appendix E.2). Our results suggest that evidence on the “inference-forecasting” gap may be at least partially driven by differences in complexity between inference and forecasting tasks (Fan et al. 2023).

distinguishable objects in terms of attentional resources, they do not notably increase representational complexity.

Including wrong direction observations. As discussed in Section 3, per our pre-registration we drop wrong direction observations from all analyses except for the asymmetric prior. Appendices C.4.2 and C.4.3 replicate our analyses including wrong direction observations. The results do not meaningfully change.

5 Testing the Attentional Mechanism

We next present direct evidence for salience-channeled attention as the proposed cognitive mechanism. First, we measure attention to see where it is directed. Second, we restrict attention and show how this generates posterior beliefs consistent with an increase in θ in our model. Third, we manipulate salience cues directly to study how belief updating changes when representativeness is removed as a salience cue and other types of salience cues are present. Importantly, in these three studies we manipulate attention while keeping the information environment constant. This rules out that any variation in beliefs across treatments is driven by differences in the information environment (for example, signal strength). Finally, we explore our model’s predictions in a signal structure where a moderate state is most salient. Together, these studies provide evidence of the causal effect of attention on belief updating.

5.1 Measuring and Restricting Attention

In the representational stage of our model, attentional constraints lead agents to focus on the most salient states. The framework predicts that agents’ attention will be channeled towards the state that is most representative of the observed signal. Any further limits on attentional resources will exacerbate this effect, leading to greater 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.⁴⁶ The Mouselab paradigm measures participants’ attention to various features of the decision environment by the timing of the objects that they click on, where the ordering of clicks corresponds to the ordering in which objects are attended to.⁴⁷ The Mouselab paradigm, which requires participants to click on attributes, has also been shown to tax attentional resources, as 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).

⁴⁶The 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).

⁴⁷For 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. The use of click data as a proxy for attention has been validated using eye-tracking tools (Meißner, Decker, and Pfeiffer 2010).

We used the Mouselab paradigm to measure which states are attended to first and how the increased tax on attentional resources impacts the level overreaction. This *Limited Attention* treatment required a participant to click on a state (e.g., Bag 5) before entering her posterior belief about the state. Once a state was clicked, the participant could enter her belief for that state as before (again, this belief had to sum to 100). The order of states was randomized so that either the bag with the most or least red balls appeared first. We ran *Limited Attention* for all 5-state information environments listed in [Table C.1](#).⁴⁸ Our comparison benchmark is the design presented in [Section 3](#) for 5-state environments, which we refer to as *Baseline*.

In the context of the Mouselab paradigm, our three predictions are as follows: (i) participants will click on the representative state first: upon observing a blue (red) ball, the most likely first-click will be on the bag with the most blue (red) balls; (ii) fixing the information environment, overreaction will be higher in *Limited Attention* relative to *Baseline*; and (iii) overreaction will be higher for participants who are more prone to salience-channeled attention—proxied by their propensity to click the representative state first—relative to other participants.

Empirical Findings. [Fig. 6](#) shows the distribution of first-clicks across all information environments. In line with prediction (i), participants were much more likely to attend to the most representative state—proxied by their first click—relative to other states. This 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 order of states was randomized rules out that this result is driven by an information-independent heuristic (e.g., always click on the left-most bag first). This finding provides evidence that representativeness does indeed drive attention.

To examine prediction (ii), [Fig. 7a](#) compares the level of overreaction in *Limited Attention* to *Baseline*. Overreaction is indeed significantly higher in the latter across nearly all information environments. Consistent with prediction (iii), [Fig. 7b](#) shows that within *Limited Attention*, overreaction is substantially higher for individuals who clicked on the representative state first relative to those who clicked on another state first. These patterns are also borne out in a regression analysis: overreaction is significantly higher in *Limited Attention* compared to *Baseline*; within *Limited Attention*, participants who clicked the representative state first exhibited greater overreaction compared to those who did not (see [Table C.11](#) in [Appendix C.5](#)). Together, these findings provide evidence that tighter attentional constraints increases overreaction.

Finally, we provide further evidence for the proposed predictions by structurally

⁴⁸ Aside from the non-good-news signal experiment, the experiments in this section focus on state spaces with five states, as limited attention is predicted to have a larger impact in these more complex environments.

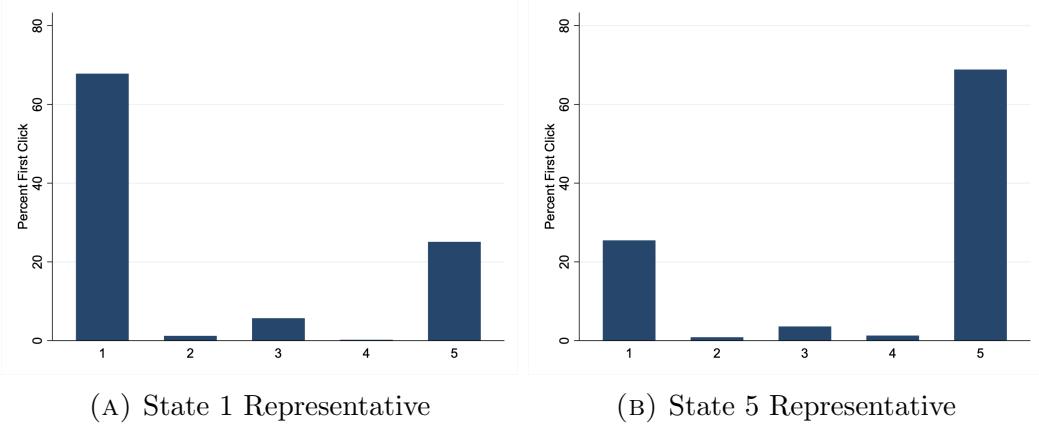


FIGURE 6. Most participants click on representative state first. Each bar aggregates both signal realizations for all 5-state environments in *Limited Attention*.

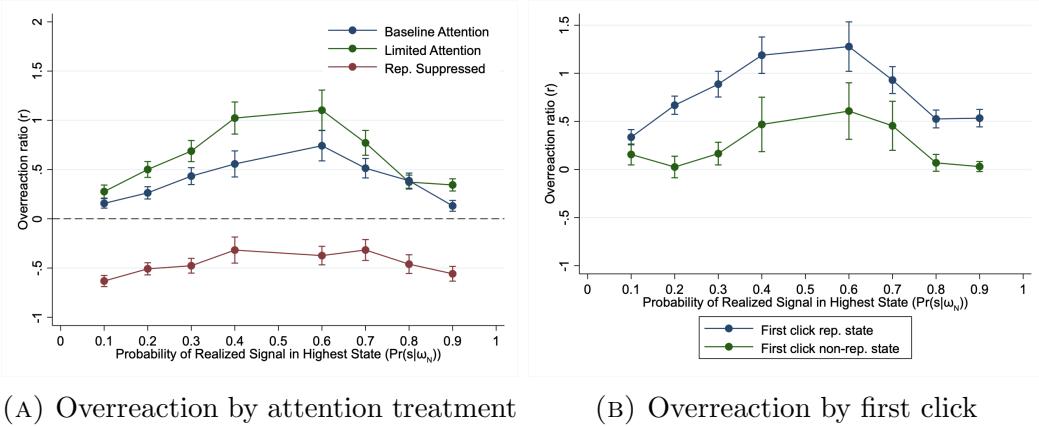


FIGURE 7. Limited attention increases overreaction and suppressing representativeness leads to underreaction. Each data point aggregates all 5-state environments in the given attention/salience treatment by diagnosticity and signal realization.

estimating the parameters for *Limited Attention* and comparing them to *Baseline* (see Appendix C.7 for a description of the estimation method). The estimate of θ increases from 0.99 to 1.26, while the estimate of λ remains similar—0.73 versus 0.74 (see Table C.12 in Appendix C.5). This lends direct support to our prediction that further restricting attentional resources through the Mouselab paradigm increases salience-driven distortion in the representational stage (higher θ), while leaving the level of cognitive imprecision unchanged.

Importantly, these results rule out insensitivity or information-independent heuristics (e.g., partition dependence (Tversky and Koehler 1994)) as alternative explanations for the findings presented in Section 4. Since the information environment was held constant across treatments, such heuristics predict that beliefs would remain the same across treatments.

5.2 The Causal Effect of Attention on Belief Updating

We next isolate the causal effect of salience-channeled attention on belief updating—specifically, how the relative attention directed to a state impacts the weight that the posterior belief places on this state. The key prediction of our framework is that more attention leads to more weight. We also show that when attention is channeled to moderate states, underreaction can emerge—even in a complex environment.

To identify the causal effect of attention, we developed an experimental setting with no representativeness-based salience cue. Specifically, we adapted the Mouse-lab variation of our paradigm so that after observing the signal, participants were required to click on each state to learn the signal distribution associated with that state (i.e., which ball composition was assigned to the bag).⁴⁹ This removed the initial representativeness salience cue, as a state’s representativeness—which is a function of the associated signal distribution—could no longer draw initial attention. Otherwise, this *Representativeness Suppressed* treatment was identical to *Limited Attention*: participants were presented with the same information about the learning environment, including the set of possible ball compositions, and the order of states was randomized as before. We ran *Representativeness Suppressed* for all 5-state environments listed in [Table C.1](#).

We conjecture that in this setting, individuals will attend to states as-if randomly since there is no salience cue to draw attention to a particular state. Applying our theoretical framework to a setting with random attention yields two key predictions: (i) limited attention leads the agent to overweigh the most-attended-to state (as indicated by first-click), in that her posterior belief places excess mass on this state relative to the objective posterior ([Prediction 8 in Appendix B.3](#)); and (ii) random attention leads to underreaction—even when the environment is complex ([Prediction 9 in Appendix B.3](#)). To see the intuition for (ii), note that the most representative state is still overweighed when it is attended to first, but when attention is randomly directed and there are five states, it is attended to first only 20% (1/5) of the time. The remaining 80% of the time, the other four less representative states are overweighed. Taken together, this pattern generates underreaction.

Empirical Findings. To test the random attention conjecture, we look at the first-click data. In stark contrast to *Limited Attention*, each of the five states was equally likely to be clicked first in *Representativeness Suppressed*, as shown in [Fig. C.2](#) in [Appendix C.5](#). In other words, attention was not associated with representativeness.

Our experimental evidence supports both theoretical predictions outlined above. In line with (i), participants’ posterior beliefs placed excess mass on the state that they attended to first and insufficient mass on the remaining states (see [Fig. 8a](#)).

⁴⁹For example, the participant had to click on the state in the first position to find out that it was associated with 90 (10) red (blue) balls.

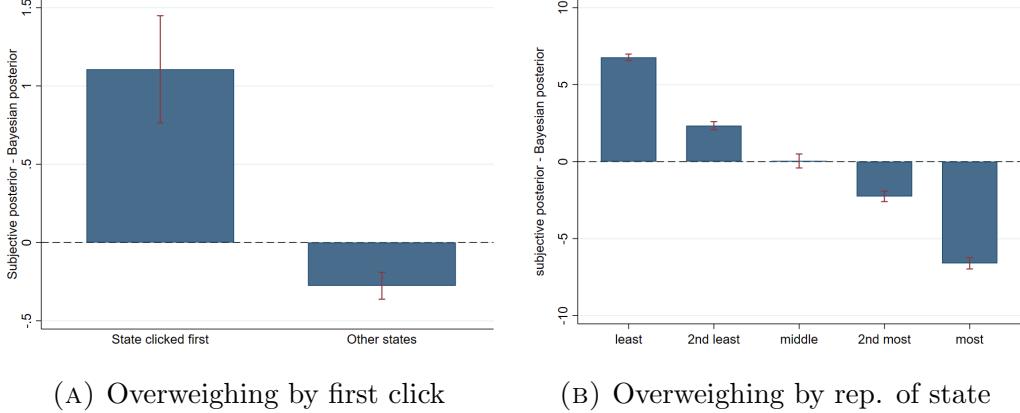


FIGURE 8. Random attention generated distinct patterns of under- and overweighing. Each bar aggregates both signal realizations for all 5-state environments in *Representativeness Suppressed*; the Other States bar averages across all states not clicked first. Beliefs are measured as a percentage from 0 to 100.

Therefore, even when attention was allocated randomly rather than via representativeness, it still led to overweighing of the most-attended-to state and underweighing of the less-attended-to states. This isolates the causal effect of attention on belief updating. In line with (ii), participants overweighed the two least representative states and underweighed the two most representative states (see Fig. 8b), leading to underreaction in all 5-state environments in our experiment (see Fig. 7a). This sharply contrasts with *Limited Attention* and *Baseline*, where participants overreacted in all 5-state environments when the representativeness cue was present.

Together, these results identify the causal effect of attention on belief updating. Importantly, they also demonstrate that underreaction can emerge in complex environments when attention is directed to non-extreme states.

5.3 Alternative Salience Cues

Salience-channelled attention plays a central role in our framework. While we focus on representativeness as the primary salience cue, other cues have also been explored in the literature. Next, we study two alternative salience cues extensively explored in prior work—namely, bottom-up visual salience and top-down goal-directed salience. We compare their impact on attention and belief updating to representativeness, both when the representativeness cue is present and when it is suppressed.

Drawing on the literature (Li and Camerer 2022; Maćkowiak et al. 2023), we generated visual salience by highlighting one state in a color (bright yellow) that starkly contrasts with the background color. We generated goal-directed salience by incentivizing participants’ beliefs about a particular state (as opposed to all states). We added these salience cues to the Mouselab variation of our paradigm and ran treatments for all 5-state information environments listed in Table C.1.

Representativeness versus Alternative Cues. We first study how attention is directed when both the representativeness and alternative salience cues are present. We ran a variation of *Limited Attention* with visual and goal-directed salience cues on the *least* representative state, as this is the state that receives the lowest attention weight when only the representativeness cue is present. If these alternative salience cues are effective in drawing attention, then the least representative state will be overweighed, and we will observe less overreaction relative to *Limited Attention* (see Appendix B.3.2 for the formal theoretical analysis).

Fig. C.5a shows that participants continued to overweigh the most representative state, but either underweighed or only slightly overweighed the least representative state. Thus, representativeness dominated the alternative salience cues in drawing attention to the associated state. As a result, the level of overreaction was similar in magnitude to *Limited Attention*, as shown in Fig. C.5b. This evidence provides support for representativeness as an important salience cue in our setting.

Alternative Cues without Representativeness. We next tested whether these alternative salience cues drew attention when the representativeness cue was not present. We ran a variation of *Representativeness Suppressed* with visual and goal-directed salience cues on the *most* representative state. If these alternative salience cues are effective at drawing attention when not competing with the representativeness cue, then we will observe attention directed towards the most representative state (as opposed to randomly, as in the baseline *Representativeness Suppressed*). This will lead to overweighing of the most representative state and more overreaction relative to *Representativeness Suppressed* (see Appendix B.3.2 for the formal theoretical analysis).

Absent the representativeness salience cue, we found that visual and goal-directed salience cues were effective in drawing attention to the associated state. Most participants clicked on the state associated with the salience cues first, as shown in Fig. C.3, and this state was overweighed, as shown in Fig. C.4a. In line with this finding, overreaction arose; in fact, the level of overreaction was similar in magnitude to *Limited Attention* where the representativeness cue was present, as shown in Fig. C.4b. This contrasts with *Representativeness Suppressed*, where underreaction emerged when there was no salience cue associated with the most representative state. Thus, while representativeness dominated other salience cues when present, other salience cues are also effective in drawing attention when this cue is absent.

5.4 Non-Good News Signal Structures

We now explore the predictions of our model beyond the good news setting. Specifically, we test whether representativeness continues to draw attention when a mod-

erate state—rather than an extreme one—is most representative.⁵⁰ As in the good news case, the belief-updating model from Section 2 predicts that the most representative state will be overweighed. But whether this overweighing translates to over- or underreaction depends on which state is most representative and other details of the information environment.

To test these predictions, we ran a variation of our experiment with three states $\Omega = \{\omega_1, \omega_2, \omega_3\}$ and three signal realizations $\mathcal{S} = \{s_1, s_2, s_3\}$ where each state was most representative of one of the signal realizations—specifically, ω_i was most representative of s_i for each i . The rest of the design was the same as in the Baseline paradigm described in Section 3. See Appendix C.5.4 for details.

[Fig. 9a](#) shows that, relative to the objective posterior, participants overweighed the most representative state following each signal realization and underweighed the other states (see [Table C.14](#) for the regression analysis). Thus, representativeness continues to draw attention even in this alternative signal structure. As in good news environments, overreaction is still predicted and observed following the low or high signal ($r = 0.38, p < .01$).⁵¹ However, participants *underreacted* to the middle signal ($r = -0.30, p < .01$). [Fig. 9b](#) illustrates these results.

Together with *Representativeness Suppressed*, these findings show that underreaction is both predicted and observed in some complex environments. This can help rationalize empirical findings of underreaction beyond the binary state settings considered in the lab—for example, to inflation and interest rate news ([Kučinskas and Peters 2022; Bordalo et al. 2020](#)).

6 Model Fit and Heterogeneity Analysis

In this section, we first examine the predictions of our model for the full belief distribution and compare them to the empirical results. We then explore model performance—in particular, how well the model is able to explain the data (completeness) and our model’s falsifiability relative to alternative models (restrictiveness). Finally, we study heterogeneity in cognitive constraints across participants.

6.1 Full Belief Distribution Analysis

To assess model fit, we first derive and test additional predictions of our two-stage model with respect to the full belief distribution. This analysis reinforces the takeaway from previous sections that the interaction between the two cognitive constraints generates distinct predictions from either constraint on its own. It also addresses potential issues related to our use of the overreaction measure $\beta(s)$ for the

⁵⁰Some important economic indicators do not have a good news structure. For example, changes in inflation often convey non-monotonic information about the state of the economy: moderate inflation may be most indicative of stable economic growth, while high inflation could reflect a supply shock (e.g., supply chain disruptions) and low inflation could signal weak demand or recession.

⁵¹As in the good news case, we set the value of the state equal to the share of red balls. See Appendix C.5.4 for further details.

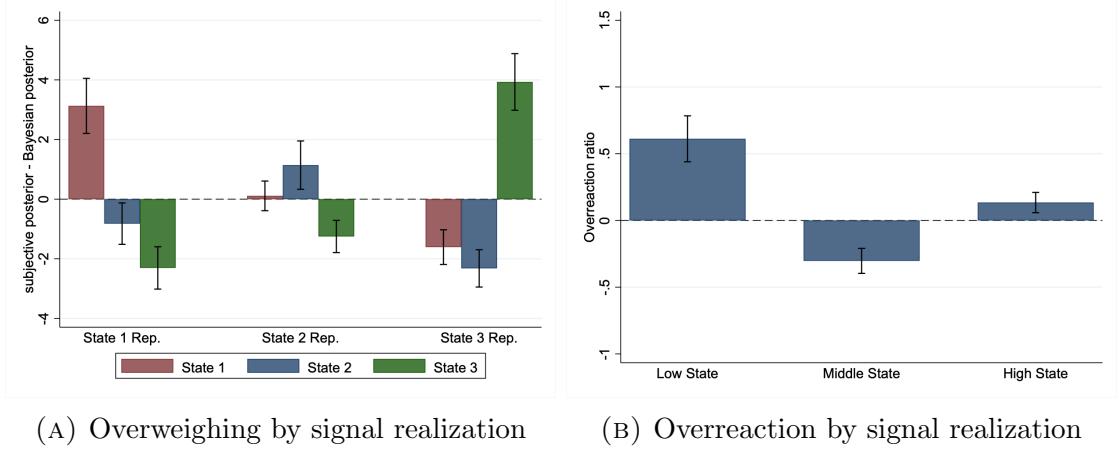


FIGURE 9. Participants overweigh the most representative state following each signal realization, underreact to the middle signal, and overreact to the low and high signal. Each bar aggregates all environments in the 3-signal treatment—by signal realization (state 1, 2, or 3 representative) and state in Panel (A), and by signal realization in Panel (B). Beliefs are measured as a percentage from 0 to 100.

analysis (see Appendix C.2 for a discussion).

Our analysis of the full belief distribution investigates which states are overweighed versus underweighed. An agent *overweights* state ω_i if her subjective posterior assigns a higher probability to it than the objective posterior, with the opposite for underweighing. The definition with respect to a set of states is analogous.

Definition 4. An agent overweights ω_i if $\hat{p}(\omega_i|s) > p_B(\omega_i|s)$ and underweights it if $\hat{p}(\omega_i|s) < p_B(\omega_i|s)$. An agent overweights set of states $\Omega_i \subset \Omega$ if $\hat{p}(\Omega_i|s) > p_B(\Omega_i|s)$ and underweights it if $\hat{p}(\Omega_i|s) < p_B(\Omega_i|s)$.

Note that since the belief distribution must sum to one, if a state is overweighed, then at least one other state must be underweighed, and vice versa.

Our first result establishes that there is a set of parameters θ and λ for which the agent overweights both the most and least representative states. This cognitive *complementarity region* stems from the interaction between the attentional constraint and cognitive imprecision: it cannot arise when only one constraint is present. In contrast, for sufficiently low cognitive imprecision, the agent overweights the most representative state and underweights the least as in the limited-attention-only model, and for sufficiently high cognitive imprecision, the agent displays the opposite belief pattern as in the cognitive-imprecision-only model. To formally state this result, let $\omega_R(s)$ denote the most representative state for signal realization s (e.g., ω_1 for b and ω_N for r) and $\omega_{NR}(s)$ denote the least representative state.

Prediction 5 (Over- and Underweighing Individual States). Consider a symmetric information environment (Ω, p_0) with a uniform prior. For each $\theta > 0$, there exist cutoffs $0 < \bar{\lambda}_1 \leq \bar{\lambda}_2 < 1$, such that:

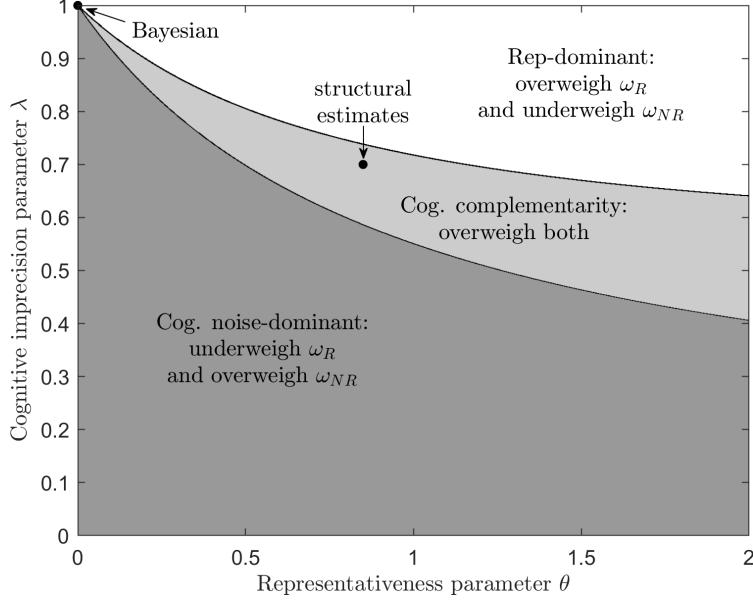


FIGURE 10. Illustration of Prediction 5 for $\Omega = (0.1, 0.3, 0.5, 0.7, 0.9)$

- (i) *Cognitive-imprecision-dominant:* for $\lambda \in [0, \bar{\lambda}_1]$, the agent underweights $\omega_R(s)$ and overweights $\omega_{NR}(s)$ for all $s \in \mathcal{S}$;
- (ii) *Cognitive interaction:* for $\lambda \in (\bar{\lambda}_1, \bar{\lambda}_2)$, the agent overweights $\omega_R(s)$ and $\omega_{NR}(s)$ for all $s \in \mathcal{S}$;
- (iii) *Limited-attention-dominant:* for $\lambda \in (\bar{\lambda}_2, 1]$, the agent overweights $\omega_R(s)$ and underweights $\omega_{NR}(s)$ for all $s \in \mathcal{S}$.

The cognitive interaction region exists, $\bar{\lambda}_1 < \bar{\lambda}_2$, if and only if $|\Omega| > 2$. Moreover, when $\theta > 0$ and $\lambda > 0$ the agent underweights the set of moderate states $\Omega \setminus \{\omega_R(s), \omega_{NR}(s)\}$ for each $s \in \mathcal{S}$.

Fig. 10 illustrates these three regions for an information environment with five states. As shown in the figure, our parameter estimates from Section 3 fall in the cognitive interaction region for this information environment.

The intuition for Prediction 5 is as follows. Salience-channeled attention prompts the agent to overweigh the most representative state and underweigh the least. Cognitive imprecision acts as an opposing force by pulling beliefs towards the uniform cognitive default, causing the agent to underweigh the most representative state (which the objective posterior assigns more mass to than the cognitive default) and overweigh the least (which the objective posterior assigns less mass to than the cognitive default).⁵² Fixing the level of the attentional constraint θ , when cognitive precision λ is sufficiently high or low, one of the two constraints dominates.

However, for intermediate levels of cognitive imprecision, the two constraints interact to generate overweighing of both the most *and* least representative states. This

⁵²This is also the case in Augenblick et al. (2022) and in a more flexible model of cognitive imprecision, as we show in Appendix D.

is because salience-channeled attention pulls mass towards the most representative state from multiple other states, and as a result, directs more mass to the most representative state than away from the least representative state. Hence, a moderate level of cognitive imprecision reverses the attention-driven underweighing of the least representative state but not the attention-driven overweighing of the most representative state. It is worth noting that the cognitive interaction region cannot exist in binary state environments, as it is not possible to simultaneously overweigh the only two states.⁵³ For three or more states, this region always exists. This insight highlights another reason why restricting attention to binary state environments provides an incomplete picture of belief updating.

Additionally, our model predicts that the attentional constraint will lead to an overall underweighing of moderate states. This also stems from salience-channeled attention overweighing the most representative state more than it underweights the least representative state. The excess mass directed to the most representative states is diverted from moderate states. In contrast, in the cognitive-imprecision-only model ($\theta = 0$), moderate states are neither under- nor overweighed on average (see [Prediction 11](#) in [Appendix B.4](#)).⁵⁴

Empirical Findings. Overall, our estimates of θ and λ fall in the cognitive interaction region of [Prediction 5](#) across the majority of information environments we consider: they are in this region for 15 of the 24 environments with 3, 4, or 5 states in the experiment (recall that the region does not exist in 2-state environments). This suggests that the interaction between salience-channeled attention and cognitive imprecision will play a key role in driving belief updating in most of our settings. Our model also predicts when one of the constraints will dominate: the parameter estimates fall in the limited-attention-dominant region for 8 of the environments we consider and in the cognitive-imprecision-dominant region for the remaining one.

We test [Prediction 5](#) by comparing participants' reported posterior beliefs to the objective posterior and the model-predicted subjective posterior at the estimated values of θ and λ . [Fig. 11a](#) plots the difference between participants' average subjective posterior and the objective posterior, aggregated across all 2, 3, 4, and 5-state uniform prior environments in the experiment. [Fig. 11b](#) plots our model's predicted differences between the subjective and objective posteriors across the same information environments. These figures show that the data is remarkably consistent with the predictions of our framework. The same pattern emerges both in predictions and the data: participants overweigh both the most and least representative states—with the most representative state overweighed to the largest extent—and underweigh the set of moderate states.

⁵³If this were the case, the belief distribution would not sum to one.

⁵⁴Predictions for individual moderate states are dependent on the details of the information environment and thus we omit them.

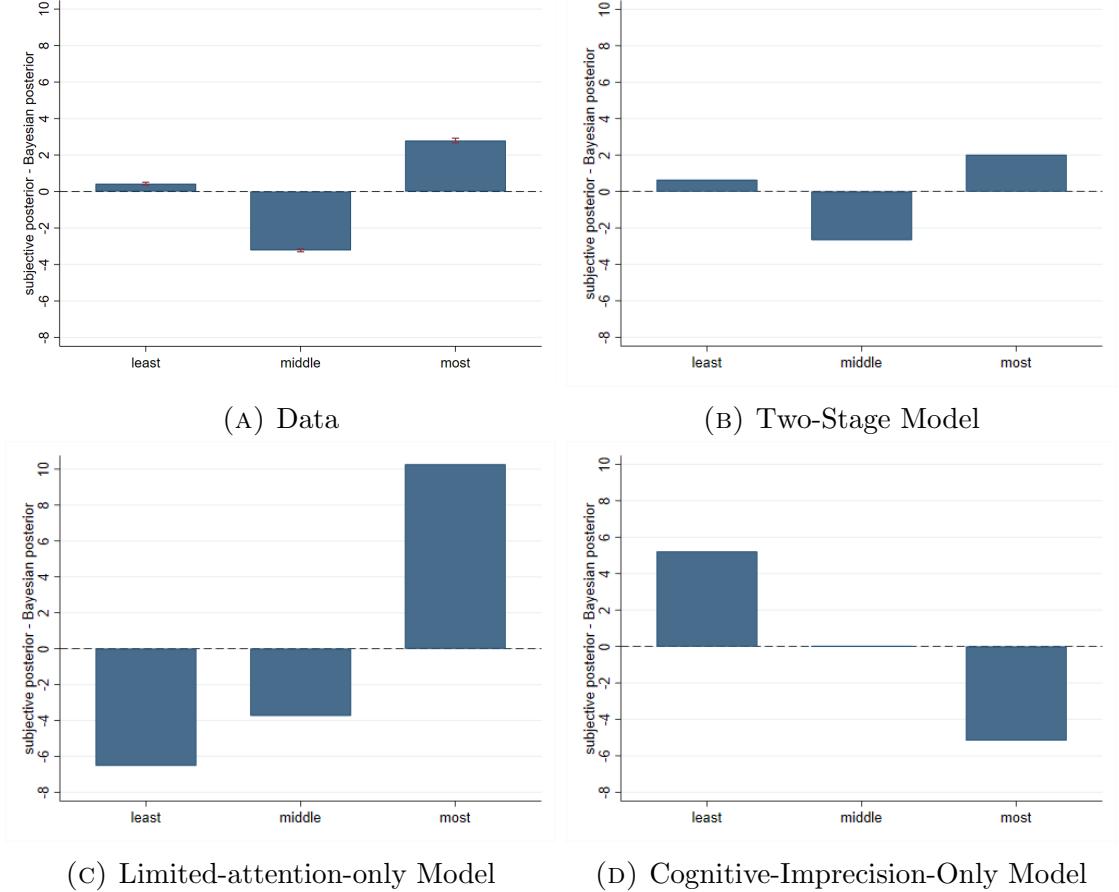


FIGURE 11. Over- and Underweighing by Representativeness of State. Each bar aggregates both signal realizations for all 2, 3, 4, and 5-state uniform prior environments; the Middle State bar aggregates all moderate states. Panels (B)-(D) are weighted to match the share of experimental observations in each environment and based on structural estimates of θ and λ : (B) $\theta = 0.85$, $\lambda = 0.7$; (C) $\theta = 0.85$, $\lambda = 1$; (D) $\theta = 0$, $\lambda = 0.7$. Beliefs are measured as a percentage from 0 to 100.

In contrast, the observed pattern of beliefs is distinct from the predicted patterns of each constraint in isolation. As shown in Figs. 11c and 11d, the limited-attention-only model predicts that the least representative state will be underweighed, while the cognitive-imprecision-only model predicts that the most representative state will be underweighed and the set of moderate states will be correctly weighed.⁵⁵ Neither pattern is consistent with our data. The observed pattern is also inconsistent with the alternative versions of cognitive imprecision explored in Appendix D.

Analytical comparative statics on how properties of the information environment impact the full belief distribution are less tractable than for the level of overreaction. This is why we focus on the latter in Section 4. In Appendix C.6 we empirically explore how over- and underweighing individual states varies with the signal structure.

⁵⁵Prediction 11 in Appendix B.4 formalizes this result for the cognitive-imprecision-only model. The limited-attention-only model is a special case of Prediction 5 at $\theta > 0$ and $\lambda = 1$.

The data exhibits interesting variation across signal diagnosticities that matches the variation in the model-predicted posteriors. Moreover, as in the other analyses presented in this subsection, the results are inconsistent with the predictions of either constraint on its own.

6.2 Model Performance

We next evaluate model performance by computing the completeness and restrictiveness of our two-stage model and comparing these measures to those of the limited-attention-only and cognitive-imprecision-only models. To do so, we follow the methodology developed by [Fudenberg et al. \(2022, 2023\)](#).

6.2.1 Completeness

Completeness is a measure of how much of the explainable variation in data a model can capture relative to a benchmark, which we take to be Bayes' rule. A model is 0% complete if it fits no better than Bayesian updating and 100% complete if it fits as accurately as the best performance achievable by a model. Completeness is distinct from the R -squared statistic typically reported for a regression analysis: it measures how well a model captures regularities in the data, while R -squared captures the overall prediction error of the model, which could stem from either the model missing regularities or the data featuring intrinsic, irreducible noise. Details of how we estimate completeness—including how we estimate the best performance achievable—can be found in [Appendix C.8.1](#).

We first estimate completeness in simple 2-state environments. As shown in [Table 1](#), both the two-stage and cognitive-imprecision-only models achieve essentially 100% completeness: accounting for attentional constraints does not improve 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. It also potentially explains the prominent role of cognitive imprecision in organizing data from primarily binary state laboratory experiments.

While the cognitive-imprecision-only model does well in predicting belief updating in simple environments, the model's explanatory power falls precipitously in more complex environments. Increasing the complexity of the state space to three or more states decreases the completeness of the cognitive-imprecision-only model to 36%. The limited-attention-only model also has little explanatory power in these more complex environments. Yet taken together, the two-stage model with both cognitive constraints achieves very high completeness: it captures 92% of the explainable variation in the data relative to Bayes' rule. As in [Sections 4](#) and [6.1](#), this finding shows that the interaction between the two constraints is key to 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 (0.15)	0.92 (0.05)	0.73 (0.00)	0.91 (0.00)
Cognitive-imprecision-only model	1.00 (0.06)	0.36 (0.02)	0.76 (0.00)	0.97 (0.00)
Channeled-attention-only model	0.00 (0.15)	0.00 (0.04)	1.00 (0.00)	1.00 (0.00)

Notes: Includes all information environments listed in [Table C.1](#) except for the 11-state complexity; includes wrong direction reactions. Restrictiveness is estimated from 1000 simulations.

6.2.2 Restrictiveness

While our two-stage model has high completeness, this good fit may be due to the model’s flexibility: it maybe flexible enough to fit any dataset. To rule out this possibility, we next estimate a measure of our model’s restrictiveness. Restrictiveness is a measure of how well a model can explain an arbitrary “synthetic” dataset relative to a benchmark, which again we take to be Bayes’ rule ([Fudenberg et al. 2023](#)). A model is 0% restrictive if it fits synthetic data perfectly and 100% restrictive if it fits synthetic data no better than Bayes’ rule. Intuitively, the model is overly flexible if it has a good fit on the synthetic data. To compute restrictiveness, we randomly generate synthetic belief data and compare the average prediction losses of the two-stage model and Bayes’ rule on this synthetic dataset. Details of the estimation procedure can be found in [Appendix C.8.2](#).

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 three or more states (0.91). Moreover, it has similar restrictiveness to the cognitive-imprecision-only model (0.76 for simple environments and 0.97 for more complex environments). Thus, the substantially higher explanatory power of the two-stage model relative to the cognitive-imprecision-only model does not come at the expense of a significant increase in flexibility.⁵⁶

While the limited-attention-only model is more restrictive than the two stage model (1.00 in both simple and complex environments), it is starkly incomplete—the model alone adds no explanatory power relative to Bayes’ rule. In contrast, the two-stage model is both highly restrictive and highly complete—it is almost as restrictive as Bayes’ rule while adding significant explanatory power. [Fig. 12](#) visualizes this

⁵⁶In [Appendix D](#) we consider a variation of the cognitive-imprecision-only model with more flexible cognitive imprecision. While this added flexibility achieves higher completeness, it is still much less complete than the two-stage model (0.65 in complex environments) and also slightly less restrictive (0.89 in complex environments).

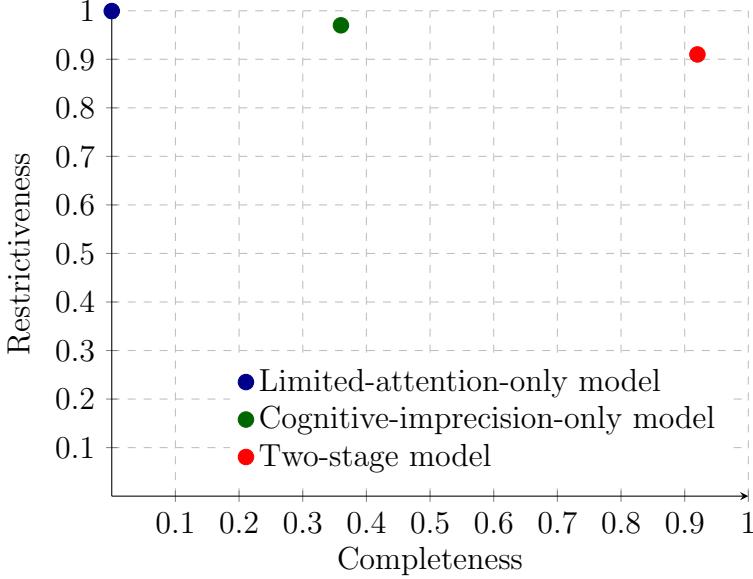


FIGURE 12. Completeness-restrictiveness tradeoff in complex environments ($|\Omega| > 2$)

trade-off between completeness and restrictiveness by plotting these measures for each model. As the figure illustrates, incorporating both cognitive constraints leads to a striking increase in explanatory power in complex environments while only minimally increasing flexibility.

6.3 Cognitive Heterogeneity

To explore individual-level heterogeneity in attention and processing constraints, we estimate the parameters separately for each participant and look at variation in the full belief distribution. 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 signal structure.

Our 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 salience-channeled attention ($\lambda = 1$), and the remaining 16% exhibit neither distortion ($\theta = 0$ and $\lambda = 1$). See Fig. C.9 in Appendix C.7 for a plot of the individual-level parameter estimates.

The estimated values of θ and λ exhibit significant negative correlation, with a correlation coefficient of -0.47 . Therefore, participants who are more prone to salience-driven distortions (higher θ) also tend to exhibit higher levels of cognitive imprecision (lower λ). This suggests a link between attention and processing capacity constraints: individual-level limits on attention coincide with cognitive imprecision, which generates both greater distortion in the representational stage and noisier evaluation in the processing stage.

Finally, we compare the full belief distribution across participants. Fixing an information environment, the distribution of subjective beliefs for a given state is generally unimodal, smooth, and centered around a mean posterior that deviates from the objective posterior. Fig. C.8 in Appendix C.6 plots these distributions for 3-state environments. Consistent with our model, participants overweigh the representative state the most and underweigh the moderate state the most. The belief distributions for other information environments are qualitatively similar; we omit them to save space.

An important implication of cognitive heterogeneity is the emergence of disagreement. Individuals with tighter attention constraints, as captured by θ , will direct more attention to the representative states; those with tighter processing constraints, as captured by λ , will exhibit more cognitive imprecision. This implies that people with different capacity constraints will systematically end up holding different posteriors even when they have the same prior and observe the same information.⁵⁷ Indeed, the cognitive heterogeneity we document generates the patterns of disagreement we observe in the subjective belief data (see Fig. C.8). A similar type of disagreement emerges in [Bordalo et al. \(2025, 2023\)](#), where contextual factors (e.g., how a problem is described) affect salience, leading to differences in which features of the information environment people attend to, and therefore, different beliefs.

7 Conclusion

This paper examines the cognitive drivers of under- and overreaction. A key contribution of our framework is the two-stage model of belief updating with multiple cognitive constraints. We theoretically and empirically show that the interaction between salience-channeled attention and cognitive imprecision plays a crucial role in explaining differences in belief updating across learning environments. While the majority of papers in psychology and behavioral economics have focused on identifying the implications of a single psychological mechanism, observed judgments and choices are likely the product of multiple interacting mechanisms. Our results show that cognitive constraints do not just operate independently but also reinforce one another. This suggests that modeling and testing more unified frameworks is a fruitful area for further exploration in other economically relevant domains.

Our framework differentiates between different forms of complexity—representational and computational—in learning environments. We focus on representational complexity as captured by the number of states and show that it leads individuals to channel attention in a way that impacts the direction and magnitude of reaction to information. Beyond state space complexity, other features of economic settings may contribute to both forms of complexity. Moreover, computational complexity may

⁵⁷The noise introduced by cognitive imprecision also implies that people with the same level of constraint will hold different posteriors, even though on average such individuals have the same posterior.

also impact belief updating in distinct and interesting ways.

Finally, as highlighted in Section 4.6, perceived complexity may differ based on how a person parses the environment into categories in her mental representation. Bordalo et al. (2025) show that categorization is determined by top-down, goal-relevant and memory-based factors, as well as bottom-up, contextual features. Goal-relevant factors related to the importance of discriminating between small differences will influence the coarseness of the mental representation, as will memory-based factors related to previous experience with similar settings. As shown in Appendix E.1, exogenously manipulating such coarseness generates differences in beliefs. Further exploring how the coarseness of the mental representation is influenced by such top-down and bottom-up factors, as well as how this coarseness impacts belief updating, is an important avenue 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.
- AWH, E., A. V. BELOPOLSKY, AND J. THEEUWES (2012): “Top-down versus bottom-up attentional control: A failed theoretical dichotomy,” *Trends in cognitive sciences*, 16, 437–443.
- 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 (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. GORGORAPTS, 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, G. LANZANI, AND A. SHLEIFER (2025): “A Cognitive Theory of Reasoning and Choice.”
- 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 (2013): “Salience and consumer choice,” *Journal of Political Economy*, 121, 803–843.
- (2022a): “Overreaction and Diagnostic Expectations in Macroeconomics,” *Journal of Economic Perspectives*, 36, 223–44.
- (2022b): “Salience,” *Annual Review of Economics*, 14, 521–544.
- BOTVINICK, M. M. AND J. D. COHEN (2014): “The Computational and Neural Basis of Cognitive Control: Charted Territory and New Frontiers,” *Cognitive Science*, 38, 1249–1285.
- 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*.
- EGLY, R., J. DRIVER, AND R. D. RAFAL (1994): “Shifting visual attention between objects and locations: evidence from normal and parietal lesion subjects.” *Journal of Experimental Psychology: General*, 123, 161.
- ENKE, B. AND T. GRAEBER (2023): “Cognitive Uncertainty,” *The Quarterly Journal of Economics*, 138, 2021–2067.
- ENKE, B., T. GRAEBER, AND R. OPREA (2023): “Confidence, Self-selection, and Bias in the Aggregate,” *American Economic Review*, 113, 1933–1966.
- ENKE, B. AND C. SHUBATT (2023): “Quantifying Lottery Choice Complexity,” Tech. rep., National Bureau of Economic Research.
- EPSTEIN, L. G., J. NOOR, A. SANDRONI, ET AL. (2010): “Non-Bayesian Learning,” *The BE Journal of Theoretical Economics*, 10, 1–20.
- FAN, T. Q., Y. LIANG, AND C. PENG (2023): “The Inference-Forecast Gap in Belief Updating,” Available at SSRN 3889069.
- FOUND, A. AND H. J. MÜLLER (1996): “Searching for unknown feature targets on more than one dimension: Investigating a “dimension-weighting” account,” *Perception & Psychophysics*, 58, 88–101.
- 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,” *Review of Economics and Statistics*, 1–50.
- 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.
- GOODHEW, S. C., R. K. LAWRENCE, AND M. EDWARDS (2017): “Testing the generality of the zoom-lens model: Evidence for visual-pathway specific effects of attended-region size on perception,” *Attention, Perception, & Psychophysics*, 79, 1147–1164.
- GOODMAN, A. AND I. PURI (2022): “Bulls and Binaries: Price Anomalies and Behavioral Biases,” Tech. rep., Working paper.
- GRAEBER, T., C. ROTH, AND M. SAMMON (2025): “Coarse Categories in a Complex World,” Tech. rep., University of Bonn and University of Cologne, Germany.
- 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.

- HANNA, R., S. MULLAINATHAN, AND J. SCHWARTZSTEIN (2014): “Learning through noticing: Theory and evidence from a field experiment,” *The Quarterly Journal of Economics*, 129, 1311–1353.
- HARNAD, S. (2003): “Categorical perception,” .
- 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.
- HUANG, L. AND H. PASHLER (2007): “A Boolean map theory of visual attention.” *Psychological review*, 114, 599.
- JOHNSON-LAIRD, P. N. (2010): “Mental models and human reasoning,” *Proceedings of the National Academy of Sciences*, 107, 18243–18250.
- 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.
- 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.
- KUČINSKAS, S. AND F. S. PETERS (2022): “Measuring Under-and Overreaction in Expectation Formation,” *Review of Economics and Statistics*, 1–45.
- 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.
- MOHR SCHLADT, 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.
- NOSOFSKY, R. M., S. E. CLARK, AND H. J. SHIN (1989): “Rules and exemplars in categorization, identification, and recognition.” *Journal of Experimental Psychology: Learning, memory, and cognition*, 15, 282.
- OBERAUER, K. (2019): “Working Memory and Attention—A Conceptual Analysis and Review,” *Journal of Cognition*, 2.
- OBERAUER, K., S. FARRELL, C. JARROLD, AND S. LEWANDOWSKY (2016): “What Limits Working Memory Capacity?” *Psychological Bulletin*, 142, 758.
- O’CRAVEN, K. M., P. E. DOWNING, AND N. KANWISHER (1999): “fMRI evidence for objects as the units of attentional selection,” *Nature*, 401, 584–587.
- OPREA, R. (2020): “What Makes a Rule Complex?” *American Economic Review*, 110, 3913–51.
- (2022): “Simplicity Equivalents,” Working paper.
- PANICELLO, M. F. AND T. J. BUSCHMAN (2021): “Shared Mechanisms Underlie the Control of Working Memory and Attention,” *Nature*, 592, 601–605.
- 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.
- POSNER, M. I. (1980): “Orienting of attention,” *Quarterly journal of experimental psychology*, 32, 3–25.
- 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.
- RABIN, M. (2002): “Inference by Believers in the Law of Small Numbers,” *The Quarterly Journal of Economics*, 117, 775–816.
- 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.
- SAENZ, M., G. BURACAS, AND G. BOYNTON (2002): “Global effects of feature-based attention to direction of motion and color,” *Journal of Vision*, 2, 588–588.
- SALANT, Y. AND J. L. SPENKUCH (2022): “Complexity and choice,” Tech. rep., National Bureau of Economic Research.
- SCHLEY, D., A. FERECHATU, H.-Y. CHAN, AND M. GUNADI (2023): “How categorization shapes the probability weighting function,” *Available at SSRN* 3959751.
- SCHWARTZSTEIN, J. (2014): “Selective Attention and Learning,” *Journal of the European Economic Association*, 12, 1423–1452.
- SHANKS, D. R. AND M. A. GLUCK (1994): “Tests of an adaptive network model for the identification and categorization of continuous-dimension stimuli,” *Connection Science*, 6, 59–89.
- 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.
- WALSH, K. S., D. P. McGOVERN, A. CLARK, AND R. G. O’CONNELL (2020): “Evaluating the neurophysiological evidence for predictive processing as a model of perception,” *Annals of the New York Academy of Sciences*, 1464, 242–268.
- 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 ($N = 100$).⁵⁸ Indeed, in contrast to a discrete state space, participants reported a cognitive default that placed substantially more weight on moderate 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. Our framework provides a complementary explanation for these results as a function of 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 signal 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.⁵⁹ [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

⁵⁸The 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.

⁵⁹Similar 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 signal 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 \(2018\)](#).

noisy signals. Our model generates the same comparative static on diagnosticity, but it stems from both representativeness and cognitive imprecision.

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

A line of work explores belief updating when agents observe multiple signals drawn from the same distribution. [Griffin and Tversky \(1992\)](#) find that people focus too much on the strength of evidence (e.g., sample proportions of each signal realization) and not enough on the weight (e.g., number of signals) in a two-state setting. [Mohrschladt, Baars, and Langer \(2020\)](#) explore the robustness of this strength/weight bias across different information environments, finding that the underinference from larger numbers of signals does not translate to more general settings. [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.

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](#)).⁶⁰

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. \(2022a\)](#) for review). A common approach is to examine the predictability of forecast errors from forecast revisions ([Coibion and Gorod-](#)

⁶⁰A similar reduced form updating rule is found in [Epstein, Noor, Sandroni et al. \(2010\)](#), which considers the implication of underreaction on asymptotic learning.

nichenko 2015).⁶¹ 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.⁶² d’Arienzo (2020) and Wang (2021) find that individual analysts’ forecasts of long-term interest rates exhibit overreaction. Bordalo et al. (2019) find overreaction in the expectations of long-term corporate earnings growth.

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 and Kučinskas and Peters (2022) find underreaction about aggregate inflation shocks. As shown and discussed in Section 5.2 and Section 5.4, this may be due to the specific signal structure and salience cues in this setting. Bouchaud, Krueger, Landier, and Thesmar (2019) document underreaction of analyst forecasts of firms’ short term earnings.⁶³ 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). Finally, other instances of documented underreaction may be due to inattention to the relevant information (DellaVigna and Pollet 2009), which is consistent with agents not attending to the signal in our model (Section 2). At the same time, we acknowledge that factors outside our model that can generate differences in over versus underreaction to information, such as frictions in the spread of information (Barberis et al. 1998) or contextual differences shifting the salience of key features (Bordalo et al. 2023).

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

⁶¹ 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).

⁶²In addition to identifying overreaction in individual forecasts, Bordalo et al. (2020) also document underreaction in consensus forecasts. They explain this underreaction with a model in which forecasters do not respond to other forecasters’ information. The underreaction we identify differs in that it stems from cognitive noise at the individual level rather than a lack of information integration across forecasters.

⁶³Kwon 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.

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

B Proofs of Theoretical Results

B.1 Proof from Section 2

Proof of Claim in Footnote 16. In our context, the representativeness-based discounting weighing function generates a distorted belief

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

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

$$p_R(\omega_i|s) = \frac{\hat{\pi}(s|\omega_i)p_0(\omega_i)}{\sum_{\omega_k \in \Omega} \hat{\pi}(s|\omega_k)p_0(\omega_k)} = \frac{\pi(s|\omega_i)R(\omega_i, s)^\theta p_0(\omega_i)}{\sum_{\omega_k \in \Omega} \pi(s|\omega_k)R(\omega_k, s)^\theta p_0(\omega_k)}$$

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

which are equal. To see the counting a signal $\theta + 1$ times property, note that $\hat{\pi}(s|\omega_i)/\hat{\pi}(s|\omega_j) := (\pi(s|\omega_i)/\pi(s|\omega_j))^{\theta+1}$, 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). \square

Proof of Lemma 1. Given that $\eta Y(s)$ follows the multinomial distribution, the cognitive prior follows the Dirichlet distribution, and the Dirichlet distribution is the conjugate prior of the multinomial distribution, it follows that the cognitive posterior also follows a Dirichlet distribution with N categories and concentration parameters $\eta Y(s) + \nu \bar{p}_d$. The cognitive posterior has mean $\mu(Y(s)) := (\eta Y(s) + \nu \bar{p}_d)/\sum_{\omega_i \in \Omega} (\eta Y(s)_i + \nu \bar{p}_{d,i}) = (\eta Y(s) + \nu \bar{p}_d)/(\eta + \nu) = \lambda Y(s) + (1 - \lambda) \bar{p}_d$, where the first equality follows from the mean of the Dirichlet distribution, the second equality follows from $\sum_{\omega_i \in \Omega} Y(s)_i = 1$ and $\sum_{\omega_i \in \Omega} \bar{p}_{d,i} = 1$, and $\lambda := \eta/(\eta + \nu) \in [0, 1]$. The mean observed posterior, which corresponds to the expectation of $\mu(Y(s))$ taken with respect to the distribution of $Y(s)$ conditional on $p_R(s)$, is equal to $\hat{p}(s) = E[\mu(Y(s))|p_R(s)] = \lambda p_R(s) + (1 - \lambda) \bar{p}_d$. \square

B.2 Proofs from Section 4

Statement and proof of Lemma 2.

Lemma 2. Consider a symmetric information environment (Ω, p_0) . Then the agent never reacts in the wrong direction, $\beta(s) \geq -1$ for all $s \in \mathcal{S}$.

Proof. As shown in Eq. (8), in a symmetric information environment, $\beta(s) = \lambda \beta_R(s) - (1 - \lambda)$, where $\beta_R(s) = (E_R(\omega|s) - E_B(\omega|s))/(E_B(\omega|s) - E_0(\omega)) > 0$. It follows that $\beta(s) \geq -1$. \square

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

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

We can write the objective expected state as

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

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

$$E_B(\omega|r) - E_0(\omega) = \frac{2}{N} \sum_{\omega_i \in \Omega} \omega_i^2 - \frac{1}{2}$$

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

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

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

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

$$\begin{aligned}
\lim_{\theta \rightarrow \infty} \beta_R(r) + 1 &= \lim_{\theta \rightarrow \infty} \frac{E_R(\omega|r) - E_0(\omega)}{E_B(\omega|r) - E_0(\omega)} \\
&= \frac{\omega_N - \frac{1}{2}}{\frac{4}{N} \left[(\omega_{K+1} - \frac{1}{2})^2 + \dots + (\omega_N - \frac{1}{2})^2 \right]}.
\end{aligned} \tag{9}$$

A similar expression to Eq. (9) with respect to Ω' holds for $\beta'(r)$. Since Ω' and Ω have the same most representative state, $\omega'_N = \omega_N$. Since Ω' is more complex than Ω and every state in $\Omega' \setminus \Omega$ is more moderate than every state in Ω ,

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

Therefore, it follows from Eq. (8) that when $\lambda > 0$, there exists $\bar{\theta} > 0$ such that $\beta'(r) > \beta(r)$ whenever $\theta > \bar{\theta}$. The proof is analogous for signal realization b . \square

Proof of Prediction 2. We prove a more general result for any size state space. Let

$$W(\Omega) := \sum_{i \in \{1, N\}} (\omega_i - 0.5)^2 - \sum_{i \notin \{1, N\}} (\omega_i - 0.5)^2. \tag{10}$$

When $W(\Omega) > 0$, the signal is more informative about extreme states and less informative about moderate states. Therefore, the objective posterior attaches higher probability to an extreme state relative to moderate states. This makes the objective movement more sensitive to the values of the extreme states, and decreasing the diagnosticity of the extreme states results in more overreaction for sufficiently large θ .⁶⁴ This leads us to the following generalization of Prediction 2.

⁶⁴When $W(\Omega) < 0$, the objective movement is less sensitive to changes in the extreme states. In this case, decreasing the diagnosticity of extreme states reduces the magnitude of the subjective movement more than the objective movement, leading to less overreaction.

Prediction 6 (Diagnosticity). Consider two symmetric information environments (Ω, p_0) and (Ω', p'_0) with the same complexity, uniform priors, and $W(\Omega) > 0$ and $W(\Omega') > 0$. Suppose Ω' is less diagnostic than Ω , $d'_i \leq d_i$ for all $i = 1, \dots, N$ with at least one inequality strict. There exists a $\bar{\theta} > 0$ such that for $\theta > \bar{\theta}$ and $\lambda > 0$, the agent overreacts more in (Ω', p'_0) than (Ω, p_0) , $\beta'(s) > \beta(s)$ for $s \in \mathcal{S}$.

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

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

The above expression is decreasing in $\omega_{K+1}, \dots, \omega_{N-1}$. Moreover, fixing $\omega_{K+1}, \dots, \omega_{N-1}$, if $W(\Omega) > 0$, then $(\omega_N - \frac{1}{2})^2 > (\omega_{K+1} - \frac{1}{2})^2 + \dots + (\omega_{N-1} - \frac{1}{2})^2$, so the above expression is decreasing in ω_N . The proof is analogous for signal realization b . It follows that the agent reacts more in (Ω', p'_0) than (Ω, p_0) for sufficiently large θ . \square

Prediction 2 follows from noting that $W(\Omega) > 0$ for all symmetric state spaces with five or fewer states.

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 $E_R(\omega'|s)$ and $E_R(\omega|s)$ converge to the highest state in the support, ω_N , when θ diverges to infinity. Thus, to show that $\beta'(s) > \beta(s)$ when θ is sufficiently large, it suffices to show that $0 < (E_B(\omega'|s) - E_0(\omega'))/(E_B(\omega|s) - E_0(\omega)) < 1$.

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

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

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, $E_B(\omega'|r) < E_B(\omega|r)$. The proof is analogous for signal realization b . \square

Proof of Prediction 4. We prove the following stronger version of Prediction 4. Prediction 4 follows as a special case.

Prediction 7 (Asymmetric Prior). *Consider an information environment (Ω, p_0) with a symmetric binary state space and an asymmetric prior. Suppose $\lambda < 1$ and let d denote the diagnosticity of Ω .*

- (i) *The level of overreaction $\beta(s)$ to a disconfirmatory signal realization s is continuously decreasing in d . There exists cutoff $c_1 \in (0.5, 1)$ such that the agent overreacts when $d \in (0.5, c_1)$ and underreacts when $d \in (c_1, 1)$.*
- (ii) *The level of overreaction $\beta(s)$ to a confirmatory signal realization s is continuous and single-peaked in d , first increasing and then possibly decreasing. There exist cutoffs $0.5 < c_2 \leq c_3 \leq c_4 \leq 1$ such that 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 $d \in (c_3, c_4)$. There exists a $\bar{\lambda} > 0$ such that for $\lambda > \bar{\lambda}$ (sufficiently low cognitive imprecision), $c_2 < 1$ and the underreaction region exists. There exists a $\bar{\theta} > 0$ such that for $\theta > \bar{\theta}$ (sufficiently limited attention), $c_3 < c_4$ and the overreaction region exists. If $c_3 < c_4$ and the overreaction region exists, then $c_2 < c_3$ and $c_4 < 1$ and the underreaction regions exist.*

Proof. For convenience, we denote the binary state space as $\Omega = \{1 - d, d\}$ where $d > 1/2$ and the prior as $(1 - p_0, p_0)$. Without loss of generality, we fix $s = r$ (the proof for $s = b$ is symmetric). This signal realization is confirmatory if $p_0 > 1/2$ and disconfirmatory if $p_0 < 1/2$. We have

$$\bar{E}(\omega) = 1/2, \quad (11)$$

$$E_0(\omega) = (1 - p_0)(1 - d) + p_0d, \quad (12)$$

$$E_B(\omega|s) = \frac{(1 - p_0)(1 - d)^2 + p_0d^2}{(1 - p_0)(1 - d) + p_0d}, \quad (13)$$

$$E_R(\omega|s) = \frac{(1 - p_0)(1 - d)^{\theta+2} + p_0d^{\theta+2}}{(1 - p_0)(1 - d)^{\theta+1} + p_0d^{\theta+1}}. \quad (14)$$

The level of overreaction is given by

$$\beta(s) = \frac{\hat{E}(\omega|s) - E_0(\omega)}{E_B(\omega|s) - E_0(\omega)} = \frac{\lambda E_R(\omega|s) + (1 - \lambda)\bar{E}(\omega) - E_0(\omega)}{E_B(\omega|s) - E_0(\omega)}.$$

By Eqs. (11) to (14), the above equation simplifies to the following,

$$\beta(s) = \frac{\lambda \ell(t^{\theta+1}) - (2p_0 - 1)}{\ell(t) - (2p_0 - 1)},$$

where $t := d/(1 - d) > 1$ and $\ell(t) := \frac{p_0 t - (1 - p_0)}{p_0 t + (1 - p_0)}$. Notice that $\ell(t)$ is increasing in t . For convenience, we omit s (since it is fixed at $s = r$) and write the level of overreaction

as a function of t instead. Further simplify,

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

Differentiate,

$$\beta'(t) = \frac{\frac{(2p_0-1)(1-\lambda)}{p_0(1-p_0)} - \frac{2\lambda}{p_0 + \frac{1}{t^{\theta+1}-1}} + \frac{2(1+p_0(t-1))(t-1)t^\theta(\theta+1)\lambda}{(1+p_0(t^{1+\theta}-1))^2}}{2(t-1)^2}. \quad (15)$$

Let $f(t)$ denote the numerator of $\beta'(t)$ in Eq. (15), and differentiate again,

$$f'(t) = -\frac{2(t-1)t^{\theta-1}(1+\theta)\lambda}{(1+p_0(t^{\theta+1}-1))^3} \cdot [-(1-p_0)^2\theta + p_0^2\theta t^{\theta+2} + p_0(1-p_0)(\theta+2)t(t^\theta-1)]. \quad (16)$$

Notice that the term inside the bracket is strictly increasing in t . When t approaches 1, the term inside the bracket approaches $(2p_0 - 1)\theta$.

Part (i). Suppose r is disconfirmatory, $p_0 < 1/2$. Then $(2p_0 - 1)\theta \leq 0$, which implies $f'(t) < 0$ when t is small and $f'(t) \geq 0$ when t is large. Observe that $\lim_{t \rightarrow 1^+} f(1) = \frac{(2p_0-1)(1-\lambda)}{p_0(1-p_0)} < 0$ and $\lim_{t \rightarrow \infty} f(t) = \frac{(2p_0-1)-\lambda}{p_0(1-p_0)} < 0$. Hence, $f(t) < 0$ for all $t \in (1, \infty)$, which immediately implies $\beta(s)$ is strictly decreasing in d .

Note that $\lim_{t \rightarrow 1^+} \beta(t) = +\infty$ and $\lim_{t \rightarrow \infty} \beta(t) = \frac{\lambda-(2p_0-1)}{2(1-p_0)} \in (0, 1)$. Therefore, there exists $c_1 \in (0.5, 1)$ such that the agent overreacts if $d \in (0.5, c_1)$ and underreacts if $d \in (c_1, 1)$.

Part (ii). Suppose r is confirmatory, $p_0 > 1/2$. Then $(2p_0 - 1)\theta \geq 0$, which implies $f'(t) < 0$ for $t > 1$. Observe that $\lim_{t \rightarrow 1^+} f(1) = \frac{(2p_0-1)(1-\lambda)}{p_0(1-p_0)} > 0$ and $\lim_{t \rightarrow \infty} f(t) = \frac{(2p_0-1)-\lambda}{p_0(1-p_0)}$. We discuss two cases below:

(a) Suppose $2p_0 - 1 \geq \lambda$, then $f(t) > 0$ for all $t > 1$, which further implies $\beta'(t) > 0$ for all $t > 1$. In this case, the level of overreaction $\beta(s)$ is strictly increasing in d for all $d \in (1/2, 1)$.

Note that $\lim_{t \rightarrow 1^+} \beta(t) = -\infty$ and $\lim_{t \rightarrow \infty} \beta(t) = \frac{\lambda-(2p_0-1)}{2(1-p_0)} \leq 0$. Therefore, there exists a cutoff $c_2 \in (0.5, 1]$ such that the agent reacts in the wrong direction when $d \in (0.5, c_2)$, and in this case $c_2 = 1$.

(b) Suppose instead $2p_0 - 1 < \lambda$, then $f(t)$ is first strictly positive then negative as t increases from 1 to ∞ . In this case, $\beta(s)$ is first strictly increasing and then strictly decreasing as d increases from $1/2$ to 1.

Note that $\lim_{t \rightarrow 1^+} \beta(t) = -\infty$ and $\lim_{t \rightarrow \infty} \beta(t) = \frac{\lambda-(2p_0-1)}{2(1-p_0)} \in (0, 1)$. In addition,

$$\lim_{t \rightarrow 1} \lim_{\theta \rightarrow \infty} \beta(t) = \lim_{t \rightarrow 1} \frac{\lambda - (2p_0 - 1)}{2p_0(1 - p_0)} \left(p_0 + \frac{1}{t-1} \right) = \infty.$$

On the other hand, if $\theta = 0$, then for any $t \in (1, \infty)$,

$$\beta(t) = \lambda - \frac{(1-\lambda)(2p_0-1)}{2p_0(1-p_0)} \left(p_0 + \frac{1}{t-1} \right) < \frac{\lambda - (2p_0-1)}{2(1-p_0)} < 1.$$

Hence, there exist cutoffs $0.5 < c_2 \leq c_3 \leq c_4 \leq 1$ such that 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 $d \in (c_3, c_4)$. Since $2p_0 - 1 < \lambda$, $c_2 < 1$ so the underreaction region exists. Moreover, for sufficiently high θ , $c_3 < c_4$ and the overreaction region exists; for θ sufficiently close to 0, $c_3 = c_4$ and the overreaction region disappears.

Combining (a) and (b) yield (ii) in [Prediction 7](#). \square

Statement and proof of Lemma 3.

Definition 5 (Posterior belief definition of over- and underreaction).

- (i) *The agent overreacts to signal realization s if $|\hat{p}(\omega_1|s) - p_0(\omega_1)| > |p_B(\omega_1|s) - p_0(\omega_1)|$ and $(\hat{p}(\omega_1|s) - p_0(\omega_1))(p_B(\omega_1|s) - p_0(\omega_1)) > 0$.*
- (ii) *The agent underreacts to signal realization s if $|\hat{p}(\omega_1|s) - p_0(\omega_1)| < |p_B(\omega_1|s) - p_0(\omega_1)|$ and $(\hat{p}(\omega_1|s) - p_0(\omega_1))(p_B(\omega_1|s) - p_0(\omega_1)) > 0$.*
- (iii) *The agent wrong direction reacts to signal realization s if $(\hat{p}(\omega_1|s) - p_0(\omega_1))(p_B(\omega_1|s) - p_0(\omega_1)) < 0$.*

Lemma 3. *When Ω is binary, [Definition 1](#) and [Definition 5](#) are equivalent. The level of overreaction simplifies to $\beta(s) = (\hat{p}(\omega_1|s) - p_B(\omega_1|s))/(p_B(\omega_1|s) - p_0(\omega_1))$ and similarly for ω_2 .*

Proof. Fix any signal realization s . Note that

$$\begin{aligned} \hat{E}(\omega|s) - E_0(\omega) &= \omega_2(\hat{p}(\omega_2|s) - p_0(\omega_2)) + \omega_1(\hat{p}(\omega_1|s) - p_0(\omega_1)) \\ &= \omega_2(p_0(\omega_1) - \hat{p}(\omega_1|s)) + \omega_1(\hat{p}(\omega_1|s) - p_0(\omega_1)) \\ &= (\omega_2 - \omega_1)(p_0(\omega_1) - \hat{p}(\omega_1|s)). \end{aligned}$$

Similarly

$$E_B(\omega|s) - E_0(\omega) = (\omega_2 - \omega_1)(p_0(\omega_1) - p_B(\omega_1|s)).$$

Therefore, comparing the magnitude of $|\hat{E}(\omega|s) - E_0(\omega)|$ and $|E_B(\omega|s) - E_0(\omega)|$ is equivalent to comparing the magnitude of $|\hat{p}(\omega_1|s) - p_0(\omega_1)|$ and $|p_B(\omega_1|s) - p_0(\omega_1)|$. Similarly, $(\hat{E}(\omega|s) - E_0(\omega))(E_B(\omega|s) - E_0(\omega)) > 0$ implies $(\hat{p}(\omega_1|s) - p_0(\omega_1))(p_B(\omega_1|s) - p_0(\omega_1)) > 0$. Hence,

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

The argument is similar for ω_2 . \square

B.3 Model of Alternative Salience Cues from Section 5

Under arbitrary attention weights captured by salience function $R(\omega_i, s)$, the mental representation is

$$\hat{\pi}(s|\omega_i) = \pi(s|\omega_i)R(\omega_i, s)^\theta$$

and

$$p_R(\omega_i|s) = \frac{\pi(s|\omega_i)R(\omega_i, s)^\theta p_0(\omega_i)}{\sum_{\omega_k \in \Omega} \pi(s|\omega_k)R(\omega_k, s)^\theta p_0(\omega_k)}.$$

As before, this generates subjective posterior belief $\hat{p}(s) = \lambda p_R(s) + (1 - \lambda)\bar{p}_d$.

B.3.1 Random Attention

Suppose that no salience cues are present and attention is directed randomly. Each state is equally likely to be attended to first, and all states that are not attended to first receive equal attention. This can be modeled as $R(\omega_i, s) = \alpha > 1$ when ω_i is attended to first and $R(\omega_i, s) = 1$ when ω_i is not attended to first.⁶⁵ Let $p_{R,l}(\omega_i|s)$ and $\hat{p}_l(\omega_i|s)$ denote the cognitively precise posterior and subjective posterior, respectively, when state ω_l is attended to first. This yields

$$p_{R,i}(\omega_i|s) = \frac{\pi(s|\omega_i)p_0(\omega_i)}{\pi(s|\omega_i)p_0(\omega_i) + (\frac{1}{\alpha})^\theta \sum_{k \neq i} \pi(s|\omega_k)p_0(\omega_k)}$$

and $\hat{p}_i(\omega_i|s) = \lambda p_{R,i}(\omega_i|s) + (1 - \lambda)\bar{p}_d(\omega_i)$ when ω_i is attended to first, and

$$p_{R,l}(\omega_i|s) = \frac{\pi(s|\omega_i)p_0(\omega_i)}{\alpha^\theta \pi(s|\omega_l)p_0(\omega_l) + \sum_{k \neq l} \pi(s|\omega_k)p_0(\omega_k)}$$

and $\hat{p}_l(\omega_i|s) = \lambda p_{R,l}(\omega_i|s) + (1 - \lambda)\bar{p}_d(\omega_i)$ when $\omega_l \neq \omega_i$ is attended to first.

We can compute the average posterior for each state ω_i across all possible attention allocations. The average subjective posterior that the state is ω_i is

$$\hat{p}(\omega_i|s) := E[\hat{p}_l(\omega_i|s)] = \frac{1}{N} \sum_{\omega_l \in \Omega} \hat{p}_l(\omega_i|s).$$

The relevant objective posterior comparison is $p_B(\omega_i|s)$ as previously defined. Analogously, the average level of overreaction across attention allocations is:

$$\beta(s) := E[r_l(s)] = \frac{1}{N} \sum_{\omega_l \in \Omega} r_l(s),$$

where $r_l(s) := (\hat{E}_l(\omega|s) - E_B(\omega|s))/(E_B(\omega|s) - E_0(\omega))$ and $\hat{E}_l(\omega|s) := \sum_{\omega_i \in \Omega} \omega_i \hat{p}_l(\omega_i|s)$.

Also relevant is the average posterior for a given level of attention (i.e., click position). Again this average is taken across all attention allocations, but now with

⁶⁵Setting the attention weight on the not-first-attended-to states equal to one is a normalization.

respect to a random variable denoting the state in a given click position. Let ω_F be a random variable that denotes the first-attended-to-state (i.e., $\omega_F = \omega_i$ when ω_i is attended to first), and $\Omega_{NF} = \Omega \setminus \{\omega_F\}$ denote the set of not-first-attended-to states. The average subjective posterior for the first-attended-to state is

$$E[\hat{p}(\omega_F|s)] = \frac{1}{N} \sum_{\omega_i \in \Omega} \hat{p}_i(\omega_i|s)$$

and the average subjective posterior for the set of remaining states is

$$E[\hat{p}(\Omega_{NF}|s)] = \frac{1}{N} \sum_{\omega_i \in \Omega} \sum_{\omega_k \neq \omega_i} \hat{p}_i(\omega_k|s),$$

where $\sum_{\omega_k \neq \omega_i} \hat{p}_i(\omega_k|s)$ is the subjective posterior of the set of not-first-attended-to states $\{\omega_k | k \neq i\}$ when ω_i is attended to first. The relevant objective posterior comparison for the first-attended-to state is the average objective posterior across all first-attended-to states (i.e., all states):

$$E[p_B(\omega_F|s)] = \frac{1}{N} \sum_{\omega_i \in \Omega} p_B(\omega_i|s) = \frac{1}{N},$$

where the second equality follows since the objective posterior sums to one across states. Analogously, the relevant objective posterior comparison for the set of remaining states is

$$E[p_B(\Omega_{NF}|s)] = \frac{1}{N} \sum_{\omega_i \in \Omega} \sum_{\omega_k \neq \omega_i} p_B(\omega_k|s) = \frac{N-1}{N}.$$

Prediction 8. Consider a symmetric information environment (Ω, p_0) with a uniform prior and no salience cues. On average across attention allocations, the agent overweights the state she attends to first and underweights the set of remaining states, $E[\hat{p}(\omega_F|s)] > E[p_B(\omega_F|s)]$ and $E[\hat{p}(\Omega_{NF}|s)] < E[p_B(\Omega_{NF}|s)]$ for $s \in \mathcal{S}$.

Proof. Since $1/\alpha < 1$, we have $p_{R,i}(\omega_i|s) > p_B(\omega_i|s)$, and similarly since $\alpha > 1$, $p_{R,l}(\omega_i|s) < p_B(\omega_i|s)$. Hence,

$$\begin{aligned} E[\hat{p}(\omega_F|s)] &= \frac{1}{N} \sum_{\omega_i \in \Omega} \hat{p}_i(\omega_i|s) = \frac{1}{N} \lambda \sum_{\omega_i \in \Omega} p_{R,i}(\omega_i|s) + \frac{1}{N} (1 - \lambda) \sum_{\omega_i \in \Omega} \bar{p}_d(\omega_i) \\ &= \frac{1}{N} \lambda \sum_{\omega_i \in \Omega} p_{R,i}(\omega_i|s) + \frac{1}{N} (1 - \lambda) \sum_{\omega_i \in \Omega} p_B(\omega_i|s) \\ &> \frac{1}{N} \sum_{\omega_i \in \Omega} p_B(\omega_i|s) = E[p_B(\omega_F|s)]. \end{aligned}$$

The third equality follows from $\sum_{\omega_i \in \Omega} \bar{p}_d(\omega_i) = \sum_{\omega_i \in \Omega} p_B(\omega_i|s) = 1$ and the last inequality follows from $p_{R,i}(\omega_i|s) > p_B(\omega_i|s)$. It is then immediate that $E[\hat{p}(\Omega_{NF}|s)] =$

$$1 - E[\hat{p}(\omega_F|s)] < 1 - E[p_B(\omega_F|s)] = E[p_B(\Omega_{NF}|s)].$$

□

Prediction 9. Consider a symmetric information environment (Ω, p_0) with a uniform prior and no salience cues. For any $\theta > 0$, on average across attention allocations:

- (i) the agent underweights the most representative state and overweights the least representative state, $\hat{p}(\omega_R(s)|s) < p_B(\omega_R(s)|s)$ and $\hat{p}(\omega_{NR}|s) > p_B(\omega_{NR}|s)$ for all $s \in \mathcal{S}$;
- (ii) the agent exhibits underreaction, $\beta(s) \in (-1, 0)$ for $s \in \mathcal{S}$.

Proof. Part (i). Assume $s = r$. Then the average salience-distorted posterior can be simplified to

$$p_R(\omega_i|r) = \omega_i \cdot \frac{1}{N} \left(\frac{\alpha^\theta - 1}{\alpha^\theta \omega_i + \sum_{\omega_l \neq \omega_i} \omega_l} + \sum_{\omega_l \in \Omega} \frac{1}{\omega_l \alpha^\theta + \sum_{\omega_k \neq \omega_l} \omega_k} \right).$$

Hence, $p_R(\omega_i|r)/\omega_i$ decreases as i increases. Since $\sum_{\omega_i \in \Omega} p_R(\omega_i|r) = \frac{2}{N} (\sum_{\omega_i \in \Omega} \omega_i) = 1$, it follows that $p_R(\omega_1|r) > \frac{2}{N} \omega_1$ and $p_R(\omega_N|r) < \frac{2}{N} \omega_N$. Therefore, $\hat{p}(\omega_1|r) = \lambda p_R(\omega_1|r) + (1 - \lambda) \frac{1}{N} > \frac{2}{N} \omega_1 = p_B(\omega_1|r)$ and $\hat{p}(\omega_N|r) = \lambda p_R(\omega_N|r) + (1 - \lambda) \frac{1}{N} < \frac{2}{N} \omega_N = p_B(\omega_N|r)$. The argument is analogous for b .

Part (ii). Again assume $s = r$. Similar reasoning as in Part (i) implies that there exists $m \in \{1, \dots, N\}$ such that $p_R(\omega_i|r) \leq p_B(\omega_i|r)$ when $i \leq m$ and $p_R(\omega_i|r) > p_B(\omega_i|r)$ when $i > m$. Moreover, $p_R(\omega_i|r) < p_R(\omega_l|r)$ for any $i < l$. So both $p_B(\omega_i)$ and $p_R(\omega_i)$ are strictly increasing in i and p_B first order stochastically dominates p_R , leading to $E_0(\omega) < E_R(\omega|r) < E_B(\omega|r)$. Note that

$$\begin{aligned} \beta(r) &= \frac{1}{N} \sum_{\omega_l \in \Omega} \frac{(\hat{E}_l(\omega|r) - E_0(\omega)) - (E_B(\omega|r) - E_0(\omega))}{(E_B(\omega|r) - E_0(\omega))} \\ &= \frac{\frac{1}{N} \left(\sum_{\omega_l \in \Omega} (\hat{E}_l(\omega|r) - E_0(\omega)) \right) - (E_B(\omega|r) - E_0(\omega))}{(E_B(\omega|r) - E_0(\omega))} \\ &= \lambda \frac{\frac{1}{N} \left(\sum_{\omega_l \in \Omega} (E_{R,l}(\omega|r) - E_0(\omega)) \right) - (E_B(\omega|r) - E_0(\omega))}{(E_B(\omega|r) - E_0(\omega))} - (1 - \lambda) \\ &= \lambda \frac{(E_R(\omega|r) - E_0(\omega)) - (E_B(\omega|r) - E_0(\omega))}{(E_B(\omega|r) - E_0(\omega))} - (1 - \lambda). \end{aligned}$$

It follows from $E_0(\omega) < E_R(\omega|r) < E_B(\omega|r)$ that $-1 < \beta(r) < 0$. The argument is analogous for b . □

B.3.2 Visual & Goal-Directed Salience Cues

Suppose a visual or goal-directed salience cue is present on state ω_l . This can be modeled as $R(\omega_l, s) = \alpha > 1$ and $R(\omega_i, s) = 1$ for $i \neq l$, yielding

$$p_R(\omega_l|s) = \frac{\pi(s|\omega_l)p_0(\omega_l)}{\pi(s|\omega_l)p_0(\omega_l) + (\frac{1}{\alpha})^\theta \sum_{k \neq l} \pi(s|\omega_k)p_0(\omega_k)},$$

$$\hat{p}(\omega_l|s) = \lambda p_R(\omega_l|s) + (1 - \lambda)\bar{p}_d(\omega_l),$$

$$p_R(\omega_i|s) = \frac{\pi(s|\omega_i)p_0(\omega_i)}{\alpha^\theta \pi(s|\omega_i)p_0(\omega_i) + \sum_{k \neq i} \pi(s|\omega_k)p_0(\omega_k)}.$$

$$\text{and } \hat{p}(\omega_i|s) = \lambda p_R(\omega_i|s) + (1 - \lambda)\bar{p}_d(\omega_i).$$

Prediction 10. Consider a symmetric information environment (Ω, p_0) with a uniform prior, suppressed representativeness, and an alternative salience cue on state ω_S . Then for any $\theta > 0$, there exists a $\bar{\lambda}(\theta)$ such that for $\lambda \in (\bar{\lambda}(\theta), 1]$:

- (i) the agent overweights the salient state ω_S and underweights each other state $\omega_i \neq \omega_S$ for $s \in \mathcal{S}$;
- (ii) if ω_S is the most representative state, the agent exhibits overreaction, $\beta(s) > 0$ for $s \in \mathcal{S}$;
- (iii) if ω_S is the least representative state, the agent exhibits underreaction or wrong direction reaction, $\beta(s) < 0$ for $s \in \mathcal{S}$.

When representativeness is not suppressed and there is an alternative salience cue on a non-representative state, then the extent to which the agent overweights the representative state versus the other salient state will depend on the relative strength of the two salience cues. If the alternative salience cue is on the least representative state, then it will reduce the extent of overreaction relative to a setting with no alternative salience cue. Whether it generates underreaction or wrong direction reaction again depends on its relative strength.

Proof. Part (i). Since $\alpha > 1$, $p_R(\omega_S|s) > p_B(\omega_S|s)$ and $p_R(\omega_i|s) < p_B(\omega_i|s)$ for all $i \neq S$. By continuity, $\hat{p}(\omega_S|s) > p_B(\omega_S|s)$ and $\hat{p}(\omega_i|s) < p_B(\omega_i|s)$ for all $i \neq S$ when λ is sufficiently close to 1.

Part (ii). Suppose $s = r$ and $i = N$ so that ω_S is the most representative state. Since p_R first-order stochastically dominates p_B , we have $E_R(\omega|s) > E_B(\omega|s)$. By continuity, for λ sufficiently close to 1, $\hat{E}(\omega|s) > E_B(\omega|s)$ and thus the agent exhibits overreaction with $\beta(s) > 0$. The argument is analogous for the case of $s = b$.

Part (iii). Finally, suppose $s = r$ and $i = 1$ so that ω_S is the least representative state. Then in this case p_B first-order stochastically dominates p_R , and thus $E_R(\omega|s) < E_B(\omega|s)$. Since $E_0(\omega) = 1/2 < E_B(\omega|s)$, it follows that the agent exhibits underreaction or wrong direction reaction, $\beta(s) < 0$. The argument is analogous for the case of $s = b$. \square

B.4 Proofs from Section 6

Proof of Prediction 5. We first prove that there exist the three regions described in the prediction and then establish the relationship between the cutoffs $\bar{\lambda}_1$ and $\bar{\lambda}_2$. For the remainder of the proof, we fix a signal realization $s \in \mathcal{S}$ and omit the argument s from ω_R and ω_{NR} for brevity. Consider the distorted posterior derived from the first stage, $p_R(s)$. Note that for all $\omega_i \in \Omega$ such that $\omega_i \neq \omega_R$,

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

Since $\sum_{\omega_i \in \Omega} p_R(\omega_i|s) = \sum_{\omega_i \in \Omega} p_B(\omega_i|s) = 1$, it must be that

$$p_R(\omega_R|s) > p_B(\omega_R|s) > \frac{1}{N}.$$

Since $\hat{p}(\omega_R|s) = \lambda p_R(\omega_R|s) + (1 - \lambda) \frac{1}{N}$, for every $\theta > 0$, there exists threshold $\bar{\lambda}_1 \in (0, 1)$ such that $\hat{p}(\omega_R|s) > p_B(\omega_R|s)$ if $\lambda > \bar{\lambda}_1$ and $\hat{p}(\omega_R|s) < p_B(\omega_R|s)$ if $0 \leq \lambda < \bar{\lambda}_1$.

Analogously, for all $\omega_i \in \Omega$ such that $\omega_i \neq \omega_{NR}$,

$$\frac{p_R(\omega_{NR}|s)}{p_R(\omega_i|s)} = \left(\frac{\pi(s|\omega_{NR})}{\pi(s|\omega_i)} \right)^{\theta+1} < 1.$$

It follows that $p_R(\omega_{NR}|s) < p_B(\omega_{NR}|s) < \frac{1}{N}$. Since $\hat{p}(\omega_{NR}|s) = \lambda p_R(\omega_{NR}|s) + (1 - \lambda) \frac{1}{N}$, for every $\theta > 0$, there exists threshold $\bar{\lambda}_2 \in (0, 1)$ such that $\hat{p}(\omega_{NR}|s) < p_B(\omega_{NR}|s)$ if $\lambda > \bar{\lambda}_2$ and $\hat{p}(\omega_{NR}|s) > p_B(\omega_{NR}|s)$ if $0 \leq \lambda < \bar{\lambda}_2$.

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

$$p_B(\omega_{NR}|s) + p_B(\omega_R|s) = \frac{\pi(s|\omega_R)p_0(\omega_R) + \pi(s|\omega_{NR})p_0(\omega_{NR})}{\sum_{\omega \in \Omega} \pi(s|\omega)p_0(\omega)} = \frac{2}{N} \quad (17)$$

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

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

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

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

Hence, $p_R(\omega_{NR}|s) + p_R(\omega_R|s) > 2/N$. It then follows from $p_R(\omega_R|s) > p_B(\omega_R|s) > \frac{1}{N}$

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

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

By definition of $\bar{\lambda}_1$,

$$\bar{\lambda}_1 p_R(\omega_R|s) + (1 - \bar{\lambda}_1) \frac{1}{N} = p_B(\omega_R|s).$$

Using the previous inequality, we have

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

Since

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

it must be that $\bar{\lambda}_1 < \bar{\lambda}_2$.

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

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

where inequality holds if $\lambda > 0$ and equality holds if $\lambda = 0$. Therefore, for each $\theta > 0$, the agent underweights moderate states $\Omega \setminus \{\omega_R, \omega_{NR}\}$ for $\lambda > 0$ and neither under- nor overweights it for $\lambda = 0$. \square

Statement and proof of Prediction 11. The following prediction characterizes the analogue of Prediction 5 for the cognitive-imprecision-only model (i.e., $\theta = 0$ and $\lambda < 1$).

Prediction 11 (Cognitive-Imprecision-Only Model). *Consider a symmetric information environment (Ω, p_0) with a uniform prior. When $\theta = 0$ and $\lambda < 1$, the agent underweights $\omega_R(s)$, overweights $\omega_{NR}(s)$, and neither under- nor overweights $\Omega \setminus \{\omega_R(s), \omega_{NR}(s)\}$ for all $s \in \mathcal{S}$.*

Proof. As in the proof of Prediction 5, we omit the dependence of ω_R and ω_{NR} on s for brevity. When $\theta = 0$, we have $p_R(\omega_{NR}|s) = p_B(\omega_{NR}|s) < p_0(\omega_{NR}) = 1/N$, $p_R(\omega_R|s) = p_B(\omega_R|s) > p_0(\omega_R) = 1/N$, and $p_R(\omega_{NR}|s) + p_R(\omega_R|s) = p_B(\omega_{NR}|s) + p_B(\omega_R|s) = 2/N$ (see Eq. (17)). It follows that $\hat{p}(\omega_{NR}|s) = \lambda p_R(\omega_{NR}|s) + (1 - \lambda) \frac{1}{N} > p_B(\omega_{NR}|s)$, $\hat{p}(\omega_R|s) = \lambda p_R(\omega_R|s) + (1 - \lambda) \frac{1}{N} < p_B(\omega_R|s)$, and $\hat{p}(\omega_{NR}|s) + \hat{p}(\omega_R|s) = \frac{2}{N} = p_B(\omega_{NR}|s) + p_B(\omega_R|s)$. \square

Statement and proof of Prediction 12.

Prediction 12 (Complexity and Cognitive Interaction). *Consider two symmetric information environments (Ω, p_0) and (Ω', p'_0) with uniform priors and the same*

most representative states, $\omega_1 = \omega'_1$ and $\omega_N = \omega'_N$. Suppose Ω' is more complex than Ω , $|\Omega'| > |\Omega|$, and every state in $\Omega' \setminus \Omega$ is more moderate than every state in Ω . For any $\theta > 0$, there exists $\bar{\lambda}_1, \bar{\lambda}_2 \in (0, 1)$ such that for $\lambda \in [0, \bar{\lambda}_1]$, the agent underreacts to (Ω, p_0) and (Ω', p'_0) and for $\lambda \in (\bar{\lambda}_2, 1]$, the agent overreacts to (Ω, p_0) and (Ω', p'_0) . For sufficiently large θ , when $\lambda \in (\bar{\lambda}_1, \bar{\lambda}_2)$ the agent underreacts to (Ω, p_0) and overreacts to (Ω', p'_0) .

Proof. Fix information environments (Ω, p_0) and (Ω', p'_0) and $\theta > 0$. The levels of overreaction under the cognitively precise posterior, $\beta_R(s)$ and $\beta'_R(s)$, are positive and finite. By Eq. (8), there exist thresholds $\bar{\lambda}_1$ and $\bar{\lambda}_2$ such that $\beta(s), \beta'(s) < 0$ when $\lambda < \bar{\lambda}_1$, and $\beta(s), \beta'(s) > 0$ when $\lambda > \bar{\lambda}_2$. Define $\bar{\lambda}_1$ and $\bar{\lambda}_2$ as the supremum and infimum, respectively, of the values of λ for which this sign change occurs, so that $\max\{\beta(s), \beta'(s)\} = 0$ at $\lambda = \bar{\lambda}_1$ and $\min\{\beta(s), \beta'(s)\} = 0$ at $\lambda = \bar{\lambda}_2$. From Eq. (8), $\beta(s)$ and $\beta'(s)$ are monotonically increasing in λ . It follows that for $\lambda \in (\bar{\lambda}_1, \bar{\lambda}_2)$, one of $\beta(s)$ and $\beta'(s)$ is negative and the other is positive. Moreover, by Prediction 1, for sufficiently large θ , the cognitively precise level of overreaction satisfies $\beta_R(s) < \beta'_R(s)$. Applying Eq. (8) again implies that in the intermediate region $(\bar{\lambda}_1, \bar{\lambda}_2)$, we have $\beta(s) < 0 < \beta'(s)$. This completes the proof. \square

Statement and proof of Prediction 13.

Prediction 13 (Diagnosticity and Cognitive Interaction). *Consider the set Ω_N of symmetric information environments with complexity N and a uniform prior. For each $\theta > 0$, there exist cutoffs $0 < \bar{\lambda}_1 < \bar{\lambda}_2 \leq 1$, with $\lambda_2(\theta) < 1$ iff N is odd, such that:*

- (i) *Cognitive-imprecision-dominant: for $\lambda \in [0, \bar{\lambda}_1]$, the agent underreacts to all information environments in Ω_N .*
- (ii) *Cognitive interaction: for each $\lambda \in (\bar{\lambda}_1, \bar{\lambda}_2)$, there exists a positive measure set of information environments in Ω_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) *Limited-attention-dominant: for $\lambda \in (\bar{\lambda}_2, 1]$, the agent overreacts to all information environments in Ω_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 at which the agent switches from overreaction to underreaction.

Proof. Part (i). Let $\bar{\beta}_R(\Omega_N, \theta)$ denote the supremum of $\beta_R(s)$ over Ω_N given parameter θ . Since $\beta(s) = \lambda\beta_R(s) - (1-\lambda)$, to show the existence of the cognitive-imprecision-dominant region, it suffices to show that $\bar{\beta}_R(\Omega_N, \theta) < \infty$ for any $\theta > 0$. Moreover,

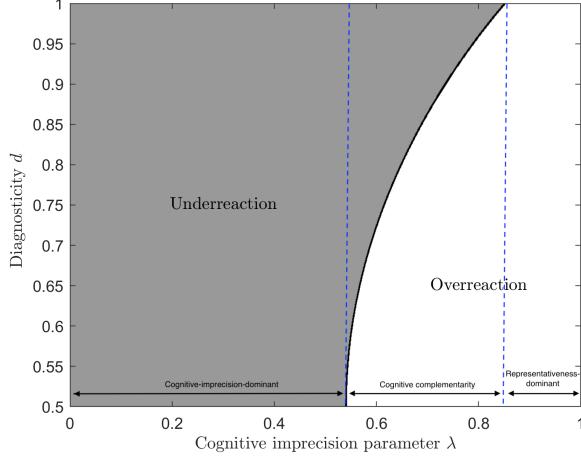


FIGURE B.1. Illustration of Prediction 13 ($\Omega_d = (1 - d, 0.5, d)$ for $d \in (0.5, 1)$, uniform prior, $\theta = 0.85$)

$\bar{\lambda}_1$ is then given by the solution to the following equation, $\lambda \bar{\beta}_R(\boldsymbol{\Omega}_N, \theta) - (1 - \lambda) = 0$. Note that

$$\begin{aligned} \beta_R(s) + 1 &= \frac{E_R(\omega|s) - E_0(\omega)}{E_B(\omega|s) - E_0(\omega)} \\ &= \frac{(\sum_{\omega_i \in \Omega} \omega_i^{\theta+2}) / (\sum_{\omega_i \in \Omega} \omega_i^{\theta+1}) - \frac{1}{2}}{(\sum_{\omega_i \in \Omega} \omega_i^2) / (\sum_{\omega_i \in \Omega} \omega_i) - \frac{1}{2}}, \end{aligned}$$

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

$$\begin{aligned} \beta_R(s) + 1 &= \frac{\sum_{k=K+1}^N (\omega_k - \frac{1}{2}) (w_k^{\theta+1} - (1 - \omega_k)^{\theta+1})}{\frac{4}{N} (\sum_{\omega_i \in \Omega} \omega_i^{\theta+1}) \left[\sum_{k=K+1}^N (\omega_k - \frac{1}{2})^2 \right]} \\ &\leq \frac{1}{4 (\frac{1}{2})^{\theta+1}} \frac{\sum_{k=K+1}^N (\omega_k - \frac{1}{2}) (w_k^{\theta+1} - (1 - \omega_k)^{\theta+1})}{\sum_{k=K+1}^N (\omega_k - \frac{1}{2})^2}. \end{aligned}$$

It suffices to show that

$$h(\omega_k) := \frac{(\omega_k - \frac{1}{2}) (w_k^{\theta+1} - (1 - \omega_k)^{\theta+1})}{(\omega_k - \frac{1}{2})^2}$$

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,

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

Part (ii). From Part (i), we know that for any $\theta > 0$ and $\lambda > \bar{\lambda}_1$, there exists a set

of signal structures where $\beta(s)$ is strictly positive and the agent overreacts. This set has a positive measure because $\beta(s)$ is a continuous function of the signal structure.

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

Suppose N is even and $N = 2K$. Then letting $\omega_{K+1}, \dots, \omega_N$ converge to 1 from below, we have $\lim_{\omega_{K+1}, \dots, \omega_N \rightarrow 1} \beta_R(s) + 1 = 1$. Hence in this case $\underline{\beta}_R(\Omega_N, \theta) = 0$. Moreover, notice that in symmetric environments, the minimum diagnosticity $\min_{\omega_i \in \Omega} d_i > 1/2$ only if N is even. Therefore, it follows from $\lim_{\omega_{K+1}, \dots, \omega_N \rightarrow 1} \beta_R(s) + 1 = 1$ and $\beta(s) = \lambda\beta_R(s) - (1 - \lambda)$ that for each $\theta > 0$ and $\lambda < 1$, there exists $c_1 \in (1/2, 1)$ such that $\beta(s) < 0$ in all environments with $\min_{\omega_i \in \Omega} d_i > 1/2$.

Suppose N is odd and $N = 2K + 1$, then

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

Fix any $\underline{\omega} \in (1/2, 1)$, then when $\max_{\omega_i \in \Omega} \omega_i = \omega_N \geq \underline{\omega}$,

$$\beta_R(s) + 1 = \xi \frac{\sum_{k=K+2}^N (\omega_k - \frac{1}{2}) (w_k^{\theta+1} - (1 - \omega_k)^{\theta+1})}{\frac{4}{N-1} \left(\sum_{i \neq K+1} \omega_i^{\theta+1} \right) \left[\sum_{k=K+2}^N (\omega_k - \frac{1}{2})^2 \right]} > \xi,$$

where $\xi := \frac{N}{N-1} \frac{\sum_{i \neq K+1} \omega_i^{\theta+1}}{(1/2)^{\theta+1} + \sum_{i \neq K+1} \omega_i^{\theta+1}} > 1$ since $\omega_i > 1/2, \forall i > K + 1$ and the strict inequality above follows from the fact that the term multiplied by ξ is equal to the value of $\tilde{\beta}_R(s) + 1$ for an alternative state space $\Omega \setminus \{\omega_{K+1}\}$ and it is strictly larger than 1 when $\theta > 0$. Furthermore, ξ is larger than and bounded away from 1 since $\omega_N \geq \omega > 1/2$. Therefore, $\beta_R(s) + 1$ is bounded below away from 1.

Meanwhile, when $\max_{\omega_i \in \Omega} \omega_i = \omega_N < \underline{\omega}$,

$$\begin{aligned} \beta_R(s) + 1 &= \xi \frac{\sum_{k=K+2}^N (\omega_k - \frac{1}{2}) (w_k^{\theta+1} - (1 - \omega_k)^{\theta+1})}{\frac{4}{N-1} \left(\sum_{i \neq K+1} \omega_i^{\theta+1} \right) \left[\sum_{k=K+2}^N (\omega_k - \frac{1}{2})^2 \right]} \\ &> \frac{\sum_{k=K+2}^N (\omega_k - \frac{1}{2}) (w_k^{\theta+1} - (1 - \omega_k)^{\theta+1})}{\frac{4}{N-1} \left(\sum_{i \neq K+1} \omega_i^{\theta+1} \right) \left[\sum_{k=K+2}^N (\omega_k - \frac{1}{2})^2 \right]} := g(\omega_{K+2}, \dots, \omega_N). \end{aligned}$$

Again, the right-hand side is equal to the value of $\tilde{\beta}_R(s) + 1$ for an alternative state space $\Omega \setminus \{\omega_{K+1}\}$ and it is strictly larger than 1 when $\theta > 0$. In addition, it is a

continuous function over $(1/2, \omega)^K$, and

$$\begin{aligned}
& \lim_{\omega_{K+2}, \dots, \omega_N \rightarrow 1/2} g(\omega_{K+2}, \dots, \omega_N) \\
&= \lim_{\omega_N \rightarrow 1/2} \frac{\left(\omega_N - \frac{1}{2}\right) (w_N^{\theta+1} - (1 - \omega_N)^{\theta+1})}{\frac{4}{N-1} \left((2K-2) \left(\frac{1}{2}\right)^{\theta+1} + (\omega_N^{\theta+1} + (1 - \omega_N)^{\theta+1}) \right) \left(\omega_N - \frac{1}{2}\right)^2} \\
&= \lim_{\omega_N \rightarrow 1/2} \frac{(\theta+1) (w_N^\theta + (1 - \omega_N)^\theta)}{\frac{4}{N-1} \left((2K-2) \left(\frac{1}{2}\right)^{\theta+1} + (\omega_N^{\theta+1} + (1 - \omega_N)^{\theta+1}) \right)} = \theta + 1 > 1,
\end{aligned}$$

where the second equality follows from L'Hopital's rule. So in this case we again have $\beta_R(s)$ bounded strictly away from 0 for any $\theta > 0$. In sum, for any $\theta > 0$ and N odd, we have $\underline{\beta}_R(\Omega_N, \theta) > 0$.

Part (iii). From Part (ii), we know that for any $\theta > 0$ and $\lambda > \bar{\lambda}_2$, $\beta(s) = \lambda\beta_R(s) - (1 - \lambda) > 0$ holds and thus the agent overreacts. \square

Statement and proof of Prediction 14.

Prediction 14 (Prior Concentration and Cognitive Interaction). *Consider the set Ω of symmetric information environments with state space Ω . For each $\theta > 0$, there exist cutoffs $0 < \bar{\lambda}_1 < \bar{\lambda}_2 < 1$ such that:*

- (i) *Cognitive-imprecision-dominant: for $\lambda \in [0, \bar{\lambda}_1]$, the agent underreacts to all information environments in Ω .*
- (ii) *Cognitive interaction: for each $\lambda \in (\bar{\lambda}_1, \bar{\lambda}_2)$, there exists a positive measure set of information environments $\Omega_O \subset \Omega$ on which the agent overreacts and a positive measure set $\Omega \setminus \Omega_O$ on which the agent underreacts. The underreaction 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) *Limited-attention-dominant: for $\lambda \in (\bar{\lambda}_2, 1]$, the agent overreacts to all information environments in Ω .*

The following figure illustrates Prediction 14 for the four 3-state spaces we consider in the experiment.

Proof. Since we consider symmetric information environments, Lemma 2 implies that the agent never wrong direction reacts. Thus, we focus on distinguishing between over- and underreaction.

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

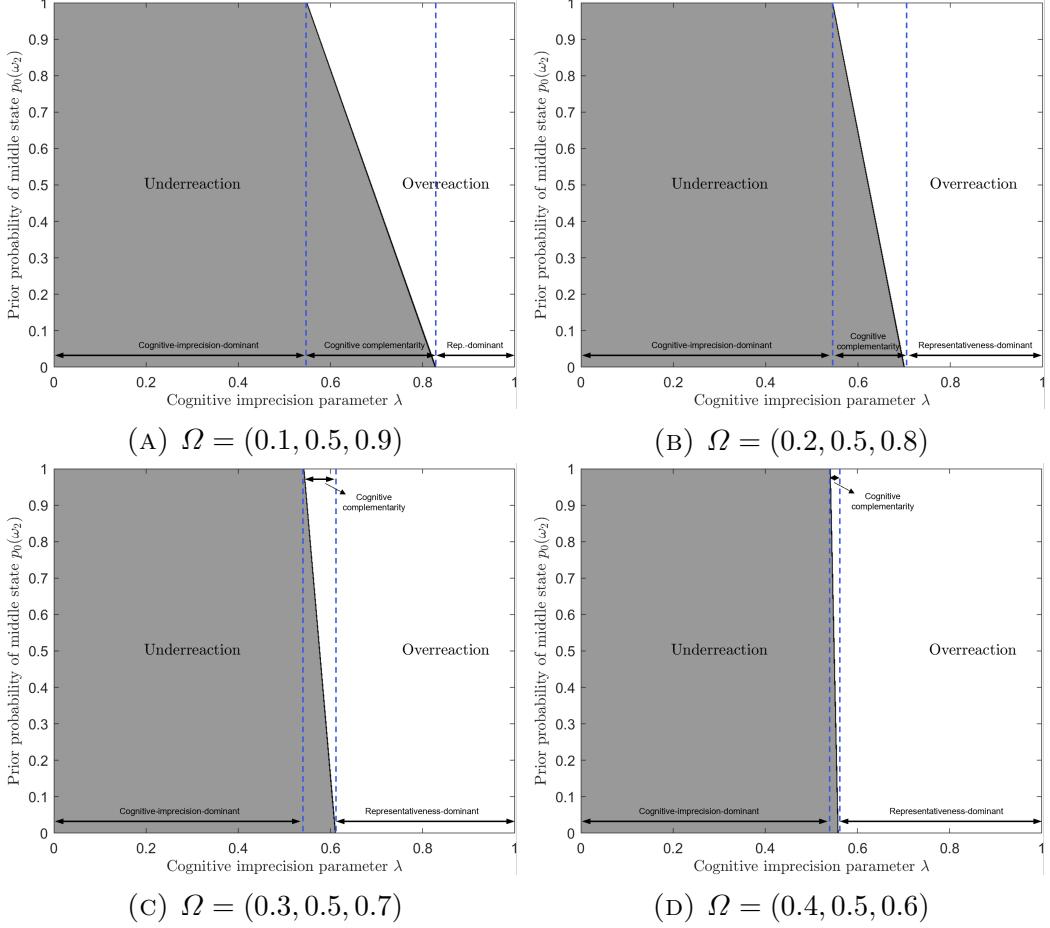


FIGURE B.2. Illustration of [Prediction 14](#) in symmetric 3-state environments ($\theta = 0.85$)

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

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

and the interim posterior is

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

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

$$\begin{aligned} \beta_R(r) + 1 &= \frac{E_R(\omega|r) - E_0(\omega)}{E_B(\omega|r) - E_0(\omega)} \\ &= \frac{\sum_{\omega_i \in \Omega} p_0(\omega_i)\omega_i^{\theta+2}/(\sum_{\omega_i \in \Omega} p_0(\omega_i)\omega_i^{\theta+1}) - \frac{1}{2}}{\sum_{\omega_i \in \Omega} 2p_0(\omega_i)\omega_i^2 - \frac{1}{2}} \\ &= \frac{\sum_{k=K+1}^N p_0(\omega_k)(\omega_k - 1/2)(\omega_k^{\theta+1} - (1 - \omega_k)^{\theta+1})}{4(\sum_{k=K+1}^N p_0(\omega_k)(\omega_k - 1/2)^2)(\sum_{k=K+1}^N p_0(\omega_k)(\omega_k^{\theta+1} + (1 - \omega_k)^{\theta+1}))}. \end{aligned} \quad (18)$$

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

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

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

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

$$\beta_R(r) + 1 \geq \min_{k=K+1, \dots, N} f(\omega_k). \quad (19)$$

This obviously holds if $N = 2$. Suppose $N > 2$ and Eq. (19) does not hold, then for any $i = K + 1, \dots, N$, we have

$$\beta_R(r) + 1 = \frac{\sum_{k=K+1}^N p_0(\omega_k) a(\omega_k)}{4(\sum_{k=K+1}^N p_0(\omega_k) b(\omega_k))(\sum_{k=K+1}^N p_0(\omega_k) c(\omega_k))} < \frac{a(\omega_i)}{2b(\omega_i)c(\omega_i)}. \quad (20)$$

We first show that Eq. (20) cannot hold for all $i = K + 1, \dots, N$ 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

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

However, this is impossible as both $b(\omega_k)$ and $c(\omega_k)$ increase in ω_k when $\omega_k > 1/2$. Therefore, Eq. (19) 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 level of overreaction 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. (19) holds for the original environment (Ω, p_0) . By continuity, Eq. (19) 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, Eq. (19) implies that $\beta_R(r) + 1 \geq f(\omega_N)$. This minimum is attained when p_0 assigns probability 1 to $\{\omega_1, \omega_N\}$. Since $\lambda \in (\bar{\lambda}_1, \bar{\lambda}_2)$, 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 $p_0(\{\omega_1, \omega_N\}) > c_1$. \square

C Additional Experimental Details and Analyses

C.1 Experimental Details

TABLE C.1. Information environments used in experiments

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

⁶⁶Note that f is decreasing in ω_k if and only if $g(x) := \frac{(x-1/2)((1-x)^{\theta+1}+x^{\theta+1})}{x^{\theta+1}-(1-x)^{\theta+1}}$ 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 $h(x) := x^{\theta+1}(x^{\theta+1} - (\theta+1)(1-x)^\theta)$. Since $h(x)$ is increasing in x , it follows that $g'(x) > 0$ for $x > 1/2$.

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 $p_0(\omega_1) \in \{0.3, 0.7\}$.

C.2 Discussion of Measurement.

Grether Regressions. Experimental studies on belief updating often measure over- and underreaction by running the so-called *Grether regression* (Grether 1980). Often applied in binary state spaces, $\Omega = \{\omega_1, \omega_2\}$, a Grether regression decomposes the logarithm of the posterior odds ratio into the logarithm of the prior ratio and the logarithm of the signal likelihood ratio,

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

The coefficients c_1 and c_2 capture the extent to which the individual incorporates the prior and the signal, respectively. Under Bayesian updating, both coefficients equal 1. Values of $c_1 < 1$ indicate base-rate neglect, while $c_2 < 1$ indicates underinference from the signal.

In a binary state space, this decomposition is sufficient because the posterior can be fully characterized by a single likelihood ratio. In multi-state settings ($|\Omega| > 2$), however, multiple likelihood ratios are needed to describe beliefs across all pairs of states. A natural extension of the Grether regression takes the form:

$$\log \frac{\hat{p}(\omega_i|s)}{\hat{p}(\omega_k|s)} = \tilde{c}_1 \log \frac{p_0(\omega_i)}{p_0(\omega_k)} + \tilde{c}_2 \log \frac{\pi(s|\omega_i)}{\pi(s|\omega_k)}, \quad (21)$$

where $i, k \in \{1, \dots, N\}$ and $i > k$.

While intuitive, this formulation imposes a strong assumption: it treats the distortion of prior odds and signal likelihood ratios as uniform across state pairs. Specifically, it implies that the agent either overreacts in all state pair comparisons ($\tilde{c}_2 > 1$) or underreacts in all of them ($\tilde{c}_2 < 1$). However, our theory and data suggest that distortions are state-specific and often non-monotone (see Section 6.1). For instance, when $|\Omega| = 3$ and the prior is uniform, participants often overweight the extreme states ω_1 and ω_3 , while underweighing the moderate state ω_2 , due to the interaction between salience-channeled attention and cognitive imprecision. As a result, participants may simultaneously overreact to $\log \frac{\pi(s|\omega_3)}{\pi(s|\omega_2)}$ and underreact to $\log \frac{\pi(s|\omega_2)}{\pi(s|\omega_1)}$.

To test this, we estimate Grether regressions in three-state environments and report the results in Table C.2. Let ω_3 denote the most representative state following each signal realization, ω_1 denote the least representative state, and ω_2 denote the middle representative state. Column (1) pools over pairs with the least representative state ω_1 as the reference (i.e., pooling (ω_3, ω_1) and (ω_2, ω_1)). We also run separate regressions for each of the three pairs of states: (ω_3, ω_2) , (ω_3, ω_1) , and (ω_2, ω_1) . In

Column (2), the log prior ratio is omitted because ω_1 and ω_3 always have equal prior probabilities in all information environments we considered.

TABLE C.2. Grether regressions for information environments with three states

	Log Posterior Ratio			
	(1) All Three States	(2) Pair (ω_3, ω_2)	(3) Pair (ω_3, ω_1)	(4) Pair (ω_2, ω_1)
Log Prior Ratio	0.2907*** (0.0251)		0.3114*** (0.0291)	0.2546*** (0.0265)
Log Signal Likelihood ratio	0.9714*** (0.0108)	1.0232*** (0.0114)	1.3617*** (0.0351)	0.8612*** (0.0132)
<i>N</i>	7,938	3,972	4,004	3,966
adj. R^2	0.754	0.809	0.532	0.640

Notes: Column (1) pools log posterior ratios where ω_1 serves as the comparison state—specifically, the pairs (ω_3, ω_1) and (ω_2, ω_1) . Columns (2) to (4) report separate regressions for each state pair as indicated. Includes all signal structures with 3 states listed in [Table C.1](#); excludes wrong direction reactions. Standard errors clustered at the individual level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

As shown in Table C.2, the estimates of c_1 are relatively stable across specifications, indicating strong base-rate neglect. Strikingly, however, the level of overreaction parameter \tilde{c}_2 varies *widely* depending on the state pair. When pooling all pairs (Column (1)), we find $\tilde{c}_2 \approx 0.97$ which is close to Bayesian. Yet splitting this analysis into each state pair (keeping ω_1 as the reference), we estimate $\tilde{c}_2 = 1.36$ for (ω_3, ω_1) , indicating overweighing of the most representative state ω_3 relative to the least representative state ω_1 , and $\tilde{c}_2 = 0.86$ for (ω_2, ω_1) , indicating underweighing of the middle representative state ω_2 relative to the least representative state ω_1 . Thus the finding that \tilde{c}_2 averaged across both state pairs is close to Bayesian masks important heterogeneity in how individual states are underweighed or overweighed—and potentially leads to the mistaken conclusion that there is no under- or overreaction. Finally, we estimate $\tilde{c}_2 = 1.02$ for (ω_3, ω_2) , indicating relatively accurate weighing of the most representative state ω_3 relative to the middle representative state ω_2 . This pattern is consistent with our state-by-state predictions (Prediction 5).

Importantly, the three estimates of \tilde{c}_2 across the three state pairs are distinct—indicating that the signal distortion varies across state pairs. This observation highlights the limitations of estimating a single parameter \tilde{c}_2 averaged across state pairs in multi-state settings, e.g., a regression specification that imposes a uniform distortion across state pairs. This limitation motivates our use of non-parametric measures of over- and underreaction in the main analysis, which more flexibly capture overall belief distortions in multi-state environments.

Changing Objective Posteriors. One potential concern with using the level of overreaction measure $\beta(s)$ is that changes in the information environment also change the objective posterior. Since the measure of overreaction used in $\beta(s)$ is defined relative to the objective posterior, 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).⁶⁷ We address this concern in several ways. First, the analysis of the full belief distribution reported in Section 6.1 is not subject to this issue as it tests the predictions of our model for each state in the information environment; for example, Fig. C.8 shows that beliefs do not follow an information-independent heuristic. Second, Section 5 presents evidence for our framework in a setting that keeps the information environment constant, which rules out mechanisms such as partition

⁶⁷Partition 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.

dependence.

C.3 Measuring the Cognitive Default

To measure the cognitive default \bar{p}_d , 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. 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. 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.”

C.4 Additional Analysis from Section 4

C.4.1 Regression Analyses excluding wrong direction reactions

TABLE C.3. Complexity increases overreaction

	Overreaction Ratio	
	(1)	(2)
4 States	0.276*** (0.0295)	0.371*** (0.0315)
5 States	0.365*** (0.0359)	0.455*** (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*** (0.0219)	0.127*** (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 C.1; excludes wrong direction reactions. Standard errors clustered at the individual level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

TABLE C.4. Overreaction increases in prior concentration

	Overreaction Ratio	
	(1)	(2)
Concentrated Prior	0.213*** (0.0547)	0.213*** (0.0547)
Diffuse Prior	-0.215*** (0.0321)	-0.214*** (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*** (0.0253)	0.603*** (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 C.1; excludes wrong direction reactions. Standard errors clustered at the individual level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

TABLE C.5. Overreaction decreases in signal diagnosticity

	Overreaction Ratio			
	(1) 2 States	(2) 3 States	(3) 4 States	(4) 5 States
$d = 0.7$	0.0450 (0.0483)	-0.218*** (0.0502)	-0.370*** (0.0655)	-0.196** (0.0863)
$d = 0.8$	-0.0268 (0.0498)	-0.421*** (0.0496)	-0.597*** (0.0692)	-0.402*** (0.0864)
$d = 0.9$	-0.0432 (0.0484)	-0.461*** (0.0505)	-0.669*** (0.0725)	-0.558*** (0.0878)
Constant	-0.110** (0.0475)	0.535*** (0.0554)	0.703*** (0.0755)	0.644*** (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 C.1 except for the 11-state complexity; excludes wrong direction reactions. 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 C.6. More overreaction to disconfirmatory realizations

	Overreaction Ratio	
	(1)	(2)
Confirmatory	-0.302*** (0.0255)	-0.208** (0.0807)
Disconfirmatory	0.443*** (0.0474)	1.255*** (0.113)
$d = .7$		0.0450 (0.0483)
$d = .8$		-0.0268 (0.0499)
$d = .9$		-0.0432 (0.0484)
Confirmatory $\times d = .7$		-0.301*** (0.0895)
Confirmatory $\times d = .8$		-0.0817 (0.0844)
Confirmatory $\times d = .9$		0.0262 (0.0837)
Disonfirmatory $\times d = .7$		-0.933*** (0.119)
Disonfirmatory $\times d = .8$		-1.233*** (0.116)
Disonfirmatory $\times d = .9$		-1.353*** (0.119)
Constant	-0.116*** (0.0219)	-0.110** (0.0476)
Observations	2432	2432
Adjusted R^2	0.148	0.304

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

C.4.2 Regression Analyses including wrong direction reactions

TABLE C.7. Complexity increases overreaction

	Overreaction Ratio	
	(1)	(2)
4 States	0.348*** (0.0369)	0.431*** (0.0398)
5 States	0.422*** (0.0393)	0.499*** (0.0419)
$d = .7$		-0.107* (0.0547)
$d = .8$		-0.271*** (0.0537)
$d = .9$		-0.401*** (0.0537)
Constant	-0.331*** (0.0268)	-0.137*** (0.0486)
N	6714	6714
adj. R^2	0.026	0.051

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 C.1; includes wrong direction reactions. Standard errors clustered at the individual level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

TABLE C.8. Overreaction increases in prior concentration

	Overreaction Ratio	
	(1)	(2)
Concentrated Prior	0.155** (0.0606)	0.155** (0.0607)
Diffuse Prior	-0.202*** (0.0361)	-0.202*** (0.0361)
$d = .7$		-0.213*** (0.0402)
$d = .8$		-0.437*** (0.0408)
$d = .9$		-0.468*** (0.0415)
Constant	0.157*** (0.0297)	0.437*** (0.0467)
N	4220	4220
adj. R^2	0.024	0.066

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

TABLE C.9. Overreaction decreases in signal diagnosticity

	Overreaction Ratio			
	(1) 2 States	(2) 3 States	(3) 4 States	(4) 5 States
$d = .7$	0.0936 (0.0713)	-0.154** (0.0650)	-0.392*** (0.0855)	-0.0872 (0.118)
$d = .8$	0.0525 (0.0699)	-0.393*** (0.0624)	-0.566*** (0.0926)	-0.279** (0.108)
$d = .9$	-0.0272 (0.0722)	-0.397*** (0.0624)	-0.656*** (0.0933)	-0.453*** (0.106)
Constant	-0.361*** (0.0572)	0.394*** (0.0635)	0.550*** (0.0979)	0.382*** (0.111)
N	986	1404	2928	2800
adj. R^2	0.001	0.038	0.048	0.027

Notes: Baseline is diagnosticity $d = 0.6$. Includes all uniform prior information environments listed in Table C.1 except for the 11-state complexity; includes wrong direction reactions. Standard errors clustered at the individual level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

TABLE C.10. More overreaction to disconfirmatory realizations

	Overreaction Ratio	
	(1)	(2)
Confirmatory	-0.870*** (0.0533)	-1.548*** (0.137)
Disconfirmatory	0.542*** (0.0516)	1.383*** (0.125)
$d = .7$		0.0936 (0.0713)
$d = .8$		0.0525 (0.0699)
$d = .9$		-0.0272 (0.0722)
Confirmatory $\times d = .7$		0.634*** (0.162)
Confirmatory $\times d = .8$		0.834*** (0.161)
Confirmatory $\times d = .9$		1.135*** (0.152)
Disconfirmatory $\times d = .7$		-0.963*** (0.134)
Disconfirmatory $\times d = .8$		-1.267*** (0.130)
Disconfirmatory $\times d = .9$		-1.358*** (0.138)
Constant	-0.331*** (0.0268)	-0.361*** (0.0572)
N	2961	2961
adj. R^2	0.192	0.273

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

C.4.3 Additional Figures

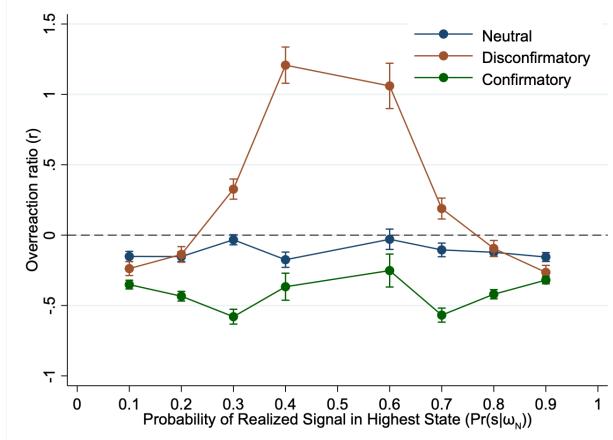


FIGURE C.1. Replication of Fig. 5a excluding wrong direction observations. Each data point corresponds to the given signal type in a 2-state environment with the given diagnosticity.

C.5 Experimental Details and Additional Analyses from Section 5

C.5.1 Measuring and Restricting 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 5.1, the paradigm itself restricts the stock of attention, while first-click is a validated measure of attention. Importantly, *Limited Attention* does not change the information environment relative to the standard Baseline condition.

The first column of Table C.11 shows that restricting attention increased overreaction significantly. The second column of the same table breaks down *Limited Attention* into tasks 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 C.12 presents the structural estimates for *Limited Attention* and *Baseline*. Consistent with our prediction, restricting attention exacerbates the distortion in the mental representation, captured by the higher θ in *Limited Attention*, while not affecting processing capacity, captured by the unchanged λ .

C.5.2 Suppressing the Representativeness Cue

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. We suppressed the representativeness cue by hiding the number of red and blue balls

TABLE C.11. Limited attention increases overreaction

	Overreaction Ratio	
	(1)	(2)
Limited Attention treatment	0.179** (0.0551)	
Click rep. state first		0.381*** (0.0526)
Constant	0.249*** (0.0284)	0.154*** (0.0464)
Observations	4379	1740
Adjusted R^2	0.012	0.038

Notes: Baseline is the Baseline Attention treatment in Column 1 and first-click on a non-representative state in Column 2. Column 1 includes the Baseline Attention and Limited Attention treatments for all 5-state information environments listed in [Table C.1](#); Column 2 includes the Limited Attention treatment for all 5-state information environments listed in [Table C.1](#); excludes wrong direction reactions. Standard errors clustered at the individual level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

 TABLE C.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 and Baseline Attention treatments. Includes all 5-state information environments listed in [Table C.1](#) for the relevant treatment; excludes wrong direction reactions. The 95% confidence intervals are obtained from 300 bootstrap samples.

associated with each state until participants clicked a “reveal” button for that state. Otherwise, this *Representativeness Suppressed* treatment was identical to *Limited Attention*. Because information on the representativeness of each state was initially not available, attention was predicted to be directed as-if randomly in *Representativeness Suppressed*. Our framework predicts that this will generate *underreaction*, despite being the same information environment as *Limited Attention* where marked overreaction was observed.

[Fig. C.2](#) shows that, in contrast to *Limited Attention*, participants’ clicking behavior was not associated with the state’s representativeness. This suggests that attention was directed as-if randomly in *Representativeness Suppressed*. As shown in [Fig. C.4b](#), consistent with our framework, we find that this leads to underreaction across all signal diagnosticities. These results highlight that the emergence of over versus underreaction depends critically on the presence of representativeness as a

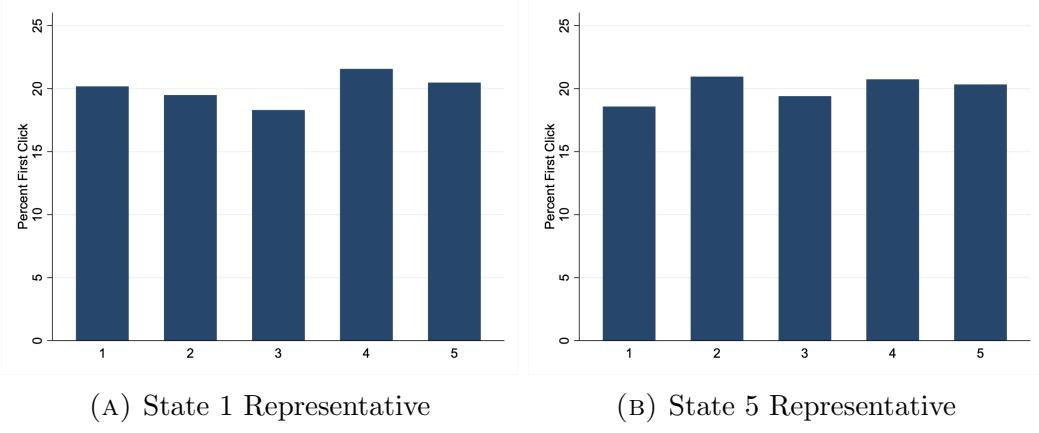


FIGURE C.2. When representativeness is suppressed and there are no salience cues, participants' first clicks are as-if random. Each bar aggregates both signal realizations for all 5-state environments in *Representativeness Suppressed*.

salience cue.

C.5.3 Alternative Salience Cues.

We next explored the impact of low-level (visual) and top-down (goal-directed) salience in drawing attention and driving belief updating. We augmented the *Representativeness Suppressed* design to place visual and goal-directed salience cues on the most representative state: it was highlighted in yellow against a neutral background, similar to the method of [Li and Camerer \(2022\)](#), and participants were told that they would be paid based on their reported beliefs for this state. We predicted that in this *Goal-Directed & Visual (most)* treatment, the alternative salience cues would draw participants' attention to the representative state, even though the representativeness cue was suppressed, and this would lead to the reemergence of overreaction. Figs. C.3 and C.4b shows that this was indeed the case.

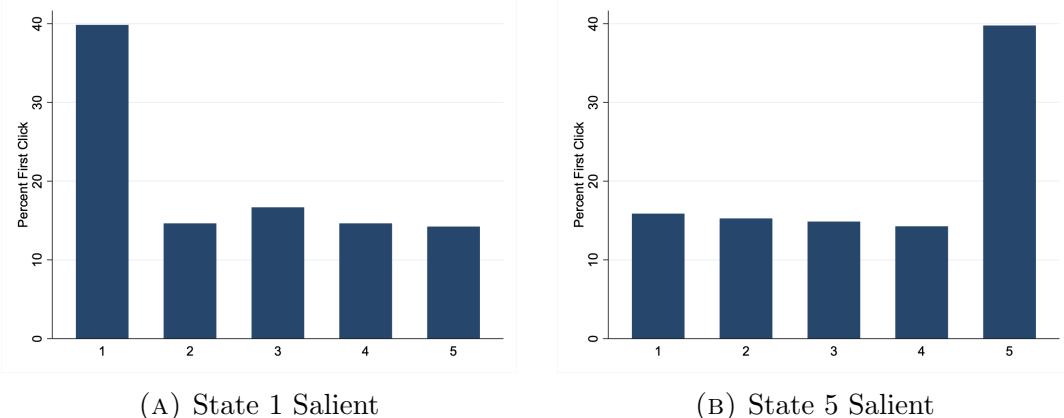


FIGURE C.3. Most participants clicked on the state associated with the alternative salience cue first when the representativeness cue is suppressed. Each bar aggregates both signal realizations for all 5-state environments in *Goal-Directed & Visual (most)*.

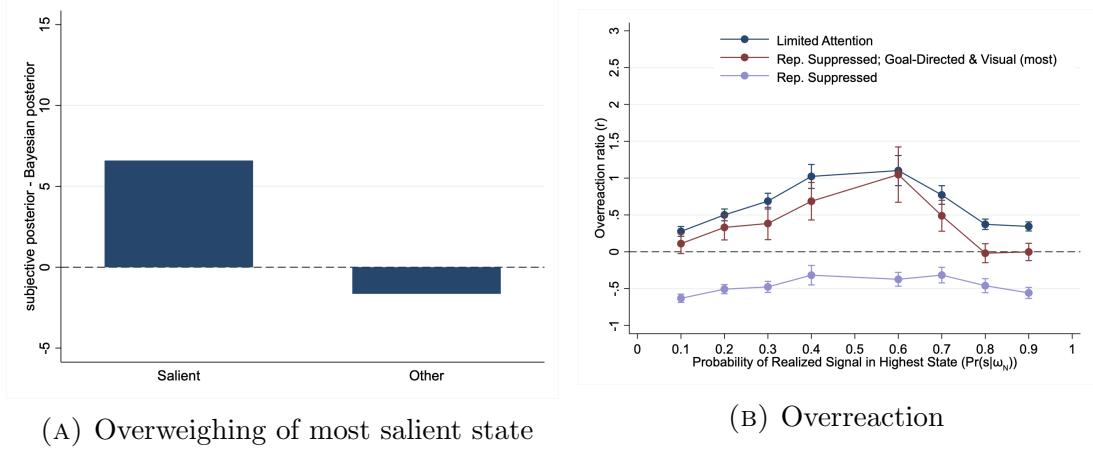


FIGURE C.4. Goal-directed and visual salience cues led to overweighing of the salient state and overreaction. In Panel (A), each bar aggregates both signal realizations for all 5-state environments in *Goal-Directed & Visual (most)*; “Other States” averages across all states besides the most salient; beliefs are measured as a percentage from 0 to 100. In Panel (B), each data point aggregates all 5-state environments in the given treatment by diagnosticity and signal realization.

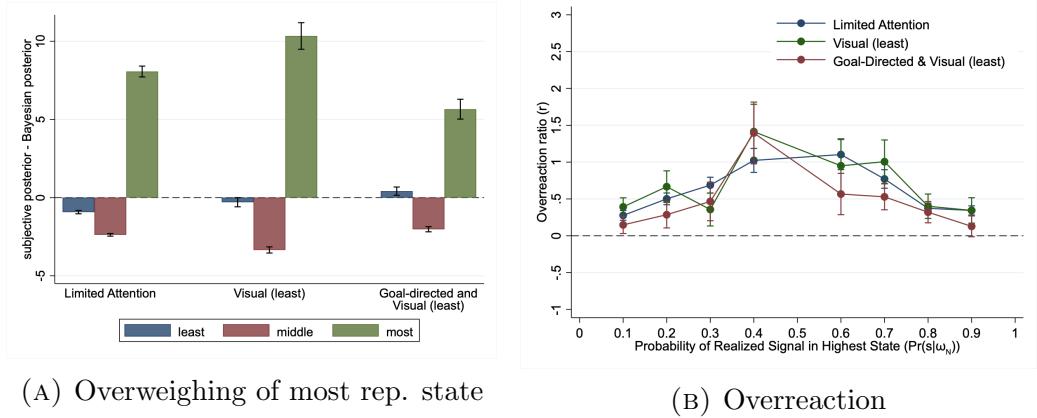


FIGURE C.5. The representativeness cue dominates alternative salience cues in driving overweighing of the most salient state and overreaction. In Panel (A), each bar aggregates both signal realizations for all 5-state environments in the given treatment; the “Middle State” bar averages across all moderate states; beliefs are measured as a percentage from 0 to 100. In Panel (B), each data point aggregates all 5-state environments in the given treatment by diagnosticity and signal realization.

To compare the impact of the representativeness salience cue to visual and top-down salience cues, we ran two treatments where representativeness was not suppressed (i.e., the *Limited Attention* design) and an alternative salience cue was placed on the least representative state: (i) *Visual (least)* had a visual salience cue on the least representative state and (ii) *Goal-Directed & Visual (least)* had both goal-directed and visual salience cues on the least representative state. Note that if goal-directed and visual cues dominate the representativeness cue, then by placing these alternative salience cues on the least representative state we should observe underreaction.

As shown in Figure C.5b, we still observe overreaction in both *Visual (least)* and *Goal-Directed & Visual (least)*. The extent of overreaction is similar to *Limited Attention*. An analysis by individual state, as depicted in Figure C.5a, shows a similar picture: when the representativeness cue is present, it dominates both goal-directed and visual cues in leading to overweighing of the associated state.⁶⁸ Together, these results highlight representativeness as an important salience cue in belief updating.

C.5.4 Non-Good News Signal Structures

We ran a variation of our experiment with three states, $\Omega = \{\text{Bag 1, Bag 2, Bag 3}\}$, and three signal realizations, $\{s_1, s_2, s_3\} = \{\text{red, blue, green}\}$. Each bag contained a combination of these three color balls such that Bag 1 was representative of the red ball, Bag 2 was representative of the blue ball, and Bag 3 was representative of the green ball. For example, Bag 1: (red, blue, green)=(40, 35, 25), Bag 2: (red,

⁶⁸In *Visual (least)* and *Goal-Directed & Visual (least)*, the alternative salience cues were associated with State 5 in Panel A and State 1 in Panel B.

blue, green)=(25, 40, 35), and Bag 3: (red, blue, green)=(35, 25, 40). The rest of the design was the same as in the paradigm in [Section 3](#). [Table C.13](#) outlines the set of information environments that we used.

TABLE C.13. Information environments used in 3-signal experiment

COMPLEXITY $ \Omega $	PRIOR p_0	SIGNAL STRUCTURE
3 states	$\begin{bmatrix} p_0(\omega_1) \\ p_0(\omega_2) \\ p_0(\omega_3) \end{bmatrix} = \begin{bmatrix} 0.33 \\ 0.33 \\ 0.34 \end{bmatrix}$	$\begin{bmatrix} .40 & .35 & .25 \\ .25 & .40 & .35 \\ .35 & .25 & .40 \end{bmatrix}, \begin{bmatrix} .40 & .25 & .35 \\ .35 & .40 & .25 \\ .25 & .35 & .40 \end{bmatrix}$ $\begin{bmatrix} .45 & .35 & .20 \\ .20 & .45 & .35 \\ .35 & .20 & .45 \end{bmatrix}, \begin{bmatrix} .45 & .20 & .35 \\ .35 & .45 & .20 \\ .20 & .35 & .45 \end{bmatrix}$

Notes: For the signal structure, each row denotes a bag (Bag 1, Bag 2, and Bag 3, respectively) and each column denotes a signal realization (red, blue, and green, respectively).

[Table C.14](#) below presents regression results on overweighing of the most representative state. Note that it includes all reactions, as in these information environments, the definition of wrong direction reaction is not clear-cut: it depends on how the numeric values of the states are chosen.

TABLE C.14. Representative state is overweighed in a non-good news setting

	Subjective - Objective Posterior (1)
Most Representative State	3.523*** (0.763)
Constant (Other States)	-0.731*** (0.164)
<i>N</i>	969
adj. R^2	0.033

Notes: Baseline is the two non-representative states. Includes all information environments listed in [C.13](#) and all reactions. Standard errors clustered at the individual level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

For the analysis of overreaction, we reordered the state space for each information environment by the share of red balls and set the numeric value of the state to be this share. The moderate state is Bag 3 with value $\omega_M = .35$ for the first and third signal structures in [Table C.13](#) and Bag 2 with value $\omega_M = .35$ for the second and fourth signal structures. The low state is Bag 2 with value $\omega_L = .25$ for the first

signal structure, Bag 3 with value $\omega_L = .25$ for the second signal structure, Bag 2 with value $\omega_L = .20$ for the third signal structure, and Bag 3 with value $\omega_L = .20$ for the fourth signal structure. The high state is always Bag 1, with value $\omega_H = .40$ for the first and second signal structures and value $\omega_H = .45$ for the third and fourth signal structures. Thus the moderate state is representative when a green ball is drawn in the first and third signal structures and when a blue ball is drawn in the second and fourth signal structures. Similarly the low state is representative when a blue ball is drawn in the first and third signal structures and a green ball is drawn in the second and fourth signal structures. The high state is always representative when a red ball is drawn. From here, we compute the level of overreaction as before.

C.6 Additional Analysis: Full Belief Distribution

We explore how the over- and underweighing of individual states varies with the signal structure. First consider the least representative state. At the estimated values of θ and λ , our model predicts underweighing in environments with low signal diagnosticities and overweighing in environments with high diagnosticities (see Fig. C.6c). When the diagnosticity is low, the objective posterior is close to the cognitive default so cognitive imprecision has little impact and the attentional constraint dominates, leading to underweighing. When the diagnosticity is high, the effect reverses: the objective posterior is close to zero so salience-channeled attention has little room for impact and cognitive imprecision dominates, leading to overweighing. As shown in Fig. C.6a, this pattern is borne out in the experimental data.

In contrast, for the most representative state, our model predicts overweighing regardless of signal diagnosticity, with the most overweighing for environments with intermediate diagnosticity (Fig. C.6d). 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 pattern matches the data (Fig. C.6b). As in the analysis in Section 6.1, the observed pattern of over- and underweighing by diagnosticity is distinct from the prediction of each stage of our model in isolation (Fig. C.7).

C.7 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. (6)). This prediction maps each information environment (Ω, p_0) and signal realization s to a subjective posterior distribution $\hat{p}_{\theta,\lambda}(s; \Omega, p_0) \in \Delta(\Omega)$. 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 tasks. We measure the distance between a reported posterior and a predicted posterior by the Kullback-Leibler (henceforth KL) divergence of the reported posterior

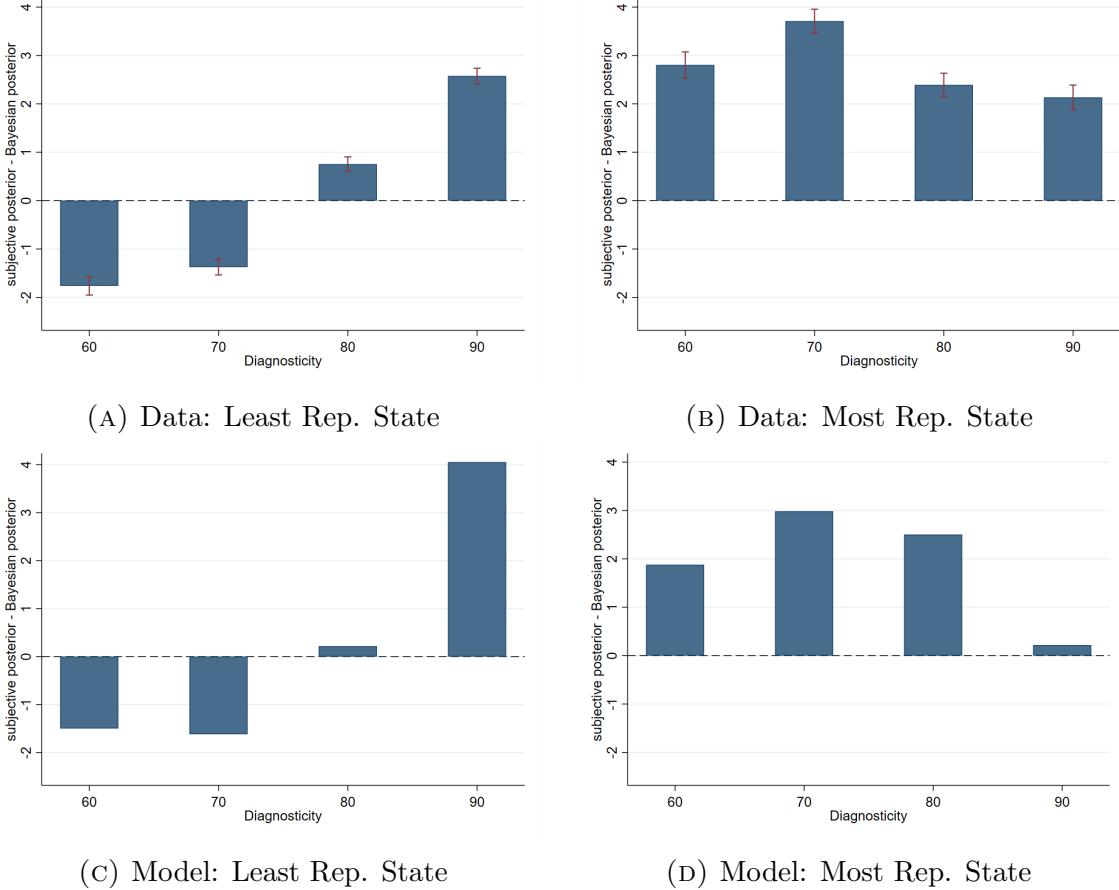


FIGURE C.6. Over- and Underweighing by Diagnosticity. Each bar aggregates both signal realizations for all uniform prior 2, 3, 4, and 5-state environments of diagnosticity d . Panels (C) and (D) are weighted to match the share of experimental observations in each environment and based on structural estimates $\theta = 0.85$ and $\lambda = 0.7$. Beliefs are measured as a percentage from 0 to 100.

from the predicted posterior.⁶⁹ This is a common measure of the statistical distance between two probability distributions. Since the KL divergence is undefined when $\hat{p}_{\theta,\lambda}(\omega_i|s; \Omega, p_0) = 0$, we restrict our analysis to information environments that generate predicted posteriors with full support on Ω . Specifically, we include tasks for all information environments listed in Table C.1 except for the 11-state complexity. The results are summarized in Table C.15.

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.⁷⁰ We chose the KL divergence as our primary measure since it is independent of the values of the states, whereas the quadratic difference places

⁶⁹The KL divergence of reported posterior $\hat{p}(s; \Omega, p_0)$ from predicted posterior $\hat{p}_{\theta,\lambda}(s; \Omega, p_0)$ is given by $\sum_{\omega_i \in \Omega} \hat{p}(\omega_i|s; \Omega, p_0) \log(\hat{p}(\omega_i|s; \Omega, p_0)/\hat{p}_{\theta,\lambda}(\omega_i|s; \Omega, p_0))$.

⁷⁰The quadratic mean difference between reported posterior $\hat{p}(s; \Omega, p_0)$ and predicted posterior $\hat{p}_{\theta,\lambda}(s; \Omega, p_0)$ is given by $(\sum_{\omega_i \in \Omega} \omega_i (\hat{p}(\omega_i|s; \Omega, p_0) - \hat{p}_{\theta,\lambda}(\omega_i|s; \Omega, p_0)))^2$.

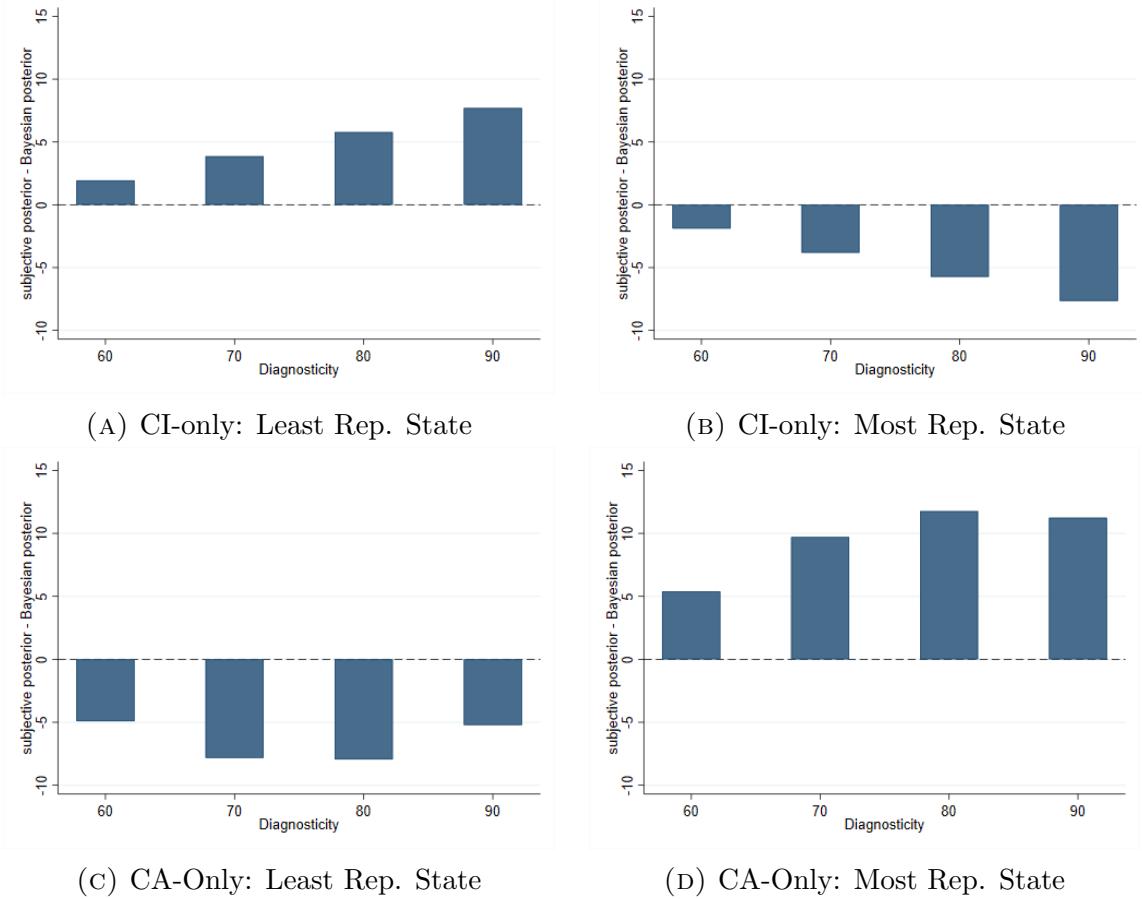


FIGURE C.7. Over- and Underweighing by Diagnosticity for One-Stage Models. Each bar aggregates both signal realizations for all uniform prior 2, 3, 4, and 5-state environments of diagnosticity d , weighted to match the share of experimental observations in each environment. Based on structural estimates of θ and λ : (A) and (B) $\theta = 0, \lambda = 0.7$; (C) and (D) $\theta = 0.85, \lambda = 1$. Beliefs are measured as a percentage from 0 to 100.

a larger weight on higher states.

TABLE C.16. 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 C.1, except for the 11-state complexity; excludes wrong direction reactions. The 95% confidence intervals are obtained from 300 bootstrap samples.

Second, we estimate the parameters for information environments with a symmetric prior. Specifically, we exclude information environments with two states and either a 30/70 or a 70/30 prior. The motivation behind this exercise stems from the model prediction that the agent may react in the wrong direction under an asymmetric prior (Prediction 4). In our main analysis, we drop wrong direction reactions.

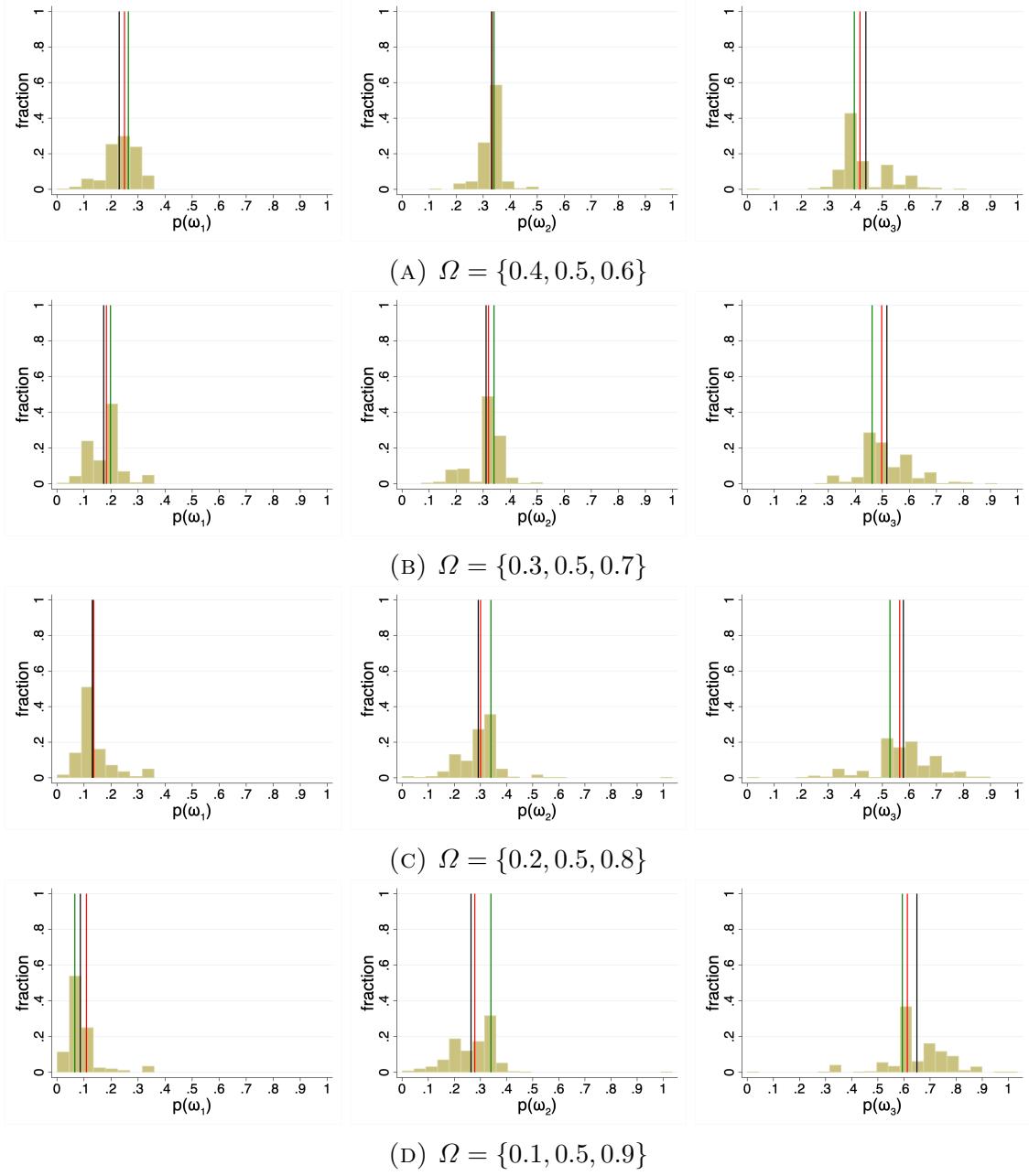


FIGURE C.8. Distribution of reported posteriors in 3-state environments with a uniform prior. Each figure aggregates both signal realizations where, in a slight abuse of notation, ω_3 denotes the most representative state for a given signal realization (i.e., ω_3 for a red ball and ω_1 for a blue ball) and ω_1 denotes the least. Green line=objective posterior, Red line=Model-predicted subjective posterior ($\theta = 0.85$, $\lambda = 0.70$), Black line=Mean reported posterior.

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 C.15. 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 C.1](#), except for the 11-state complexity; excludes wrong direction reactions. The 95% confidence intervals are obtained from 300 bootstrap samples.

TABLE C.17. 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 C.1](#), except for the 11-state complexity; excludes wrong direction reactions. 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 tasks. The results are presented in [Fig. C.9](#). Each point in the figure represents the parameter estimates for one participant.

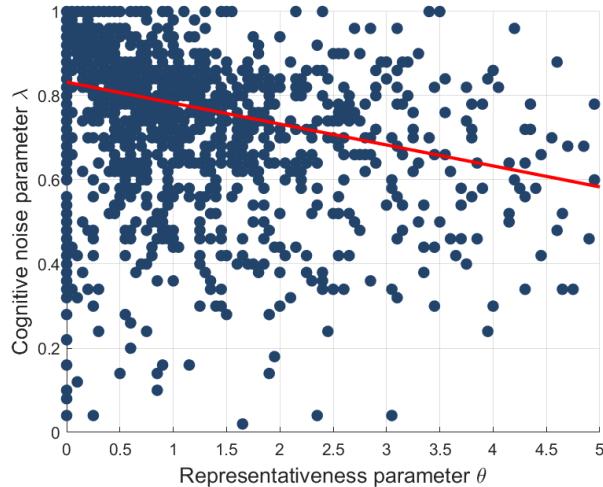


FIGURE C.9. Individual level parameter estimates.

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

C.8 Evaluating Model Performance

C.8.1 Model Completeness

We define model completeness as follows. Similar to the structural estimation in Section 3, 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 under the Bayesian prediction. Let e^M denote the minimum expected loss under model $M \in \{T, P, R\}$, where $M = T$ corresponds to our two-stage model, $M = P$ corresponds to the cognitive-imprecision-only model ($\theta = 0$), $M = R$ corresponds to the limited-attention-only model ($\lambda = 1$), and the minimum is taken with respect to all feasible values of the model parameter(s) within each model. Finally, let e^* denote the expected loss achieved by the “best” model, i.e., the lowest possible loss attainable by any mapping from features of the information environment and signal realization to posteriors. The completeness of model M is given by

$$\kappa^M \equiv \frac{e^B - e^M}{e^B - e^*} \in [0, 1]. \quad (22)$$

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 estimates of the “best” model and the KL-minimizing parameters within each model M . Since prediction loss is evaluated using KL divergence, the best model corresponds to the one that predicts, for each environment-signal pair, the empirical average of the reported posteriors. Estimates of the KL-minimizing parameters within each model M are straightforward to compute from the model and data. As Fudenberg et al. (2023), we use ten-fold cross-validation for these estimates: for each fold, we estimate the “best” model and model parameters for each M using the training set (90% of the data), and then compute e^B , e^M , and e^* using the test set (the remaining 10%). Final estimates are averaged across the ten folds. For this analysis, we do not exclude tasks in which participants react in the wrong direction so as to capture the full extent of model fit to the data.

C.8.2 Model Restrictiveness

Following Fudenberg et al. (2023), we randomly generate 1,000 mappings, where each mapping assigns a posterior distribution over the state space to each information environment from our experimental set (see Table C.1) and each signal realization $s \in \{b, r\}$. We draw mappings uniformly from an admissible set of mappings that satisfy basic directional and monotonicity properties.⁷¹ These properties hold for Bayes’ rule and other common models of belief updating. We impose such properties

⁷¹For 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.

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 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,

$$\rho^M \equiv \frac{d^M}{d^B} \in [0, 1]. \quad (23)$$

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.

C.9 Alternative Design: Reporting Signal Likelihood

This experiment mirrored the baseline design in Section 3 with one modification. As before, participants were presented with the information environment and told that one of several potential bags would be chosen (as specified by the cards in the deck). After a bag was chosen, they observed a ball drawn from this bag. But rather than reporting their belief about the probability of each bag as the main dependent variable, they reported their belief about the probability of drawing a red ball from the chosen bag.

The probability of drawing a red ball from the chosen bag corresponds to the expected state $E(\omega|s)$ in our model. To match the structure of the baseline design, participants were instructed that this belief corresponds to the expected state (bag), which is a function of their full belief distribution (likelihood of each bag being chosen). Participants thus considered the likelihoods of the potential states that would generate a signal, as in the baseline design, before recording their belief as a single number (the probability of drawing a red ball). They were also given questions that allowed them to practice understanding how this probability depended on the full belief distribution. Note that considering each state first decreased the chance that the participant responded to the task as if the red ball was drawn without replacement (Rabin 2002).

We ran this design on the same 2-state, 3-state, and 5-state information environ-

ments as in our baseline design. We calculated the level of overreaction $\beta(s)$ using participants' beliefs about the probability of drawing a red ball for $E(\omega|s)$ in the formula. As can be seen in Fig. C.10 and Table C.18, all of the main results replicate. We observe significant underreaction in 2-state environments and significant overreaction in the more complex 3-state and 5-state environments. We also observe the predicted relationship with signal diagnosticity, with more overreaction to noisier signals.

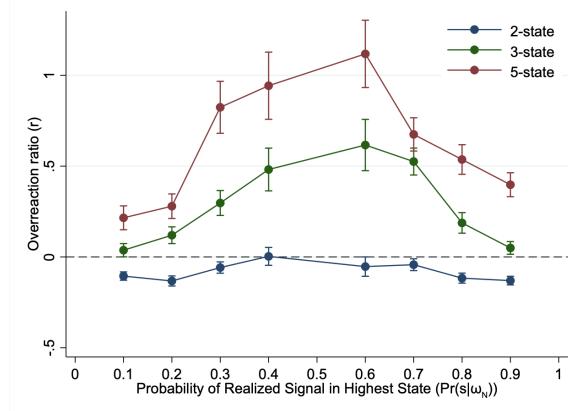


FIGURE C.10. Level of overreaction by complexity and diagnosticity. Each data point aggregates all uniform prior environments of a given complexity by diagnosticity and signal realization.

TABLE C.18. Overreaction increases in complexity

	Overreaction Ratio	
	(1)	(2)
3 States	0.365*** (0.0458)	0.368*** (0.0459)
5 States	0.466*** (0.0518)	0.547*** (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*** (0.0232)	0.150*** (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 C.1; excludes wrong direction reactions. Standard errors clustered at the individual level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

D Alternative models of cognitive imprecision

D.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 $\mathbb{S} = \left| \log \left(\frac{P(s|\omega=\omega_1)}{P(s|\omega=\omega_0)} \right) \right|$, 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 2, 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 4 that the agent updates 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 salience-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.

D.2 Flexible Cognitive Imprecision Model

We now explore a more general version of the processing stage of our model, where we isolate it and allow the cognitive default to deviate from the ignorance prior. This mirrors the core idea in Augenblick et al. (2022), where the agent is biased toward a moderate signal strength.⁷² To capture this structure, we allow the cognitive default to vary with the signal realization—e.g., setting it to the objective belief that

⁷²ALT 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.

would result from observing the same signal under moderate signal strength—so that, like ALT, the agent always interprets the signal in the correct direction. In multi-state environments, where the support of beliefs changes with the state space, we additionally allow the cognitive default to vary with state space complexity. We derive the predictions of this 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 $\bar{p}_d(s, N) \in \Delta(\Omega)$, which may vary according to the signal realization s and the complexity of the state space, $N := |\Omega|$. The agent combines this prior belief with a noisy representation extracted from the information environment, resulting in an average subjective posterior given by

$$\hat{p}(s) := \lambda p_B(s) + (1 - \lambda)\bar{p}_d(s, N). \quad (24)$$

When $\bar{p}_d(s, N) := \bar{p}_d$ for all s 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 $\bar{p}_d(s, N)$ takes the value of a objective posterior derived from a *default information environment* with a symmetric *default state space* $\bar{\Omega}(N) := \{\bar{\omega}_1, \dots, \bar{\omega}_N\}$ and a uniform *default prior* \bar{p}_d . That is, $\bar{p}_d(s, N) = \mathcal{B}(s, \bar{\Omega}(N), \bar{p}_d)$, where \mathcal{B} denotes the Bayesian operator.⁷³ We assume $0 < \bar{\omega}_1 \leq \dots \leq \bar{\omega}_N < 1$, 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 aligns with the direction of the signal realization relative to a uniform prior, $(\bar{E}(\omega|s, N) - 1/2)(E(\omega|s) - 1/2) \geq 0$. For example, suppose the agent’s default state space for binary information environments is $\bar{\Omega}(2) = \{0.3, 0.7\}$. Upon observing r , he compresses his posterior towards a cognitive default with $\bar{p}_d(\omega_1|r, N) = 0.3$; upon observing b , he biases towards $\bar{p}_d(\omega_1|b, N) = 0.7$. Compared to the cognitive-imprecision-only model, this model has six additional parameters for the set of information environments we considered.⁷⁴

Similar to ALT and our two-stage model, the flexible cognitive imprecision model

⁷³For any information environment (Ω, p_0) and signal realization s , let $\mathcal{B}(s, \Omega, p_0)$ represent the implied objective posterior.

⁷⁴This 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 information environments with higher complexities.

predicts that the agent tends to overreact to precise signals and underreact to noisy signals ([Predictions 2](#) and [13](#)). However, the flexible cognitive imprecision model does not predict our key result ([Prediction 1](#)) 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 full subjective belief distribution provides the simplest test to distinguish the flexible cognitive imprecision model and the two-stage model. [Prediction 15](#) 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 *underweights* the most representative state and *overweights* the least representative state since cognitive imprecision pulls his posterior back to the moderate cognitive default. The proof of [Prediction 15](#) is straightforward.⁷⁵

Prediction 15 (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 moderate 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 underweights ω_R and overweights ω_{NR} .*
- (ii) *If $\omega_R = 1 - \omega_{NR} < d$, the agent overweights ω_R and underweights ω_{NR} .*

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

[Fig. D.1](#) depicts the predictions of the flexible noise model, aggregating across uniform prior information environments used in experiments. In contrast, as shown in [Prediction 5](#), 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

⁷⁵Suppose the signal realization is r such that $\omega_R = \omega_N$ and $\omega_{NR} = \omega_1$. Note that $\hat{p}(\omega_i|r) = \lambda p_B(\omega_i|r) + (1 - \lambda)\bar{p}_d(\omega_i|r, N) = \frac{2}{N}(\lambda\omega_i + (1 - \lambda)\bar{\omega}_i)$. Letting $d = \bar{\omega}_N = 1 - \bar{\omega}_1$, then the agent overweights ω_{NR} and underweights ω_R if and only if $\omega_R = 1 - \omega_{NR} > d$, and the opposite holds if and only if $\omega_R = 1 - \omega_{NR} < d$.

diagnosticity at the extreme states. Comparing Fig. D.1 and Fig. 11, we observe that the data is consistent with the two-stage model and inconsistent with the flexible cognitive imprecision model.

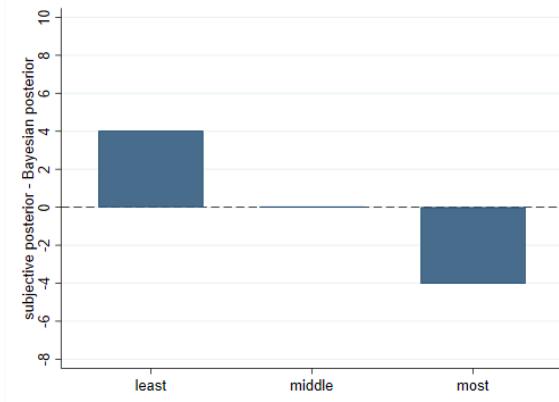


FIGURE D.1. The predictions of the flexible cognitive imprecision model on the difference between subjective and objective posterior beliefs for the least and most representative states and states that are in between, aggregating across uniform prior information environments used in the experiment. Structural 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 D.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 the cognitive-imprecision-only model 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 cognitive-imprecision-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 and the cognitive-imprecision-only models in all environments.

TABLE D.1. Completeness and Restrictiveness

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

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

Finally, because the flexible cognitive imprecision model does not incorporate attention, it predicts the same beliefs whenever the information environment (and thus computational complexity) is fixed. This is inconsistent with the results in Section 5, where exogenous variation in salience cues—holding the environment constant—produces stark differences in belief updating. These patterns are predicted by our two-stage model but not by cognitive imprecision alone.

E Additional Experiments

E.1 Forecasting Price Growth

The analysis in the body of the paper was focused largely on the inference domain, where people are tasked with inferring the likelihood of states after observing a noisy signal. In forecasting, the relevant representational objects are typically either the same as the signal (e.g., predicting a future price based on today’s price) or a direct function of it (e.g., the future payoff of an option, which is a direct function of the future price, based on today’s price). Here, we compare these domains directly, demonstrating that our model can be applied to both settings.

To do this, we build on the setting used in Fan et al. (2023), which featured a binary state space (a firm was either good or bad) and a discretized normal signal distribution (the firm’s monthly stock price growth). In their study, all participants observed a price drawn from the chosen firm’s signal distribution. Over half of them underreacted when asked to report their posterior about the firm’s state, similar to our inference experiment, but over half *overreacted* when asked to report their prediction of the next signal, i.e., forecasting the stock price growth next month. The authors refer to this as the “inference-forecasting” gap in belief updating.

In the context of our framework, attention is channeled based on the number of objects or outcomes that one must form beliefs over. For the inference task, participants needed to update their beliefs about two objects—the firm is good or bad—as in our two-state paradigm. But in the forecasting task, participants needed to form a belief over many objects—the 11 potential price outcomes. Given the good news structure of both environments, our model predicts that the higher representational complexity of the forecast task will lead participants to overreact more than in the inference task, which has lower representational complexity.

We test this conjecture by manipulating the number of objects that participants need to form beliefs over in their forecasts. We presented participants with a distribution of prices for each type of firm (good or bad). Each participant observed a price signal and reported a forecast in one of two conditions, *Simple* or *Complex*. In *Complex*, the participant reported a forecast about the likelihood of each of 11 potential prices, as in Fan et al. (2023), whereas in *Simple*, the price space was partitioned into two bins and the participant reported a forecast on the likelihood of each bin. The two conditions were identical except for this difference in the number of

objects over which the forecast was reported; specifically, the underlying information environment was exactly the same.

Specifically, participants were first shown the stock price growth distribution for good and bad firms as in Fig. E.1. Note that there are 11 potential stock price growths (signal realizations). Participants were told that the average stock price growth of a Good (Bad) firm was +100 (-100). Across all treatments, participants were told that a firm would be selected at random, with Good and Bad firms equally likely to be selected, they would observe 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.

As in Fan et al. (2023), each participant made forecasts by reporting their beliefs about the likelihood of future price growth realizations (for the next month) after observing the price growth in the current month. They did so in one of two conditions that differed in representational complexity: *Complex* or *Simple*. In *Complex*, participants forecasted the likelihood that the selected firm would experience each of the possible eleven stock price growths next month. *Simple* sought to change representational complexity without affecting the underlying information environment. It was the same as *Complex* but partitioned the price growth space into the negative domain (less than 0) versus the weakly positive domain (more than or equal to 0). After observing the current month’s price growth, participants forecast whether next month’s price growth would be positive or negative.

Despite the same underlying information environment, the change in representational complexity significantly affected belief updating. Comparing the share of participants who overreacted versus underreacted—the same measure as in Fan et al. (2023)—*Complex* replicates their results that more participants overreacted when forming a forecast ($r = 0.24$, $p < .01$). However, participants underreacted in *Simple* ($r = -0.26$, $p < .01$).⁷⁶

As shown in Appendix E.1, we replicate the predominant finding of overreaction when participants form a forecast over a large number of objects (*Complex*). However, despite facing the same information environment, marked underreaction arose when the forecast was over two objects (*Simple*). These results demonstrate that our model can also help explain over- versus underreaction in the forecasting domain. We view this as complementary to the mechanism for forecasting versus inference discussed in Fan et al. (2023), which proposes that people use different simplifying heuristics across the two domains.

⁷⁶Note that Fan et al. (2023) elicited forecasts as an expectation while we elicited forecasts as the likelihood of each potential price growth. As shown in Appendix C.9, this difference should not affect our results. We chose the latter approach as it provides richer data to explore our model’s predictions.

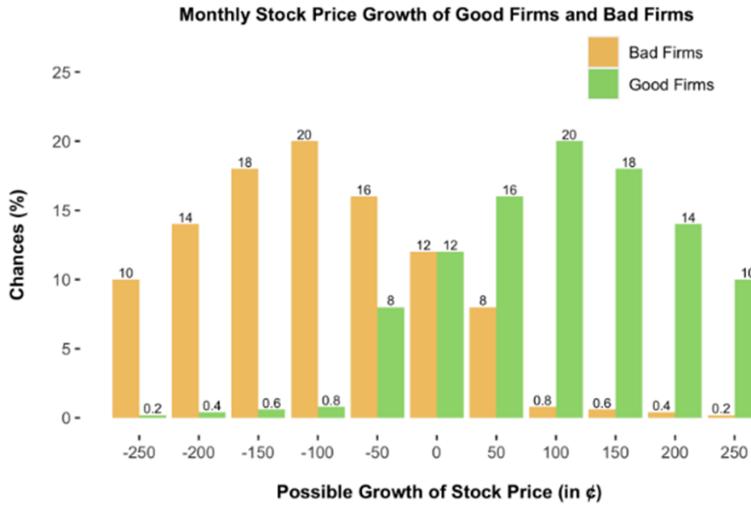


FIGURE E.1. Signal structure in the inference-forecast problem

E.2 Forecasting Financial Instruments

We now proceed to explore the implications of our framework for forecasting the payoffs of financial instruments—specifically, stock options. 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, i.e., the number of objects one needs to consider when making a judgment. Goodman and Puri (2022) show that attitudes towards complexity can explain the preference for binary options over bull-spreads on the same asset, even when the latter dominates the former.

For the purposes of our investigation, the main substantive difference between the two is that a binary option has 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 on the same asset has a larger number of potential outcomes (also based on the price of the asset).⁷⁷ Our framework predicts that the difference in complexity between the two instruments will generate a difference in belief updating. Specifically, fixing the underlying asset and information environment, people will underreact when predicting the payoff of a binary option but overreact for a bull spread.

Specifically, participants were told that there was a pool of Good and Bad firms with respective stock price growth distributions shown in Fig. E.3. 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 option whose payoff space is either simple—the binary option—or more complex—the bull spread—while keeping the structure of

⁷⁷For example, consider a binary option that returns a pre-determined payoff if the price of an asset is greater than S in a pre-determined period and zero otherwise. A dominating bull spread would generate the same payoff if the price is greater than S , but also generate a series of smaller payoffs when the price is between S' and S for some $S' < S$.

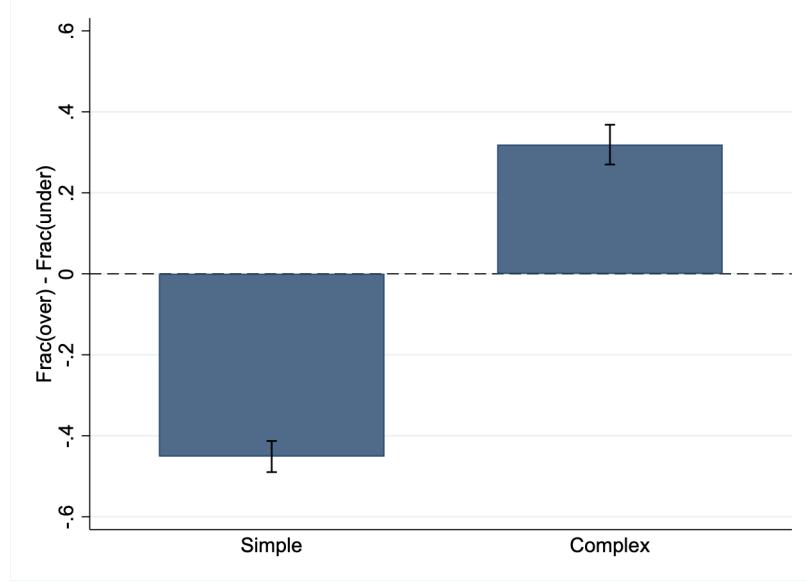


FIGURE E.2. Participants overreact in complex forecasting tasks and underreact in simple ones. Each bar aggregates all signal realizations in the relevant condition.

the underlying asset fixed (i.e., the firm). Participants were endowed with a financial option based on the randomly selected firm. The firm's monthly price increase was drawn from the distribution in Fig. E.3. In the Bull Spread condition, participants were told that they would receive the following payoff based on the price increase 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 increase was less than \$3 (i.e., \$0 or \$2) and \$6 if the price increase 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 both conditions. Each participant was shown the price increase of the selected firm in the current month and asked to forecast the likelihood of the potential payoffs of their asset based on the price increase 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.

We tested this prediction experimentally by endowing participants with either a binary option (simple) or bull spread (complex) on the same underlying asset. They then observed a signal of the asset's performance (its price increase this month) and forecasted the likelihood of potential payoffs for their option next month. Fig. E.4a shows that indeed, the majority of participants underreacted to the price signal in the case of the binary option, while the majority of participants overreacted to the same information in the case of the bull spread. This pattern is robust across signal realizations and also holds for the level of overreaction measure (see Fig. E.4).



FIGURE E.3. Signal structure in the financial asset problem

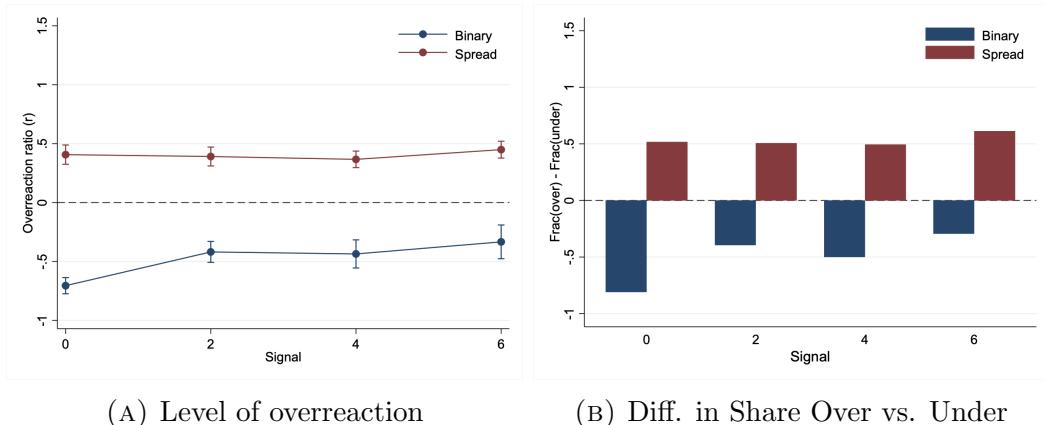


FIGURE E.4. Participants underreact to a simple binary option and overreact to a complex bull spread across all signal realizations. Each data point or bar corresponds to a single signal realization in the relevant financial options condition.

F 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.

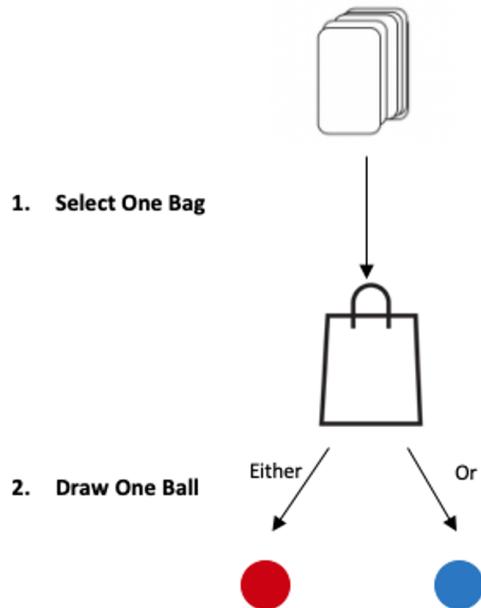
Your task is to **guess the probability that each bag was selected** based on the available information. The exact procedure is described below.

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**.