Underreaction and Overreaction in Inference and Forecasting*

Francesca Bastianello† Alex Imas‡

February 4, 2025

Abstract

We study variation in belief formation both within and across different decision domains as a function of context and the relevant features. We begin by mapping the commonly-studied inference and forecasting domains to the same underlying framework. In the model, the decision context cues a mental representation of the information environment, and limited attention leads to more sensitivity to salient features compared to neglected ones. A series of experiments show that variation in over and underreaction *across* domains is due to inference and forecasting cueing different mental representations of the same information environment, leading to a reliance on different cognitive defaults. Variation in over and underreaction *within* domains is instead explained by the allocation of attention across features, which generates differences in sensitivity away from the cognitive defaults. When varying a single feature, we recover the standard result of overreaction to weaker signals and underreaction to stronger signals. However, once we allow for attention to interact with multiple features, we show that this comparative static is modulated by attention, and can even be fully reversed. Empirical tests further identify the mechanism by directly manipulating attention across features and by introducing exogenous variation in cued representations.

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

*We thank Thomas Graeber, Andrei Shleifer, Alex Todorov, George Wu, seminar participants at Chicago Booth, Harvard, Cornell, and Claremont McKenna, as well as conference participants at the Cognitive Foundations of Finance Conference, Yale Junior Finance Conference, and Emrging Voices in Finance Conference for helpful comments. Zlata Krasic and Aria Liu provided expert research assistance.

†Booth School of Business, University of Chicago: francesca.bastianello@chicagobooth.edu

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

1 Introduction

Learning from information is a central feature of dynamic environments. It can take the form of inference, where a person uses information to infer the underlying state. For example, after eating a meal at a new restaurant, how does a consumer update her beliefs about the restaurant's quality? Learning can also take the form of forecasting, where a person interprets a signal to determine the likelihood of future signals. For example, when observing a stock price increase, how does a trader update her beliefs about future price movements?

While both inference and forecasting are relevant in a wide array of domains, historically there has been little overlap in the literature studying them. Inference questions have been widely studied by a large body of experimental work, which presents people with a signal about a binary state, and asks them how likely they believe each state to be. In these types of settings, the literature has mostly found that people underreact to information (see Augenblick et al., 2021, Fan et al., 2021, and Ba et al., 2022 for important qualifications). Moreover, variation in over and underreaction within this environment has been associated with various notions of signal strength, with people underreacting more to stronger signals and overreacting more to weaker signals (see Benjamin, 2019 for a review).

On the other hand, forecasting tasks require people to make predictions about future signals. In this context, much of the evidence from professional forecasting of macro and financial variables points towards a greater prevalence of overreaction to information Bordalo et al. (2020b) (though underreaction is observed in certain important domains, e.g., interest rates). Moreover, variation in over and underreaction in forecasting has been associated with a different set of features relative to inference, finding less overreaction at shorter horizons and for more persistent series, and more overreaction at longer horizons and for more transitory processes (Bordalo et al., 2020b, Afrouzi et al., 2023).

This heterogeneity of evidence both within and across domains raises the question of what drives variation in over and underreaction. Is variation in belief formation domain-specific, or is there a common psychological mechanism that can rationalize the findings? In other words, can the insights uncovered in the inference literature be applied to forecasting? And vice versa, can the richer forecasting environment help in refining theories that have been developed in simpler inference settings? In principle, both sets of results are a function of how people update their beliefs. However, understanding the findings within a common framework is challenging because studies of belief formation in inference and

forecasting have traditionally focused on different sets of features. For example, what does it mean for a signal to be stronger in forecasting? Is signal strength only related to the amount of unexpected news in an observed realization (e.g. shock size), or do the underlying properties of the data generating process also play a role in determining signal strength?

To answer these questions, this paper starts by developing a common framework that nests both inference and forecasting within the same information environment. This allows us to map features that are relevant in forecasting (persistence, horizon, and shock size) to features that are relevant in inference (signal strength). We then study how mental representations and limited attention interact in shaping belief formation. We show that variation in over and underreaction both within and across inference and forecasting domains can be understood through this common psychological mechanism.

In our framework, the decision context cues a mental representation of the information environment. A mental representation has two key components. The first component is a vector of meta priors on the values of decision-relevant features. This pins down the cognitive defaults that people approach a given problem with, e.g., approaching a forecasting problem may cue beliefs that persistence will be high. The second component is a vector of attention weights across features. This determines how sensitive people's beliefs are to objective variation of these features, and how much they instead rely on their cognitive defaults.[1] We show that variation in over and underreaction across domains is driven by inference and forecasting cueing a different mental representation of the same information environment. At the same time, variation in belief updating within a domain is driven by the allocation of limited attention across features and how those features interact. Importantly, we show that the traditional comparative static in the inference literature of more underreaction to weaker signals and more overreaction to stronger signals is modulated by attention, and can even be reversed.

In developing the common framework, we introduce the features that are associated with belief formation in forecasting (persistence and horizon) into the canonical inference paradigm. Inference has typically been studied using versions of a balls-and-urns experiment (Edwards, 1968). In these experiments, respondents are faced with two states (urns with different compositions of colored balls) that produce iid signals (drawn balls).

[1]Specifically, full attention implies that people internalize the objective value of the feature in the information environment; full neglect implies complete reliance on the cognitive default; partial attention implies a weighted average between the objective value and the cognitive default. This generates insensitivity to objective variation that is a function of the attention weight. See Enke and Graeber 2023 and Ba et al. 2022 for discussion of cognitive defaults in the case of inference problems.

One of the states is randomly realized according to a prior probability. People do not observe the realized state, but instead observe a noisy signal of its realization. Given this information, an inference question elicits beliefs about the likelihood of the realized state, while a forecasting question elicits people's predictions of future signals. While these experiments lend themselves to asking questions in both types of domains, the iid nature of the observed signals means that there is no room for either persistence or horizon—two important features in real-world forecasting—to play a meaningful role. To overcome this issue, we change the data generating processes that people are faced with. Instead of presenting participants with two urns with different compositions of different color balls, we let the two data generating processes be AR(1)s with different unconditional means. Given this common framework, we then show that there is a very tight mapping between persistence, horizon, and signal strength in each domain. Specifically, increasing persistence makes current signals more informative about future signals, and less informative about the underlying state, therefore leading to a stronger signal in forecasting and to a weaker signal in inference. Similarly, increasing the horizon being forecasted intuitively decreases signal strength in forecasting: since the signal is persistent, a new observation is more informative about the next observation than it is about an observation at a longer horizon.

We use our framework to design a series of empirical studies that test its predictions. In the study, there is a pool of firms that can be either good or bad. Firm profits evolve according to an AR(1). The profits of good firms fluctuate around a higher unconditional mean than the profits of bad firms. One of the firms is randomly selected from the pool. Participants do not know the firm type, but instead they observe a sequence of monthly profits generated by the selected firm. Given this information, we either elicit inference questions on the likelihood that the firm is good or bad, or forecasting questions on expectations of future profits. This paradigm has several advantages. First, it is straightforward to introduce variation in the autocorrelation of signals, which allows both persistence and horizon to play a meaningful role in forecasting and inference problems. Second, the paradigm also nests prior inference and forecasting experiments within a common framework.[2]

[2]Importantly, the common framework allows us to nest the information environments used in prior work. For example, setting persistence to zero corresponds to the environment considered in Fan et al., 2021, Ba et al., 2022 and Augenblick et al., 2021, while allowing for a continuous state space reproduces the environment studied in Afrouzi et al., 2023—see Appendix D for a detailed discussion on the exact mappings. The ability to nest different information environments within the same setting allows us to study whether the same psychological mechanism may be driving variation in belief formation both across and within inference and forecasting domains.

Our first set of results examine variation in belief formation *across* inference and forecasting domains. We replicate the finding that people overreact more in forecasting than in inference. Using survey data, structural estimation, and exogenous variation in context, we show that the difference in contextual cues in inference and forecasting problems generate different mental representations of the same information environment. For example, people have a higher cognitive default and pay more attention to persistence in forecasting problems than in inference problems. Crucially, our framework shows that the estimated differences in representation generate the observed differences in over versus underreaction across domains.[3]

Second, our framework predicts that persistence interacts with the strength of a given signal in both inference and forecasting tasks, but the comparative static goes in the *opposite* direction: increasing persistence makes a given signal less informative about the underlying mean of the data generating process (inference), but more informative about the next observation (forecasting). In other words, increasing persistence weakens the signal in inference and strengthens the signal in forecasting. Since insensitivity to features leads to overreaction to weaker signals and underreaction to stronger signals, our framework predicts that increasing persistence should lead to more overreaction in inference and less overreaction in forecasting. Our empirical results confirm this reversed comparative static. This provides support for the conjecture that variation in persistence and horizon impacts belief-updating through their effect on signal strength.

Our second set of results examine variation in belief formation *within* inference and forecasting domains. We begin by varying a single feature (e.g. persistence or the horizon being forecasted) while holding other features fixed at the cognitive default. Consistent with prior findings in the inference literature, we find that varying a single feature recovers the traditional result of overreaction to weak signals and underreaction to strong signals in both inference and forecasting. However, attention begins to play a key role once we consider the interaction of multiple features. Specifically, limited attention generates a wedge between actual signal strength and *perceived* signal strength. Variation in features such as persistence and the horizon shift the strength of the perceived signal only to the extent to which these features are attended to. The more a feature is neglected, the greater the wedge between actual and perceived signal strength. This generates the

[3]Intuitively, the location of the cognitive default plays an important role in determining the overall extent of over or underreaction in a given domain. If the cognitive default is such that respondents excessively anchor on a very high level of signal strength, we will predominantly observe overreaction. Conversely, if the cognitive default is such that respondents anchor on a very low level of signal strength, we will observe a greater prevalence of underreaction. Differences in cognitive defaults across inference and forecasting can then translate in different extents of over and underreaction across domains.

testable prediction that variation in over and underreaction will be much larger for features that are neglected (e.g., persistence, horizon) than for features that are attended to (e.g., size of the most recent shock).

Our data shows that people indeed attend to some features over others: attention to the size of the shock is substantially higher in both inference and forecasting problems than to persistence. This generates greater variation in over and underreaction when the latter is varied compared to the former. To examine the psychological mechanism more directly, we designed an experimental condition that exogenously manipulated the allocation of attention across features. Channeling more attention to persistence decreased the associated variation in over and underreaction. Importantly, the bounds on attentional capacity also generate a critical link across features: paying more attention to persistence decreased attention to the shock size, which exacerbated variation in over and underreaction with respect to the latter.

Second, we show the importance of considering the interaction of attention across *multiple* features. Specifically, we show that this interaction can completely reverse the standard result of overreaction to weak signals and underreaction to strong signals. Greater attention to shock size over persistence in people's mental representations can lead them to overreact more to *stronger* signals.[4] Notably, we show that *excess sensitivity* in belief movement to one feature can actually be driven by *insensitivity* to another feature. And this is indeed what we document in the data. Together, these results highlight the need to consider predictions for 'behavioral attenuation' at the feature level rather than directly on statistics that are functions of multiple features, e.g., signal strength (Enke et al., 2024a).

Our paper contributes to a large body of work on belief formation. We provide a more detailed review of the literature that documents variation in over and underreaction in inference and forecasting environments in Section 7. Here, we draw the connection to recent work on belief formation that attempts to explain deviations from the rational benchmark through a cognitive lens. The first set of papers study how attention and memory constraints generate distorted mental representations of features. In this line of work, features are either fully attended to or fully neglected, which leads people to

[4]For example, when persistence approaches zero (near-iid process) and the unconditional mean is known, a rational agent forecasts using the unconditional mean so that the true signal strength is *unresponsive to shock size*. However, if agents anchor on a higher persistence level through their cognitive default—as they do in our experiments—then their perceived signal strength will be *increasing in shock size*. This makes perceived signal strength excessively sensitive to shock size, leading to overreaction to stronger signals and reversing traditional comparative statics observed in inference studies (Augenblick et al., 2021).

use heuristics such as reporting the signal diagnosticity in inference problems (Bordalo et al., 2024c, 2023a). Other work has also considered the possibility that inference and forecasting problems may cue different heuristics through a similarity-based mechanism, which generates different patterns of belief-updating Fan et al. (2021). This work does not consider how multiple features such as persistence and horizon—nor their mental representation—factor into the belief formation process. It is therefore unable to speak to variation in over and underreaction within inference and forecasting domains. Another line of research looks at how noisy cognition impacts belief formation. There, people have a cognitive default for a specific feature such as the base rate and do not fully adjust from it with respect to objective variation in the parameter. This generates insensitivity and attenuation in beliefs (Enke and Graeber, 2023; Afrouzi et al., 2023; Enke et al., 2024a). The papers on noisy cognition do not consider the role of the mental representation, neither in determining what defaults are used or how multiple features are linked through attention. Finally, Ba et al. (2022) incorporates both representation and noisy cognition to study belief updating, but considers the mental representation only with respect to one feature—the state space.

Our paper connects the representational and cognitive noise literatures by modeling how the decision context—inference versus forecasting domains—cues different mental representations of the same problem. This mental representation includes cognitive defaults of the relevant features and the attention channeled to these features. In this way, the framework is able to explain both variation in belief formation *across* domains—as a function of context cueing different cognitive defaults and attention weights—and *within* domains—by linking attenuation or lack-thereof across multiple features through bounded attention.

The rest of the paper proceeds as follows. Section 2 presents the common framework that we use to study both inference and forecasting within the same information environment. Section 3 maps features that are relevant in forecasting (persistence and horizon) to features that are relevant in inference (signal strength). Section 4 introduces representation and limited attention to features, and highlights predictions on the variation in under and overreaction both within and across inference and forecasting domains. Section 5 presents the results from our baseline experiments. Section 6 introduces additional interventions that allow us to directly test our mechanism. Section 7 discusses how our results relate to the literature, and Section 8 concludes.

2 Common Framework

In this section we introduce a framework that integrates features from both inference and forecasting into a single belief-updating model. To do so, we incorporate forecasting elements such as persistence and horizon into the canonical paradigm traditionally used to study inference problems.

2.1 Baseline Inference Paradigm

Figure 1 shows a graphical representation of the canonical paradigm that has been used to study inference in much of the literature (Edwards, 1968). There are two urns that differ in their composition of orange and green balls. One of the two urns is randomly selected according to prior probabilities. People do not know which urn was selected. Instead, they observe one or more balls drawn from the selected urn. In light of this information, participants are asked about the likelihood that the selected urn was one or the other.

Formally, there are two states, $S \in \{G(ood), B(ad)\}$, that differ in their likelihood of generating one of two signals, $y \in \{g(reen), o(range)\}$. State G generates a green signal with probability $\pi_{g|G}$ and an orange signal with probability $1-\pi_{g|G}$, while state B generates a green signal with probability $\pi_{g|B}$ and an orange signal with probability $1 - \pi_{g|B}$. One of the two states is realized according to prior probabilities π_G and $\pi_B = 1 - \pi_G$. People do not observe the realized state. Instead, they are shown one or more signals drawn from the selected state, $y = \{y_1, y_2, ..., y_N\}$, where $y_i \in \{o, g\}$. Given these signals, participants are asked how likely it is that the selected state is Good or Bad.

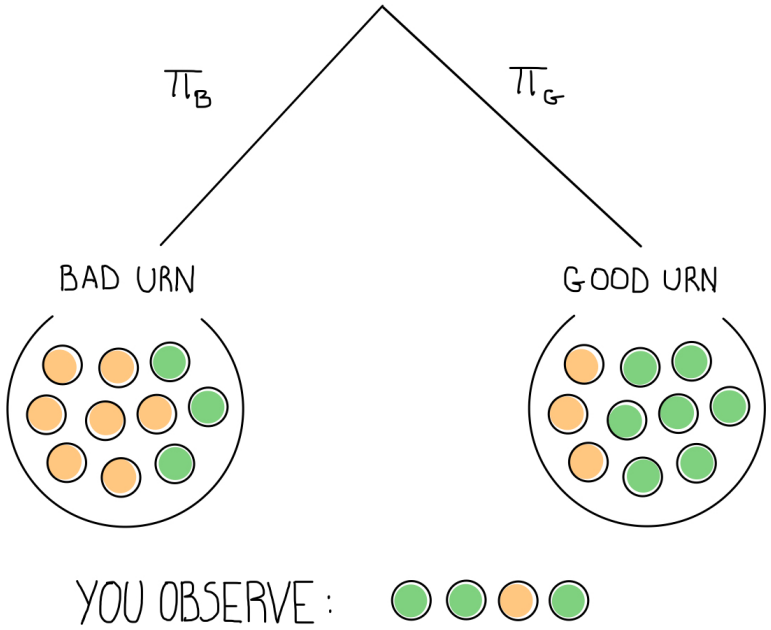
Bayesian updating leads to the following log odds ratio, which is informative of how much people should update their posterior beliefs:

$$\log\left(\frac{\pi_{G|y}}{\pi_{B|y}}\right) = \log\left(\frac{\pi_{y|G}}{\pi_{y|B}}\right) + \log\left(\frac{\pi_G}{\pi_B}\right) \tag{1}$$

where the first term on the right hand side corresponds to the likelihood ratio and the second term to a ratio of the base rates. The former is typically referred to as *signal diagnosticity* because it captures how diagnostic a given signal is of one state over the other, or simply how much more likely you are to observe that signal in one state or the other.

Within this information environment, prior work has found that people are insensitive

Figure 1: Canonical Inference Experiment.



to variation in both the likelihood and the base rates (Benjamin, 2019), and that this can translate into both over- and underreaction in belief updating. For example, if people anchor on a signal of intermediate strength, they overreact when the signal is weaker than their anchor, and underreact when the signal is stronger than their anchor (Augenblick et al., 2021). To understand how these lessons can be used to understand variation in belief formation across the domains of inference and forecasting, we start by formally defining signal strength, and we then augment this paradigm to study variation in features that are relevant in forecasting, namely persistence and horizon.

2.1.1 Signal Strength

Signal strength refers to the informativeness of the signal in shifting prior beliefs, with a stronger signal leading to a greater update in beliefs, meaning it provides more certainty or a clearer indication about the realized state. Consider the balls and urns setup we just introduced. Fixing prior beliefs, the expression in (1) makes clear that the amount by which participants should update their posterior beliefs is increasing in the log likelihood. Assume that N_g green balls and N_o orange balls were drawn from the selected urn. If we let the two states be symmetric, such that $\pi_{g|G} = 1 - \pi_{g|B}$, the log-likelihood ratio can be

written as (Griffin and Tversky, 1992):

$$\log\left(\frac{\pi_{y|G}}{\pi_{y|B}}\right) = \underbrace{(N_g - N_o)}_{\text{strength of evidence}} \underbrace{\left(\frac{\pi_{g|G}}{\pi_{g|B}}\right)}_{\text{diagnosticity}} = (N_g + N_o)\left(\frac{N_g - N_o}{N_g + N_o}\right)\left(\frac{\pi_{g|G}}{\pi_{g|B}}\right) \tag{2}$$

The first equality in the above expression makes clear that the strength of the signal depends on two components. The first term in (2) captures the relative number of green and orange balls drawn from the selected urn. In the example in Figure 1, green balls are signals in the direction of urn G, as the likelihood of drawing a green ball is higher if urn G is used than if urn B is used. Conversely, orange balls are signals in the direction of urn B. The greater the number of green balls relative to orange balls, the stronger the signal in the direction of urn G. The second term in (2) captures how diagnostic a green ball is of urn G relative to urn B. The more the two urns differ along this dimension, the more informative will the signal be. For example, if urn G only had green balls and urn B only had orange balls, then a green ball would be perfectly informative about the urn. Conversely if both urns had the same fraction of green balls, a green ball would be completely uninformative about the urn.

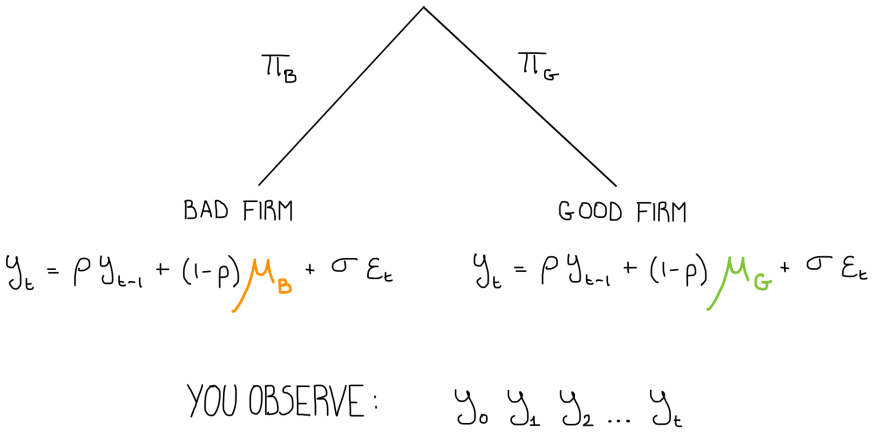
As well as providing a clear notion of signal strength, this setting also lends itself to asking both inference and forecasting questions. Inference tasks involve asking participants about the likelihood that the selected state is $G(ood)$ or $B(ad)$, while forecasting tasks involve predicting what signal will materialize in future draws (Fan et al., 2021). However, since all draws are iid, the current setup leaves no room for persistence or horizon to play a role—two features which we know to be important in forecasting. We now proceed to augment the framework to address this.

2.2 Introducing Persistence and Horizon

To introduce persistence in the canonical inference paradigm, we vary the data generating processes outlined above. Instead of asking people about iid signals generated by urns with different compositions of orange and green balls, we let signals be generated by AR(1)s with different unconditional means, as shown in Figure 2.

Specifically, consider a setting with a pool of firms, that can be either good or bad, $S \in \{G, B\}$ (where S now denotes the type of firm). Firm type is distributed according to prior probabilities π_G and $\pi_B = 1 - \pi_G$. The firm generates signals in the form of profits, y_t, which evolve according to an AR(1). Good firms' profits have an unconditional mean

Figure 2: Augmenting the Inference Paradigm with Persistence



of μ_G, while bad firm profits have an unconditional mean of $\mu_B < \mu_G$:

- Good firm profits evolve as an AR(1) with unconditional mean μ_G:

$$y_t = (1-\rho)\mu_G + \rho y_{t-1} + \sigma\epsilon_t \tag{3}$$

- Bad firm profits evolve as an AR(1) with unconditional mean μ_B:

$$y_t = (1-\rho)\mu_B + \rho y_{t-1} + \sigma\epsilon_t \tag{4}$$

where $\epsilon_t \sim^{iid} N(0,1)$, $\rho \in [0,1)$ is the autocorrelation coefficient, and $\sigma > 0$ is the conditional standard deviation.

Table 1 and the discussion in Appendix D.2 shows how this setting can nest previous empirical studies on how people form beliefs in both inference and forecasting. For example, when $\rho = 0$ this setup nests the one considered in Fan et al. (2021), while allowing for a continuous state further nests Afrouzi et al. (2023).[5] Most importantly, this framework allows us to ask both inference and forecasting questions in a context where both persistence and horizon play a meaningful role, therefore allowing us to study how multiple features interact in driving belief formation.

[5]See Appendix D.2 for a detailed discussion of these mappings.

Table 1: "Nesting" Experiments in Previous Work

		AKLMT	FLP Baseline	Possible Bridge	FLP Binary	BBI
True DGP	AR(1)	$\rho \in \{0, 0.2, 0.4, 0.6, 0.8, 1\}$ $\epsilon_t \sim N(0, \sigma^2)$, continuous $\sigma^2 = 20$	$\rho = 0$ $\epsilon_t \sim N(0, \sigma^2)$, continuous $\sigma^2 \in \{50, 60, 70, 80, 90, 100\}$	n.a.	n.a.	n.a.
	BBPC	n.a.	n.a.	Bernoulli(p), binary numerical states numerical signals	Bernoulli(p), binary non-numerical states non-numerical signals	Bernoulli(p), binary non-numerical states non-numerical signals
Prior	# of states	$\times^*$	2: $\mu_A = 100$, $\mu_B = 0^{***}$	2: value = 100, value = 0	2: Good, Bad	$\{2, 3, 5, 11\}$: Bags A, B,...
	diagnosticity		$\mu_A - \mu_B = 100$	$\frac{p(signal=1\|value=100)}{p(signal=1\|value=0)}$	$\frac{p(up\|Good)}{p(up\|Bad)}$	$\frac{p(G\|\text{Bag i})}{p(R\|\text{Bag j})}$
	prob. distribution		$p(A) = \{0.2, 0.5, 0.8\}$	$p(A) = \{0.2, 0.5, 0.8\}$	$p(A) = \{0.2, 0.5, 0.8\}$	$J = 2$ differ in symmetry, $J = 3$ differ in concentration, $J = 5, 11$ uniform
Signals	# simultaneous	$\underline{\mathrm{T}} = 40$	1	$\underline{\mathrm{T}}^{*****}$	1	1
	# sequential	$T - \underline{\mathrm{T}} = 40$	0	0	0	0
Inference	value	$\times^{**}$	$\times^{****}$	✓	×	×
	probability		✓	✓	✓	✓
Forecast	horizon	$h \in \{1, 2, 5, 10\}$	$h = 1$	$h = 1$	$h = 1$	×
	value	✓	✓	✓	×	
	probability	×	$\times^{****}$	✓	✓	

* In Experiment 1 participants are told that the data is generated by a "stable random process," and in Experiment 3 participants are told it is a "fixed and stationary AR(1) process: $x_t = \mu + \rho x_{t-1} + \epsilon_t$, with a given μ, a given ρ in the range $[0, 1]$, and ϵ_t is an iid random shock.
** While they do not elicit beliefs about the underlying mean, they conjecture that biased beliefs about the mean are central to explain their results.
*** Notice that the state can be numerical or non-numerical.
**** The "more similar" and "less similar" treatments get close.
***** This case could provide a bridge between FLP Binary/BBPCs when $\underline{\mathrm{T}} = 1$ (and signal/state are non-numerical), and FLP Baseline when $\underline{\mathrm{T}}$ is large enough so that the # of "success" signals becomes normally distributed.

After observing a sequence of firm profits, an inference task asks people how likely they think the firm is good or bad. Instead, a forecasting task asks people to make $h-$step ahead predictions about the profit of a company in month $t+h$, where h corresponds to the horizon being forecasted. In both tasks, participants have information about the following features: prior probabilities π_G and $\pi_B = 1 - \pi_G$, the unconditional means μ_G and μ_B, the persistence of the data generating process ρ, the horizon being forecasted h, and the path of previous period profits $y = \{y_{t-i}\}_{i=0}^{t}$.

3 Rational Benchmark

In this section we map persistence and horizon to signal strength in both inference and forecasting.

3.1 Inference

We start building intuition from the case where signals are iid, and then allow for the process to be persistent.

No persistence (iid series). When $\rho = 0$ profits are iid, and section B.1 in the Appendix shows that the log likelihood function can be decomposed in the same two terms that we encountered in the standard inference setup in (2):

$$\log\left(\frac{\pi_{y|G}}{\pi_{y|B}}\right) = \sum_{i=0}^{t}\left(\frac{y_i - \frac{\mu_G+\mu_B}{2}}{\sigma}\right)\left(\frac{\mu_G - \mu_B}{\sigma}\right) \tag{5}$$

$$= \sum_{i=0}^{t}\underbrace{\left(\epsilon_i + \frac{\mu_G - \mu_B}{2\sigma}\right)}_{\text{strength of evidence}}\underbrace{\left(\frac{\mu_G - \mu_B}{\sigma}\right)}_{\text{diagnosticity}} \tag{6}$$

where the second equality simply writes the log likelihood in terms of the iid standard normal shocks. This allows us to decompose the log likelihood ratio into the more basic features of our problem, with a tight connection to the decomposition in (2). The first bracketed term in (6) captures the strength of the evidence, and is equivalent to the relative number of orange and green signals in (2). The second bracketed term in (6) captures how diagnostic a given signal is of either hypothesis, and is equivalent to diagnosticity in (2).

Specifically, the first term in these expressions depends on the size of the shock (ϵ_i), on the difference in the unconditional means ($\mu_A - \mu_B$) and on the conditional standard deviation (σ). It captures how strongly the signal points in favor of one state or the other: when $y_i = \frac{\mu_G+\mu_B}{2}$ the signal is uninformative, as would have been the case if we had drawn the same number of orange and green signals in the standard inference paradigm. Increasing $\left|y_i - \frac{\mu_G+\mu_B}{2}\right|$ is instead equivalent to increasing the relative number of orange to green signals associated with a given state, which translates into a stronger signal. The second term in the above expressions depends on the difference in the unconditional means, and on the conditional standard deviation. It captures how different the states under consideration are: the greater the difference between the states, the more diagnostic is a given signal about the state, and this also contributes to a stronger signal.

Persistent series. Next, we consider the case where signals are no longer iid, and instead allow for $\rho \in (0, 1)$. When this is the case, Appendix B.1 shows that we can write the log likelihood as:

$$\log\left(\frac{\pi_{y|G}}{\pi_{y|B}}\right) = \sum_{i=1}^{t}\left(\frac{y_i - \left(\rho y_{i-1} + (1-\rho)\frac{\mu_G+\mu_B}{2}\right)}{\sigma}\right)\left(\frac{(1-\rho)(\mu_G-\mu_B)}{\sigma}\right) \tag{7}$$

$$= \sum_{i=1}^{t}\underbrace{\left(\epsilon_i + \frac{\mu_G-\mu_B}{2\left(\frac{\sigma}{1-\rho}\right)}\right)}_{\text{strength of evidence}}\underbrace{\left(\frac{\mu_G-\mu_B}{\frac{\sigma}{1-\rho}}\right)}_{\text{diagnosticity}} \tag{8}$$

The strength of the evidence and diagnosticity now depend on the same set of features as in the iid case in (6), as well as on one additional feature: the persistence of the process (ρ).

Increasing the persistence of the process decreases the log-likelihood by affecting both the strength of the evidence and signal diagnosticity. Intuitively, increasing persistence decreases the amount of independent information contained in every new observation, therefore decreasing the normalized strength of the signal. Moreover, increasing persistence also decreases the diagnosticity as it increases the volatility of the process, therefore increasing the overlap between the profits of good and bad firms, and making it harder to tell the two states apart.

Remark 1. In inference problems, the strength of a given signal is increasing in the size of the shock (ϵ_t) and in the difference between the two means ($\mu_G - \mu_B$), while it is decreasing in the persistence of the process (ρ) and in the conditional variance (σ^2).

In our empirical investigation, we consider instances where $\log\left(\frac{\pi_{y_1,\ldots,y_{t-1}|G}}{\pi_{y_1,\ldots,y_{t-1}|B}}\right) = 0$. This allows us to focus on how agents update their beliefs in light of a single new signal (one new month of profits), while abstracting from base-rate neglect. Moreover, we set $\left(\frac{\mu_G - \mu_B}{\sigma}\right) = 1$, meaning that signal strength reduces to:

$$\log\left(\frac{\pi_{y|G}}{\pi_{y|B}}\right) = \left(\epsilon_T + \frac{1-\rho}{2}\right)(1-\rho) \tag{9}$$

3.2 Forecasting

Next, we can look at what determines signal strength in forecasting. The $h-$step ahead Bayesian forecast is given by:

$$\mathbb{E}[y_{t+h}|\{y_i\}_{i=0}^{t}] = \rho^h y_t + (1-\rho^h)\mathbb{E}[\mu|\{y_i\}_{i=0}^{t}] \tag{10}$$

and the revision from observing an additional signal is:

$$\mathbb{E}[y_{t+h}|\{y_i\}_{i=0}^{t}] - \mathbb{E}[y_{t+h}|\{y_i\}_{i=0}^{t-1}] = \rho^h\sigma\epsilon_t + (1-\rho^h)\left(\mathbb{E}\left[\mu|\{y_i\}_{i=0}^{t}\right] - \mathbb{E}[\mu|\{y_i\}_{i=0}^{t-1}]\right) \tag{11}$$

This expression has two terms. The first term captures how much a rational agent would update their forecast if they knew the unconditional mean of the process. The second term captures the fact that, given our information environment, people do not know the mean of the process, and they revise their beliefs about that too when they receive a new signal. Starting from the first term, $\rho^h\epsilon_t$ is increasing in the persistence of the process and decreasing in the horizon. On the other hand, the second term, $(1-\rho^h)\left(\mathbb{E}\left[\mu|\{y_i\}_{i=0}^{t}\right] - \mathbb{E}[\mu]\right)$, is decreasing in both persistence and horizon. Intuitively, since the data generating process is an AR(1), increasing persistence makes the current observation more relevant for next period (increasing persistence makes y_t and y_{t+1} closer), but it also makes the current observation less informative about the underlying mean (increasing persistence means that y_t is closer to y_{t-1} than it is to μ, therefore making any given observation a weaker signal for the unconditional mean).

When the first term in (11) dominates, increasing persistence and lowering the horizon both increase the strength of the signal in the forecasting task. In what follows, we

consider cases where this comparative static holds. Appendix B.2 shows how this is the same comparative static that would arise if people were learning the mean via maximum likelihood when the true mean has a continuous support; we view this as the empirically relevant case.

Remark 2. In the cases we consider, signal strength in forecasting is increasing in the size of the shock (ϵ_t), in the difference between the unconditional means ($\mu_G - \mu_B$), in the persistence of the process (ρ), and in the conditional variance (σ^2), while it is decreasing in the horizon being forecasted (h).

4 Representation and Attention across Features

The previous sections outlined how features of the information environment impact belief-updating in the fully rational model.

A large literature in cognitive psychology has shown that people have limited attention and cannot attend to multiple features at the same time (Luck and Vogel, 1997; Oberauer, 2019; Loewenstein and Wojtowicz, 2023). Moreover, they approach a problem with a meta prior about the values that these features can take, which we term cognitive defaults (Friston, 2008; Enke and Graeber, 2023). Here, we introduce the concept of *mental representation* of the information environment to study how limited attention across features interacts with the cognitive defaults associated with those features.[6] Our main proposition is that inattention to a feature leads to a greater reliance on the cognitive default relative to the objective parameter value in the information environment. Less attention to a feature generates insensitivity to variation in that feature in how it impacts belief formation. We show that the wedge between the objective information environment and people's perceptions of the information environment—their mental representation—generates a rich set of predictions for variation in belief formation both across and within inference and forecasting domains.

[6] See Markman (2013) for a review on how people's perceptions of the information environment—their mental representations—can differ from the objective information environment.

4.1 Defining the Mental Representation

An agent's mental representation of the information environment is characterized by a vector of attention weights and a vector of cognitive defaults.

Attention to features. Formally, consider an n-dimensional vector of features $\mathbf{x} = (x_1, x_2, \ldots, x_n)$, where each $x_i \geq 0$ represents the magnitude of a distinct feature. The vector of features corresponds to the objective information environment $\mathbf{x}$. We follow Nosofsky et al. (1992) in representing the allocation of bounded attention across features as a set of feature-specific weights. Specifically, the attention allocated to each feature can be represented by a set of weights $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_n)$, which are subject to a task specific budget constraint on attention c_k, and need not add up to one:

$$0 \leq \alpha_i \leq 1 \quad \text{and} \quad \sum_{i=1}^{n} \alpha_i = c_k \tag{12}$$

A higher value of α_i indicates that more attention is allocated to feature x_i.

Cognitive Defaults. Now let $\bar{\mathbf{x}} = \{\bar{x}_1, \bar{x}_2, ..., \bar{x}_n\} \geq 0$ correspond to the set of cognitive defaults for each feature in $\mathbf{x}$. Specifically, the cognitive default for a given feature is determined by an agents' past experiences with that feature in similar problems faced in the past.[7] Specifically, the current problem q_c cues an agents' mental representation of the information environment, with its corresponding cognitive defaults:

$$\bar{x}_i = \sum_{j \in q_c} P(x_j | q_c) x_j \tag{13}$$

Importantly, as shown in Ba et al. (2022), the cognitive default is not a free parameter but can be explicitly elicited and verified in the data.

For the purpose of the current analysis, it is sufficient to presume that if agents have differential experiences of persistence levels across inference and forecasting tasks, they will approach the problems with different cognitive defaults. As shown later in this section, the difference in cued defaults generates differences in the extent of over and underreaction to information across the two domains.

[7] This ties in with a large literature on experience effects (Malmendier and Nagel, 2011, Malmendier and Nagel, 2016), which has also documented that the experiences that matter the most are those that are cued by the current context via associative recall (Enke et al., 2024b, Jiang et al., 2022).

Mental Representation and Belief-Updating. We define an agent's mental representation of the information environment as a vector of *perceived* features $\widetilde{\mathbf{x}} = (\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_n)$. The mental representation of the information environment $\widetilde{\mathbf{x}}$ is a function of the objective information environment $\mathbf{x}$, as well as vectors of cognitive defaults and attention weights, $(\bar{\mathbf{x}}, \boldsymbol{\alpha})$.

Specifically, an agent's perception $\tilde{x}_i$ of feature x_i is obtained by anchoring on a default, and adjusting from there as a function of attention to that feature (Gabaix, 2019):

$$\tilde{x}_i = \bar{x}_i + \alpha_i(x_i - \bar{x}_i) = (1 - \alpha_i)\bar{x}_i + \alpha_i x_i \tag{14}$$

where full attention $\alpha_i = 1$ corresponds to the objective variation in that feature, while $\alpha_i < 1$ corresponds to some degree of neglect. Neglecting a feature means that agents are not sensitive enough to objective variation in that feature, which allows us to identify channeled attention in the data.

Bounded attention implies that some features are attended to over others as a function of their bottom-up or top-down salience.[8] It also implies a budget constraint that generates an important link across multiple perceived features: directing more attention to one feature decreases attention to another. In later sections we show that the interaction of differential attention across multiple features can generate non-monotonic comparative statics on how signal strength translates into over versus underreaction.

The distinction between the mental representation $\widetilde{\mathbf{x}}$ and the objective information environment $\mathbf{x}$ creates a wedge between true and perceived signal strength.

Remark 3. Objective signal strength is a function of the objective features: $S(\mathbf{x})$. Instead, perceived signal strength is a function of the perceived features: $S(\tilde{\mathbf{x}})$.

Discussion. Our framework is related to several existing approaches in the literature. Gabaix (2019) similarly models inattention as an anchoring and adjustment process where neglect leads the agent to insufficiently adjust their perception of a feature from its default value. Our model builds on this by formalizing the notion of a mental representation that captures the allocation of limited attention across multiple features. The analysis that

[8] See Ba et al. (2022) and Bordalo et al. (2024c) for a discussion of the factors that determine top-down (goal-oriented) and bottom-up (visual/payoff-based) attention in belief updating. As discussed further below, we take a data-driven approach to attention, estimating which features are attended to versus neglected as a function of their relationship with over and underreaction.

follows shows the advantage of this approach: the interaction of multiple features at different levels of neglect explains key aspects of the variation in belief updating that we observe in the empirical results. The allocation of bounded attention across features also generates a link in sensitivities across features that is also supported in the data.

Our framework is also related to Augenblick et al. (2021) who model insensitivity at the level of signal strength. We depart from this approach in studying attention—and therefore insensitivity—at the feature level. Differential attention to features means that not all variations in signal strength lead to the same variation in over and underreaction. As we show in the analysis that follows, one implication of considering insensitivity at the feature level is that the comparative static on over and underreaction as a function of signal strength may be reversed.

Finally, we present a reduced-form approach for how context can cue different representations of the same information environment, focusing specifically on how the domains of inference or forecasting may cue different cognitive defaults. This is related to Fan et al. (2021) who argue that inference and forecasting may lead agents to adopt different heuristics through similarity-based matching. It is also related to Bordalo et al. (2024c) who model different aspects of context as determining which features are attended to versus neglected.

Appendix C presents a richer model that microfounds different possible mental representations of the same information environment being cued by different aspects of context. This generates heterogeneity in mental models as well as the heuristics used. Specifically, we consider a two-step process of belief updating, akin to Ba et al. (2022). In the first step, the context the DM is faced with cues a given (potentially distorted) representation of the problem (Bordalo et al., 2024c), together with the relevant cognitive defaults. In the second step, cognitive imprecision introduces insensitivity to variation in the features that are part of the agent's mental representation.

Here we focus on a more parsimonious framework, which ties together key elements of previous models in a transparent fashion, while also being able to comprehensively capture key patterns in variation in inference and forecasting that we observe in the data.

4.2 Over/Underreaction vs Insensitivity/Excess Sensitivity

We follow the prior literature and define insensitivity and excess sensitivity in terms of slopes, and we define over and underreaction in terms of levels.

Insensitivity and Excess Sensitivity. We define insensitivity and excess sensitivity in terms of slopes. Whenever $\frac{\partial S(\tilde{\mathbf{x}})}{\partial x_i} < \frac{\partial S(\mathbf{x})}{\partial x_i}$, subjective signal strength is insensitive to feature x_i. Conversely, when $\frac{\partial \tilde{S}(\tilde{\mathbf{x}})}{\partial x_i} > \frac{\partial \tilde{S}(\mathbf{x})}{\partial x_i}$ subjective signal strength is excessively sensitive to variation in feature x_i.

Over and Underreaction. We define over and underreaction in terms of levels. Whenever $S(\tilde{\mathbf{x}}) > S(\mathbf{x})$ people perceive the signal to be stronger than it is, and update too much, leading to overreaction. Conversely, whenever $S(\tilde{\mathbf{x}}) < S(\mathbf{x})$, people do not update enough, leading to underreaction.

While the definitions we provided above apply at the outcome level (signal strength), the same definitions can be applied at the feature level, ($\partial\tilde{\mathbf{x}}/\partial\mathbf{x}$ and $\tilde{\mathbf{x}} - \mathbf{x}$).

Feature level and Outcome level. Limited attention as in (14) implies that variation in subjective features is always insensitive to variation in objective features. However, as we further explore in Section 4.3, insensitivity at the feature level does not necessarily translate into insensitivity at the outcome level.

4.3 A Simple Example

Since the functional form that determines signal strength differs for inference and forecasting, we start by conveying the intuition behind our results through a simple example. We then show how to apply the intuition behind these results to inference and forecasting.

To understand how limited attention across features can generate over and underreaction, assume for simplicity that there are only two features $\mathbf{x} = \{x, y\}$, that people are only insensitive to variations in x, such that $\tilde{\mathbf{x}} = \{\tilde{x}, y\}$, and that true and perceived signal strength are determined by the following functional form:

$$S(\mathbf{x}) = yx \tag{15}$$

$$\tilde{S}(\tilde{\mathbf{x}}) = y\tilde{x} \tag{16}$$

The difference between true and perceived signal strength is then given by:

$$\tilde{S}(\tilde{\mathbf{x}}) - S(\mathbf{x}) = y\,(\tilde{x} - x) = y(1 - \alpha_1)(\bar{x} - x) \tag{17}$$

Given this framework, we can study how the location of the cognitive default, as well as variation in both the neglected and attended to features contribute to belief formation.

4.3.1 Variation Across Domains

We start by studying how differences in cognitive defaults can contribute to explaining variation in over and underreaction *across* domains.

Context and Cognitive Defaults. As discussed in Section 4.1, people approach an information environment with a meta prior for each feature.[9] These priors are based on people's experiences in approaching similar problems in the past. Importantly, if people's experience of a given feature is different across two domains, this can translate into different mental representations of the same underlying information environment.

Fixing the variation in x, the second equality in (17) shows how different cognitive defaults translate into differences in the prevalence of over or underreaction. Specifically, agents overreact whenever $x < \bar{x}$, and underreact whenever $x > \bar{x}$. Fixing the distribution of x, the lower $\bar{x}$ is, the greater the prevalence of overreaction. Conversely, increasing $\bar{x}$ increases the prevalence of underreaction.

Remark 4. Fixing the distribution of x, the position of the cognitive default matters in determining the prevalence of over and underreaction. When S is increasing (decreasing) in x, a higher (lower) cognitive default $\bar{x}$ increases the prevalence of overreaction.

As further explored in Section 5.4.1, if inference and forecasting cue different mental representations of the information environment—with different cognitive defaults—then this will generate different levels of over and underreaction across the two domains.

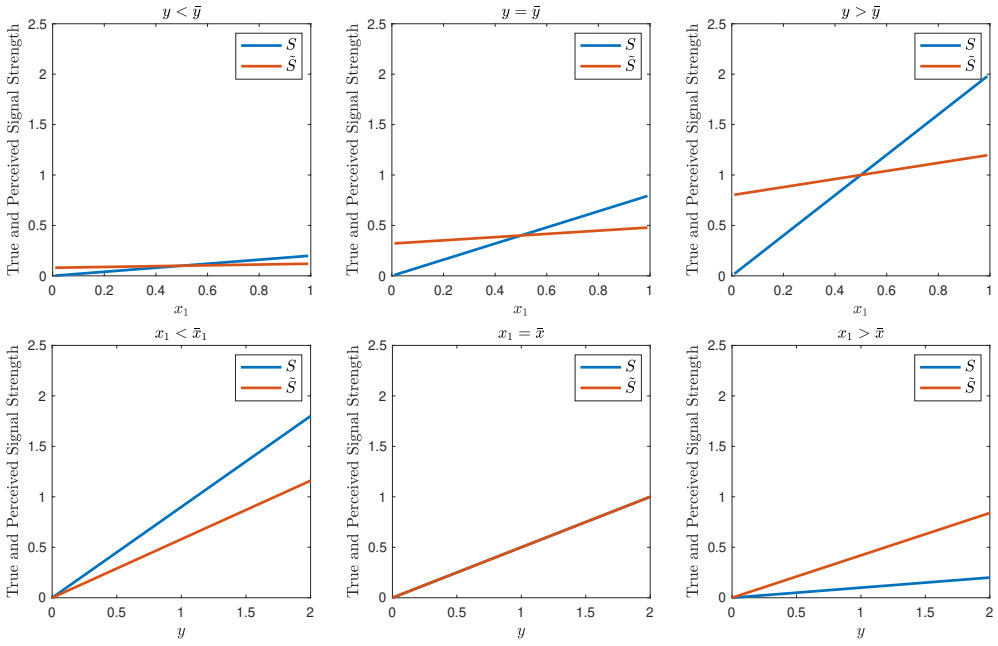
4.3.2 Variation within Domains

Next, given a cognitive default, we turn to studying how variation in signal strength contributes to explaining variation in over and underreaction *within* domains.

[9] A large literature in psychology and economics provides support for the notion that agents approach information environments with meta priors about the relevant parameter values (see discussion in Friston, 2008).

Variation in the Neglected Feature. Here we consider belief updating in response to variation of the more neglected feature, x, while other features are set to their cognitive default. The expression in (17) illustrates two additional points on the extent of over and underreaction when varying the *neglected* feature. First, as shown in the top panel of Figure 3, varying the neglected feature leads to overreaction to weak signals ($x < \bar{x}$) and underreaction to strong signals ($x > \bar{x}$). This is studied in Augenblick et al. (2021), who show that neglect of signal diagnosticity can generate *both* under and overreaction.[10] Second, decreasing the sensitivity to a given feature (smaller α) amplifies the wedge between $\tilde{S}$ and S. This allows for identification of inattention in the data: greater neglect of a feature leads to greater variation in under and overreaction in response to objective variation in that feature. We summarize these results in the following remarks.

Figure 3: True and perceived signal strength when neglecting a single feature. These figures plot variation in true signal strength (blue line) and perceived signal strength (red line), as a function of either the attended or neglected feature, while fixing the value of the other feature. The top row shows variation in true and perceived signal strength when varying the neglected feature, and fixing the value of the attended to feature at $y < \bar{y}$, $y = \bar{y}$, or $y > \bar{y}$, from left to right respectively. The bottom panel shows variation in true and perceived signal strength when varying the attended to feature, and fixing the value of the neglected feature at $x < \bar{x}$, $x = \bar{x}$ or $x > \bar{x}$, from left to right respectively.



[10] In Augenblick et al. (2021), signal strength is determined by a single feature, and they find that people are not sensitive enough to its variation. Instead, they tend to attenuate the signal strength towards the cognitive default. An over-reliance on the cognitive default leads people to overreact when signals are weaker than the default and underreact when signals are stronger than the default.

Remark 5. Insensitivity to a feature can generate both under and overreaction. Specifically, setting other features to their cognitive default values, varying the more neglected feature x leads to overreaction to weak signals and underreaction to strong signals.

Remark 6. The lower the attention to a single feature, the more does objective variation in that feature generate variation in observed under and overreaction.

Interaction Across Multiple Features. Next, we look at how variation in signal strength due to the *attended to* feature y alters the extent of over and underreaction. Equation (17) and the bottom panel of Figure 3 show how increasing y amplifies any wedge in perceived and objective signal strength due to misspecified perceptions of the neglect feature x. Whenever $x > \bar{x}$ increasing the value of the attended to feature translates into more underreaction—i.e. greater underreaction to stronger signals, the standard comparative static. However, when $x < \bar{x}$, increasing the attended to feature leads to more overreaction—i.e. greater *overreaction* to stronger signals, the reverse of the comparative static. This latter case clearly illustrates the importance of considering the interaction of attention and variation across multiple features.

Remark 7. The interaction of attended to and neglected features can *reverse* the comparative static of more overreaction to weaker signals and more underreaction to stronger signals.

Taken together, these results show how the prediction of more overreaction to weak signals and more underreaction to strong signals that is prevalent in inference studies (Benjamin, 2019, Augenblick et al., 2021, Enke and Graeber, 2023) is not only modulated by attention, but can also be reversed. In what follows, we apply this framework to make predictions for both inference and forecasting.

4.4 Predictions in Inference and Forecasting

Table 2 outlines the features that are relevant in determining signal strength in both inference and forecasting.[11] Throughout the analysis, we focus on variation in persistence (ρ), in the size of the shock (ϵ_t), and in the horizon being forecasted (h). We now proceed

[11] We define features over the most basic primitives of the problems. This allows us to have an agnostic view and model-free definition.

to derive predictions for belief-updating as a function of attention and variation in these features.

Table 2: Features in Inference and Forecasting

	Features we vary			Features we hold constant		
	Persistence	Shock	Horizon	Diff in Means	Priors	Volatility
Inference	ρ	ϵ_t		$\mu_A - \mu_B$	π_G, π_B	σ
Forecasting	ρ	ϵ_t	h	$\mu_A - \mu_B$	π_G, π_B	σ

Since signal strength in inference and forecasting take on different and more complex functional forms than the one considered in Section 4.3, in this section we make predictions across domains by simulating the extent of over and underreaction predicted by our model. To do so, we need to fix the cognitive defaults and attention weights. Rather than making ad-hoc assumptions, we take a data-driven approach and pick the values that best fit the data we gather in the experiment in Section 5. Table 3 summarizes these values.

Table 3: Cognitive Defaults and Attention to Features in Inference and Forecasting

	$\bar{\rho}$	$\bar{\epsilon}$	α_ρ	α_ϵ
Inference	0.6	medium shock	0	0.4
Forecasting	0.9	medium shock	0.4	0.8

4.4.1 Variation in Over and Underreaction Across Domains

To study variation in over and underreaction across domains, we start by exploring the implications of the differences in cognitive defaults documented in Table 2.

Cognitive defaults and sensitivity to features. Table 2 shows how the default level of persistence varies across the two domains. These values imply that, despite facing the same set of information environments, inference and forecasting are cueing different mental representations. According to our model this means that people are more likely to recollect less persistent series when cued with an inference problem than when they are cued with a forecasting problem.[12] Given the same level of variation in persistence, this difference in defaults translates into more underreaction in inference and more overreaction in forecasting.

[12]We provide direct support for this claim in Section 5.4.1, where we show that, fixing the information environment, people associate forecasting with more persistent processes from their everyday lives than inference.

Prediction 1. Forecasting tasks are more likely to cue the recall of less transitory (more persistent) processes than inference tasks. This generates mental representations with a larger persistence default in forecasting than inference, which translates to more overreaction in forecasting and more underreaction in inference.

4.4.2 Variation in Over and Underreaction Within Domains

Table 2 also shows that persistence is neglected more than the size of the shock in both inference and forecasting. In what follows, we explore the implications of these differences in explaining variation in belief formation across domains.

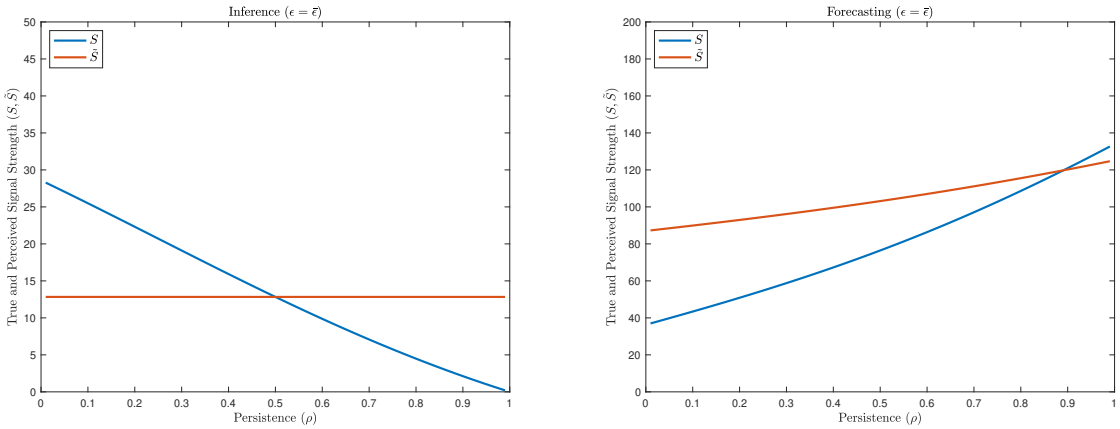
Variation in a single feature. We begin by considering the implications of varying the most neglected feature ($x_i = \rho$) while setting all other features at their cognitive defaults, such that $\mathbf{x}_{-i} = \bar{\mathbf{x}}_{-i}$. Figure 4 shows how true and perceived signal strength vary in inference (left panel) and in forecasting (right panel). If we first focus on the true signal strength (blue lines), we see that signal strength in inference is decreasing in persistence, while signal strength in forecasting is increasing in persistence (as discussed in Sections 3.1 and 3.2). Turning to the perceived signal strength (red lines), we see that $\tilde{S}$ is less sensitive to changes in persistence in both inference and forecasting. This insensitivity translates into more overreaction when variation in persistence leads to a weaker signal, and to more underreaction when variation in persistence leads to a stronger signal. We thus obtain an opposite comparative static in inference and forecasting: increasing persistence leads to more overreaction in inference and more underreaction in forecasting (Remarks 1 and 2).

Figure 4 also shows how the position of the cognitive default matters in determining whether over or underreaction prevails, on average. Since the default persistence level in forecasting is $\bar{\rho} = 0.9$, and most persistence levels are below that value, we should expect to observe predominant overreaction in forecasting. Instead, the lower default of $\bar{\rho} = 0.5$ in inference should translate into greater levels of underreaction relative to forecasting (Remark 4).

Finally, the more a given feature is neglected, the greater the variation in over and underreaction that its objective variation generates (Remark 6).

A similar analysis on the size of the shock and on the horizon being forecasted leads to the following set of predictions.

Figure 4: True and perceived signal strength as a function of persistence. These figures plot variation in true signal strength (blue lines) and perceived signal strength (red lines) as a function of persistence, while holding shock size fixed at its cognitive default. The left panel depicts these relationships in inference, and the right panel depicts these relationships for forecasting.



Prediction 2. When $\mathbf{x}_{-i} = \bar{\mathbf{x}}_{-i}$, insensitivity to a *single feature* x_i as in (14) leads to overreaction to weak signals and underreaction to strong signals:

- Increasing persistence (ρ) leads to less underreaction/more overreaction in inference and more underreaction/less underreaction in forecasting.
- Increasing the size of the shock (ϵ_t) leads to more underreaction/less overreaction in both inference and forecasting.
- Increasing the horizon (h) leads to more overreaction/less underreaction in forecasting.

Prediction 3. The more a feature is neglected, the less responsive will people be to its variation. When $\mathbf{x}_{-i} = \bar{\mathbf{x}}_{-i}$, lowering attention to feature x_i translates into a larger variation in under/over-reaction.

Prediction 3 implies that the same variation in signal strength does not always generate the same variation in over and underreaction. Instead, a given variation in signal strength leads to greater variation in over and underreaction when the more neglected feature is varied. In our setting, changes in persistence lead to more variation in over and underreaction than changes in the size of the shock.

This result also allows us to illustrate why, in the presence of multiple features, predictions on attenuation should only be applied at the feature level rather than to statistics that are

a function of multiple features, such as signal strength. If we were to follow the approach in Augenblick et al. (2021) and consider attenuation at the signal strength level, then the same variation in signal strength should always yield the same variation in over and underreaction. We show that this is not the case—in fact, depending on which features are generating the variation in signal strength, the comparative static can actually reverse. By extending the framework to multiple features and considering the relevant mental representation, we consider a richer set of predictions with respect to which features are generating variation in signal strength.[13] Importantly, this also allows us to rationalize the results in the forecasting literature, which documents more variation in under and overreaction when varying persistence than when unconditionally varying the size of the shock (Afrouzi et al., 2023).

Interaction of neglected and attended to features. So far we have considered predictions when considering variation in a single feature x_i at a time, while keeping all other features set to $\mathbf{x}_{-i} = \bar{\mathbf{x}}_{-i}$. However, as noted in Remark 6, the interaction of neglected and attended to features can break the relationship of more overreaction to weaker signals and more underreaction to stronger signals.

Figures 5 and 6 show how this plays out in inference and forecasting, respectively. We begin by focusing on inference. Moving from the left most to the right most panel in Figure 5, we see that the sensitivity of true signal strength to the size of the shock decreases as we increase persistence (the blue lines become flatter as we move from left to right). Intuitively, in the extreme case when the process is close to being a random walk, any new observation is uninformative about the underlying mean, so the size of the shock does not have a large impact on signal strength. Conversely, when the process is close to being iid, any new observation is very diagnostic of the underlying mean, and true signal strength is more sensitive to the size of the shock.

Turning our attention to perceived signal strength (red lines), we see that its sensitivity to the size of the shock is unchanged as we vary persistence. This is because persistence is completely neglected in inference, and people perceive it to be 0.5 regardless of its true level. Combining these two observations, we see that the sensitivity of perceived

[13] Augenblick et al. (2021) do consider an extension where signal strength is determined by multiple features. However, in their model, multiple features affect signal strength [14]. This linearity assumption rules out the type of important interaction across features that we show to be highly relevant for signal strength in both inference and forecasting tasks. By considering the full scope of interactions between features, our framework allows us to be more explicit on what features determine signal strength and how exactly they contribute to explaining variation in under and overreaction in both inference and forecasting.

signal strength goes from being lower than the true sensitivity in the left most panel, to being greater than the true sensitivity in the right most panel. Therefore, we have a clear example where insensitivity to variation in persistence and shock size leads to *excess sensitivity* of perceived signal strength to the size of the shock.

Figure 6 shows how this also applies to forecasting. The sensitivity of true signal strength to the size of the shock varies much more than the sensitivity of perceived signal strength. This allows the latter to go from being lower than the former when $\rho = 0.9$, to being higher than the former when $\rho = 0.1$. Intuitively, when the process is close to being iid, the size of the shock should not matter in forecasting: people should always forecast the unconditional mean for an iid series regardless of the size of the shock. Instead, if people perceive the process to be persistent even when it is not, they will mistakenly associate a larger shock with a stronger signal, which again translates into excess sensitivity.

Figure 5: True and perceived signal strength in inference, fixing persistence, and varying the size of the shock. These figures plot variation in true signal strength (blue lines) and perceived signal strength (red lines) in inference as a function of shock size, while fixing persistence $\rho = 0.1$, $\rho = 0.5$, or $\rho = 0.9$, from left to right respectively.

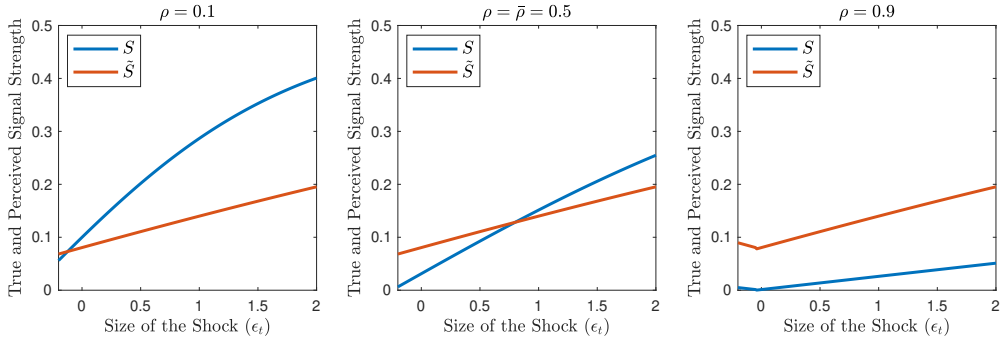
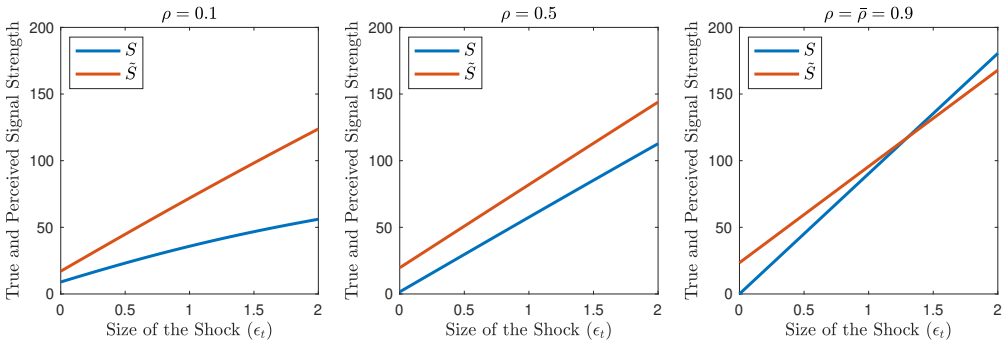


Figure 6: True and perceived signal strength in forecasting, fixing persistence, and varying the size of the shock.



Prediction 4. The interaction of attended to and neglected features can reverse the comparative static of more overreaction to weaker signals and more underreaction to stronger signals. In inference, increasing the size of the shock leads to more overreaction when objective persistence is high. In forecasting, increasing the size of the shock leads to more overreaction when objective persistence is low.

In both examples, we see how insensitivity to features can translate into *excess sensitivity* in perceived signal strength. This highlights how insensitivity and excess sensitivity can be thought of as two sides of the same coin. These results also allow us to draw a closer connection between the inference and forecasting literature. In inference, we often associate variation in over and underreaction to insensitivity. In forecasting, we often associate variation in over and underreaction to excess sensitivity to shocks. These results show that we do not need to model separate biases to capture both these phenomena: both phenomena can be generated by the same psychological mechanism.

5 Empirical Investigation

We test the predictions of our framework in a series of experimental studies. Participants ($N = 578$) were recruited from an online subject pool (Academic Prolific). All participants had to first pass an attention check before accessing the experimental instructions. Those failing this preliminary check were excluded from further participation and their data was not included in the study. Participants who passed this initial screen were informed that on top of a \$5 base payment, they could earn an extra \$10 if their answers to a randomly selected belief elicitation question fell within 3% of the actual objective posterior.[15]

5.1 Design

The design of the study closely matched the framework outlined in Section 2.2. Participants were told that there was a pool of 20 firms: 10 of these firms were Good firms, 10 of these firms were Bad firms. One firm would be randomly selected from this pool and used for the study; all firms had an equal chance of being selected, which pins down a uniform prior. The average profit of Good firms was higher than the average profit of

[15]This incentive structure was chosen over more complex methods like quadratic or binarized scoring rules. Recent research, including a study by Danz et al. (2022), suggests that simpler incentivization mechanisms encourage more honest reporting and reduce cognitive load, as opposed to more complex methods that tend to encourage conservative responses and higher error rates.

bad firms: while the profits of good firms fluctuated around a mean of 100, the profits of bad firms fluctuated around a mean of 0. Firm profits evolve as AR(1). To convey this information, we showed participants the formula for an AR(1) process, as depicted at the top of Figure 8, and explained each component of the formula in detail.[16] For example, we explained how current firm profits depended on previous period profits, on the firm's average profit, as well as a random shock that usually stays within ±200. Moreover, we explained that the weight on previous period profits is determined by the persistence of profits over time, and that this parameter can change across rounds. Importantly, each participant had to go through a large battery of comprehension questions about each aspect of the data generating process before proceeding. Responses to these questions were accompanied by explanations for why a certain answer was either correct or incorrect.

Once the questions were passed, participants were shown 30 months of profits from the selected firm, and were tasked with either an inference or forecasting problem.[17] We refer to this as a "Base Trial," and we show one such example in Figure 7. Based on the available information, the inference question asks participants how likely they think the firm is Good and how likely they think the firm is Bad. Beliefs are reported in the form of point estimates.[18] The forecasting questions ask participants their predictions for firm profits in months 32 and 33. Importantly, both forecasting and inference tasks also ask participants one other question: what they expect profits to be over the long run, on average.

Each respondent completed 9 rounds of Base trials: 3 of these trials have an objective posterior that favors Good firms, 3 of these trials have an objective posterior that favors Bad firms, and 3 of these trials have an objective posterior of 50:50. For the 3 trials that resulted in an objective posterior of 50:50, we then showed participants one more month of profits from the selected firm, and ask them to update their answers. We refer to these as "Shock Trials." An example of this type of trial is depicted the right column of Figure 8. Thus, each respondent completed a total of 12 rounds: 9 Base trials and 3 Shock Trials.

[16] All instructions, including the training, is available in the Appendix.

[17] Each participant either answered forecasting or inference questions. Randomization into domain was done between subject.

[18] Note that these point estimates amounts to eliciting beliefs about the mode of the distribution. Our comparative static predictions both across and within domains are designed to accommodate this type of data.

Figure 7: Firm profits over time – Base Trial. We elicited inference and forecasting questions after showing participants 30 months of profits.

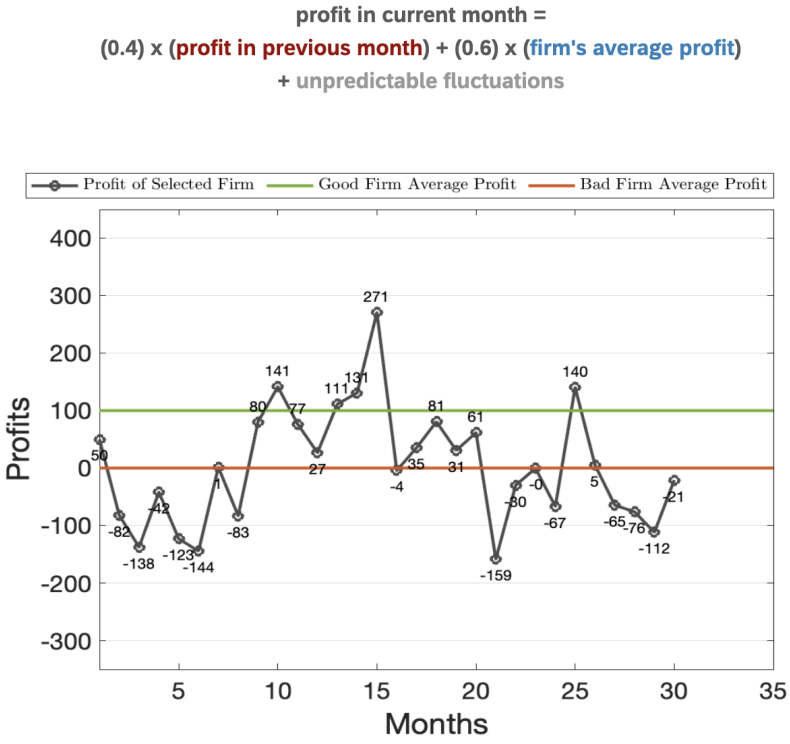
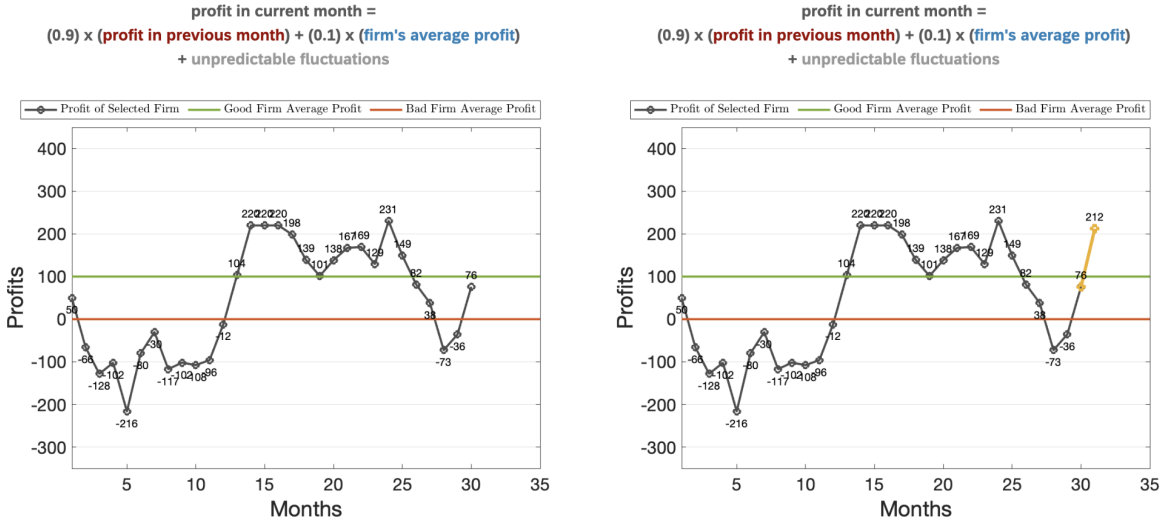


Figure 8: Firm profits over time – Shock Trial. For the trials that have a Bayesian posterior of 50 after 30 months of profits, we showed each participant an additional month of profits in month 31, and asked them to revise their beliefs.



5.2 Sources of Variation

Table 4 contains the conditions in the experiment. The first 6 trials generate an objective posterior of 0.5 in the Base condition. This allows us to have all Shock conditions starting from an uninformative base rate, so that the results we capture after the shock are less likely to be affected by base-rate neglect. We vary both persistence and shock size across trials.[19]

Table 4: Conditions for Base and Shock Trials for both inference and forecasting.

Base					**Shock Inference**			**Shock Forecasting**		
Trial	μ_{true}	ρ	σ	$\bar{\epsilon}_t$	$\epsilon_{t+1,1}$	$\epsilon_{t+1,2}$	$\epsilon_{t+1,3}$	$\epsilon_{t+1,1}$	$\epsilon_{t+1,2}$	$\epsilon_{t+1,3}$
1	100	0.9	100	-0.05	2.00	0.80	-0.22	2.00	1.34	0.62
2	100	0.5	100	-0.25	2.00	0.80	-0.22	2.00	1.34	0.62
3	100	0.1	100	-0.45	2.00	0.80	-0.22	2.00	1.34	0.62
4	0	0.9	100	0.05	-2.00	-0.80	0.22	-2.00	-1.34	-0.62
5	0	0.5	100	0.25	-2.00	-0.80	0.22	-2.00	-1.34	-0.62
6	0	0.1	100	0.45	-2.00	-0.80	0.22	-2.00	-1.34	-0.62
7	100	0.8	100	0.00	-	-	-	-	-	-
8	100	0.6	100	0.00	-	-	-	-	-	-
9	100	0.4	100	0.00	-	-	-	-	-	-
10	0	0.8	100	0.00	-	-	-	-	-	-
11	0	0.6	100	0.00	-	-	-	-	-	-
12	0	0.4	100	0.00	-	-	-	-	-	-

We convey information on variation in persistence in two ways. First, we explicitly provide participants with the DGP, and highlight changes in persistence. Second, we ensure that the in sample autocorrelation of the data we present them with matches the true autocorrelation by choosing trials that fit this criteria. This ensures that variation in persistence is also visible from from the graphs themselves.

[19]The reason why the second and third shocks are not the same across inference and forecasting is due to the fact that different shocks are needed to obtain combinations of persistence and shock size that have the same posterior in inference and forecasting (see Result 2).

5.3 Analysis

In analyzing our results, we define under and overreaction by comparing participants' revisions to the revisions under the rational benchmark.

$$\theta = (\tilde{\pi}_{A|y} - \tilde{\pi}_A) - (\pi^*_{A|y} - \pi^*_A) \tag{18}$$

Whenever $\theta > 0$, it means that participants react to new information more than a rational agent would, yielding overreaction. Conversely, whenever $\theta < 0$, participants react to new information less than a rational agent would, yielding underreaction. We use the same definition to define over and underreaction in the forecasting task.

5.4 Results

We begin by examining variation in over and underreaction across domains and then proceed to examine variation within domain.

5.4.1 Variation Across Domains

We first look at the overall level of over versus underreaction in the forecasting and inference tasks. Pooling across all parameter values, we see that people overreact significantly more in forecasting than in inference. In the Base portion, $\theta_{forecasting} = 28.32$ while $\theta_{inference} = -6.37$ ($p < .01$); that is, people predominantly overreact in forecasting and predominantly underreact in inference. In the Shock portion, $\theta_{forecasting} = 13.53$ versus $\theta_{inference} = 1.67$ ($p < .01$).

Eliciting Cognitive Defaults. Prediction 1 explains the observed variation across inference and forecasting through a difference in cued mental representations—namely, the cognitive defaults. We test this prediction in two ways. First, we structurally estimate both the attention and default parameters in the data. These results are presented in Table 3. There we indeed see that people pay little attention to persistence and that the cued default is higher for forecasting than it is for inference. Second, we examine what type of processes "come to mind" when people approach forecasting and inference problems; this allows us to study whether the type of processes that come to mind—and, in turn, generate the cognitive default—have higher or lower persistence across the two

environments. We do so by presenting a separate group of participants ($N = 101$) with two potential data generating processes that vary in their persistence level, and asking them which was more "representative" of forecasting versus inference problems in their every day lives (Bordalo et al., 2021).

Specifically, participants were presented with two potential data generating processes. The first consists of *iid* flips of a biased coin. The coin can either be biased towards Heads or Tails. The coin is flipped and the outcome revealed. The participant's task would then be to either guess the outcome of the next flip (forecasting) or guess whether the coin is biased towards Heads or Tails (inference). The second data generating process corresponds to the price of a company's stock. The underlying 'quality' of the company could either be good or bad, such that the chances of observing a price increase (decrease) are higher (lower) in the former case than the latter. The participant sees a price increase or decrease and has to make a guess about the next price movement (forecasting) or a guess about the quality of the underlying company (inference). Importantly, the *iid* coin process has a lower persistence than the company processes.

Both data generating processes were described to the participants, but some were randomized into the forecasting domain and others into the inference domain. They were then asked, when you encounter these types of problems in your every day life, "are the scenarios that come to mind more similar to the coin or company scenario described above?" They indicated the extent to which each scenario was representative of the type of problem they were presented with on a scale of 0 = 'most similar to coin' to 10 = 'most similar to company'. Participants rated the company scenario as significantly more representative of forecasting problems than inference problems (6.42 versus 4.37; $p = .01$), suggesting that the former will cue higher persistence defaults than the latter.

Result 1. Forecasting and inference cue different mental representations of the underlying information environments. As shown in Section 4.3.1, the difference in cognitive defaults can explain greater overrreaction in forecasting and greater underreaction in inference.

It is important to note that while we replicate the difference in over versus underreaction in forecasting and inference documented in Fan et al. (2021), this difference is driven by a conceptually different—but complementary—mechanism. Fan et al. (2021) explain the difference between domains through the differential use of exact heuristics across forecasting and inference in the case of an iid data generating process. In our setting, the difference

in belief formation is driven primarily by forecasting and inference cueing different mental representations—specifically, cognitive defaults—which generate greater overreaction (underreaction) in the former (latter) domain. Appendix C shows that the differences in over and underreaction across inference and forecasting are still large and significant even when accounting for the exact heuristics documented in Fan et al. (2021) Moreover, the use of heuristics remains unchanged as we vary different features, meaning that exact heuristics cannot explain variation in over and underreaction *within* the inference and forecasting domains. Finally, Appendix C also shows that differences in heuristics do not explain the variation in over and underreaction that we obtain from manipulating attention, which we further discuss in Section 6. Together, these results suggest that variation in belief updating in our setting is primarily driven by the interaction of limited attention and differences in cued mental representations.

5.4.2 Variation Within Domains

Next, we test the set of predictions on the variation in over and underreaction within inference and forecasting.

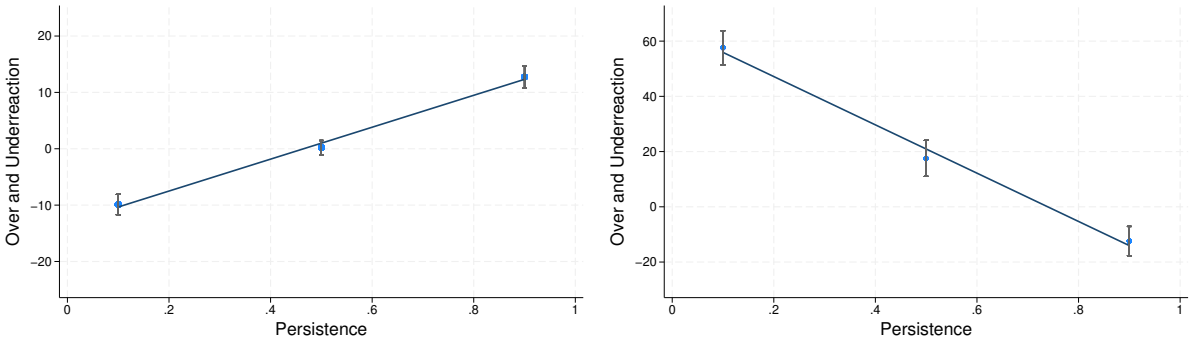
Varying a single feature. We begin by testing Prediction 2. This requires examining belief formation as a function of varying a single feature—e.g. persistence ($x_i = \rho$) or the horizon being forecasted ($x_i = h$)—while fixing other features at their cognitive defaults, $x_{-i} = \bar{x}_{-i}$.

Figure 9 shows how varying persistence leads to an opposite comparative static with respect to over and underreaction in inference and forecasting. Increasing persistence weakens signal strength in inference and strengthens signal strength in forecasting. Therefore, the patterns in Figure 9 are both consistent with people overreacting more to weaker signals (higher persistence in inference and lower persistence in forecasting) and underreacting more to stronger signals (lower persistence in inference and higher persistence in forecasting).

Observing this opposite comparative static in the data suggests that persistence does indeed drive variation in over and underreaction through its effect on signal strength.

Result 2. Results in Figure 9 illustrate the opposite comparative static between overreaction and persistence in inference versus forecasting. The extent of overreaction has

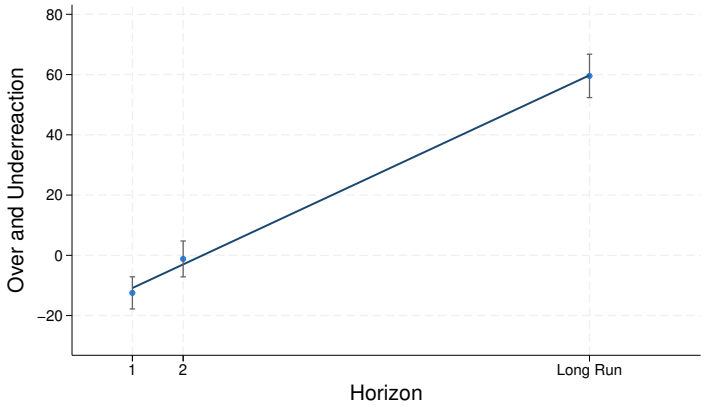
Figure 9: Over and underreaction in as a function of persistence. The left panel plots over and underreaction over persistence in inference. The right panel plots over and underreaction over persistence in forecasting. Increasing persistence decreases signal strength in inference and increases signal strength in forecasting. Both plots are therefore consistent with more overreaction to weaker signals and more underreaction to stronger signals.



a positive relationship with persistence in inference and a negative one in forecasting. These results are consistent with people overreacting to weak signals and underreacting to strong signals.

Turning to variation in over and underreaction with respect to changes in the horizon being forecasted, Figure 10 shows that people underreact more at shorter horizons, and overreact more at longer horizons. Since signal strength is decreasing with the horizon being forecasted, these results are also consistent with more overreaction to weaker signals.

Figure 10: Over and underreaction in forecasting as a function of horizon. Increasing the horizon being forecasted, decreases signal strength. We see less overreaction when the signal is strong and more overreaction when the signal is weak.



Together, these results illustrate how mapping features that are relevant in forecasting

(paeristence and horizon) to features that are relevant in inference (signal strength) allows us to understand variation in over and underreaction within both domains through a common psychological mechanism.

The Role of Attention. So far we have focused on how variation in signal strength through a single (neglected) feature leads to overreaction to weak signals and underreaction to strong signals. Next, we test Prediction 3 by studying how the *amount* of variation in over and underreaction can be explained by variation in different features as a function of attention. This allows us to explore which features (if any) are likely to drive more or less variation in over and underreaction. To do so, we focus on two features that differ in how much attention is allocated to each: persistence, which mostly neglected, and shock size, which attracts more attention.

To draw a direct comparison between the impact of persistence and shock size on signal strength, we focus on combinations of persistence and shock size that generate the same variation in signal strength.[20] Figure 11 illustrates one such example in inference. Both of these shocks lead to the same Bayesian belief revision. Trial 1 and Shock 1 (left panel) has a greater shock size, but the underlying data generating process is more persistent, meaning that the influence of shock size on signal strength is tempered by higher persistence. Conversely, Trial 3 and Shock 3 (right panel) has a smaller shock size, but lower persistence, so that the influence of shock size on signal strength is more pronounced.

The left panel of Figure 12 shows that, fixing the Bayesian revision, people revise their beliefs more when the size of the shock is larger, compared to the case where the same variation in signal strength is achieved via lower persistence. In other words, the amount by which people revise their beliefs is more sensitive to variation in the size of the shock than it is to variation in persistence.[21]

The right panel of Figure 12 shows how this then translates into a larger variation in over and underreaction when variation in signal strength is due to persistence (which is a more neglected feature) than when that same variation in signal strength is due to the size of the shock (the more attended-to feature).

[20] When we designed the experiment, we did not know the cognitive default levels of persistence and shock size. Therefore, we picked combinations of persistence and shock size that generated the same variation in signal strength, while at the same time maximizing the variation we could obtain in signal strength itself. Therefore, for this exercise, we vary persistence while setting shock size to its largest value, and we vary shock size while fixing persistence at 0.1 in inference and 0.9 in forecasting.

[21] All differences in slopes discussed in the text are highly significant using statistical tests, with $p < .001$ for both inference and forecasting domains.

Figure 11: Same posterior but different combinations of shock size and persistence. The left panel corresponds to Trial 1 Shock 1, with a larger shock, and higher persistence of $\rho = 0.9$. The right panel corresponds to Trial 3 Shock 1, with a smaller shock, and a lower persistence of $\rho = 0.1$. Both shocks lead to the same Bayesian revision.

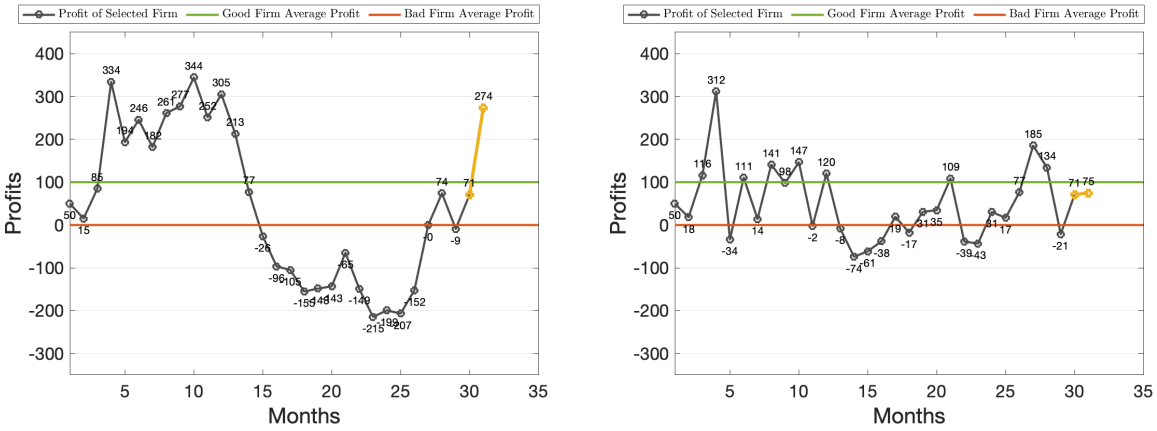
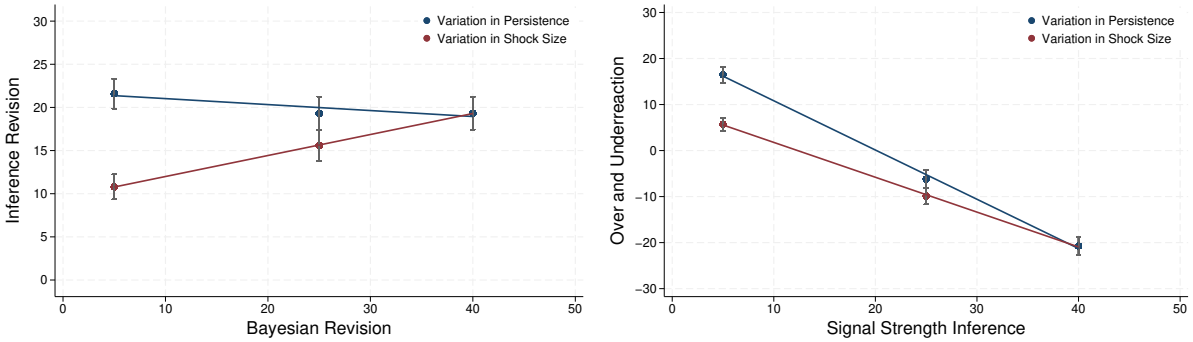


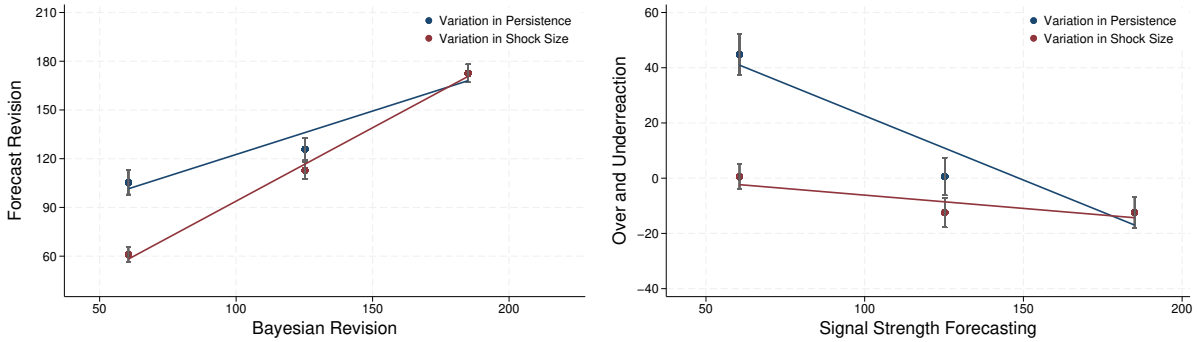
Figure 12: Variation in inference revision and over versus underreaction as shock size and persistence are varied.



The same patterns are even more accentuated in forecasting. The left panel of Figure 13 shows that forecast revisions are more sensitive to variation in shock size rather than the persistence of the process. The right panel of Figure 13 shows that variation in persistence drives generates more variation in under and overreaction compared to shock size, given the same impact on signal strength.

These results are consistent with Prediction 3, which states that variation in under and overreaction due to signal strength is modulated by attention. Given the same variation in signal strength, there will be more variation in over and underreaction if the former is coming from changes in the more neglected feature. Notably, this allows for measuring the allocation of attention across features just from observing belief revisions in response

Figure 13: Variation in forecast revision and in over and underreaction as shock size and persistence are varied.



to their variation, holding signal strength constant.

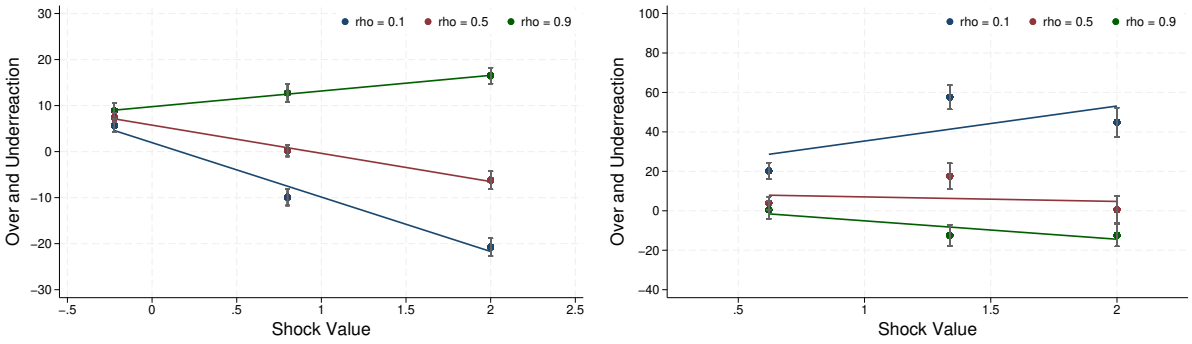
Result 3. The more a feature is neglected, the more it drives variation in under and overreaction. Persistence is neglected more than the size of the shock, and Figures 12 and 13 show that variation in persistence leads to more variation in under and overreaction than does variation in the size of the shock.

Interaction of multiple features. We now explore how variation in belief-updating depends on the interaction of attended to and neglected features. The left panel of Figure 14 plots the extent of over and underreaction in inference for all trials as a function of shock size. Fixing persistence, a larger shock size increases signal strength. When objective persistence is relatively high ($\rho = 0.9$ and $\rho = 0.5$), stronger signals lead to more underreaction, similar to before. However, consistent with Prediction 4, when objective persistence is low ($\rho = 0.1$)—lower than the cognitive default—stronger signals result in *more* overreaction.

Similarly, the right panel of Figure 14 illustrates a similar set of patterns for forecasting. However, consistent with Prediction 4 and the intuitions we provided in Section 4.4.2, more overreaction to stronger signals now arises for low levels of persistence ($\rho = 0.1$).

Result 4. The relationship between belief-updating and signal strength is modulated by attention. The standard comparative static of greater over (under) reaction to weaker (stronger) signals can be reversed due to the interaction of attended to and neglected features. Given the greater neglect of persistence relative to shock size, in inference

Figure 14: Over and underreaction as a function of shock size, for different persistence levels. The left left panel shows results in inference: when ρ is high we have more overreaction for stronger signals. The right panel shows results in forecasting: when ρ low we have more overreaction for stronger signals.



people overreact more to stronger signals when persistence is high, and in forecasting they overreact more to stronger signals when persistence is low.

6 Testing the Mechanism

In this section we more directly test the mechanisms of our framework. We do so by exogenously manipulating the relevant psychological parameters of our model (attention and cognitive defaults), while keeping the underlying structure and decision environment constant. Specifically, the first study exogenously manipulates the allocation of attention across features to explain within-domain variation in belief formation. The second study exogenously manipulates context to identify its role in the cueing of different cognitive defaults.

6.1 Manipulating Attention

We manipulate the allocation of attention across features by modifying the design in the main experiment to make persistence more top of mind. Specifically, participants ($N = 295$) first read the information about the design and passed the same comprehension checks as in the baseline study. We then gave these participants examples of data generating processes with low and high persistence, and asked them to write down a brief description of processes from their everyday lives that either had low or high persistence. All participants then answered the same set of forecasting questions as in the

main study.[22] This intervention was meant to draw participant's attention to variation in persistence. Notably, if the stock of attention is limited as our framework presumes, then greater attention to persistence should lead to less attention to other features of the problem—in our case, the size of the shock.

As outlined in Prediction 3, greater attention to a given feature is predicted to generate *less* variation in over versus underreaction as a function of objective changes in that feature. The results that follow compare variation in θ as a function of either persistence or shock size. To test Prediction 3 directly, in all cases that follow, we examine variation in over and underreaction in the parameter that was more neglected in the Baseline study ($x_i = \rho$) while setting the value of the other parameter (e.g., shock size) closest to the estimated cognitive default, $x_{-i} = \bar{x}_{-i}$.

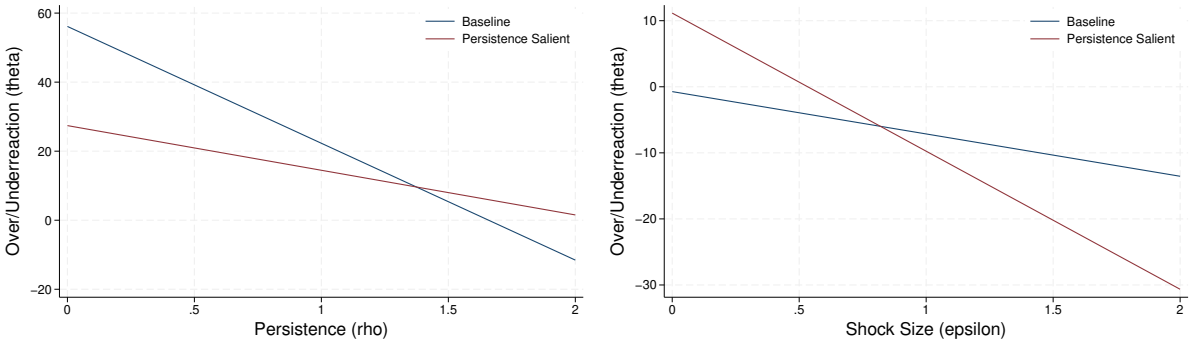
We first look at variation in θ as a function of persistence, and compare results from the Persistence Salient condition to the Baseline case. The left panel of Figure 15 shows that variation in over and underreaction with respect to persistence decreased substantially when more attention was allocated to it. We can compare the slopes across the baseline and Persistence Salient treatment by regressing θ on the underlying persistence parameter, a treatment variable, and their interaction. As shown in Column 1 of Table A.1, the positive coefficient on the interaction suggests that increasing attention to persistence flattened the slope relative to the baseline case. This lends support to our core prediction that variation in over and underreaction is driven by limited attention to variation in objective features.

Next, the right panel of Figure 15 looks at variation in θ across treatments, as a function of shock size. In this case, we see that variation in over and underreaction *increased* when more attention was allocated to persistence; the negative interaction coefficient in Column 2 of Table A.1 confirms the steeper slope between shock size and our measure of over and underreaction, meaning that the same variation in shock size now leads to more variation in over and underreaction. This provides support for the limited-stock-of-attention assumption in our framework: more attention allocated to persistence means that less attention was allocated to variation in shock size.

Consistent with our mechanism, we find that increasing attention to persistence decreases attention to the size of the shock. Relative to our baseline experiment, this decreases

[22] We focused on forecasting for this study because, as shown in Table 3, attention to persistence was fully neglected in inference. Our manipulation was meant to shift attention across features on the intensive margin. A fully neglected feature may not be part of the agent's mental representation at all, so a salience manipulation would have no scope to manipulate its impact on belief formation.

Figure 15: Comparing variation in over and underreaction in the Baseline and Persistence Salient experiments. The left panel plots variation in over versus underreaction as a function of persistence. The right panel plots variation in over and udnerreaction as a function of shock size. The blue lines plot these relationships for the Baseline experiment, while the red lines plot these relationships for the Persistence Salient experiment. Increasing attention to persistence decreases variation in over and underreaction with respect to persistence, and increases variation in over and underreaction with respect to shock size.



variation in under and overreaction when varying persistence, and increases variation in under and overreaction when varying the shock size.

6.2 Manipulating the Cognitive Default

To test the impact of context on cognitive defaults—which determines variation in belief formation across domains—we asked participants to report the 'expected long run profits' $\mathbb{E}_t[\mu]$ of the selected firm both in the inference and in the forecasting survey.

When answering this question, people may use the context of the problem to represent it as either an inference task or as a forecasting task. Specifically, when the question about $\mathbb{E}_t[\mu]$ is included in the inference survey, it can cue an inference question, leading people to frame it as "what are the firm's expected profits?" Instead, when the same question is included in the forecasting survey, it can cue a forecasting question, leading people to frame it as "what will profits be in the very long run?" This leads to the prediction that the estimated default level of persistence from answers to the same question will be higher when it is included in the forecasting survey, than when it is included in the inference survey.

This is indeed what we find. When we include the expected profits question in the inference survey we recover a default persistence level of $\bar{\rho}_{\text{inference}} = 0.1$, while when we

include it in the forecasting survey we recover a higher estimate of $\bar{\rho}_{\text{inference}} = 0.9$. These results offer complementary evidence for the importance of context—here in the form of the inference versus forecasting domains—for cueing mental representations, which can then explain variation in belief formation across domains.

7 Discussion

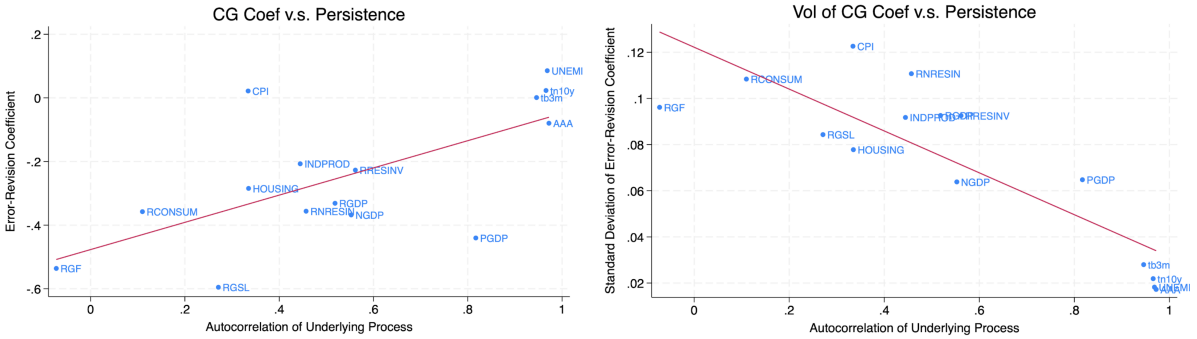
Besides the research reviewed in Section 1, we build on a large body of work on belief-updating across different environments. Starting from forecasting, empirical evidence points towards predominant overreaction to information, with more over-reaction at longer horizons, and for more transitory variables. Bordalo et al. (2020b) find evidence of over-reaction in most macro and financial variables, with the exception of short-term interest rates and unemployment. Afrouzi et al. (2023) find that the following patterns are robust in experiments as well: i) over-reaction to most recent observations; ii) more persistent series have less over-reaction; iii) more under-reaction at short horizons and over-reaction at long horizons. Da Silveira et al. (2020) rationalize this evidence with a model of optimal decision making subject to a memory constraint. Giglio and Kelly (2018) compare prices of claims with the same cash flow stream but with different maturities. They find that long-maturity claims are significantly more variable than justified by the behavior of short maturities, with an implied persistence in short maturity claims that is lower than the implied persistence in long maturity claims. Similarly, Wang (2021) finds underreaction for interest rates at short-maturity and overreaction for long-maturity interest rates, with the switch from under to overreaction showing up around the two year maturity. Finally, **?** show that analyst short-run earnings per share expectations (1 or 2 years ahead) exhibit underreaction, while their expectations of long-run earnings growth (3-5 years) exhibit overreaction.[23]

Beyond the predictions of these papers on the *relative level* of over and underreaction, our framework also predicts that *variation* in over and underreaction should change with the underlying parameters that drive signal strength. The left panel of Figure 16 plots

[23]For more evidence of over and underreaction in expectations in the context of macro and financial variables, see De Bondt and Thaler (1990), La Porta (1996), Coibion and Gorodnichenko (2012), Amromin and Sharpe (2014), Greenwood and Shleifer (2014), Gennaioli et al. (2016), Bordalo et al. (2019, 2020a), Bouchaud et al. (2019), Kwon and Tang (2020), Angeletos et al. (2021), De La O and Myers (2021), Jiang et al. (2022), Kohlhas and Walther (2021), Nagel and Xu (2022), Bordalo et al. (2024b,a,d), among others. For evidence of over and underreaction in forecasting experiments, see Asparouhova et al. (2009), Reimers and Harvey (2011), Assenza et al. (2014), Beshears et al. (2013), Frydman and Nave (2017), among others.

Coibion-Gorodnichenko (CG) regression coefficients on forecasts of macro and financial variables, over the persistence of the corresponding objective series. This replicates results in Bordalo et al. (2020b) and shows that professional forecasters overreact to information more when the series they are forecasting is more transitory. In other words, there is a relationship between the *level* of over versus underreaction and the persistence of the underlying series. The right panel of Figure 16 extends this analysis and plots the volatility of CG regression coefficients computed over 8 quarter rolling window regressions. We see that more transitory series (for which persistence is further from people's cognitive default level of 0.9) exhibit more *variation* in over and underreaction. We leave a more formal empirical investigation of the relationship between variation in over versus underreaction and shock size to future research.

Figure 16: Over and Underreaction in the Survey of Professional Forecasters. The left panel plots Coibion-Gorodnichenko (CG) regression coefficients on forecasts of macro and financial variables, over the persistence of the corresponding objective series. This replicates results in Bordalo et al. (2020b) and shows that professional forecasters overreact to information more when the series they are forecasting is more transitory. The right panel extends this analysis and plots the volatility of CG regression coefficients computed over 8 quarter rolling window regressions. We see that more transitory series (for which persistence is further from people's cognitive default level of 0.9) also exhibit more *variation* in over and underreaction.



The evidence on inference in surveyed in Benjamin (2019) highlights that most prior work has found underreaction to information. Griffin and Tversky (1992) argue that agents put too much emphasis on the strength of signals and too little emphasis on the weight of signal. They find over-reaction when the signal strength is high and weight is low (and also when diagnosticity and base rate are low). Massey and Wu (2005) put forward a "system neglect hypothesis," and consistent with Griffin and Tversky (1992) argue that individuals react primarily to the signals they observe and secondarily to the environment that produces the signal, which may be unstable. This is related conceptually to our finding that people tend to neglect the persistence of a process. Prat-Carrabin and Woodford (2024) find that when participants are shown signals sequentially, they

underreact to the first few signals, but then overreact once a sufficient number of signals have been observed. Finally, Enke and Graeber (2023) show that people with higher cognitive uncertainty are more insensitive to parameters of the decision problem.

8 Conclusions

We develop a common framework that unifies the traditionally separate domains of learning in inference and forecasting tasks. By incorporating forecasting elements like persistence and horizon, and inference elements such as signal strength, our model shows how different features of data generating processes influence people's responses to information. The experimental results provide support for the theoretical predictions, demonstrating that underinference and overreaction are not isolated phenomena but are interconnected through common psychological mechanisms. This integration helps to reconcile inconsistencies found in previous studies and enhances our understanding of the psychology behind belief-updating across different contexts.

Looking forward, a significant avenue for future research will involve a deeper investigation into the role of attention in belief updating across different environments. Attention to various features of information, such as the salience and prominence of information, plays a crucial role in how individuals process and react to it. Future studies could explore how manipulating the salience of certain features affects belief updating, potentially using eye-tracking technology or neuroimaging to assess which elements capture attention and how they influence the subsequent judgment process.

Additionally, further research could examine the effects of attention in more varied belief updating environments, such as those involving multiple interacting signals or longer time horizons. Experiments could be designed to vary the presentation of information to determine how these factors influence the selective attention to features that are critical for forecasting and inference tasks. By systematically studying these interactions, researchers can better understand how cognitive processing limitations and attentional biases contribute to the deviations from rational behavior observed in studies of learning.

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.

Amromin, G. and S. A. Sharpe (2014): "From the horse's mouth: Economic conditions and investor expectations of risk and return," *Management Science*, 60, 845–866.

Angeletos, G.-M., Z. Huo, and K. A. Sastry (2021): "Imperfect macroeconomic expectations: Evidence and theory," *NBER Macroeconomics Annual*, 35, 1–86.

Asparouhova, E., M. Hertzel, and M. Lemmon (2009): "Inference from streaks in random outcomes: Experimental evidence on beliefs in regime shifting and the law of small numbers," *Management Science*, 55, 1766–1782.

Assenza, T., T. Bao, C. Hommes, and D. Massaro (2014): "Experiments on expectations in macroeconomics and finance," in *Experiments in macroeconomics*, Emerald Group Publishing Limited, vol. 17, 11–70.

Augenblick, N., E. Lazarus, and M. Thaler (2021): "Overinference from weak signals and underinference from strong signals," *arXiv preprint arXiv:2109.09871*.

Ba, C., J. A. Bohren, and A. Imas (2022): "Over-and underreaction to information," *Available at SSRN 4274617*.

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

Beshears, J., J. J. Choi, A. Fuster, D. Laibson, and B. C. Madrian (2013): "What goes up must come down? Experimental evidence on intuitive forecasting," *American Economic Review*, 103, 570–574.

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

Bordalo, P., J. J. Conlon, N. Gennaioli, S. Y. Kwon, and A. Shleifer (2023a): "How people use statistics," Tech. rep., National Bureau of Economic Research.

——— (2023b): "How people use statistics," Tech. rep., National Bureau of Economic Research.

Bordalo, P., N. Gennaioli, R. La Porta, M. OBrien, and A. Shleifer (2024a): "Long-Term Expectations and Aggregate Fluctuations," *NBER Macroeconomics Annual*, 38, 311–347.

Bordalo, P., N. Gennaioli, R. La Porta, and A. Shleifer (2020a): *Expectations of fundamentals and stock market puzzles*, National Bureau of Economic Research.

——— (2024b): "Finance without exotic risk," Tech. rep., National Bureau of Economic Research.

Bordalo, P., N. Gennaioli, G. Lanzani, and A. Shleifer (2024c): "A Cognitive Theory of Reasoning and Choice," .

Bordalo, P., N. Gennaioli, Y. Ma, and A. Shleifer (2020b): "Overreaction in macroeconomic expectations," *American Economic Review*, 110, 2748–2782.

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

——— (2024d): "Belief overreaction and stock market puzzles," *Journal of Political Economy*, 132, 1450–1484.

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

Coibion, O. and Y. Gorodnichenko (2012): "What can survey forecasts tell us about information rigidities?" *Journal of Political Economy*, 120, 116–159.

Da Silveira, R. A., Y. Sung, and M. Woodford (2020): "Optimally imprecise memory and biased forecasts," Tech. rep., National Bureau of Economic Research.

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

De Bondt, W. F. and R. H. Thaler (1990): "Do security analysts overreact?" *The American economic review*, 52–57.

De La O, R. and S. Myers (2021): "Subjective cash flow and discount rate expectations," *The Journal of Finance*, 76, 1339–1387.

Edwards, W. (1968): "Conservatism in human information processing," *Formal representation of human judgment*.

Enke, B. and T. Graeber (2023): "Cognitive uncertainty," *The Quarterly Journal of Economics*, 138, 2021–2067.

Enke, B., T. Graeber, R. Oprea, and J. Yang (2024a): "Behavioral attenuation," Tech. rep., National Bureau of Economic Research.

Enke, B., F. Schwerter, and F. Zimmermann (2024b): "Associative memory, beliefs and market interactions," *Journal of Financial Economics*, 157, 103853.

Fan, T., Y. Liang, and C. Peng (2021): "The inference-forecast gap in belief updating," *Unpubished Manuscript*.

Friston, K. (2008): "Hierarchical models in the brain," *PLoS computational biology*, 4, e1000211.

Frydman, C. and G. Nave (2017): "Extrapolative beliefs in perceptual and economic decisions: Evidence of a common mechanism," *Management Science*, 63, 2340–2352.

Gabaix, X. (2019): "Behavioral inattention," in *Handbook of behavioral economics: Applications and foundations 1*, Elsevier, vol. 2, 261–343.

Gennaioli, N., Y. Ma, and A. Shleifer (2016): "Expectations and investment," *NBER Macroeconomics Annual*, 30, 379–431.

Giglio, S. and B. Kelly (2018): "Excess volatility: Beyond discount rates," *The Quarterly Journal of Economics*, 133, 71–127.

Greenwood, R. and A. Shleifer (2014): "Expectations of returns and expected returns," *The Review of Financial Studies*, 27, 714–746.

Griffin, D. and A. Tversky (1992): "The weighing of evidence and the determinants of confidence," *Cognitive psychology*, 24, 411–435.

Jiang, Z., H. Liu, C. Peng, and H. Yan (2022): "Investor memory and biased beliefs: Evidence from the field," *Available at SSRN*.

Kohlhas, A. N. and A. Walther (2021): "Asymmetric attention," *American Economic Review*, 111, 2879–2925.

Kwon, S. Y. and J. Tang (2020): "Extreme events and overreaction to news," *Available at SSRN 3724420*.

La Porta, R. (1996): "Expectations and the cross-section of stock returns," *The Journal of Finance*, 51, 1715–1742.

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.

Malmendier, U. and S. Nagel (2011): "Depression babies: do macroeconomic experiences affect risk taking?" *The quarterly journal of economics*, 126, 373–416.

——— (2016): "Learning from inflation experiences," *The Quarterly Journal of Economics*, 131, 53–87.

Markman, A. B. (2013): *Knowledge representation*, Psychology Press.

Massey, C. and G. Wu (2005): "Detecting regime shifts: The causes of under-and overreaction," *Management Science*, 51, 932–947.

Nagel, S. and Z. Xu (2022): "Asset pricing with fading memory," *The Review of Financial Studies*, 35, 2190–2245.

Nosofsky, R. M., J. K. Kruschke, and S. C. McKinley (1992): "Combining exemplar-based category representations and connectionist learning rules." *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18, 211.

Oberauer, K. (2019): "Working Memory and Attention–A Conceptual Analysis and Review," *Journal of Cognition*, 2.

Prat-Carrabin, A. and M. Woodford (2024): "Imprecise probabilistic inference from sequential data." *Psychological Review*.

Reimers, S. and N. Harvey (2011): "Sensitivity to autocorrelation in judgmental time series forecasting," *International Journal of Forecasting*, 27, 1196–1214.

Wang, C. (2021): "Under-and overreaction in yield curve expectations," *Available at SSRN 3487602*.

A Additional Figures and Tables

Table A.1 show results from regressing our measure of over and underreaction over persistence (column 1) or shock size (column 2) across our Baseline and Persistence Salient experiments.

Table A.1: Allocating Attention Across Features

	(1)	(2)
ρ	-33.843***	
	(5.157)	
ϵ		-6.406
		(4.541)
Persistence Salient (dummy)	-28.730**	
	(13.522)	
Interaction	20.906**	-14.493*
	(9.095)	(8.552)
Constant	56.139***	-0.732
	(7.840)	(5.330)
Observations	260	279
R-squared	0.146	0.042

Notes: Standard errors in parentheses. Column (1) regresses θ on ρ, the Persistence Salient treatment dummy, and their interaction. The value of ϵ is held constant at the cognitive default (medium shock). Column (2) regresses θ on ϵ, the Persistence Salient treatment dummy, and their interaction. The value of ρ is held constant at the cognitive default (high persistence). *** p<0.01, ** p<0.05, * p<0.1.

B Proofs and Derivations

B.1 Rational Benchmark - Log Likelihood Derivations

Finite Number of States — iid shocks with continuous support

Assume that $y_t \sim^{iid} N(\mu, \sigma^2)$, or similarly, we can think of y_t as being generated by the

following DGP:

$$y_t = (1-\rho)\mu + \rho y_{t-1} + \epsilon_t \tag{19}$$

where $\rho = 0$ and $\epsilon_t \sim^{iid} N(0, \sigma^2)$.

The log likelihood function can be written as:

$$\log\left(\frac{P(\mu=\mu_A|y_0,y_1,\ldots,y_t)}{P(\mu=\mu_B|y_0,y_1,\ldots,y_t)}\right) = \log\left(\frac{P(y_0,y_1,\ldots,y_t|\mu=\mu_A)}{P(y_0,y_1,\ldots,y_t|\mu=\mu_B)}\right) + \log\left(\frac{P(\mu=\mu_A)}{P(\mu=\mu_B)}\right) \tag{20}$$

so we need to understand how to compute $P(y_0, y_1, \ldots, y_t|\mu = \mu_A)$, which we compute as follows.

$$P(y_0,y_1,\ldots,y_t|\mu=\mu_A) = P(y_t|y_{t-1},\mu=\mu_A) \times P(y_{t-1}|y_{t-2},\mu=\mu_A) \times \ldots \\ \times P(y_1|y_0,\mu=\mu_A) \times P(y_0|\mu=\mu_A) \tag{21}$$

$$P(y_0,y_1,\ldots,y_t|\mu=\mu_A) = N(\mu_A,\sigma^2;y_t) \times N(\mu_A,\sigma^2;y_{t-1}) \times \ldots \\ \times N(\mu_A,\sigma^2;y_1) \times N(\mu_A,\sigma^2;y_0) \tag{22}$$

$$P(y_0,y_1,\ldots,y_t|\mu=\mu_A) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(y_t-\mu_A)^2}{2\sigma^2}\right) \times \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(y_{t-1}-\mu_A)^2}{2\sigma^2}\right) \times \ldots \\ \times \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(y_1-\mu_A)^2}{2\sigma^2}\right) \times \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(y_0-\mu_A)^2}{2\sigma^2}\right) \tag{23}$$

$$P(y_0,y_1,\ldots,y_t|\mu=\mu_A) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^{t+1}\exp\left(-\frac{(y_t-\mu_A)^2+(y_{t-1}-\mu_A)^2+\ldots+(y_0-\mu_A)^2}{2\sigma^2}\right) \tag{24}$$

$$= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^{t+1}\exp\left(-\frac{\sum_{i=0}^{t}(y_i-\bar{y})^2+(t+1)(\bar{y}-\mu_A)^2}{2\sigma^2}\right) \tag{25}$$

We can then substitute this into the log-likelihood to obtain:[24,25]

$$\log\left(\frac{P(\mu=\mu_A|y_0,y_1,\ldots,y_t)}{P(\mu=\mu_B|y_0,y_1,\ldots,y_t)}\right) = -\frac{(t+1)}{2\sigma^2}\left((\bar{y}-\mu_A)^2-(\bar{y}-\mu_B)^2\right)+\log\left(\frac{P(\mu=\mu_A)}{P(\mu=\mu_B)}\right) \tag{30}$$

$$= (t+1)\left(\bar{y}-\frac{\mu_A+\mu_B}{2}\right)\left(\frac{\mu_A-\mu_B}{\sigma^2}\right)+\log\left(\frac{P(\mu=\mu_A)}{P(\mu=\mu_B)}\right) \tag{31}$$

which gives us the expression we are after:

$$\log\left(\frac{P(\mu=\mu_A|y_0,y_1,\ldots,y_t)}{P(\mu=\mu_B|y_0,y_1,\ldots,y_t)}\right) = \underbrace{(t+1)}_{\text{weight}}\underbrace{\left(\bar{y}-\frac{\mu_A+\mu_B}{2}\right)}_{\text{strength}}\underbrace{\left(\frac{\mu_A-\mu_B}{\sigma^2}\right)}_{\text{diagnosticity}}+\underbrace{\log\left(\frac{P(\mu=\mu_A)}{P(\mu=\mu_B)}\right)}_{\text{base rate}} \tag{32}$$

Finite Number of States — persistent shocks with continuous support

$$P(y_0,y_1,\ldots,y_t|\mu=\mu_A) = N((1-\rho)\mu_A+\rho y_{t-1},\sigma^2;y_t)\times N((1-\rho)\mu_A+\rho y_{t-2},\sigma^2;y_{t-1})\times\ldots$$
$$\times N((1-\rho)\mu_A+\rho y_0,\sigma^2;y_1)\times N\left(\mu_A,\frac{\sigma^2}{1-\rho^2};y_0\right) \tag{33}$$

$$P(y_0,y_1,\ldots,y_t|\mu=\mu_A) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(y_t-((1-\rho)\mu_A+\rho y_{t-1}))^2}{2\sigma^2}\right)$$
$$\times\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(y_{t-1}-((1-\rho)\mu_A+\rho y_{t-2}))^2}{2\sigma^2}\right)\times\ldots$$
$$\times\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(y_1-((1-\rho)\mu_A+\rho y_0))^2}{2\sigma^2}\right)\times\frac{1}{\sqrt{2\pi\frac{\sigma^2}{1-\rho^2}}}\exp\left(-\frac{(y_0-\mu_A)^2}{2\frac{\sigma^2}{(1-\rho^2)}}\right) \tag{34}$$

$$P(y_0,y_1,\ldots,y_t|\mu=\mu_A) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{\left(\frac{y_t-\rho y_{t-1}}{1-\rho}-\mu_A\right)^2}{2\frac{\sigma^2}{(1-\rho)^2}}\right)$$

[24]The second equality in the above expression comes from adding and subtracting $\bar{y}$ in each term and then rearranging:

$$(y_t-\mu_A)^2+\ldots+(y_0-\mu_A)^2 = ((y_t-\bar{y})+(\bar{y}-\mu_A))^2+\ldots+((y_0-\bar{y})+(\bar{y}-\mu_A))^2 \tag{26}$$

$$= (y_t-\bar{y})^2+\ldots(y_0-\bar{y})^2+2(\bar{y}-\mu_A)(\underbrace{(y_t+\ldots+y_0)}_{n\bar{y}}-n\bar{y})+(t+1)(\bar{y}-\mu_A)^2 \tag{27}$$

[25]The second equality in the following expression uses the following:

$$(\bar{y}-\mu_A)^2-(\bar{y}-\mu_B)^2 = (\bar{y}^2-2\bar{y}\mu_A+\mu_A^2)-(\bar{y}^2-2\bar{y}\mu_B+\mu_B^2) \tag{28}$$

$$= -2\bar{y}(\mu_A-\mu_B)+(\mu_A-\mu_B)(\mu_A+\mu_B) = -2(\mu_A-\mu_B)\left(\bar{y}-\frac{\mu_A+\mu_B}{2}\right) \tag{29}$$

$$\times \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\left(\frac{y_{t-1}-\rho y_{t-2}}{1-\rho} - \mu_A\right)^2}{2\frac{\sigma^2}{(1-\rho)^2}}\right)$$
$$\times \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\left(\frac{y_1-\rho y_0}{1-\rho} - \mu_A\right)^2}{2\frac{\sigma^2}{(1-\rho)^2}}\right) \times \frac{1}{\sqrt{2\pi\frac{\sigma^2}{1-\rho^2}}} \exp\left(-\frac{(y_0-\mu_A)^2}{2\frac{\sigma^2}{(1-\rho^2)}}\right) \quad (35)$$

Let $s_i \equiv \frac{y_i-\rho y_{t-1}}{1-\rho} = \mu + \frac{\epsilon_i}{1-\rho}$ be the unbiased signal of the underlying mean that respondents observe in each period. Using similar steps to those we used to solve for the case where $\rho = 0$, we have that:[26]

$$\log\left(\frac{P(\mu=\mu_A|y_0,y_1,\ldots,y_t)}{P(\mu=\mu_B|y_0,y_1,\ldots,y_t)}\right) = \underbrace{t}_{\text{weight}} \underbrace{\left(\frac{(1-\rho)\,\bar{s} - (1-\rho)\,\frac{\mu_A+\mu_B}{2}}{\sigma}\right)}_{\text{strength}} \underbrace{\left(\frac{\mu_A-\mu_B}{\frac{\sigma}{1-\rho}}\right)}_{\text{diagnosticity}}$$
$$+ \underbrace{1}_{\text{weight}} \underbrace{\left(y_0 - \frac{\mu_A+\mu_B}{2}\right)}_{\text{strength}} \underbrace{\left(\frac{\mu_A-\mu_B}{\frac{\sigma^2}{1-\rho^2}}\right)}_{\text{diagnosticity}} + \underbrace{\log\left(\frac{P(\mu=\mu_A)}{P(\mu=\mu_B)}\right)}_{\text{base rate}} \quad (36)$$

where $\bar{s} = \sum_{i=1}^{t} \frac{y_i-\rho y_{t-1}}{1-\rho}$. We can also re-write this in terms of the underlying observations rather than in terms of the signals:

$$\log\left(\frac{P(\mu=\mu_A|y_0,y_1,\ldots,y_t)}{P(\mu=\mu_B|y_0,y_1,\ldots,y_t)}\right) = \underbrace{\left(\frac{y_t - \frac{\mu_A+\mu_B}{2}}{\sigma}\right)}_{\text{strength}} \underbrace{\left(\frac{\mu_A-\mu_B}{\frac{\sigma}{1-\rho}}\right)}_{\text{diagnosticity}}$$
$$+ \sum_{i=1}^{t-1} \underbrace{(1-\rho)\left(\frac{y_i - \frac{\mu_A+\mu_B}{2}}{\sigma}\right)}_{\text{strength}} \underbrace{\left(\frac{\mu_A-\mu_B}{\frac{\sigma}{1-\rho}}\right)}_{\text{diagnosticity}}$$
$$+ \underbrace{\left(\frac{y_0 - \frac{\mu_A+\mu_B}{2}}{\sigma}\right)}_{\text{strength}} \underbrace{\left(\frac{\mu_A-\mu_B}{\frac{\sigma}{1-\rho}}\right)}_{\text{diagnosticity}} + \underbrace{\log\left(\frac{P(\mu=\mu_A)}{P(\mu=\mu_B)}\right)}_{\text{base rate}} \quad (37)$$

Notice how persistence affects both signal strength and diagnosticity.

[26] Where we also used the fact that $\frac{\sigma^2}{(1-\rho)^2} = \left(\frac{\sigma^2}{1-\rho^2}\right)\left(\frac{1+\rho}{1-\rho}\right)$ so we can write diagnosticity in terms of the unconditional variance.

B.2 Continuous Support and Maximum Likelihood

While the binary state is helpful in connecting to the balls and urns experiments used in the inference literature, forecasters are usually not constrained by this binary state characterization. In the next section, we look at how signal strength varies in inference and forecasting when we allow for a continuous state, and the optimal forecast involves maximum likelihood estimation.

Signal Strength in Inference. Instead of asking whether the mean is μ_A or μ_B, an inference task in this scenario could involve asking whether the mean is above or below some threshold M. For example, it could be that good firms have a mean higher than M and bad firms have a mean lower than M.

$$\frac{\pi_{G|y}}{\pi_{B|y}} = \frac{\Phi\left(\frac{\hat{\mu}_{MLE}-M}{\frac{1}{\sqrt{T}}\frac{\sigma}{1-\rho}}\right)}{1-\Phi\left(\frac{\hat{\mu}_{MLE}-M}{\frac{1}{\sqrt{T}}\frac{\sigma}{1-\rho}}\right)} = \frac{\Phi\left(\frac{\mu-M}{\frac{1}{\sqrt{T}}\frac{\sigma}{1-\rho}}+\frac{\bar{\epsilon}_t}{\frac{1}{\sqrt{T}}}\right)}{1-\Phi\left(\frac{\mu-M}{\frac{1}{\sqrt{T}}\frac{\sigma}{1-\rho}}+\frac{\bar{\epsilon}_t}{\frac{1}{\sqrt{T}}}\right)} \tag{38}$$

where the last equality uses the expression for maximum likelihood estimation:

$$\hat{\mu}_{MLE} = \frac{1}{T}\sum_{i=1}^{t}\left(\frac{y_i-\rho y_{i-1}}{1-\rho}\right) = \mu + \frac{\sigma\bar{\epsilon}_t}{1-\rho} \tag{39}$$

Therefore, *signal strength in inference is decreasing in persistence.* This holds in the shock condition as well if we start from the same base rate.

Signal Strength in Forecasting. The optimal forecast is given by:

$$\mathbb{E}_t[y_{t+h}] = \rho^h y_t + (1-\rho^h)\mathbb{E}_t[\mu] = \rho^h y_t + (1-\rho^h)\left(\mu + \frac{\sigma\bar{\epsilon}_t}{1-\rho}\right) \tag{40}$$

Forecast revision in the shock condition is given by:

$$\mathbb{E}_t[y_{t+h}] - \mathbb{E}_{t-1}[y_{t+h}] = \rho^h\sigma\epsilon_t + (1-\rho^h)\left(\mathbb{E}_t[\mu] - \mathbb{E}_{t-1}[\mu]\right) \tag{41}$$

$$= \rho^h\sigma\epsilon_t + \left(\frac{1-\rho^h}{1-\rho}\right)\left(\frac{\sigma\epsilon_t}{t} - \frac{\sigma\bar{\epsilon}_{t-1}}{t}\right) \tag{42}$$

So we see that *signal strength in forecasting is increasing in persistence.*

C Categories and Attention

In this section we consider a model of categorization to account for the exact heuristics that are prevalent in our data. To do so, we build on Bordalo et al. (2024c), and adapt it to our setting.

C.1 Setup

A firm with average profits $\mu_\theta \in \mathbb{R}^+$ is selected according to a prior $\mathbb{P}_0$ in a set of firms Θ, and generates sequences of up to $T \in \mathbb{N}$ draws of monthly profits that evolve according to an AR(1) process: $y_t = \rho y_{t-1} + (1-\rho)\mu_\theta + \epsilon_t$, where $\epsilon_t \sim^{iid} N(0,1)$. Hypotheses $o \in O$ are events in the sampling space Ω, whose probability can be estimated in $(\Omega, \mathcal{F}, \mathbb{P})$. A hypothesis is characterized by an elementary event $\omega \in \Omega$ with choice features $y(\omega)$, together with a vector of context features k_t, as in Bordalo et al. (2024c).

Choice Features. Choice features $i \in M_O$ capture both features that relate to the data generating process, as well as relevant realizations:[27]

$$y(\omega) = (e_\theta, e_{\epsilon_T|\theta}, \rho, \epsilon_T) \tag{43}$$

where the first three features relate to properties of the DGP: e_θ is the event of selecting firm θ, $e_{\epsilon_T|\theta}$ is the event of the firm's profits in month T having a shock of ϵ_T, ρ is the autocorrelation of the firm's profits. Instead, the fourth feature relates to observed realizations: ϵ_T is the realized shock component of the selected firm's profits in month T. In general, realized features are more prominent than features that relate to the DGP, which tend to be more in the background and less salient.

Attention to choice features is given by a vector of weights $\alpha_O \in [0,1]^{M_O}$, that are subject to a budget constraint that is task specific, and need not add up to one. Feature $i \in \{e_\theta, e_{\epsilon_T|\theta}, \rho, \epsilon_T\}$ is fully attended to if $\alpha_i = 1$, it is partially attended to when $0 < \alpha_i < 1$, and it is edited out entirely if $\alpha_i = 0$. When a feature is edited out it takes the value associated with the DM's cognitive default for that feature, which in turn, is determined by the DM's categorization of a problem, as further discussed below.

[27] Connect to PO comment and draw clear distinction between features that relate to the DGP, features that relate to reliazations, features that relate to hypothesis under consideration, and features that relate to non-payoff relevant context.

The DM's mental representation of the choice feature vector is given by:

$$y(\alpha_O) = (e_\theta(\alpha_O), e_{\epsilon_T|\theta}(\alpha_O), \rho(\alpha_O), \epsilon_T(\alpha_O)) \tag{44}$$

where $e_\theta(\alpha_O)$ and $e_{\epsilon_T|\theta}(\alpha_O)$ can take any value if they are edited out, and where the attention-based perception of ρ and ϵ_T is given by:

$$\rho(\alpha_O) = \alpha_\rho \rho + (1 - \alpha_\rho)\bar{\rho} \tag{45}$$

$$\epsilon_T(\alpha_O) = \begin{cases} \alpha_\epsilon \epsilon_T + (1 - \alpha_\epsilon)\bar{\epsilon}_T & \text{if } \pi_{\theta|\epsilon_T} \geq \pi_\theta \\ \alpha_\epsilon \epsilon_T - (1 - \alpha_\epsilon)\bar{\epsilon}_T & \text{if } \pi_{\theta|\epsilon_T} < \pi_\theta \end{cases} \tag{46}$$

so that when evaluating the shock, the DM is correctly specified about its direction, but misperceives its magnitude, as in Augenblick et al. (2021). Moreover, the cognitive defaults are determined by

$$\bar{\rho} = \sum_{j=1}^{J} \left(\frac{S(n_j, N) \times Pr(n_j|Q)}{\sum_k S(n_k, N) \times Pr(n_k|Q)} \right) \rho_j \tag{47}$$

$$\bar{\epsilon}_T = \sum_{j=1}^{J} \left(\frac{S(n_j, N) \times Pr(n_j|Q)}{\sum_k S(n_k, N) \times Pr(n_k|Q)} \right) \epsilon_{T,j} \tag{48}$$

where $Pr(n_j|Q)$ captures how likely you are to encounter series n_j when tackling a problem Q, and $S(n_j, N)$ captures the similarity between a series n_j and the series named in the question N.

Attention to features and the cognitive defaults are determined by the DM's categorization of the problem, which depends on the current context.

Context Features. Context features $i \in M_K$ capture choice features, as well as additional features that relate to the hypothesis under consideration, and non payoff-relevant context features:

$$k = (y, q, n) \tag{49}$$

where y are the choice features, q captures features related to the problem under consideration such as its focus (e.g. whether it relates to firm type, or firm profits) its unit (e.g. probability or profit value) or other details (e.g. the horizon being forecasted, or the number of states under consideration), and n captures non payoff-relevant features such as the name of the sequence under consideration (e.g. firm profits). We can re-write this

as:

$$k = (\pi_\theta, \pi_{\epsilon_T|\theta}, \rho, \epsilon_T, q_{focus}, q_{unit}, q_\Theta, n) \tag{50}$$

Current context is given by k_t, and attention to context features is given by $\alpha_K \in [0,1]^{M_K}$. Context features affect categorization, which in turn determines the DM's attention to choice features.

Categories. Context features cue categorization. At time t, the DM's has access to a set of categories C. Each category is pinned down by its attention and context vectors (α_c, k_c), and by how frequently they acre encountered, $F_c \in \mathbb{R}_+$. We consider five categories, which have been documented empirically in prior work (Fan et al., 2021, Bordalo et al., 2023b). Three of these categories are relevant across all tasks: no update, base-rate neglect, and full inference. The remaining two categories are relevant in forecasting: exact representativeness, and naive extrapolation.

No Update. This category corresponds to answering the question: "given a randomly selected firm, what is the probability that the firm is good?" or "given a randomly selected firm, what do you expect average profits to be?" The context associated with it is given by:

$$k_{ai} = (\pi_\theta, \varphi, \varphi, \varphi, q_u \in \{\pi, y\}, q_f = \theta, q_\Theta = \{G, B\}, n) \tag{51}$$

where φ indicates that the corresponding feature in (58) is edited out by the DM.

Base-rate Neglect. This category corresponds to answering the question: "given a firm of type θ, what is the probability of observing a given sequence of profits?" or "given a firm of type θ, what do you expect next period profits to be?" The context associated with it is given by:

$$k_{ss} = (\varphi, \pi_{\epsilon_T|\theta}, \rho, \epsilon_T, q_u \in \{\pi, y\}, q_f = y, q_\Theta = \{\theta\}, n) \tag{52}$$

where φ indicates that the corresponding feature in (58) is edited out by the DM.

Full Inference. This category corresponds to answering the full question, taking both the base-rate and the provided data into account. The context associated with it is given by:

$$k_{ss} = (\pi_\theta, \pi_{\epsilon_T|\theta}, \rho, \epsilon_T, q_u \in \{\pi, y\}, q_f = y, q_\Theta = \{\theta\}, n) \tag{53}$$

Exact Representativeness. This category corresponds to answering the question: "is the firm good or bad?" The relevant context vector is:

$$k_{er} = (\pi_\theta, \pi_{\epsilon_T|\theta}, \rho, \epsilon_T, \varphi, q_u = y, q_f = \theta, q_\Theta \subseteq \mathbb{R}, n) \tag{54}$$

where π_θ is characterized by an ignorance prior over $\mathbb{R}$, and $\rho = 0$.

Naive Extrapolation. This category corresponds to answering the same forecasting question elicited in the experiment, but with a different set of features:

$$k_{ne} = (\varphi, \varphi, \rho = 1, \epsilon_T, \varphi, q_u = y, q_f = y, \varphi, n) \tag{55}$$

Categorization is determined based on the similarity of current context k_t, with the context of each of the possible categories. The more similar a category is to current context, the more likely it is to be used to solve the problem at hand.

In what follows, we first specify the context associated with the three tasks we elicit in our experiment: inference, forecasting, and expected long run profits. Next, compare how similar a given category is to each task. This allows us to make predictions as to the use of a given category across tasks: if a category is more similar to inference than forecasting, we expect the associated heuristic to be used more in inference than forecasting, and so on. To do so we use a simple distance metric: two contexts are more similar to each other, if they agree over a greater number of features.

C.2 Inference

In our setting, a firm's type can take one of two values, $\theta \in \{G, B\}$. An inference problem in this environment corresponds to eliciting the probability the firm is good, after observing one additional month of firm profits:

$$\pi_G|y \propto \left(\frac{\pi_G}{\pi_B}\right)\left(\frac{\pi_{y|G}}{\pi_{y|B}}\right) \tag{56}$$

Assuming without loss of generality that $\pi_G|y > \pi_{B|y}$, we can write the likelihood as:[28]

$$\frac{\pi_{y|G}}{\pi_{y|B}} = \exp\left(\left(\epsilon_T + \frac{1-\rho}{2}\right)(1-\rho)\right) \tag{57}$$

The relevant context vector for this problem is given by:

$$k_t = \left(\underbrace{\pi_\theta, \pi_{\epsilon_T|\theta}, \rho, \epsilon_T}_{y\text{: event features}}, \quad \underbrace{q_{\text{unit}} = \pi, q_{\text{focus}} = \theta, q_\Theta = \{G, B\}}_{q\text{: question features}}, \underbrace{n}_{n\text{: non payoff-relevant features}} \right) \tag{58}$$

where there are three question features q that relate to this task: q_u captures the unit of the problem, in this case a probability (as opposed to e.g. a profit value); q_f captures the focus of the problem, in this case the realized state (as opposed to e.g. realized profits); q_Θ captures the set of possible states considered in the problem, in our case good or bad (as opposed to e.g. many possible types of firms).

There are three categories of similar problems that the DM might attend to when solving this problem: agnostic inference, simple sampling, and full inference.

No Update. When agents fit the inference problem to this category, they use the following heuristic, which results in no update of prior beliefs:

$$\pi_{G|\theta} = \pi_G \tag{59}$$

Base-rate Neglect. When agents fit the inference problem to this category, they use the following heuristic.

$$\pi_G|y \propto \frac{\pi_{y|G}}{\pi_{y|B}} \tag{60}$$

where:

$$\frac{\pi_{y|G}}{\pi_{y|B}} = \exp\left(\left(\epsilon_T(\alpha_O) + \frac{1-\rho(\alpha_O)}{2}\right)(1-\rho(\alpha_O))\right) \tag{61}$$

[28]The full expression for the likelihood is given by:

$$\frac{\pi_{y|G}}{\pi_{y|B}} = \exp\left(\sum_{t=1}^{T}\left(\epsilon_t + \left(\frac{\mu_A - \mu_B}{2\sigma}\right)(1-\rho)\right)\left(\frac{\mu_A - \mu_B}{\sigma}\right)(1-\rho)\right).$$

To simplify, we use the fact that in our experiment $\frac{\mu_A - \mu_B}{\sigma} = 1$. Since these quantities are provided to the DM at every stage of the experiment, and they do not change, we assume that the DM knows them. Moreover, since in our experiment it is always the case that $\pi_{G|y_0,\ldots,y_{T-1}} = 0.5$, we subsume this into the prior, and focus on how the DM updates in light of a single new observation. The model can be extended to multiple observations as well, but we start with this case for simplicity.

and:

$$\alpha_\rho \geq 0 \qquad \bar{\rho} = \sum_{j=1}^{J} \left(\frac{S(n_j, N) \times Pr(n_j|Q)}{\sum_k S(n_k, N) \times Pr(n_k|Q)} \right) \rho_j \tag{62}$$

$$\alpha_\epsilon \geq 0 \qquad \bar{\epsilon}_T = \sum_{j=1}^{J} \left(\frac{S(n_j, N) \times Pr(n_j|Q)}{\sum_k S(n_k, N) \times Pr(n_k|Q)} \right) \epsilon_{T,j} \tag{63}$$

and $S(n_j, N)$ captures the similarity between a series n_j and the series named in the question N. For simplicity, we assume that the only series that come to mind when solving this problem are the most representative series in the population (given by the average value), and the series named in the question. This implies that $J = 2$. In our setting, people are used to solving these problems with iid draws ($\bar{\rho} = 0$), but they know profits are persistent ($\bar{\rho} > 0$), which leads to an intermediate default persistence.

Full Inference. Finally, the DM might attend to both the base rate and the likelihood, and answer the question: "what is the probability that the selected firm is good, given the observed sequence of profits?" Since in our experiment, we always ensure that the prior is uniform over states, full inference yields the same heuristic as simple sampling.

C.3 Forecasting

A forecasting problem in our environment corresponds to eliciting future period profits:

$$E[y_{t+1}|y] = \rho(\alpha_O)^{h(\alpha_O)} y_T + \left(1 - \rho(\alpha_O)^{h(\alpha_O)}\right) \mathbb{E}[\mu|y] \tag{64}$$

The relevant context vector for this problem is given by:

$$k_t = \left(\pi_\theta, \pi_{\epsilon_T|\theta}, \rho, \epsilon_T, h, q_u = y, q_f = y, q_\Theta = \{G, B\}, n\right) \tag{65}$$

where h is an additional feature, which captures the horizon being forecasted.

There are a number of categories of similar problems that the DM might attend to when solving this problem. The first three set of heuristics closely match those in inference. There are then two more that we consider, and which relate to other salient quantities the DM might anchor their answers on.

No Update. When agents fit the forecasting problem to this category, they use the fol-

lowing heuristic, which results in no update of prior beliefs:

$$E[y_{t+1}|y] = \pi_G \times \mu_G \tag{66}$$

Base-rate Neglect. When agents fit the forecasting problem to this category, they use the following heuristic:

$$\mathbb{E}_t[y_{t+1}|y] = \rho(\alpha_O)^{h(\alpha_O)}(\rho y_{T-1} + (1-\rho)\mu_\theta + \epsilon_T(\alpha_O)) + \left(1 - \rho(\alpha_O)^{h(\alpha_O)}\right)\left(\pi_{G|\theta} \times \mu_G\right) \tag{67}$$

where:

$$\pi_G|y \propto \frac{\pi_{y|G}}{\pi_{y|B}} \tag{68}$$

with:

$$\frac{\pi_{y|G}}{\pi_{y|B}} = \exp\left(\left(\epsilon_T(\alpha_O) + \frac{1-\rho(\alpha_O)}{2}\right)(1-\rho(\alpha_O))\right) \tag{69}$$

and:

$$\alpha_\rho \geq 0 \qquad \bar{\rho} = \sum_{j=1}^{J}\left(\frac{S(n_j,N) \times Pr(n_j|Q)}{\sum_k S(n_k,N) \times Pr(n_k|Q)}\right)\rho_j \tag{70}$$

$$\alpha_\epsilon \geq 0 \qquad \bar{\epsilon}_T = \sum_{j=1}^{J}\left(\frac{S(n_j,N) \times Pr(n_j|Q)}{\sum_k S(n_k,N) \times Pr(n_k|Q)}\right)\epsilon_{T,j} \tag{71}$$

and $S(n_j, N)$ captures the similarity between a series n_j and the series named in the question N. When facing forecasting problems, the DM is used to doing so for persistent series where past information is relevant to predict the future (high $\bar{\rho} > 0$). Moreover, their perception of firm profits is that these also tend to be highly persistent. This leads to a high default persistence level when tackling forecasting problems.

Full Inference. Finally, the DM might attend to both the base rate and the likelihood. Since in our experiment, we always ensure that the prior is uniform over states, full inference yields the same heuristic as simple sampling.

Exact Representativeness. When the DM fits the problem to this category, the esulting heuristic is given by:

$$\mathbb{E}[y_{t+1}|y] = \mathbb{1}\left(\int(\pi_\theta \pi_{y|\theta}) \times \mu_\theta \, d\theta > 50\right)\mu_G \tag{72}$$

Naive Extrapolation. When the DM fits the problem to this category, the resulting heuris-

tic is given by:

$$\mathbb{E}[y_{t+1}|y] = y_T \tag{73}$$

There are technically even more heuristics one could easily construct with forecasting (e.g. forecasting as weighted average of the final observation and the sample mean/exact representativeness mean, etc). By increasing the number of ways the DM might solve the problem, the added complexity in forecasting means that we are less likely to observe strong modal answers. That said, instability is still possible, as is illustrated in our long run expected profits analysis.

C.4 Long Run Expected Profits

When solving this problem, the DM might represent it as either an inference task, or as a forecasting task. If they represent it as an inference task, they frame it as "what are the firm's expected profits?" If instead they represent it as a forecasting task, they frame it as "what will profits be in the very long run?"

$$E[y_{t+1}|y] = \rho(\alpha_O)^{h(\alpha_O)} y_T + \left(1 - \rho(\alpha_O)^{h(\alpha_O)}\right) \mathbb{E}[\mu|y] \tag{74}$$

Context features given by:[29]

$$k_t = (\pi_\theta, \pi_{y_T|\theta}, \pi_{\theta|y_T}, \rho, \epsilon_T, h, q_u = y, q_f \in \{y, \theta\}, q_\Theta \in \{G, B\}, n) \tag{75}$$

When this question is included in the inference survey, it cues an inference question. While when it is included in the forecasting survey, it cues a forecasting question. This leads to $\bar{\rho}_{\text{inference}} < \bar{\rho}_{\text{forecasting}}$ for the same problem.

C.5 Calibration Results

Table C.1 documents the prevalence of each exact heuristic in our experiments.

Comparing Columns (2) and (3) of Table C.1, we see that the prevalence of exact heuristics is not significantly different across the baseline and Salient Persistence conditions. Moreover, in unreported results we show that the use of exact heuristics also does not

[29]Sort out notation – avoid using y for features given that y already represents outcomes.

Table C.1: Exact Heuristics – Baseline Experiment

	$\pi_{G\|y}$ (1)	$\mathbb{E}_t[y_{t+1}]$ (2)	$\mathbb{E}_t[y_{t+1}]_{\text{Salient }\rho}$ (3)	$\mathbb{E}_t[\mu]_{\text{inference}}$ (4)	$\mathbb{E}_t[\mu]_{\text{Forecasting}}$ (5)
Bayesian	6%	1%	1%	3%	1%
50	8%	1%	1%	7%	5%
No update	11%	4%	6%	19%	20%
Exact representativeness	5%	2%	3%	21%	22%
Sample average	3%	0%	0%	3%	1%
Naïve extrapolation	0%	2%	2%	2%	2%
Total	22%	8%	11%	36%	35%

vary across different persistence levels. These results show that the variation in over and underreaction that we document is not driven by differences in exact heuristics, but rather to changes in the attention and cognitive defaults vectors, as claimed in the paper.

To be conservative, our main calibration removes exact heuristics from our sample, therefore only capturing the "full inference" category defined above. However, results are unchanged if we apply our model to the full sample, which includes exact heuristics as well. Table C.2 reports the calibration results for both the baseline and the Salient Persistence conditions. Starting from the baseline condition, Columns (1) and (2) show that forecasting is associated with a higher cognitive default level of persistence, and with a greater attention to its variation relative to inference. Moreover, Column (3) shows that the intervention influences attention to persistence and shock size in forecasting, increasing attention to persistence and decreasing attention to shock size relative to the baseline condition in Column (2).

Table C.2: Calibration for Baseline and Salient Persistence Experiments

	$\pi_{G\|y}$ (1)	$\mathbb{E}_t[y_{t+1}]$ (2)	$\mathbb{E}_t[y_{t+1}]_{\text{Salient }\rho}$ (3)
α_ρ	0	0.4	0.5
α_ϵ	0.4	0.8	0.6
$\bar{\rho}$	0.6	0.9	0.9
$\bar{\epsilon}_T$	1.4	1.6	1.5
h	-	1	1
R^2	0.4020	0.8048	0.7533

D Nesting Previous Experiments

D.1 Features we can vary experimentally

Given this setup, we can experimentally control the following sets of features.

The first set relates to the statistical properties of the DGP and data shown:

1. True DGP:
   (a) persistence $\rho \in \{0, 0.2, 0.4, 0.6, 0.8, 1\}$[30]
   (b) unconditional variance $\frac{\sigma^2}{1-\rho^2} \in \{20, \ldots, 50, 60, 70, 80, 90, 100\}$[31]
   (c*) if we relax normality of ϵ_t, then there are more features (among others):
      i. number of possible values ϵ_t can take $\in \{2, \ldots, \text{continuous}\}$[32]
      ii. whether ϵ_t is numerical or non-numerical (e.g. color).
2. Prior knowledge over possible DGPs:
   (a) number of possible states/bags $\in \{1, 2, 3, 5, 11, \text{continuous}\}$[33]
   (b) differences across states (diagnosticity) – normalized difference in means
   (c) probability distribution over states – variance, concentration, symmetry
   (d*) if respondents are shown more than one signal ($T > 1$), then one can ask inference and forecast questions without prior knowledge of the DGP
3. Number of observations shown $\in \{1, \ldots, T\}$
   (a) with $T > 1$, signals can be shown contemporaneously or sequentially
   (b*) can also have $\underline{\text{T}}$ signals shown contemporaneously, and the remaining $T - \underline{\text{T}}$ signals then shown sequentially

The second set relates to the quantities we wish to elicit and how to elicit them (e.g. in value or probability terms):

[30] These are the parameters used in AKLMT.

[31] AKLMT use $\sigma^2 = 20$ and scale screen in experiment. FLP use $\sigma^2 = \{50, 60, 70, 80, 90, 100\}$.

[32] Standard BBPC experiments have 2 (the two different color balls), while FLP has a continuous normal distribution of signals. FLP also consider the case with a binary state and signal, but forecasting problem is then elicited in probabilistic terms instead of a point estimate. To what extent do coarsedness and distribution of *signals* (and not just of states) also matter?

[33] AKLMT may be closest to 1; FLP and ALT use 2; BBI use $\{2, 3, 5, 11\}$; continuous may be relevant empirically.

1. Inference:
   (a) probability distribution over states/means
   (b) expected value of underlying process
2. Forecast – horizon, $h \in \{1, 2, 5, 10\}$[34]:
   (a) expected value of $h-$step ahead forecasts
   (b) probability distribution over $h-$step ahead forecast

It might also be interesting to elicit a measure of cognitive uncertainty eventually.

D.2 Nesting experiments in previous work

Table D.1 summarizes the following discussion, where papers are ordered from forecasting experiments to experiments that are increasingly about inference.

[34]These are the horizons used in AKLMT.

Table D.1: "Nesting" Experiments in Previous Work

		AKLMT	FLP Baseline	Possible Bridge	FLP Binary	BBI
True DGP	AR(1)	$\rho \in \{0, 0.2, 0.4, 0.6, 0.8, 1\}$ $\epsilon_t \sim N(0, \sigma^2)$, continuous $\sigma^2 = 20$	$\rho = 0$ $\epsilon_t \sim N(0, \sigma^2)$, continuous $\sigma^2 \in \{50, 60, 70, 80, 90, 100\}$	n.a.	n.a.	n.a.
	BBPC	n.a.	n.a.	Bernoulli(p), binary numerical states numerical signals	Bernoulli(p), binary non-numerical states non-numerical signals	Bernoulli(p), binary non-numerical states non-numerical signals
Prior	# of states	$\times^*$	2: $\mu_A = 100$, $\mu_B = 0^{***}$	2: value = 100, value = 0	2: Good, Bad	$\{2, 3, 5, 11\}$: Bags A, B,...
	diagnosticity		$\mu_A - \mu_B = 100$	$\frac{p(signal=1\|value=100)}{p(signal=1\|value=0)}$	$\frac{p(up\|Good)}{p(up\|Bad)}$	$\frac{p(G\|\text{Bag i})}{p(R\|\text{Bag j})}$
	prob. distribution		$p(A) = \{0.2, 0.5, 0.8\}$	$p(A) = \{0.2, 0.5, 0.8\}$	$p(A) = \{0.2, 0.5, 0.8\}$	$J = 2$ differ in symmetry, $J = 3$ differ in concentration, $J = 5, 11$ uniform
Signals	# simultaneous	$\underline{T} = 40$	1	$\underline{T}^{*****}$	1	1
	# sequential	$T - \underline{T} = 40$	0	0	0	0
Inference	value	$\times^{**}$	$\times^{****}$	✓	×	×
	probability		✓	✓	✓	✓
Forecast	horizon	$h \in \{1, 2, 5, 10\}$	$h = 1$	$h = 1$	$h = 1$	×
	value	✓	✓	✓	×	
	probability	×	$\times^{****}$	✓	✓	

* In Experiment 1 participants are told that the data is generated by a "stable random process," and in Experiment 3 participants are told it is a "fixed and stationary AR(1) process: $x_t = \mu + \rho x_{t-1} + \epsilon_t$, with a given μ, a given ρ in the range $[0, 1]$, and ϵ_t is an iid random shock.
** While they do not elicit beliefs about the underlying mean, they conjecture that biased beliefs about the mean are central to explain their results.
*** Notice that the state can be numerical or non-numerical.
**** The "more similar" and "less similar" treatments get close.
***** This case could provide a bridge between FLP Binary/BBPCs when $\underline{T} = 1$ (and signal/state are non-numerical), and FLP Baseline when $\underline{T}$ is large enough so that the # of "success" signals becomes normally distributed.

D.2.1 AKLMT

True DGP. The true DGP is an AR(1) with unconditional mean μ, autocorrelation coefficient ρ, and $\epsilon_t \sim N(0, \sigma^2)$. They vary the persistence of the process across experiments.

Prior knowledge about the DGP. They *do not give information about the prior distribution of the state* (including information about ρ). Instead, participants are told that the process is a "stable random process," and initially "observe 40 past realizations.[35]

Signals. Then, in each round, participants make forecasts and observe the next realization, for 40 rounds."

Inference and Forecast. Their focus is purely on forecasting problems. They elicit forecasts at a $1, 2, 5, 10$ horizon in value terms. While they do not elicit the beleifs about the mean, they conjecture that respondents have biased beliefs about the mean of the underlying process in order to explain their results.

D.2.2 Nesting FLP

4.2.2.1 Baseline Experiment

True DGP. True DGP can be thought of as an "AR(1)" with $\rho = 0$ (so that realizations are in fact iid), and $\epsilon_t \sim N(0, \sigma^2)$. They frame this DGP as the monthly stock price growth of a firm.

Prior knowledge about the DGP. Participants know all aspects about of the DGP other than the underlying mean, which can take on one of two value, depending on the state ($N = 2$, Good and Bad). Specifically, $\mu_G = 0$ or $\mu_B = 100$, and they are given the prior distribution over these two possible states.

Signals. They observe a single draw from one of the DGPs.

Inference and Forecast. They elicit both inference questions (posterior beliefs of the firm being good or bad conditional on the stock price growth observed), and forecasting $h = 1$ (expected stock price growth tomorrow given the stock price growth today). In the Baseline Experiment the inference question is elicited in probabilistic terms, and the

[35]Experiment 3 gives slightly more information. Participants are told the process is a "fixed and stationary AR(1) process: $x_t = \mu + \rho x_{t-1} + \epsilon_t$, with a given μ, a given ρ in the range $[0, 1]$, and ϵ_t is an iid random shock. But they still aren't told the parameters of the model.

forecasting question is elicited in value terms.[36]

4.2.2.2 Binary Experiment

True DGP. The true DGP is now Bernoulli(p), and generates an "up" signal with probability p, and a "down" signal with probability $1-p$. The framing is that the signal captures the direction of the firm's stock price movement, which can be either up or down, so that signals are binary, and non-numerical.

Pior knowledge about the DGP. Participants know that there are two possible states ($N = 2$, Good and Bad). The probability of an up signal is higher in the good state than in the bad state, $p_G > p_B$, and p_G and p_B are known. They are also given the prior distribution over these two states.

Signal. Respondents observe a single signal drawn from one of the two DGPs.

Inference and Forecast. They elicit inference questions (posterior probability that the firm is Good or Bad, conditional on the signal they observe), and a forecasting question (probability that the next signal is up, given today's signal). Notice that both questions are elicited in probabilistic terms.

Therefore, there are two main differences relative to the Baseline experiment. First, the signal follows a binary distribution instead of a continuous distribution. Second, the signal is non-numerical, so they can only ask the forecasting question in probabilistic terms.

We discuss how to more directly nest this type of experiment into our setup in Section D.2.4.

D.2.3 Nesting BBI

True DGP. The true DGP is Bernoulli(p), and generates an Red ball with probability p, and a Green ball with probability $1-p$. The framing is the same as standard BBPC experiment with balls and urns/bags (so e.g. signals are binary and non-numerical).

[36] In the "less similar" treatment, the setup is the same (but all framed in terms of revenue rather than stock price growth), and they frame *both the inference and the forecasting question in probabilistic terms* (the latter is achieved by asking respondents what is the probability that revenue will go up vs down next period). *Note:* to differentiate this more from the inference question, it would be interesting to elicit this for bins that are finer than just up and down.

In the "more similar" treatment, the signal is reframed as the firm's profit in the current month, and the state is framed as the profitability of the firm, defined as the long run mean of the firm's monthly profit. Both the signal and the state are numerical, and they elicit *both the inference and the forecasting question in value terms.*

Prior knowledge about the DGP. Participants know that there are N possible states (where $N = \{2, 3, 5, 11\}$ is varied across experiments), and each of these possible N states differ in the probability of drawing a Red vs Green ball, with p_j being the probability of drawing a red ball from bag j. Participants know the prior probability distribution over bags (which they vary in symmetry/concentration across experiments).

Signals. Participants are shown a single draw from one of the bags.

Inference and Forecast. They elicit inference questions in probabilistic terms.

D.2.4 Possible Bridge Experiment

The standard BBPC experiment considered in FLP's Binary Experiment and in BBI, cannot *directly* be mapped into the setup we explored in (3) because the DGP has a different distribution. The following "Bridge Experiment" might help in bridging this gap, and in nesting standard BBPC experiments in our baseline framework, with $\rho = 0$, $\mu = p$ and the number of draws in the BBCP experiment being (a large enough) T.

Specifically, we can go from one distribution to the other by varying the number of signals shown, and exploiting the fact that a binomial distribution is asymptotically normal. Given T success/fail draws from a Binomial distribution Bionmial(T, p), the fraction of success draws becomes asymptotically normal as T grows large, $N\left(p, \frac{p(1-p)}{T}\right)$. Therefore, if respondents are shown T draws, these T draws will represent a normally distributed signal of p.

True DGP. The true DGP is Bernoulli(p), and generates a signal equal to 1 with probability p, and a signal of 0 with probability $1 - p$. Signals are binary and numerical.

Pior knowledge about the DGP. Participants know that there are two possible states ($N = 2$, Good with value = 100 and Bad with value = 0). The probability of a signal=1 is higher in the good state than in the bad state, $p_G > p_B$, and p_G and p_B are known. They are also given the prior distribution over these two states.

Signal. Respondents observe T simultaneous signals drawn from one of the two DGPs.

Inference and Forecast. Can elicit inference and forecast questions in both value and probabilistic terms (because of having numerical state definitions and numerical signals).

This then allows us to nest both set of experiments by varying a single parameter, the number of draws shown in the BBPC experiment. When $T = 1$ we are back to the

Binary experiment in FLP. Instead, the case for T large enough is nested in our baseline framework as follows.

True DGP. The true DGP is as in (3) with $\rho = 0$, $\mu = p$, and $\epsilon_t \sim N(0, \sigma^2)$.

Pior knowledge about the DGP. Participants know that there are two possible states ($N = 2$, Good with value = 100 and Bad with value = 0), with $p_G > p_B$, and p_G and p_B are known. They are also given the prior distribution over these two states.

Signal. Respondents observe a single draw $y_t = p + \epsilon_t$, generated from one of the two DGPs.

Inference and Forecast. Can elicit inference and forecast questions in both value and probabilistic terms (because of having numerical state definitions and numerical signals).

D.2.5 Nesting BCGKS

The baseline balls and urns experiment is equivalent to FLP binary (with only inference).

The coin flip experiment is most similar to the Possible Bridge experiment with the prior providing exact knowledge of the DGP, and with 2 (or 6) drawn balls from the urn. However, notice that in the coin flip experiment respondents are asked about the probability distribution of the signals, while in standard balls and urns problems respondents are asked about the DGP itself.[37]

True DGP. The true DGP is Bernoulli(1/2), and generates a signal equal to h with probability 1/2 and a signal equal to t with probability 1/2.

Prior knowledge of the DGP. Complete knowledge of the DGP.

Signal. In 100 sequences generated by the DGP, the order of heads vs tails was (h,t) or (h,h).

"Inference and Forecast." Elicit how many of the sequences described in *Signal* were of each type.[38]

[37] This difference is somewhat subtle. In one case, you are asked about the cause of what you observed (properties of the DGP). In the other case you are asked about the properties of what you observed, knowing its cause.

[38] Somewhat of an unnatural question – given info about DGP, we are more used to thinking about what is the share of h vs t in any given sample.

E Experiment Instructions

Introduction

Welcome to the study!

If you read the following instructions carefully, you will be able to earn additional money. The actual amount you will earn depends on your decisions. We will also test your understanding of them later.

There is a base fee of $5 for completing the study. **To complete the study, you will need to read all instructions carefully and answer the corresponding comprehension questions correctly.**

In addition, you can earn a potential bonus of **$10.00**. At the end of the study, one of the rounds will be randomly selected and your decision in this round will determine your bonus.

Important information

- You should think about each round **independently** of all other rounds in this study.
- Whenever a round involves a random draw, then this random draw is **actually implemented by the computer** in exactly the way it is described to you in the instructions.

The Experiment

In this study you will be asked to complete **12 rounds of guessing tasks.**

The survey has multiple parts.

- **Part 1 (9 rounds):** In each of the first 9 rounds, there is a new pool of 20 firms belonging to two types: some firms are good firms, while others are bad firms.
- **Part 2 (3 rounds):** In the final 3 rounds, you will revisit some of the firms you studied in the first 9 rounds.

Your task: Guess the probability that a firm is good or bad depending on the available information.

The information you will have: Profits of the firm over time, and information on how profits evolve over time.

A firm's profits vary over time. The profits of good firms are generally higher than those of bad firms.

- For example, the average profit of **good** firms is **100**. This means that the profits of good firms tend to fluctuate around 100.
- The average profit of **bad** firms is **0**. This means that the profits of bad firms tend to fluctuate around 0.
- At the same time, there are some random fluctuations in the monthly profit for any given firm.
- In any given month, a good firm's profit may increase or decrease, as can a bad firm's.

Understanding a Firm's Profits

The **black line** in the figure below describes the profits of a given firm over the last **30 months.**

- The **current profits** of any given firm depend on its **profit in the previous month**, on the **firm's average profit**, and on **unpredictable fluctuations**:

profit in current month =
(0.4) x (profit in previous month) + (0.6) x (firm's average profit)
+ unpredictable fluctuations

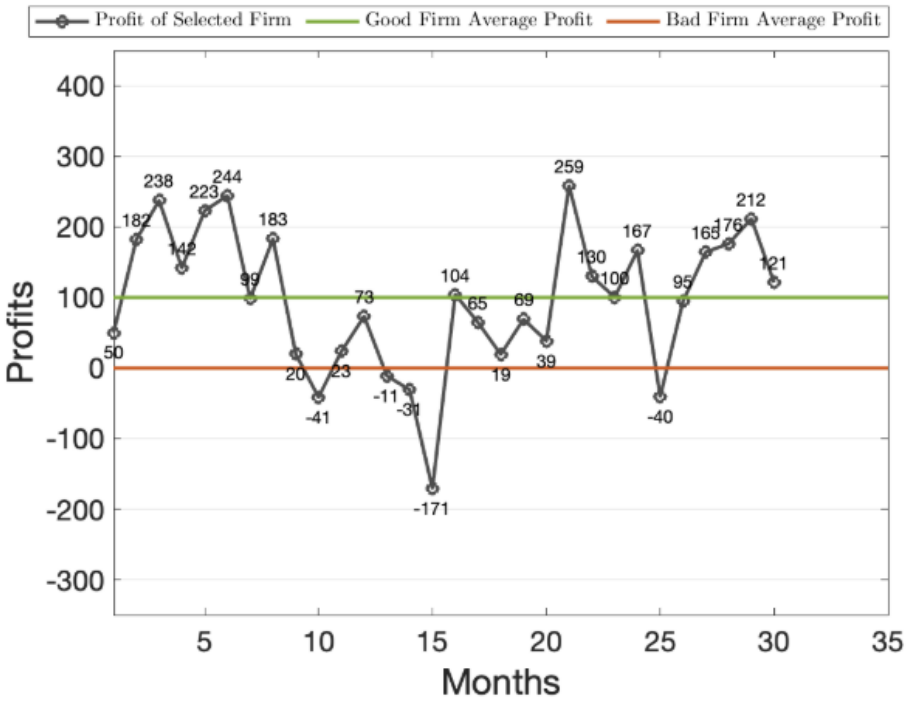
- The **firm's average profit** is **100** if the firm is **good**. This means that a good firm's profits fluctuate around 100.
- The **firm's average profit** is **0** if the firm is **bad**. This means that a bad firm's profits fluctuate around 0.
- The **unpredictable fluctuations** in each period capture random fluctuations in the firm's profits. They average out to zero and usually stay within the range of -200 to +200.
- The **number in front** of "**profit in previous month**" (0.4 in this case) **captures how strongly past profits influence current profits (the persistence of profits over time)**. A higher number (closer to 1) means that profits in the previous month have a greater influence on current profits. So when this number is higher, current profits will be closer to the previous month's profits.

The **green line** at 100 in the figure below corresponds to the average profit of a **good** firm. If the selected firm is good, we should expect its profits to fluctuate around this line.
The **orange line** at 0 in the figure below corresponds to the average profit of a **bad** firm. If the selected firm is bad, we should expect its profits to fluctuate around this line.

profit in current month =

(0.4) x (profit in previous month) + (0.6) x (firm's average profit)

+ unpredictable fluctuations



Note: The numbers in front of profit in previous month (in this case 0.4) and in front of firm's average profit (in this case 0.6) may vary across rounds. Please pay attention to this number when answering questions.

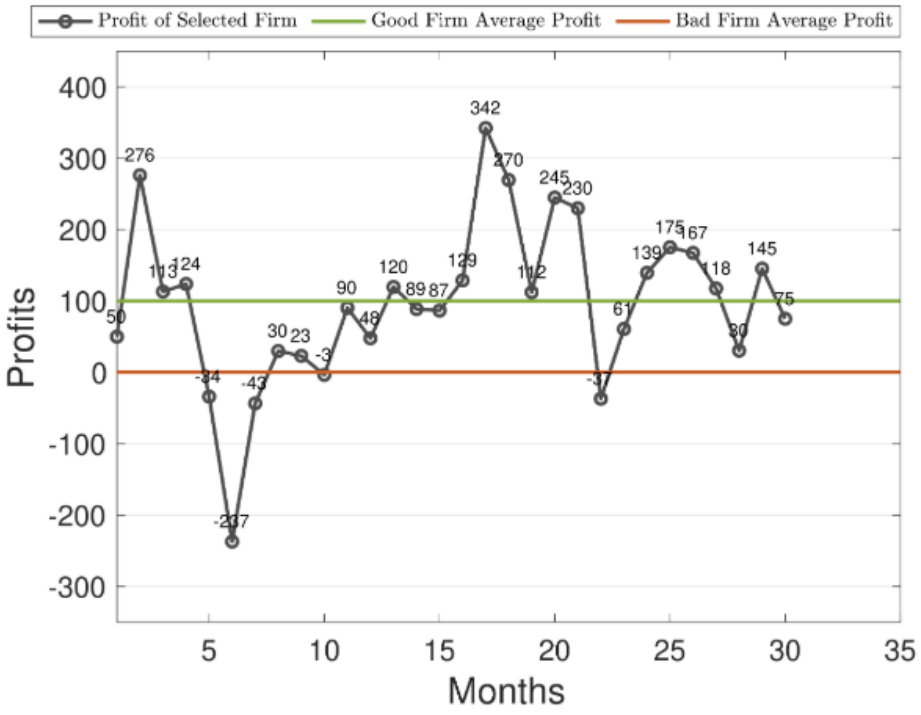
CHECK YOUR UNDERSTANDING on the difference between good and bad firms.

In the example above, is the selected firm more likely to be a good firm or a bad firm?

The selected firm is more likely to be a GOOD firm

The selected firm is more likely to be a BAD firm

Correct. This firm is more likely to be **GOOD, as its profits seem to fluctuate more around 100 (the green line)** than 0 (the orange line). And remember that good firms have an average profit of 100, while bad firms have an average profit of 0.

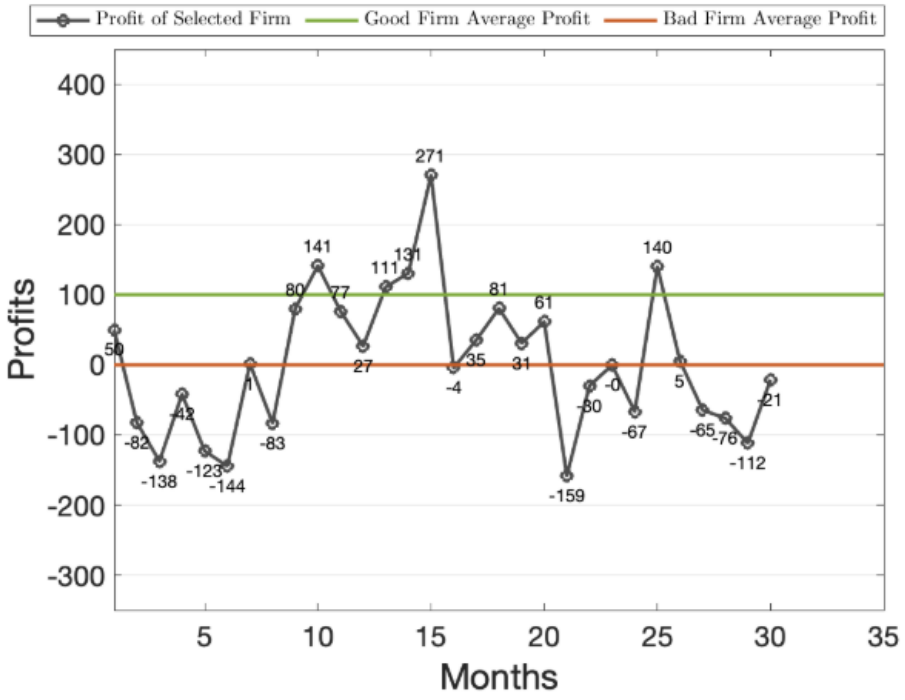


CHECK YOUR UNDERSTANDING on the difference between good and bad firms.

profit in current month =

(0.4) x (profit in previous month) + (0.6) x (firm's average profit)

+ unpredictable fluctuations

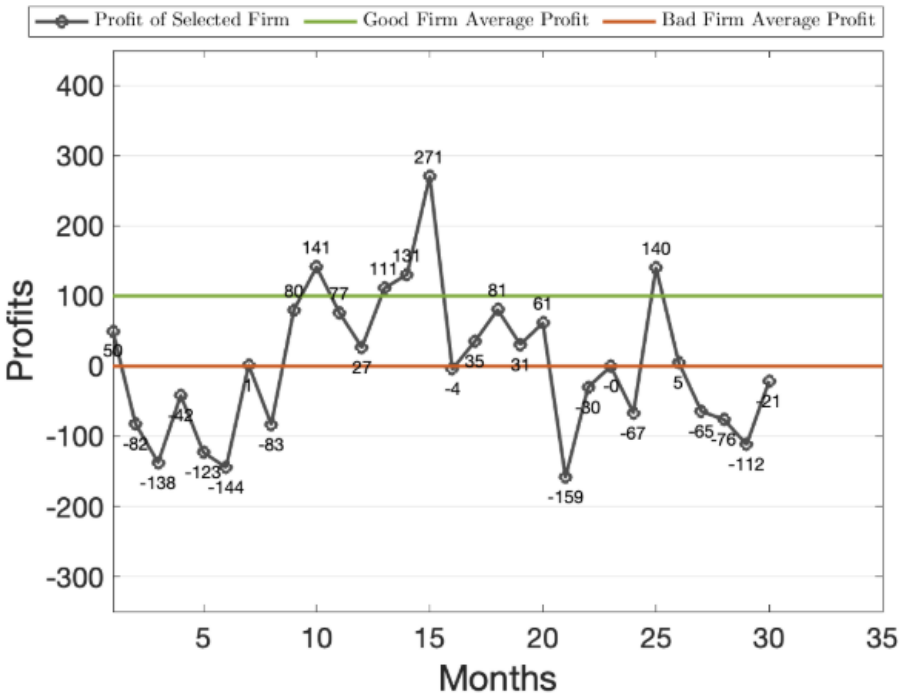


In the example above, is the selected firm more likely to be a good firm or a bad firm?

The selected firm is more likely to be a GOOD firm

The selected firm is more likely to be a BAD firm

Correct. This firm is more likely to be **BAD, as its profits seem to fluctuate more around 0 (the orange line)** than 100 (the green line). And remember that good firms have an average profit of 100, while bad firms have an average profit of 0.



CHECK YOUR UNDERSTANDING on how profits evolve over time.

Let's say that the profit in current period of a firm were calculated according to the following equation:

profit in current month =
(0.9) x (profit in previous month) + (0.1) x (firm's average profit)
+ unpredictable fluctuations

Do the profits in the current period mostly depend on the profit in the previous period or the firm's average profit?

- Previous Period
- No way to tell
- Average profit

If you were to predict the firm's profit in the next period, would you expect it to be closer to today's profit, or to the firm's average profit?

- Today's profit
- No way to tell
- Average profit

profit in current month =

(0.9) x (profit in previous month) + (0.1) x (firm's average profit)

+ unpredictable fluctuations

Correct. The firm's profit in the current period mostly depends on the profit in the previous period. This is because the weight attached to the previous period's profit is equal to 0.9, which is greater than the weight attached to the firm's average profit, which is 0.1.

For the same reason, the firm's profit next period is more likely to be closer to the current profit than to the firm's average profit.

CHECK YOUR UNDERSTANDING on how profits evolve over time.

Let's say that the profit in current period of a firm were calculated according to the following equation:

profit in current month =
(0.2) x (profit in previous month) + (0.8) x (firm's average profit)
+ unpredictable fluctuations

Do the profits in the current period mostly depend on the profit in the previous period or the firm's average profit?

- Previous period
- No way to tell
- Average profit

If you were to predict the firm's profit in the next period, would you expect it to be closer to today's profit, or to the firm's average profit?

- Today's profit
- No way to tell
- Average profit

profit in current period =
(0.2) x (profit in previous period) + (0.8) x (firm's average profit)
+ unpredictable fluctuations

Correct. The firm's profit in the current period mostly depends on the firm's average profit. This is because the weight attached to the previous period's profit is equal to 0.2, which is smaller than the weight attached to the firm's average profit, which is 0.8.

For the same reason, the firm's profit next period is more likely to be closer to the firm's average profit than to the current profit.

This study has multiple parts, and the questions we ask you will differ across parts.

Your task: Guess the probability that a firm is good or bad depending on the available information.

In Part 1 (the first 9 rounds) we will ask you the following questions.

EXAMPLE ROUND - PART 1 (the first 9 rounds)

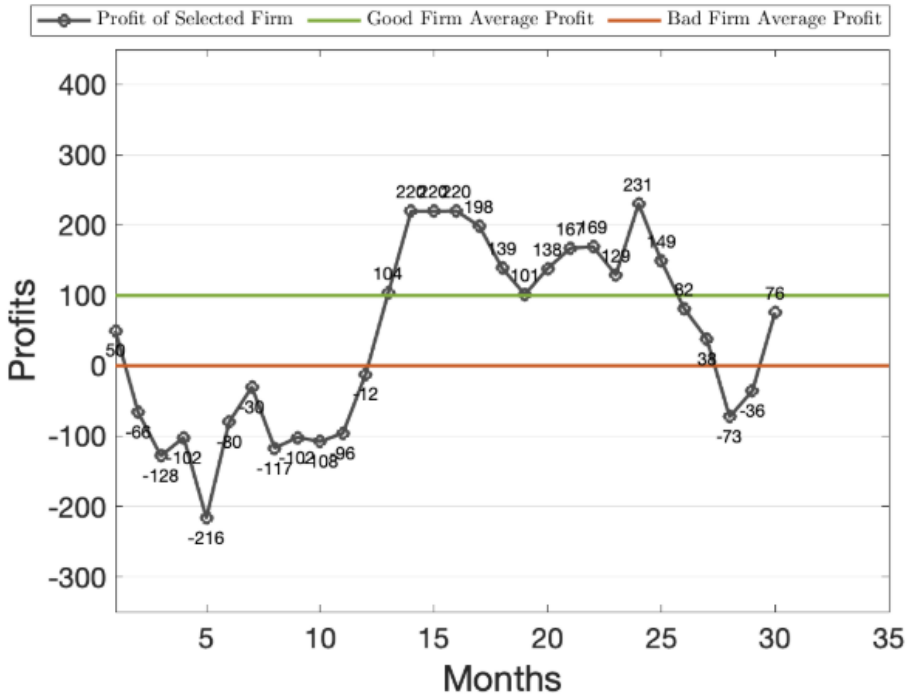
There is a new pool of 20 firms. The pool of firms has the following composition:



Each of the 20 firms is equally likely to be picked. **In each round, one firm is randomly picked from this pool.** You will not know whether a firm is good or bad, but we will give you information about the profits of the firm over time. You will then make a guess as to whether the firm is good or bad.

The **green line** at 100 in the figure below corresponds to the average profit of a **good** firm. If the selected firm is good, we should expect its profits to fluctuate around this line.
The **orange line** at 0 in the figure below corresponds to the average profit of a **bad** firm. If the selected firm is bad, we should expect its profits to fluctuate around this line.

profit in current month =
(0.9) x (profit in previous month) + (0.1) x (firm's average profit)
+ unpredictable fluctuations



Given the information above how likely do you think the selected firm is good or bad? (Note: your answers must sum to 100.)

% chance the selected firm is GOOD	0
% chance the selected firm is BAD	0
Total	0

What do you expect the firm's profit to be, on average, over the long run? For this question, try to consider the potential profit movements over a very long horizon and predict what the average of those profit movements will be.

This study has multiple parts, and the questions we ask you will differ across parts.

Your task: Guess the probability that a firm is good or bad depending on the available information.

In Part 2 (the final 3 rounds) we will ask you the following questions.

NEW MONTH 31 OBSERVATION

In Part 2 (the final 3 rounds), we will revisit some of the firms you studied in Part 1, and we will show you the selected firm's profits.

Additionally, we will show you the selected firm's profits in **month 31** (in **yellow** on the Figure). Here is an example.

EXAMPLE ROUND - PART 2 (the final 3 rounds)

The firm's profit in **month 31** is **212** (in **yellow** on the Figure below).

Given the **new information** on the firm's profit in month 31, how do you think you should **update** how likely it is that the selected firm is is good or bad?

Month 31 profits INCREASE the chances that the selected firm is good

Month 31 profits DECREASE the chances that the selected firm is good

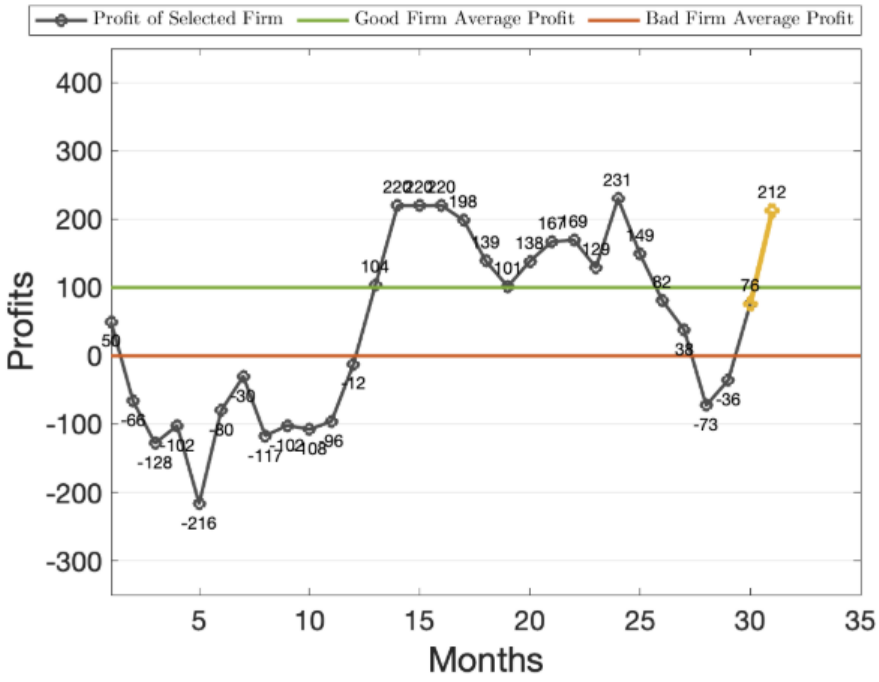
---

The **green line** at 100 in the figure below corresponds to the average profit of a **good** firm. If the selected firm is good, we should expect its profits to fluctuate around this line.
The **orange line** at 0 in the figure below corresponds to the average profit of a **bad** firm. If the selected firm is bad, we should expect its profits to fluctuate around this line.

profit in current month =

(0.9) x (profit in previous month) + (0.1) x (firm's average profit)

+ unpredictable fluctuations



QUESTIONS IN PART 2. We will ask you the following questions.

Before seeing that the selected firm's profit in **Month 31 is 212,** you reported that you thought there was a **20% chance that the firm was good** and a **80% chance that the firm was bad.**

Given the **new information** on the firm's profit in month 31, **update your answer**, and report how likely you now think the selected firm is good or bad. (Note: your answers must sum to 100.)

% chance the selected firm is GOOD	0
% chance the selected firm is BAD	0
Total	0

Before seeing that the selected firm's profit in **Month 31 is 212**, you reported that you expected the firm's average profit over the long run to be **0**, on average.

Given the **new information** on the firm's profit in month 31, **update your answer**, and report what you expect the firm's profit to be, on average, over the long run? For this question, try to consider the potential profit movements over a very long horizon and predict what the average of these profit movements will be.

Your Payment (click here to save this message in the new window for later)

There is a best answer to every question. Using the laws of probability, the computer determines a **statistically correct statement of the probability that a particular company is good or bad**, based on all of the information available to you. This **optimal guess** does not rely on information that you do not have. It is just the best possible (this means: payoff maximizing) estimate given the available information. In technical terms, this guess is based on a statistical rule called Bayes' Law. If your guess is within 3% of the optimal guess, you will earn the additional $10.00.

All this means is that, in order to earn as much money as possible, you should try to give your best estimate of the probability that each company is good or bad.

You have completed all the instructions and comprehension questions. You will now proceed to the actual task.