Biases in Belief Updating Within and Across Domains

November 19, 2025

Abstract

Why do people sometimes overreact and other times underreact to new information? We develop a model in which signal strength—how much beliefs should change in response to new information—depends on multiple features of the information environment. People misperceive signal strength because they are insensitive to variation in these features due to limited attention. Instead, they approach problems with an experience-based perception of the information environment, which they only partially adjust as a function of how much attention they allocate across features. This mechanism explains a wide range of belief-updating patterns both within and across different learning domains. A series of experiments provides support for the model and underlying mechanism.

Keywords: overreaction, underreaction, beliefs, learning, attention, insensitivity, mental representations, forecasting, inference, behavioral economics.

1 Introduction

How do people update their beliefs in light of new information? Rational models of learning assume that people use Bayes rule. However, studies on belief formation document substantial departures from this benchmark: not only are people's expectations often systematically biased, but these biases differ both within and across domains.

Take the inference and forecasting learning domains. A person sees the price of a stock increase. They can use this information to either *forecast* future prices, or *infer* whether the quality of the company is good or bad. Previous work has found substantial differences in belief updating depending on whether people are tasked with either the former or the latter. Not only do we observe predominant overreaction in forecasting and underreaction in inference (Benjamin, 2019, Bordalo et al., 2020), but there is also just as much variation in bias within each domain. In forecasting, we observe overreaction at long horizons and for transitory series, but underreaction at short horizons and for highly persistent processes (Afrouzi et al., 2023). In inference, we observe underreaction when signals are diagnostic, but overreaction when signals are weak (Augenblick et al., 2025). This heterogeneity in evidence raises the question of what drives variation in over and underreaction. Are its determinants domain specific, or is there a common psychological mechanism that can rationalize these findings?

Recent work in behavioral economics has attempted to explain both over and underreaction to information by proposing models where people's perceptions of signal strength—the amount by which one should update their beliefs in response to new information—deviate from the objective value.[1] In these class of models, people anchor their responses on a default value of signal strength and adjust only partially to new information; anchoring on an intermediate level of signal strength generates underreaction to strong signals and overreaction to weak ones (Augenblick et al., 2025). This prediction lends itself to explaining the evidence in inference, as we do observed underreaction when signals are highly diagnostic, and overreaction when signals are weak. But applying this framework directly to forecasting generates a puzzle. While shock size is the most natural proxy for signal strength, there is no systematic evidence that people underreact when shocks are large and overreact when shocks are weak. Instead, we observe substantial variation in over and underreaction along other dimensions, such as the persistence of the underlying data generating process, and the forecast horizon. Similarly, attempts to reconcile differ-

[1] See, for example, Gabaix (2019), Enke and Graeber (2023), Bordalo et al. (2024a), Enke et al. (2024), Augenblick et al. (2025), Ba et al. (2025), Fan et al. (2025).

ences in over and underreaction across domains (Fan et al., 2025) cannot explain the rich variation in belief updating *within* domains; as shown in Bordalo et al. (2020), there is nearly as much variation in over and underreaction within forecasting as there is between forecasting and inference.

To what extent then can we apply lessons in belief updating across domains? In particular, can we use predictions developed in inference to explain the evidence in forecasting? And vice-versa, can the richer environment in forecasting help us refine the theories that have been useful in explaining the evidence in other domains?

To answer these questions, we provide a common framework that nests both inference and forecasting within the same information environment, and show how signal strength depends on multiple interacting features (e.g., persistence, horizon, shock size). We then propose a model where people misperceive signal strength due to insensitivity to variation in its features. This insensitivity is driven by two factors. First, people approach problems with an experience-based perception of the information environment; second, limited attention leads them to only partially adjust from this default, where the extent of adjustment depends on the attention allocated across features.

Applying the model to inference and forecasting, we show how it can explain variation in belief updating both within and across domains. *Across* domains, inference and forecasting lead to different default perceptions of signal strength. *Within* domains, attention modulates the prediction of underreaction to strong signals and overreaction to weak signals. First, more neglected features generate greater variation in over and underreaction. Second, insensitivity to one feature can create excess sensitivity to another, leading to overreaction to *both* weak and strong signals. We provide support for the model's predictions in a series of experimental studies.

In our framework, we adapt the canonical balls and urns inference paradigm (Edwards, 1968) to allow for autocorrelated signals. This allows persistence and horizon to play a meaningful role in belief updating. There is a pool of firms with an equal number of good and bad firms. Firm profits evolve according to an AR(1), and 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. People do not know the selected firm's type, but instead they observe a sequence of monthly profits generated by the selected firm. An inference question then asks how likely it is that the selected firm is good or bad, while a forecasting question elicits people's beliefs of profits at future horizons.

This framework allows us to map features of the information environment (shock size, persistence, horizon) to signal strength, the amount by which people should update their beliefs in light of new information. While signal strength is increasing in shock size in both inference and forecasting, the relationship between signal strength and persistence differs across inference and forecasting. Increasing the autocorrelation of signals makes the current signal more informative about future signals, but less informative about the underlying mean.[2] This implies that a larger persistence leads to a stronger signal in forecasting, but to a weaker signal in inference. Finally, increasing the horizon being forecasted decreases signal strength in forecasting, as today's observation is more informative about the near future than the distant future.

Under rational expectations, these mappings would be irrelevant, as people would perfectly attend to all features and always have the correct perception of signal strength. However, processing costs and limited attention may lead people to misperceive their information environment, generating a biased perception of signal strength. To incorporate these cognitive constraints into the model, we assume that people approach a problem with an experience-based perception of the information environment. This generates a default level for each feature. People then adjust their default perceptions of each feature depending on how much attention they pay to them. Features that are more relevant and easier to process receive greater attention, meaning that perceived signal strength is more sensitive to objective variation in those features.

We show that differences in over and underreaction across domains are due to inference and forecasting leading to different default perceptions of signal strength. Differences in over and underreaction within domains is due to variation in attention across features. When varying a single feature (e.g., persistence or forecasting horizon), we recover the comparative static of underreaction to strong signals and overreaction to weak signals (Augenblick et al., 2025). However, this comparative static is modulated by attention when signal strength is determined by multiple interacting features. First, more neglected features drive more variation in over and underreaction. Second, insensitivity to a feature can lead to excess sensitivity to another, leading to overreaction to both weak and strong signals, and breaking the comparative static altogether.

We test these results in a series of experiments that directly mirror the setup of our theoretical framework. Our first set of results examine variation in belief formation *across* the inference and forecasting domains. By calibrating our model to the experimental

[2]When the data generating process is highly persistent, today's observation is closer to tomorrow, meaning that today's observation constitutes a stronger signal in forecasting. However, high persistence also implies that each new signal contains a lower amount of new independent information about the realized state, therefore weakening the signal in inference.

responses, we show that people's default perceptions of the information environment are associated with high default levels of persistence across both inference and forecasting: on average, people approach information environments with the perception that the data generating process is highly autocorrelated. This then leads to very different anchoring points across the two tasks. Implicitly, a high default level of persistence anchors people on a view of the world where today's observation is very close to tomorrow's (strong signal in forecasting) and does not contain much new information about the underlying mean relative to what you already knew (weak signal in inference). This translates into a default perception of signal strength that is high in forecasting and low in inference. This explains the prevalence of overreaction in forecasting and underreaction in inference that has been documented in prior work and which we replicate in our experiment.

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 horizon) while holding other features fixed at their cued default values. Varying signal strength in this way leads to patterns that are consistent with overreaction to weak signals and underreaction to strong signals in both inference and forecasting. Crucially, this is validated by the fact that persistence interacts with signal strength in both inference and forecasting, but the comparative static goes in the *opposite* direction. Increasing persistence strengthens the signal in forecasting but weakens the signal in inference. Therefore, our framework predicts that increasing persistence should lead to more underreaction in forecasting and more overreaction in inference. Our empirical results confirm this reversed comparative static, which provides support for the conjecture that variation in persistence impacts belief-updating through its effect on signal strength. Similarly, varying signal strength through the forecast horizon also leads to underreaction to strong signals (short horizons), and overreaction to weak signals (long horizons).

However, once we allow for signal strength to depend on *multiple features*, we show that the comparative static of overreaction to weak signals and underreaction to strong signals is modulated by attention in two key ways. First, not all changes in signal strength now lead to the same variation in over and underreaction. Instead, our model allows us to make predictions regarding *which* features are more strongly associated with variation in over and underreaction. Second, the interaction of multiple features can break the comparative static altogether. This analysis further highlights the importance of modeling insensitivity at the feature level, rather than at a more aggregate notion of signal strength.

Starting from the first prediction, our model states that features that are less relevant, harder to process, and less salient are neglected more. Changes in the values of these neglected features then generates a bigger wedge between true and perceived signal strength, leading to greater variation in over and underreaction. Consistent with these predictions, we show that attention to the size of the most recent shock in firm profits—the most salient feature—is substantially higher than attention to persistence—a feature that is harder to process and map into signal strength. As a consequence, changes in persistence generate substantially more variation in over and underreaction than changes in shock size.

To test the role of attention more directly, we exogenously manipulate salience across features. Specifically, we exploit the fact that bounds on attentional capacity generate a critical *link* across features, where increasing attention to one necessarily decreases attention to another. Consistent with our model, exogenously increasing attention to persistence decreases attention to the size of the most recent shock. This then decreases variation in over and underreaction with respect to persistence, while simultaneously exacerbating variation in over and underreaction with respect to shock size.

Turning to the second prediction, we show how the *interaction* of attended to and neglected features can fully reverse the comparative static with respect to signal strength. For this to be the case, we show that insensitivity to a feature can lead to *excess sensitivity* in belief movement to another. We confirm this in the data: insensitivity to persistence can lead to excess sensitivity to shock size, and to more overreaction to strong signals when shock size is increased. Besides illustrating the important role of attention across multiple features in belief updating, these results highlight the need to consider predictions for belief updating at the feature level rather than with respect to statistics that are functions of multiple features, e.g., signal strength.

Contribution to the literature. Our paper builds on a large body of research on belief formation, and contributes to recent work that attempts to explain deviations from the rational benchmark through a cognitive lens.

A first line of research looks at how noisy cognition impacts belief formation to explain variation in belief updating *within* domains.[3] In these models people have a cognitive default for a specific feature and do not fully adjust from it with respect to objective variation in the parameter. The lack of adjustment generates insensitivity and attenuation in beliefs (Augenblick et al., 2025, Enke and Graeber, 2023, Enke et al., 2024, Ba et al., 2025), which then delivers underreaction to strong signals and overreaction to

[3] These models are silent on what drives differences in belief updating *across* domains.

weak signals (Augenblick et al., 2025). However, in order to apply this prediction across domains, we need to understand what determines signal strength. 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 (i.e., shock size), or do the underlying properties of the data generating process also play a role in determining signal strength? We operationalize this prediction by showing that signal strength in forecasting is determined by multiple features, including the persistence of the process and the horizon being forecasted. Moreover, by studying the role of multiple interacting features we show how the richer environment in forecasting can help us refine the predictions that have been made in univariate environments (Augenblick et al., 2025, Enke et al., 2024). With multiple features, not all changes in signal strength lead to the same variation in over and underreaction. Instead, our model predicts that more neglected features generate bigger differences in over and underreaction (Massey and Wu, 2005),[4] and insensitivity to a feature can generate excess sensitivity with respect to another, leading to overreaction to both weak and strong signals.

These models require taking a stance on the location of the cognitive defaults. In this respect, a second line of papers study how attention and memory constraints generate distorted mental representations of the information environment through categorization (Hanna et al., 2014, Schwartzstein, 2014).[5] Specifically, the task at hand cues a category of similar problems (Bordalo et al., 2023a, 2024a, Fan et al., 2025). Each category is then associated with a different default frame of the information environment, where features are either fully attended to or fully neglected.

There are two key differences relative to our model. First, these papers are silent on what generates variation in over and underreaction within forecasting. Second, our framework differs in how we model the default frames that are cued by the problem at hand: instead of modeling them as exact heuristics in reduced form (e.g., reporting the base rate in inference, or naive extrapolation in forecasting), we model them in terms of implied feature-values.[6]

[4]Frydman and Jin (2022), Frydman and Jin (2023) further show that people's perceptions are more accurate for payoffs that occur most frequently.

[5]Bohren et al. (2024) study how constraints on memory and attention interact to generate distorted representations of the information environment in the context of risky choice.

[6]For example, a default level of persistence of $\bar{\rho} = 0.9$ in forecasting means that people's default way of solving a forecasting problem is to anchor on the most recent observation and adjust a small amount from there. A default level of $\bar{\rho} = 1$ would instead capture the heuristic of naive extrapolation that is explored empirically in Fan et al., 2025. Our way of modeling categories provides greater nuance, allowing for heuristics to embed some adjustment from a given anchor.

Moreover, instead of adopting exact heuristics in each problem, our model allows for people to partially adjust to objective changes in the information environment. This approach has two advantages. First, it is parsimonious and requires relatively few adjustments to neoclassical models. Second, it also allows us to study variation in over and underreaction *within* domains, where the context that cues the relevant heuristic is held fixed and as such cannot drive the observed differences in belief updating.[7,8,9]

Our framework brings together the approaches outlined here: people's perceptions of the information environment shape the relevant anchoring point, and top-down and bottom-up attention then determine how much people adjust their responses (Ba et al., 2025).[10] The way people solve problems is then determined by their attempt to optimally allocate attention across features (Loewenstein and Wojtowicz, 2023), based on the salience, relevance, and volatility of each feature, as well as their accumulated experiences of solving similar problems in the past (Bastianello et al., 2025).[11] By bringing these approaches together, our framework rationalizes observable variation in over and underreaction *both* within and across domains.

This allows us to reconcile many patterns in belief updating. In forecasting, empirical and experimental evidence shows that there is more over-reaction at long horizons and for transitory series, and underreaction at short horizons and highly persistent processes (Giglio and Kelly, 2018, Bordalo et al., 2020, Da Silveira et al., 2020, Wang, 2021, Afrouzi et al., 2023, Bordalo et al., 2024b). Also consistent with our model, Guo, 2025 and Guo and Wachter, 2025 show that earnings announced in the second month of a quarter are similar to those announced in the first month, meaning that second month earning announcements constitute weaker signals. Investors neglect the difference in autocorrelation across announcements, which leads them to anchor on an intermediate level of signal strength, generating overreaction to predictably repetitive earnings news.[12] Finally, Kwon and Tang

[7]For example, Fan et al., 2025 and Ba et al., 2025 are silent on what drives variation in over and underreaction *within* forecasting, which is instead part of the key focus of this paper.

[8]Appendix C shows how our feature-based approach maps onto the framework of Bordalo et al. (2024a) and how it can capture the exact heuristics studied in prior work through categorization.

[9]Finally, another distinction with this line of work lies in the implicit assumption that individuals are able to correctly categorize problems. One way to interpret the multimodality in responses documented by Bordalo et al. (2023b) and Bordalo et al. (2024a) is that, when faced with *unfamiliar* problems, individuals may misclassify them. Since multimodality is not a first-order property of our data, we abstract from this dimension and leave its exploration for future research.

[10]Ba et al., 2025 incorporate both attention-based distortions in perception and noisy cognition to explain variation in over and underreaction in inference.

[11]For the role of top-down attention, see Sims, 2003, Woodford, 2001, Veldkamp, 2011, Gabaix, 2014, Caplin and Dean, 2015, Caplin et al., 2019, Gabaix, 2019, Kohlhas and Walther, 2021.

[12]Similarly, Wu (2023) show that people misperceive cross-variable relations, which also contribute to

(2020), Bastianello (2025), de Silva et al. (2025), and Graeber et al. (2025) analyze how the extent of over and underreaction varies with shock size. The evidence is mixed, and consistent with people underreacting to strong signals and overreacting to weak signals because of insensitivity to various properties of the underlying data generating process.

In inference, people underreact to highly diagnostic signals but overreact when signals are weak (Griffin and Tversky, 1992, Augenblick et al., 2025). Prat-Carrabin and Woodford (2024) consider a setting where participants are learning about a fixed state. They show that when people respond to sequential signals, they underreact to the first few signals, but then overreact once a sufficient number of signals have been observed. Since the first signals contain more information than the later signals, this pattern is also consistent with underreaction to strong and overreaction to weak signals. Finally, Massey and Wu (2005) put forward a "system neglect hypothesis," and consistent with Griffin and Tversky (1992), they 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, and more so in inference compared to forecasting.

Organization of the paper. 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, and map features that are relevant in forecasting (persistence and horizon) to features that are relevant in inference (signal strength). Section 3 introduces our model and shows how constraints on memory and limited attention interact in generating over and underreaction in an information environment with multiple features. Section 4 describes the experiments and presents the results from our empirical investigation. Section 5 discusses how our results relate to patterns in the Survey of Professional Forecasts, and Section 6 concludes.

2 A Common Setting

To study belief updating across domains, we start by developing a setting that integrates features from both inference and forecasting within the same information environment. To do so, we first introduce the canonical paradigm that has been used to study inference problems, and then extend it to allow for key features of forecasting (persistence and

---

observed patterns in over and underreaction to information.

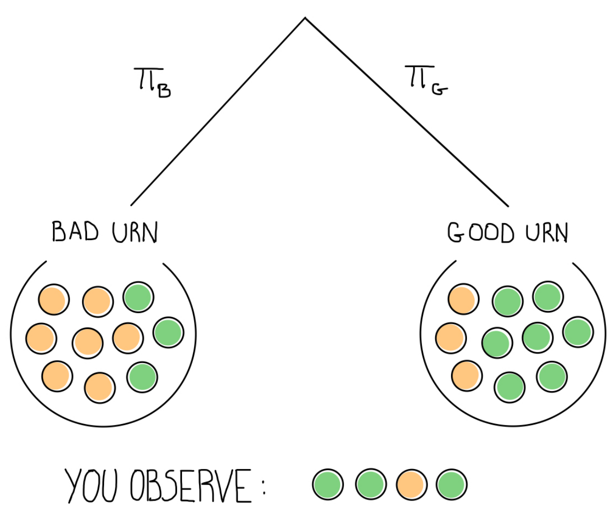
horizon) to play a meaningful role.

2.1 Canonical 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 (urns), $s \in \{G(ood), B(ad)\}$, that differ in their likelihood of generating one of two signals (colored balls), $y \in \{g(reen), o(range)\}$. 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, people are asked how likely it is that the selected state is Good or Bad.

Figure 1: Balls and Urns Inference Paradigm



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

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

The first term on the right hand side corresponds to the log-likelihood ratio, and the second term captures the base rate. Signal strength—the amount by which a Bayesian agent would update their prior beliefs—is increasing in the log-likelihood ratio. Let N_g and N_o be the number of green and orange balls drawn from the selected urn, respectively. If we let the two states be symmetric, such that $\pi_{g|G} = 1 - \pi_{g|B}$, we can then write the log-likelihood as:

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

The first term on the right hand side of this expression captures the relative number of signals in favor of one hypothesis or the other, and is sometimes further decomposed into signal extremeness $\left(\frac{N_g - N_o}{N_g + N_o}\right)$ and signal weight $(N_g + N_o)$ (Griffin and Tversky, 1992). The second term instead captures how diagnostic a given signal is of one state over the other. The larger the difference in the two urns along the observed dimension, the easier it is to tell them apart. Therefore, greater diagnosticity leads to a stronger signal.

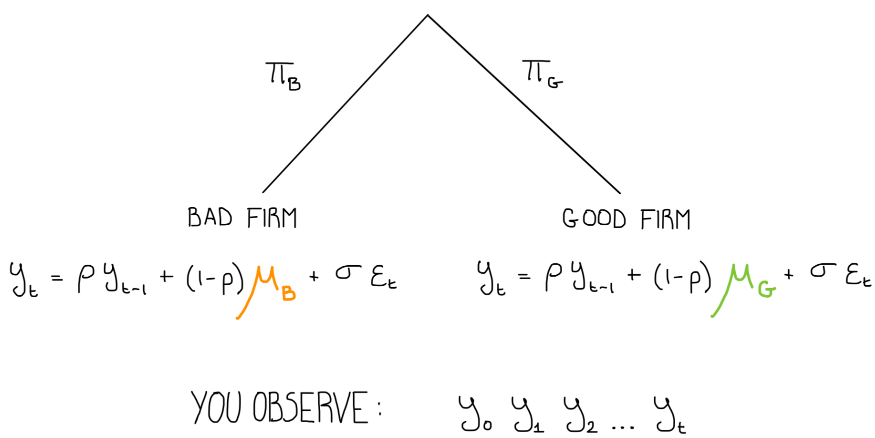
This setting has been used to study inference problems both theoretically and empirically, as reviewed in Benjamin (2019). Prior work has found that people are insensitive to variation in various notions of signal strength, and that this can translate into both over and underreaction in belief updating. For example, Griffin and Tversky (1992) found that people tend to be insensitive to the weight of the evidence (and excessively sensitive to how extreme it is), and a number of papers have shown that people are insensitive to variation in objective signal diagnosticity (Benjamin, 2019). In this paper we are interested in understanding whether these lessons on belief updating in inference generalize to other domains. For example, is it always the case that people underreact to strong signals and overreact to weak signals (Augenblick et al., 2025)? What does it even mean for a signal to be stronger and more diagnostic in other domains, such as forecasting? What common forces can generate the observed variation in over and underreaction both across and within domains?

The current setting provides us with a good starting point to tackle these questions as it lends itself to studying both inference and forecasting problems. 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., 2025). 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 driving variation in over and underreaction in forecasting. We now proceed to augment the setting to address this.

2.2 Introducing Persistence and Horizon

To allow for persistence and horizon to play a meaningful role in belief updating, we change the data generating processes to allow for autocorrelated signals. 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.

Figure 2: Information Environment with Autocorrelated Signals



Specifically, consider a setting with a pool of firms, that can be either good or bad, $s = \{G, B\}$. 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 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 $\epsilon_t \overset{iid}{\sim} N(0,1)$, $\rho \in [0,1)$ is the autocorrelation coefficient, and $\sigma > 0$ is the conditional standard deviation. After observing a sequence of firm profits (as in Figure 3), an inference task asks people how likely they think the firm is good or bad, while 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.[13] As further discussed in Appendix D.2, this setting nests previous empirical studies on how people form beliefs in both inference and forecasting.[14]

Given this information environment, we can derive expressions for signal strength in both inference and forecasting. Starting from inference tasks, Appendix B.1 shows that signal strength is increasing in the following log likelihood ratio:

$$\log\left(\frac{\pi_{y|G}}{\pi_{y|B}}\right) = \sum_{i=0}^{t} \underbrace{\left(\frac{y_i - \mathbb{E}_{i-1}[y_i]}{\sigma}\right)}_{\text{strength of evidence}} \underbrace{\left(\frac{\mu_G - \mu_B}{\frac{\sigma}{1-\rho}}\right)}_{\text{diagnosticity}} \tag{5}$$

The two terms in this expression have a very tight connection to the decomposition we encountered in the canonical balls and urns experiment in (2). The first term captures how strongly the signal points in favor of one state over the other: when $y_i = \mathbb{E}_{i-1}[y_i]$ the signal is uninformative, as would have been the case if we had drawn the same number of orange and green signals in the canonical inference paradigm. Instead, increasing $|y_i - \mathbb{E}_{i-1}[y_i]|$ increases signal strength, and is equivalent to increasing the relative number of orange to green signals. The second term in the above expression instead captures signal diagnosticity: the further away the two means are, or the smaller the variance around those means, the easier it is to tell the two hypothesis apart, and the stronger the signal. Assuming that the true state is μ_G, and simplifying the expression in (5), we can

[13] 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 $y = \{y_{t-i}\}_{i=0}^{t}$.

[14] For example, when $\rho = 0$ our setup nests the one considered in Fan et al. (2025), while allowing for a continuous state further nests the forecasting experiments in Afrouzi et al. (2023). Most importantly, this setting allows us to ask both inference and forecasting questions in a context where both persistence and horizon play a meaningful role, therefore permitting us to study how multiple features interact in driving belief formation.

then write it in terms of the primitives of the model:

$$\log\left(\frac{\pi_{y|G}}{\pi_{y|B}}\right) = \underbrace{\left(\epsilon_i + \frac{\mu_G - \mu_B}{2\left(\frac{\sigma}{1-\rho}\right)}\right)}_{\text{strength of evidence}} \underbrace{\left(\frac{\mu_G - \mu_B}{\frac{\sigma}{1-\rho}}\right)}_{\text{diagnosticity}} \tag{6}$$

While the two terms in (6) have a clear mapping to concepts that have been studied in the inference literature (strength of the evidence and diagnosticity, Griffin and Tversky, 1992), this expression also makes clear how it is useful to go beyond this decomposition once we consider richer environments where signal strength is determined by multiple interacting features. For example, persistence affects both terms in (6), making it hard to think of the strength of the evidence and of diagnosticity as two separate entities.[15] Instead, we can move to thinking about comparative statics with respect to more primitive features such as persistence of the data generating process, while still linking those comparative statics to the broader notion of signal strength.

The importance of this feature-based perspective becomes even clearer once we turn to signal strength in forecasting, which takes the following form:

$$\mathbb{E}[y_{t+h}|\{y_i\}_{i=0}^{t}] - \mathbb{E}[y_{t+h}|\{y_i\}_{i=0}^{t-1}] = \rho^h \sigma \epsilon_t + (1-\rho^h)\left(\mathbb{E}\left[\mu|\{y_i\}_{i=0}^{t}\right] - \mathbb{E}[\mu|\{y_i\}_{i=0}^{t-1}]\right) \tag{7}$$

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.

There are two points worth highlighting when analyzing the expressions (6) and (7) for signal strength in inference and forecasting, respectively. First, as also outlined in Table 1, signal strength in both inference and forecasting depends on *multiple* features that interact in non-additive ways.[16] 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).

[15] 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.

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

Table 1: 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

Second, the expressions in (6) and (7) make clear how to map shock size, persistence and horizon to signal strength. While signal strength in both inference and forecasting is increasing in shock size, as long as the first term in (7) dominates, the persistence of the process influences signal strength in inference and in forecasting in *opposite* directions: signal strength in inference is decreasing in persistence, while signal strength in forecasting is increasing in persistence.[17] Intuitively, increasing persistence makes the current observation less informative about the underlying mean (inference) but more informative about future observations (forecasting). For example, as we approach a random walk, the signal becomes completely uninformative about the mean, while being very informative about the next observation. Conversely, as the process becomes iid, any new observation contains more independent information, increasing signal strength in inference, but decreasing signal strength in forecasting.

Finally, as long as the first term in (7) dominates, signal strength in forecasting is decreasing in the horizon being forecasted. Intuitively, the current observation is more informative about the next observation than an observation further into the future.

3 Introducing Cognitive Constraints

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 bounds on cognitive resources such as attention and memory. These cognitive constraints create distortions in how people perceive the features in their information environment. Memory constraints anchor people to (potentially biased) feature values through the recall of previous experiences while limited attention prevents them from fully attending to multiple features at the same time (Luck and Vogel, 1997, Loewen-

[17]When the first term in (7) 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.3 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.

stein and Wojtowicz, 2023). We consider a framework that incorporates these constraints into a multi-feature model of belief-updating: memory leads the current problem to cue past experiences which generate a default perception of feature values, while limited attention determines how much people anchor on these default values when responding to new information.

These behavioral distortions have been studied in univariate settings (Augenblick et al., 2025, Enke et al., 2024, Ba et al., 2025). We show that introducing the constraints into a model where multiple features interact in determining signal strength (Gabaix, 2014) generates new predictions that rationalize the evidence on belief updating both within and across domains.

3.1 Experience-Based Perceptions and Insensitivity

Since the functional form that determines signal strength differs for inference and forecasting, we start by conveying the intuition of our model by adopting a simpler functional form. We then apply our model to the inference and forecasting domains in Section 4, and present empirical support for the model.

Signal Strength and Information Structure. The functional form we use to develop intuition is designed to embed two key characteristics of the expressions for signal strength we derived in (6) and (7): first, signal strength depends on multiple features; second, these features interact in non-additive ways. Specifically, let signal strength be given by:

$$S(\mathbf{x}) = \prod_{k=1}^{K} x_k \tag{8}$$

where $\mathbf{x} = (x_1, ..., x_K)$ is a vector of K mutually independent latent features (e.g., persistence, shock size, horizon), that are normally distributed, $x_t \overset{iid}{\sim} N\left(\mu_k, (\tau_{0k})^{-1}\right)$. While true signal strength depends on the objective value of all features, perceived signal strength depends on the DM's perceptions of those features:

$$\tilde{S}(\tilde{\mathbf{x}}) = \prod_{k=1}^{K} \tilde{x}_k \tag{9}$$

Specifically, we assume that the DM does not observe the latent features directly. Instead, she approaches a problem with an experience-based prior about the value of each feature, $x_k^{prior} \overset{iid}{\sim} N(\tilde{\mu}_k, (\tau_{0k})^{-1})$. She then observes a noisy estimate about the objective value of that feature in the current information environment, $\hat{x}_k = x_k + u_k$, where the noise terms $u_k \overset{iid}{\sim} N(0, (\tau_{sk})^{-1})$ are independent across k.[18] Given these priors and noisy estimates, the DM updates her beliefs about each feature using Bayesian inference (Woodford, 2020). This yields the following posterior mean and precision expressions: $\mathbb{E}[x_k] = \alpha_k \hat{x}_k + (1 - \alpha_k)\tilde{\mu}_k$ and $\mathbb{V}_t[x_{kt}] = (\tau_{sk} + \tilde{\tau}_{0k})^{-1}$, where $\alpha_k \equiv \frac{\tau_{sk}}{\tau_{sk}+\tilde{\tau}_{0k}}$. On average, posterior beliefs for feature x_k are then centered on:

$$\tilde{x}_k = \alpha_k x_k + (1 - \alpha_k)\tilde{\mu}_k \tag{10}$$

In what follows we endogenize agents' prior beliefs ($\tilde{\mu}_k$), as well as the weight they put on their noisy estimates (α_k).

Experience-based perceptions. Past experiences shape the priors that the DM approaches the problem with, i.e., her default perception of the information environment:

$$\tilde{\mu}_k = \sum_{\tau \in \mathcal{E}_k} w_{k\tau} \hat{x}_{k\tau} \tag{11}$$

where $w_{k\tau}$ are non-negative recall weights that sum to one over the DM's set of experiences with those features, $\mathcal{E}_k$. Importantly, as shown in Ba et al. (2025) and further discussed in our empirical investigation, the default is not a free parameter but can be explicitly elicited and verified in the data.

Attention Allocation. Next, we endogenize the DM's attention allocation across features. To do so, we assume that by paying more attention to feature x_k, the DM obtains a more precise estimate of that feature (i.e., greater precision τ_{sk}).

While increasing attention to feature x_k improves the accuracy of agents' beliefs, attention is also costly and limited. We denote c_k to be the marginal cost of attending to feature k. Importantly, c_k may capture cognitive costs (e.g., cognitive control in parsing complex information), as well as bottom-up attention factors. For example, this cost is lower when a feature is cognitively less costly to use (e.g., if it's easier to understand how it maps into signal strength), or when it is more easily available (e.g., it is more prominent and salient in the information environment). Finally, total attention is bounded, and agents face a linear budget constraint on attention, $\sum_{k=1}^{K} c_k \tau_{sk} \leq C$. This budget constraint on attention

[18] See also Frydman and Jin (2022) and Frydman and Jin (2023) for evidence on imperfect perception of features in the case of the prospect theory value function and probability weighting function, respectively.

implies an important link across multiple perceived features: directing more attention to one feature decreases the amount of attention left to evaluate all other features.

To endogenize the attention weights, we assume that attention is channeled to minimize mean squared errors subject to the attention budget. Appendix B.4 shows that ceteris paribus, this results in more attention α_k being allocated to feature k when it is more relevant, cheaper to acquire information about (lower c_k), and more volatile (lower $\tilde{\tau}_{0k}$), or when the attention budget is higher (higher C). For example, the model predicts that all else equal people will allocate greater attention to features that are more salient, as these features are more available and easier to process (lower c_k).

3.2 Belief Updating Terminology

The distinction between the perceived features ($\tilde{\mathbf{x}}$) and the objective information environment ($\mathbf{x}$) creates a wedge between true and perceived signal strength. When this difference is positive (negative), agents perceive signal strength to be greater (smaller) than it actually is.

This difference in perceived versus objective signal strength can generate both excess sensitivity and insensitivity to objective variation in the feature, as well as over and underreaction in beliefs. We define insensitivity and excess sensitivity in terms of slopes, and define over and underreaction in terms of levels.

Insensitivity and Excess Sensitivity. Whenever $\frac{\partial S(\tilde{\mathbf{x}})}{\partial x_k} < \frac{\partial S(\mathbf{x})}{\partial x_k}$, subjective signal strength is insensitive to feature x_k. Conversely, when $\frac{\partial S(\tilde{\mathbf{x}})}{\partial x_k} > \frac{\partial S(\mathbf{x})}{\partial x_k}$ subjective signal strength is excessively sensitive to variation in feature x_k.

Over and Underreaction. 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, ($\partial \tilde{x}/\partial x$ and $\tilde{x} - x$). It is important to note that limited attention implies that variation in subjective features is always insensitive to variation in objective features. However, as we illustrate in the next section, insensitivity at the feature level does not necessarily translate into insensitivity at the outcome level.

3.3 Over and Underreaction Across and Within Domains

We now derive predictions on how the cognitive constraints introduced in the previous subsection generate variation in over and underreaction both across and within domains.

Default Perceptions. We start by studying how cued prior experiences can contribute to explaining variation in over and underreaction across domains. Ceteris paribus, when signal strength is increasing in a feature x_k, a higher experience-based prior $\tilde{\mu}_k > \mu_k$ induces the DM to anchor on a higher level of signal strength, therefore contributing to more overreaction (less underreaction). At the same time, $\tilde{\mu}_k < \mu_k$ leads the DM to anchor on a lower level of signal strength, therefore contributing to more underreaction (less overreaction).

Prediction 1. The location of experienced-based priors pin down the default level of signal strength, and determine the overall prevalence of over or underreaction to information. An excessively high prior contributes to overreaction when signal strength is increasing in that feature, and to underreaction when signal strength is decreasing in that feature.

Variation in a Single Feature. Next, we turn to studying how variation in signal strength contributes to explaining variation in over and underreaction *within* domains. First, we consider belief updating in response to variation in a single feature, x_k, while fixing all other features to their cognitive defaults ($x_j = \tilde{\mu}_j$ for all $j \neq k$). The difference between perceived and true signal strength is then given by:

$$\tilde{S}(\tilde{x}_k; \tilde{\mu}_{j \neq k}) - S(x_k; \tilde{\mu}_{j \neq k}) \propto (\tilde{x}_k - x_k) = (1 - \alpha_k)(\tilde{\mu}_k - x_k) \tag{12}$$

This makes clear that limited attention ($\alpha_k < 1$) leads to underreaction to strong signals ($x_k > \tilde{\mu}_k$) and overreaction to weak signals ($x_k < \tilde{\mu}_k$). This is consistent with the results in Augenblick et al. (2025), who show that neglect of signal diagnosticity can generate *both* under and overreaction.

Prediction 2. Insensitivity to a feature can generate both under and overreaction. Varying a single feature while setting all other features at their experience-based prior values leads to overreaction to weak signals and underreaction to strong signals.

Moreover, the expression in (12) shows that decreasing the sensitivity to feature x_k (smaller α_k) amplifies the wedge between $\tilde{S}$ and S. This allows for identification of inattention in the data: for a given variation in signal strength, greater neglect of a feature leads to greater variation in under and overreaction in response to objective changes in that feature. In our model, the allocation of attention across features is determined by how relevant they are, how costly it is to acquire information about them, and how volatile they are. This means that not all features will lead to the same variation in over and underreaction. Ceteris paribus, features that are more salient and easier to process receive greater attention, and drive less variation in over and underreaction.

Our model also predicts that exogenously increasing attention to a feature—e.g., making feature k cheaper to process by increasing its salience—will decrease variation in under and overreaction with respect to that feature. This allows for direct tests of the attention mechanism on observables.

Prediction 3. The lower the attention to a given feature, the more does objective variation in that feature generate variation in observed under and overreaction. Fixing the relevance and volatility of a given feature, making it easier to process (e.g., by making it more salient, or easier to map into signal strength) will draw greater attention to it, and lower the associated variation in over and underreaction.

Interaction of Multiple Features. So far we have considered variation in a single feature while fixing all other features at their default values. We now look at how the interaction of neglected features can break the comparative static of underreaction to strong signals and overreaction to weak signals. To see how this might arise, we focus on the interaction of two features (k and m) while fixing all other features at their default values ($x_j = \tilde{\mu}_j$ for all $j \neq k, m$). We study variation in over and underreaction as we vary feature k, while fixing feature m *away* from its default value.

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

$$\tilde{S}(\tilde{x}_k; \tilde{x}_m, \tilde{\mu}_{j\neq k,m}) - S(x_k; x_m, \tilde{\mu}_{j\neq k,m}) \propto \tilde{x}_m \tilde{x}_k - x_m x_k \tag{13}$$

$$= (\alpha_k \tilde{x}_m - x_m) x_k + \tilde{x}_m (1 - \alpha_k) \tilde{\mu}_k \tag{14}$$

When we evaluate this expression at $x_k = 0$, we see that it is positive, meaning that people overreact to weak signals. Since the first term in (14) is positive and independent of x_k, to obtain underreaction as we increase signal strength via x_k, we need the first term in (14) to be negative, which can only occur if $(\alpha_k \tilde{x}_m - x_m) < 0 \iff \frac{x_m}{\tilde{\mu}_m} > \frac{\alpha_k - \alpha_k \alpha_m}{1 - \alpha_k \alpha_m}$. When this condition is not satisfied, insensitivity to a feature ($\alpha_m < 1$) leads to excess sensitivity with respect to another feature ($\frac{\partial \tilde{S}}{\partial x_k} > \frac{\partial S}{\partial x_k}$).[19] Varying signal strength through x_k then leads to overreaction to both weak and strong signals, therefore breaking the original comparative static.

In Section 4 we provide examples where this condition is satisfied in both inference and forecasting. For example, there are common environments where persistence differs from respondents' experience-based priors ($x_m \neq \tilde{\mu}_m$), and varying shock size (x_k) leads to overreation to both weak and strong signals.

Prediction 4. Once we allow signal strength to depend on multiple interacting features, insensitivity to one feature can lead to excess sensitivity to another. This can break the comparative static of underreaction to strong signals and overreaction to weak signals.

This last result shows how the prediction of overreaction to weak signals and underreaction to strong signals in inference (Augenblick et al., 2025) need not always hold. Section 4 provides additional intuition for this result in the context of the inference and forecasting domains, and further illustrates how there are common environments where the original comparative static breaks down. This illustrates the importance of considering the interaction between *multiple* features in understanding belief formation.

4 Empirical Investigation

We design an experimental setting to test the predictions of our framework while independently varying the relevant features outlined in Table 1. Nesting features of inference and forecasting within the same paradigm allows us to study whether a common mechanism can potentially explain variation in belief formation both within and across domains.

4.1 Experimental Design

Participants ($N = 579$) 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

[19] True signal strength is given by: $S \propto x_m x_k$, while perceived signal strength is given by $\tilde{S} \propto \tilde{x}_m \alpha_k x_k + \tilde{x}_m (1 - \alpha_k)\tilde{\mu}_k$. If we plot S and $\tilde{S}$ against x_k, we find that S has a zero intercept, while $\tilde{S}$ has a positive intercept. This yields overreaction to weak signals (when $x_k = 0$). Moreover, the slope of S is given by x_m, while the slope of $\tilde{S}$ is given by $\tilde{x}_m \alpha_k$. If $\tilde{S}$ is steeper than S, then the wedge $\tilde{S} - S$ is increasing in x_k, and it will never turn negative, and we won't get underreaction to strong signals.

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.[20]

The design of the study closely matched the framework outlined in Section 2.2, and screenshots with instructions are included in Appendix E. 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 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 an AR(1). To convey this information, we showed participants the formula for an AR(1) process, as depicted at the top of Figure 4, and explained each component of the formula in detail. 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. 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.[21] 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.[22] We refer to this as a "Base Trial," and we show one such example in the left panel of Figure 3. 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.[23] The forecasting questions ask participants their predictions for

[20]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.

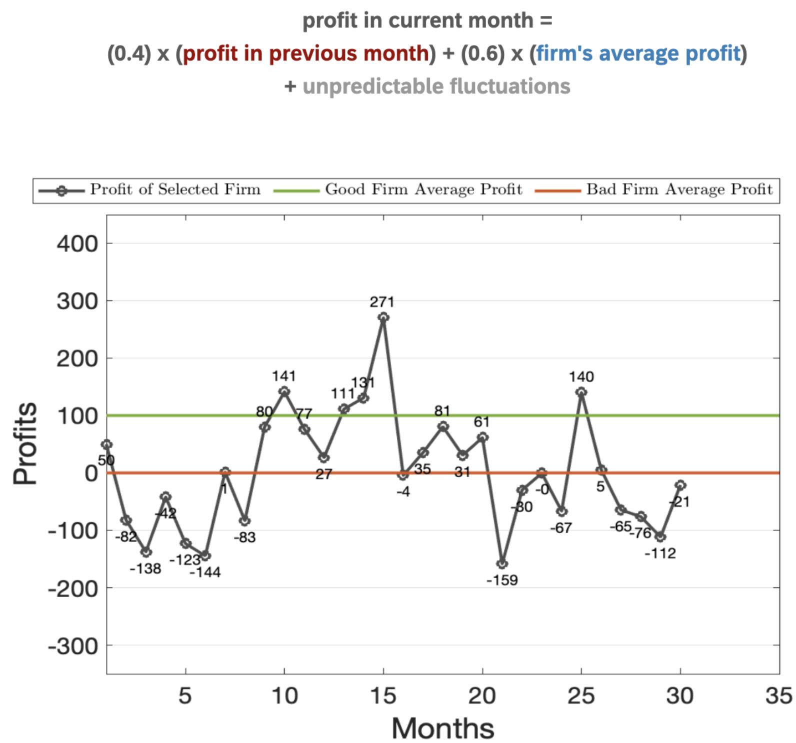
[21]We convey information on variation in persistence in two ways. First, we explicitly provide participants with the DGP, and highlight changes in persistence. Second, we ensure that the in sample autocorrelation of the data we present them with matches the true autocorrelation by choosing trials that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

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

[23]Note that these point estimates amounts to eliciting beliefs about the mode of the distribution. Our

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.

Figure 3: Firm profits over time – Base Trial. We elicited inference and forecasting questions after showing participants 30 months of profits.

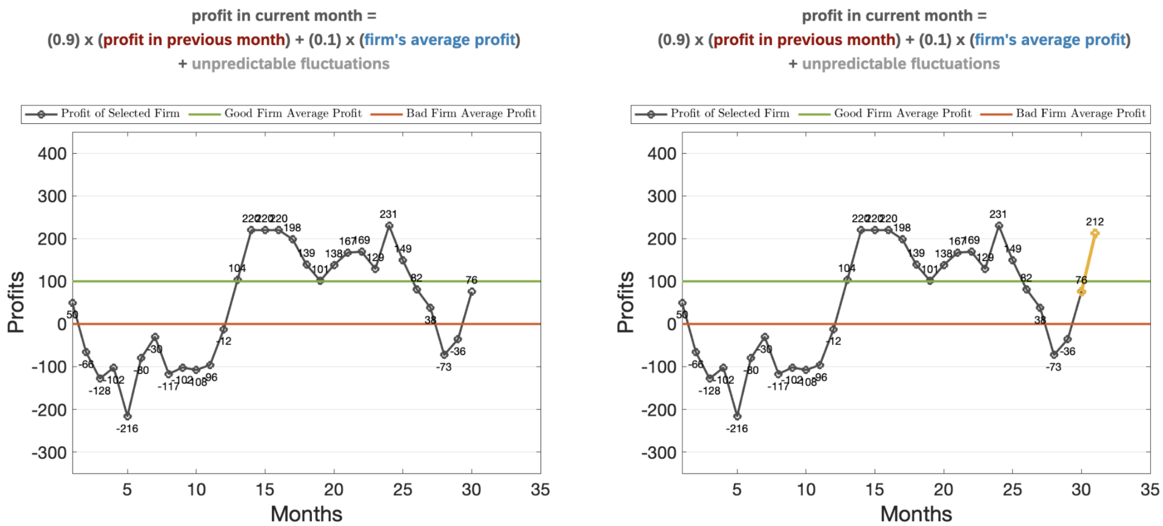


Each respondent completed 9 rounds of Base trials: 3 of these trials have an objective posterior that favors Good firms, 3 of these trials have an objective posterior that favors Bad firms, and 3 of these trials have an objective posterior of 50:50. For the 3 trials that resulted in an objective posterior 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 4. Thus, each respondent completed a total of 12 rounds: 9 Base trials and 3 Shock Trials.

Table 2 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

comparative static predictions both across and within domains are designed to accommodate this type of data.

Figure 4: 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.



trials.[24]

Table 2: 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	-	-	-	-	-	-

Note: Participants saw a total of 12 trials either in the forecasting or inference domains: 9 Base trials and 3 Shock trials, with the Shock trials being a function of the Base trials.

[24]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.

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 the graphs themselves.

Measuring Over and Underreaction. In analyzing our results, we define over and underreaction by comparing participants' revisions to the revisions under the rational benchmark.

$$\theta = (\tilde{\pi}_{A|y} - \tilde{\pi}_A) - (\pi^*_{A|y} - \pi^*_A) \tag{15}$$

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.

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

4.2 Variation in Over and Underreaction Across Domains

Pooling across all parameter values, we see that people overreact significantly more in forecasting than in inference, despite the fact that people are presented with the same information environment across both tasks. In the Base portion, $\theta^{base}_{forecasting} = 36.68$ while $\theta^{base}_{inference} = -6.37$ ($p < .01$); that is, people predominantly overreact in forecasting and predominantly underreact in inference. In the Shock portion, $\theta^{shock}_{forecasting} = 11.88$ versus $\theta^{shock}_{inference} = 1.67$ ($p < .01$).

Prediction 1 explains this variation through different anchoring levels of signal strength. To test this, we calibrate our model to obtain experience-based priors and attention profiles that best fit the data. We then show how these profiles generate the observed variation in over and underreaction across inference and forecasting.

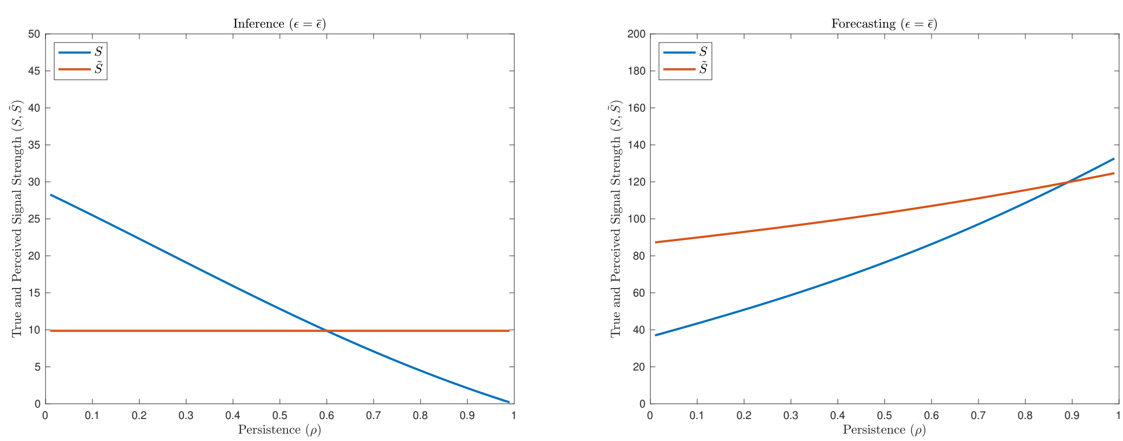
Estimating the Model. Table 3 shows the estimates we obtain from calibrating our model to best fit the data. Figure 5 then uses these parameter values to simulate our model's predicted variation in true (blue lines) and perceived (red lines) signal strength both in inference and forecasting. The first point to notice is that signal strength is decreasing in persistence in inference and increasing in persistence in forecasting. As discussed in Section 3, increasing the persistence of the process makes today's observation less informative about the underlying mean (inference), and more informative about future observations (forecasting).

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

	μ_ρ	μ_ϵ	α_ρ	α_ϵ
Inference	0.6	medium shock	0	0.4
Forecasting	0.9	medium shock	0.4	0.8

Second, and most importantly for this section, Table 3 shows that across both inference and forecasting, people's experience-based priors are associated with high levels of persistence. This generates a greater prevalence of overreaction in forecasting than in inference. Crucially, because signal strength in forecasting increases with persistence, a high default level of persistence anchors individuals to a high perceived signal strength, leading to predominant overreaction. In contrast, since signal strength in inference decreases with persistence, the same default level anchors individuals to a low perceived signal strength, resulting in predominant underreaction. Consistent with this pattern, Figure 5 shows that in forecasting, there are more values of persistence for which perceived signal strength (the red line) exceeds true signal strength (the blue line), whereas the opposite holds for inference.

Figure 5: 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.



The different default levels of signal strength that we document are also consistent with reduced-form heuristics identified in prior work across inference and forecasting (Fan et al., 2025). To build intuition, consider the case where $\rho \to 1$. Evaluating signal strength in inference using (5) at $\rho = 1$ yields an uninformative signal with a log-likelihood of 1. This, in turn, is consistent with people anchoring on the base rate, leading to predominant underreaction. In contrast, evaluating signal strength in forecasting using (7) at $\rho = 1$ leads to anchoring on the most recent observation, which generates predominant overreaction. The fact that we recover a default level of persistence lower than 1 suggests that people's default approach to these problems is to anchor on those values and adjust away from them slightly.[25]

Result 1. People perceive stochastic processes to be highly autocorrelated. High default levels of perceived persistence lead to predominant overreaction in forecasting and predominant underreaction in inference.

4.3 Variation Within Domains

We now proceed to examining variation in belief updating within domains. We start by varying a single feature while fixing all other features at their cued default values. Having established how features map into signal strength, we show that this single feature environment generates results that are consistent with the comparative static of underreaction to strong signal and overreaction to weak signals (Augenblick et al., 2025). We then consider variation in multiple features and show how this comparative static can break down once we consider the interaction of attended to and neglected features.

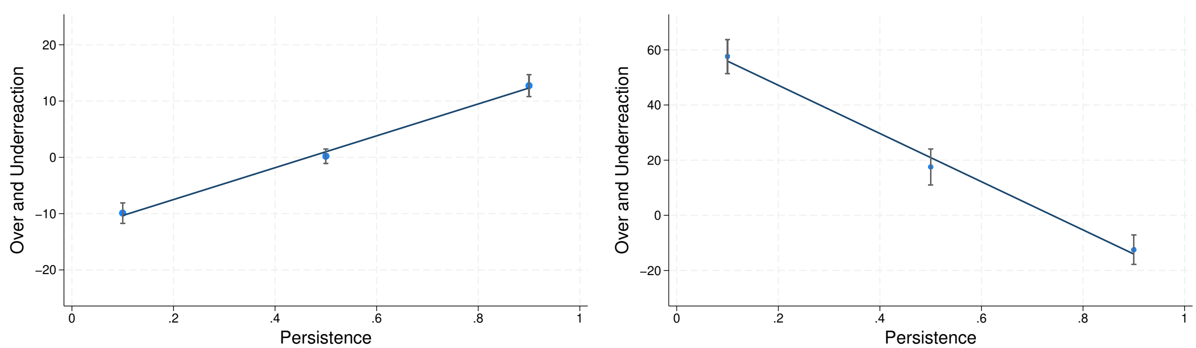
4.3.1 Varying a Single Feature: Underreaction to Strong Signals and Overreaction to Weak Signals

As discussed in Section 2.2, signal strength in both inference and forecasting is a function of many different features in our information environment. In this section, we begin by testing the results in Prediction 2, which considers the implications of varying a single feature while setting all other features at their default values.

[25] Appendix C shows that 22% of responses in inference, and 8% of responses in forecasting are captured by the exact heuristics classified in Fan et al. (2025). The patterns of over and underreaction that we document are robust to excluding these observations.

Persistence. We first consider variation in persistence. Figure 6 shows that varying persistence leads to an *opposite comparative static* with respect to over and underreaction in inference and forecasting. Importantly, both of these patterns are consistent with more underreaction to strong signals and more overreaction to weak signals. The reason for this is that increasing persistence lowers signal strength in inference (the current observation is less informative about the mean) and increases signal strength in forecasting (the current observation is more informative about future observations). Therefore, the patterns in Figure 6 are also both consistent with Prediction 2: people overreact more to weaker signals (higher persistence in inference and lower persistence in forecasting) and underreact more to stronger signals (lower persistence in inference and higher persistence in forecasting).

Figure 6: 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.



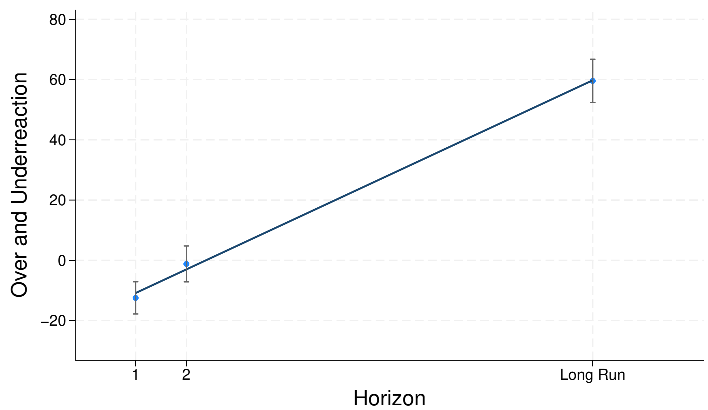
Observing this *opposite comparative static* in the data suggests that persistence does indeed drive variation in over and underreaction through its effect on signal strength.

Finally, notice that the case when $\rho = 0$ nests the one considered in Fan et al., 2025, and in that instance we also find overreaction in forecasting and underreaction in inference. However, by combining the role of noisy cognition with distorted mental representations, our paper is able to explain why we do not always observe overreaction in forecasting and underreaction in inference. Indeed, when $\rho = 0.9$, we observe the opposite, with overreaction in inference and underreaction in forecasting, despite the non-payoff relevant contextual cues being the same across the two cases.[26]

[26] 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. (2025). Moreover, the use of heuristics remains unchanged as we vary different features, meaning that

Horizon. Turning to variation in over and underreaction with respect to changes in the horizon being forecasted, Figure 7 shows that people underreact at short horizons, and overreact at longer horizons. Since signal strength is decreasing with the horizon being forecasted, these results are also consistent with underreaction to strong signals and overreaction to weak signals.

Figure 7: 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 inference and forecasting domains through a common psychological mechanism.

Result 2. When varying a single features while fixing all other features at their default values, people underreact to strong signals and overreact to weak signals. As a result, increasing persistence leads to an opposite comparative static in inference and forecasting, with more overreaction in inference and more underreaction in forecasting. Moreover, increasing the horizon being forecasted leads to more overreaction in forecasting.

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 at the end of this section . 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.

4.3.2 Attention and Variation in Over/Underreaction

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. However, not all variation in signal strength leads to the same variation in over and underreaction. In what follows, we test Prediction 3 by studying how the *amount* of variation in over and underreaction can be explained by how much attention is allocated to a given feature. To do so, we exploit the results in Table 3 to focus on two features that differ in how much attention is allocated to each: persistence, which is mostly neglected, and shock size, which attracts more attention.

The Role of Attention in Over and Underreaction. 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.[27] Figure 8 illustrates one such example in inference. Both of the month 31 shocks depicted in yellow 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 should be discounted more. 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 9 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.[28] The right panel of Figure 9 shows how the same change in signal strength then translates into a greater change in over and underreaction when variation in signal strength is due to persistence (the more neglected feature) than when it is due to the size of the shock (the more attended-to feature).

The same patterns are even more accentuated in forecasting. The left panel of Figure 10 shows that forecast revisions are more sensitive to variation in shock size than to the

[27]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.

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

Figure 8: 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.

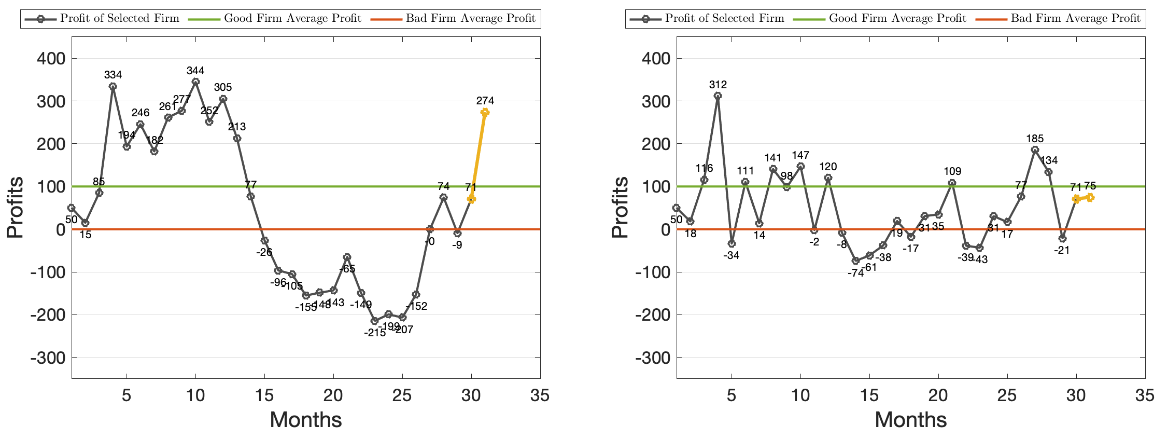
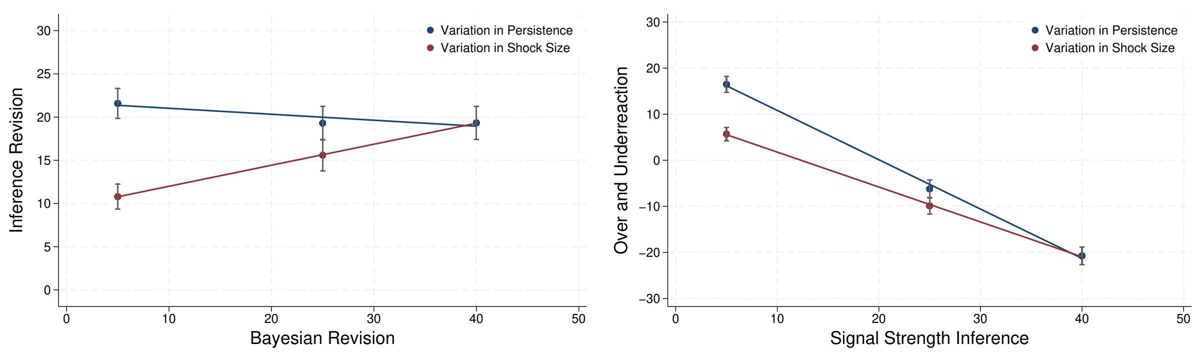


Figure 9: Variation in inference revision and over versus underreaction as shock size and persistence are varied.

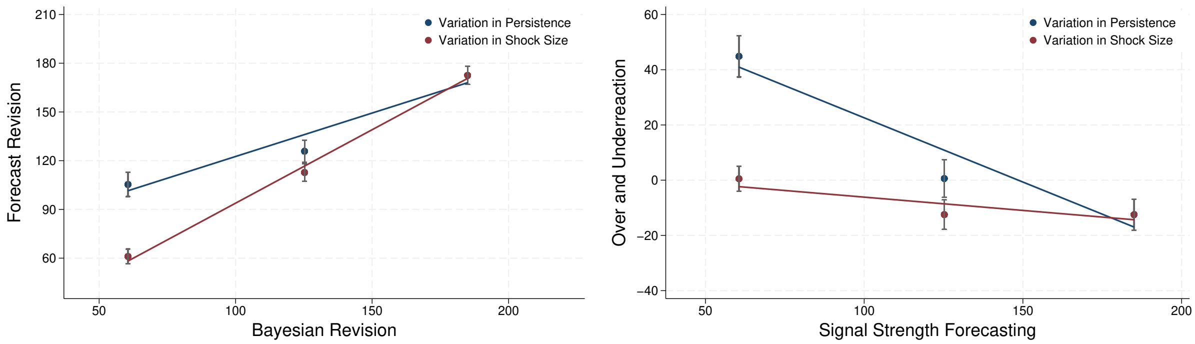


persistence of the process. The right panel of Figure 10 shows that variation in persistence drives more variation in under and overreaction compared to shock size, given the same impact on signal strength. These results are consistent with Prediction 3 of our model.

Result 3. Given the same variation in signal strength, 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 in both inference and forecasting, and as such it generates more variation in over and underreaction.

These results show that variation in under and overreaction with respect to signal strength is modulated by attention. Notably, the relationship between attention and over and

Figure 10: Variation in forecast revision and in over and underreaction as shock size and persistence are varied.



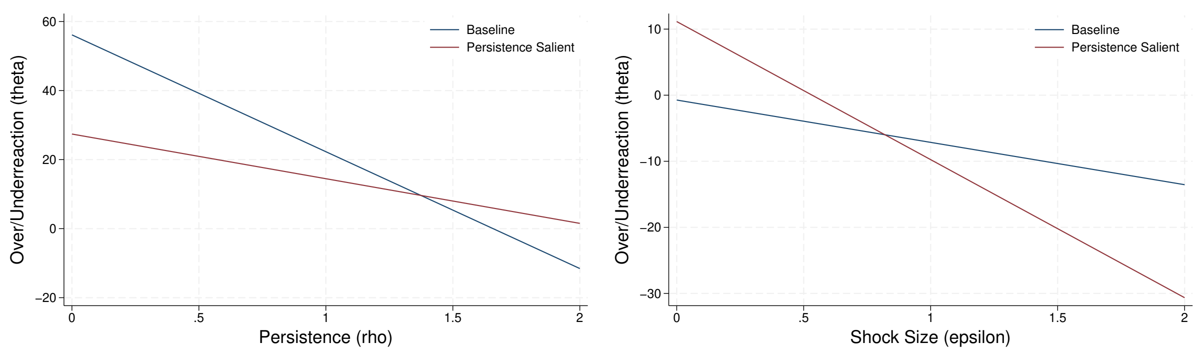
underreaction allows for identifying the allocation of attention—which is often difficult to measure directly—from observable variation in belief-updating. The results also allow us to make progress in understanding what features are associated with variation in over and underreaction. For example, our analysis 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).

Exogenous Variation in Attention. To more directly test our mechanism, we now proceed to test Prediction 3 by exogenously manipulating attention across features. We do so by modifying the design in the main experiment to make persistence more salient. Given a finite attention budget, our model predicts that people will shift attention to persistence and away from shock size, which will decrease the variation in over and underreaction with respect to the former and increase it with respect to the latter.

In the experiment, 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. Writing down descriptions of persistence from their own lives is used to shift attention to persistence while not giving participants any extra information that could be used in the experiment. All participants then answered the same set of forecasting questions as in the main study.

The results that follow compare variation in over and underreaction (θ) as a function of either persistence or shock size. The left panel of Figure 11 shows that variation in over and underreaction with respect to persistence decreased substantially when more attention was allocated to it (the red line associated with more salient persistence is less steep, as also confirmed in Column 1 of Table A.1). This lends support to our core prediction that variation in over and underreaction is driven by limited attention to features: the more a feature is neglected, the greater reliance on the cognitive default and the greater is the discrepancy between changes in true and perceived signal strength. This results in larger variation in over and underreaction.

Figure 11: 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.



Next, the right panel of Figure 11 shows how variation in over and underreaction with respect to shock size differed across treatments. In this case, we see that variation in over and underreaction *increased* when more attention was allocated to persistence (the red line associated with more salient persistence is steeper, as also confirmed by Column 2 of Table A.1). In other words, 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 is allocated to variation in shock size, which then translates into more variation in over and underreaction with respect to the latter.

Importantly, these results highlight how bounded attention generates a critical link between how different features impact belief-updating: increasing attention to one feature will decrease over and underrecation with respect to it, while simultaneously increasing this variation with respect to the feature that is now neglected more.

Result 4. The more a given feature is neglected, the more it drives variation in over and underreaction. Moreover, limited attention generates a key link across features: increasing attention to one feature reduces attention to another. Therefore, shifting attention to one feature and away from another decreases variation in over and underreaction with respect to the former, but increases variation in over and underreaction with respect to the latter.

4.3.3 Interaction of Multiple Features

So far we have analyzed predictions when considering variation in a single feature at a time, while keeping all other features fixed at their default values. However, as highlighted in Prediction 4, the interaction of neglected and attended to features can break the relationship of overreaction to weak signals and underreaction to strong signals. When this is the case, people's perception of signal strength can be excessively sensitive to variation in a given feature, yielding overreaction to all values of signal strength.

Simulations. To gain intuition for this result, Figures 12 and 13 simulate our model to 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 12, 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 informative 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 almost completely neglected in inference, and people perceive it to be 0.6 regardless of its true level. Combining these two observations, we see that the sensitivity of perceived 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. This illustrates how the allocation of attention across features—where one feature is attended to more than another—can lead to insensitivity to the latter (persistence) and *excess sensitivity* to the other (shock size).

To help with intuition, we can think back at the random walk example. As the persistence of the process approaches one, each observation becomes uninformative about the underlying mean, and people should not be very responsive to differences in shock size. However, if people believe the process to be less persistent than it is (because they are anchoring on a lower level of persistence), then they will be more responsive to shock size than they should be. This example makes clear how insensitivity to persistence leads to excess sensitivity to shock size. Ultimately, this then leads to overreaction to both weak and strong signals, therefore breaking the original comparative static of overreaction to weak signals and underreaction to strong signals, as shown in the right-most panel of Figure 12.

Figure 12: 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.

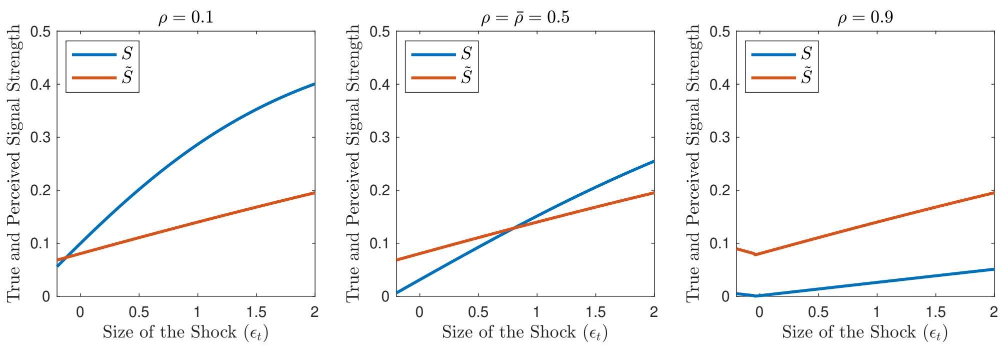
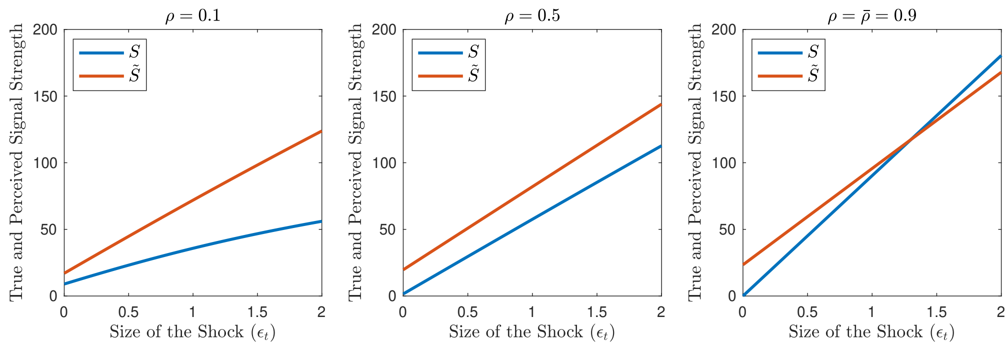


Figure 13 shows how this also applies to forecasting. The sensitivity of true signal strength to the size of the shock varies much more than the sensitivity of perceived signal strength. This allows the latter to go from being lower than the former when $\rho = 0.9$, to being higher than the former when $\rho = 0.1$. 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, and overreaction to both weak and strong signals.

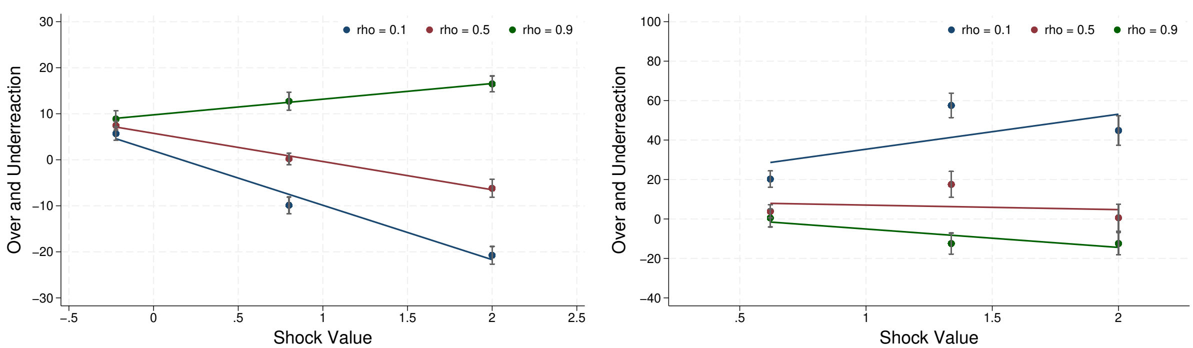
Experimental Evidence. We now proceed to test these predictions in the data. The left panel of Figure 14 plots the extent of over and underreaction in inference for all trials as a function of shock size. Fixing persistence, a larger shock size increases signal strength. When objective persistence is relatively high, stronger signals lead to underreaction, similar to before. However, consistent with Prediction 4, once objective persistence

Figure 13: True and perceived signal strength in forecasting, fixing persistence, and varying the size of the shock.



is low enough ($\rho = 0.1$), stronger signals result in more overreaction, and we never observe underreaction as we vary shock size. Similarly, the right panel of Figure 14 illustrates a similar set of patterns for forecasting. However, consistent with the intuitions we provided above, overreaction is prevalent across both weak and strong signals for low levels of persistence ($\rho = 0.1$).

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.



Result 4. The comparative static of over reaction to weak signals and underreaction to strong signals is modulated by attention. This comparative static can break down once we consider the interaction of attended to and neglected features. The allocation of attention across multiple features can generate excess sensitivity to variation in one due to insensitivity to variation in another.

Result 4 highlights how insensitivity to features can translate into excess sensitivity in perceived signal strength.[29] This illustrates 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 Discussion

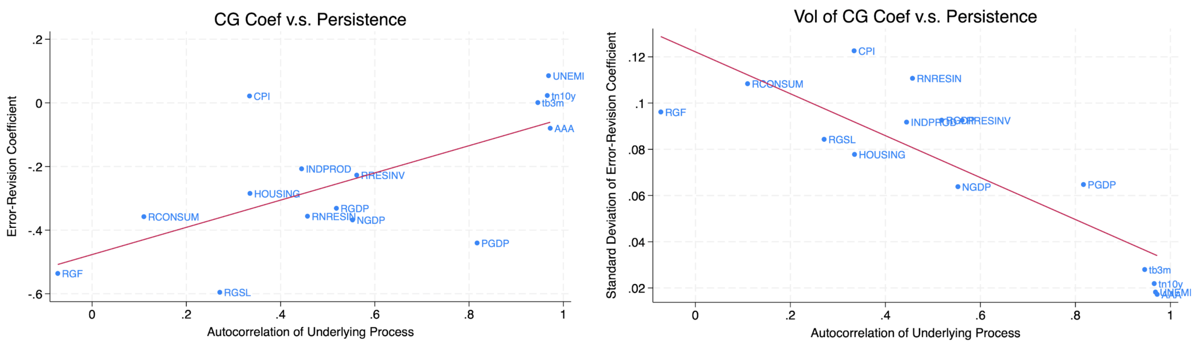
An important aspect of our framework is the prediction that *variation* in over and underreaction should change with the underlying parameters that drive signal strength. The left panel of Figure 15 plots Coibion-Gorodnichenko (CG) regression coefficients on forecasts of macro and financial variables from the Survey of Professional Forecasters, over the persistence of the corresponding objective series. This replicates results in Bordalo et al. (2020) 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 15 extends this analysis and plots the volatility of CG regression coefficients computed over 8 quarter rolling window regressions. Consistent with our model's predictions, we see that more transitory series (for which persistence is further from people's default value of 0.9) exhibit more *variation* in over and underreaction. We view a more formal investigation of these relationships as a fruitful avenue for future research.

6 Conclusions

We propose a framework where constraints on attention and working memory mean that people's perception of the information environment are anchored on past experiences, and are not sensitive enough to objective variation in its features. Variation in over and underreaction across domains can be understood by different tasks cueing different

[29]The fact that insensitivity or neglect can lead to excess sensitivity is also reminiscent of prior work showing that neglecting general equilibrium effects can lead to overreaction (Bastianello and Fontanier, 2025a, Bastianello and Fontanier, 2025b, Mei and Wu, 2024).

Figure 15: 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. (2020) 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.



sets of experiences, leading to different default perceptions of what values each feature should take. Variation in over and underreaction within domains is instead driven by insensitivity away from their cued defaults. When varying a single feature, we show that evidence in both inference and forecasting is consistent with overreaction to weak signals and underreaction to strong signals. However, once we allow for multiple features, we show that this comparative static is modulated by attention. First, not all variation in signal strength leads to the same variation in over and underreaction. Instead, our model predicts that more neglected features lead to greater variation in observed departures from the rational benchmark. Second, the interaction of attended to and neglected features can break the comparative static altogether. When this is the case, insensitivity to a feature can generate excess sensitivity to another.

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.

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

Ba, C., J. A. Bohren, and A. Imas (2025): "Over-and underreaction to information," *Available at SSRN 4274617*.

Bastianello, F. (2025): "Bias and Predictability in Analysts' Beliefs," *Working paper*.

Bastianello, F. and P. Fontanier (2025a): "Expectations and learning from prices," *Review of Economic Studies*, 92, 1341–1374.

——— (2025b): "Partial equilibrium thinking, extrapolation, and bubbles," *Available at SSRN: https://ssrn.com/abstract=4666338*.

Bastianello, F., P. H Décaire, and M. Guenzel (2025): "Mental Models and Financial Forecasts," *Available at SSRN: https://ssrn.com/abstract=5004839*.

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

Bohren, J. A., J. Hascher, A. Imas, M. Ungeheuer, and M. Weber (2024): "A cognitive foundation for perceiving uncertainty," Tech. rep., National Bureau of Economic Research.

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, G. Lanzani, and A. Shleifer (2024a): "A Cognitive Theory of Reasoning and Choice," .

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

Bordalo, P., N. Gennaioli, R. L. Porta, and A. Shleifer (2024b): "Belief overreaction and stock market puzzles," *Journal of Political Economy*, 132, 1450–1484.

Caplin, A. and M. Dean (2015): "Revealed preference, rational inattention, and costly information acquisition," *American Economic Review*, 105, 2183–2203.

Caplin, A., M. Dean, and J. Leahy (2019): "Rational inattention, optimal consideration sets, and stochastic choice," *The Review of Economic Studies*, 86, 1061–1094.

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 Silva, T., E. Larsen-Hallock, A. Rej, and D. Thesmar (2025): "Expectations Formation with Fat-Tailed Processes: Evidence and Theory," *Working Paper*.

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 (2024): "Behavioral attenuation," Tech. rep., National Bureau of Economic Research.

Fan, T., Y. Liang, and C. Peng (2025): "The inference-forecast gap in belief updating," *Working Paper*.

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

——— (2023): "On the source and instability of probability weighting," Tech. rep., National Bureau of Economic Research.

Gabaix, X. (2014): "A sparsity-based model of bounded rationality," *The Quarterly Journal of Economics*, 129, 1661–1710.

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

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

Graeber, T., C. Roth, and M. Sammon (2025): "Coarse Categories in a Complex World," *Working Paper*.

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

Guo, H. (2025): "Earnings extrapolation and predictable stock market returns," *The Review of Financial Studies*, 38, 1730–1782.

Guo, H. and J. A. Wachter (2025): "Correlation neglect in asset prices," *Available at SSRN.*

Hanna, R., S. Mullainathan, and J. Schwartzstein (2014): "Learning through noticing: Theory and evidence from a field experiment," *The Quarterly Journal of Economics*, 129, 1311–1353.

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.*

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

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

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

Mei, P. and L. Wu (2024): "Thinking about the economy, deep or shallow?" Tech. rep., Working Paper.

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

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

Sims, C. A. (2003): "Implications of rational inattention," *Journal of monetary Economics*, 50, 665–690.

Veldkamp, L. L. (2011): *Information choice in macroeconomics and finance*, Princeton University Press.

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

Woodford, M. (2001): "Imperfect common knowledge and the effects of monetary policy," .

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

Wu, L. (2023): "Mental Macro-Finance Models: Evidence and Theory," *Available at SSRN 4332258*.

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:

$$y_t = (1-\rho)\mu + \rho y_{t-1} + \epsilon_t \tag{16}$$

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

The log likelihood function can be written as:

$$\log\left(\frac{P(\mu=\mu_A|y_0,y_1,\ldots,y_t)}{P(\mu=\mu_B|y_0,y_1,\ldots,y_t)}\right)=\log\left(\frac{P(y_0,y_1,\ldots,y_t|\mu=\mu_A)}{P(y_0,y_1,\ldots,y_t|\mu=\mu_B)}\right)+\log\left(\frac{P(\mu=\mu_A)}{P(\mu=\mu_B)}\right) \tag{17}$$

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

$$\begin{aligned}P(y_0,y_1,\ldots,y_t|\mu=\mu_A)&=P(y_t|y_{t-1},\mu=\mu_A)\times P(y_{t-1}|y_{t-2},\mu=\mu_A)\times\ldots\\&\times P(y_1|y_0,\mu=\mu_A)\times P(y_0|\mu=\mu_A)\end{aligned} \tag{18}$$

$$\begin{aligned}P(y_0,y_1,\ldots,y_t|\mu=\mu_A)&=N(\mu_A,\sigma^2;y_t)\times N(\mu_A,\sigma^2;y_{t-1})\times\ldots\\&\times N(\mu_A,\sigma^2;y_1)\times N(\mu_A,\sigma^2;y_0)\end{aligned} \tag{19}$$

$$\begin{aligned}P(y_0,y_1,\ldots,y_t|\mu=\mu_A)&=\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(y_t-\mu_A)^2}{2\sigma^2}\right)\times\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(y_{t-1}-\mu_A)^2}{2\sigma^2}\right)\times\ldots\\&\times\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(y_1-\mu_A)^2}{2\sigma^2}\right)\times\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(y_0-\mu_A)^2}{2\sigma^2}\right)\end{aligned} \tag{20}$$

$$P(y_0,y_1,\ldots,y_t|\mu=\mu_A)=\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^{t+1}\exp\left(-\frac{(y_t-\mu_A)^2+(y_{t-1}-\mu_A)^2+\ldots+(y_0-\mu_A)^2}{2\sigma^2}\right) \tag{21}$$

$$=\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^{t+1}\exp\left(-\frac{\sum_{i=0}^{t}(y_i-\bar{y})^2+(t+1)(\bar{y}-\mu_A)^2}{2\sigma^2}\right) \tag{22}$$

We can then substitute this into the log-likelihood to obtain:[30,31]

$$\log\left(\frac{P(\mu=\mu_A|y_0,y_1,\ldots,y_t)}{P(\mu=\mu_B|y_0,y_1,\ldots,y_t)}\right)=-\frac{(t+1)}{2\sigma^2}\left((\bar{y}-\mu_A)^2-(\bar{y}-\mu_B)^2\right)+\log\left(\frac{P(\mu=\mu_A)}{P(\mu=\mu_B)}\right) \tag{27}$$

$$=(t+1)\left(\bar{y}-\frac{\mu_A+\mu_B}{2}\right)\left(\frac{\mu_A-\mu_B}{\sigma^2}\right)+\log\left(\frac{P(\mu=\mu_A)}{P(\mu=\mu_B)}\right) \tag{28}$$

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

$$(y_t-\mu_A)^2+\ldots+(y_0-\mu_A)^2=((y_t-\bar{y})+(\bar{y}-\mu_A))^2+\ldots+((y_0-\bar{y})+(\bar{y}-\mu_A))^2 \tag{23}$$

$$=(y_t-\bar{y})^2+\ldots(y_0-\bar{y})^2+2(\bar{y}-\mu_A)(\underbrace{(y_t+\ldots+y_0)}_{n\bar{y}}-n\bar{y})+(t+1)(\bar{y}-\mu_A)^2 \tag{24}$$

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

$$(\bar{y}-\mu_A)^2-(\bar{y}-\mu_B)^2=(\bar{y}^2-2\bar{y}\mu_A+\mu_A^2)-(\bar{y}^2-2\bar{y}\mu_B+\mu_B^2) \tag{25}$$

$$=-2\bar{y}(\mu_A-\mu_B)+(\mu_A-\mu_B)(\mu_A+\mu_B)=-2(\mu_A-\mu_B)\left(\bar{y}-\frac{\mu_A+\mu_B}{2}\right) \tag{26}$$

which gives us the expression we are after:

$$\log\left(\frac{P(\mu=\mu_A|y_0,y_1,\ldots,y_t)}{P(\mu=\mu_B|y_0,y_1,\ldots,y_t)}\right) = \underbrace{(t+1)}_{\text{weight}}\underbrace{\left(\bar{y}-\frac{\mu_A+\mu_B}{2}\right)}_{\text{strength}}\underbrace{\left(\frac{\mu_A-\mu_B}{\sigma^2}\right)}_{\text{diagnosticity}}+\underbrace{\log\left(\frac{P(\mu=\mu_A)}{P(\mu=\mu_B)}\right)}_{\text{base rate}} \tag{29}$$

Finite Number of States — persistent shocks with continuous support

$$\begin{aligned} P(y_0,y_1,\ldots,y_t|\mu=\mu_A) = N((1-\rho)\mu_A+\rho y_{t-1},\sigma^2;y_t)\times N((1-\rho)\mu_A+\rho y_{t-2},\sigma^2;y_{t-1})\times\ldots \\ \times N((1-\rho)\mu_A+\rho y_0,\sigma^2;y_1)\times N\left(\mu_A,\frac{\sigma^2}{1-\rho^2};y_0\right) \end{aligned} \tag{30}$$

$$\begin{aligned} P(y_0,y_1,\ldots,y_t|\mu=\mu_A) &= \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(y_t-((1-\rho)\mu_A+\rho y_{t-1}))^2}{2\sigma^2}\right) \\ &\times\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(y_{t-1}-((1-\rho)\mu_A+\rho y_{t-2}))^2}{2\sigma^2}\right)\times\ldots \\ &\times\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(y_1-((1-\rho)\mu_A+\rho y_0))^2}{2\sigma^2}\right)\times\frac{1}{\sqrt{2\pi\frac{\sigma^2}{1-\rho^2}}}\exp\left(-\frac{(y_0-\mu_A)^2}{2\frac{\sigma^2}{(1-\rho^2)}}\right) \end{aligned} \tag{31}$$

$$\begin{aligned} P(y_0,y_1,\ldots,y_t|\mu=\mu_A) &= \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{\left(\frac{y_t-\rho y_{t-1}}{1-\rho}-\mu_A\right)^2}{2\frac{\sigma^2}{(1-\rho)^2}}\right) \\ &\times\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{\left(\frac{y_{t-1}-\rho y_{t-2}}{1-\rho}-\mu_A\right)^2}{2\frac{\sigma^2}{(1-\rho)^2}}\right) \\ &\times\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{\left(\frac{y_1-\rho y_0}{1-\rho}-\mu_A\right)^2}{2\frac{\sigma^2}{(1-\rho)^2}}\right)\times\frac{1}{\sqrt{2\pi\frac{\sigma^2}{1-\rho^2}}}\exp\left(-\frac{(y_0-\mu_A)^2}{2\frac{\sigma^2}{(1-\rho^2)}}\right) \end{aligned} \tag{32}$$

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:[32]

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

[32]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.

$$+\underbrace{1}_{\text{weight}}\underbrace{\left(y_0 - \frac{\mu_A + \mu_B}{2}\right)}_{\text{strength}}\underbrace{\left(\frac{\mu_A - \mu_B}{\frac{\sigma^2}{1-\rho^2}}\right)}_{\text{diagnosticity}} + \underbrace{\log\left(\frac{P(\mu = \mu_A)}{P(\mu = \mu_B)}\right)}_{\text{base rate}} \quad (33)$$

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:

$$\log\left(\frac{P(\mu = \mu_A | y_0, y_1, \ldots, y_t)}{P(\mu = \mu_B | y_0, y_1, \ldots, y_t)}\right) = \underbrace{\left(\frac{y_t - \frac{\mu_A+\mu_B}{2}}{\sigma}\right)}_{\text{strength}}\underbrace{\left(\frac{\mu_A - \mu_B}{\frac{\sigma}{1-\rho}}\right)}_{\text{diagnosticity}}$$
$$+\sum_{i=1}^{t-1} \underbrace{(1-\rho)\left(\frac{y_i - \frac{\mu_A+\mu_B}{2}}{\sigma}\right)}_{\text{strength}}\underbrace{\left(\frac{\mu_A - \mu_B}{\frac{\sigma}{1-\rho}}\right)}_{\text{diagnosticity}}$$
$$+\underbrace{\left(\frac{y_0 - \frac{\mu_A+\mu_B}{2}}{\sigma}\right)}_{\text{strength}}\underbrace{\left(\frac{\mu_A - \mu_B}{\frac{\sigma}{1-\rho}}\right)}_{\text{diagnosticity}} + \underbrace{\log\left(\frac{P(\mu = \mu_A)}{P(\mu = \mu_B)}\right)}_{\text{base rate}} \quad (34)$$

Notice how persistence affects both signal strength and diagnosticity.

B.2 Signal Strength in Inference and Forecasting

In this section we use the above derivations to explain the key features which affect signal strength in inference and forecasting.

B.2.1 Inference

We start by providing intuition for the determinants of signal strength in inference when the process is iid. We then extend the discussion to the case where the process is persistent.

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

$$\log\left(\frac{\pi_{y|G}}{\pi_{y|B}}\right) = \sum_{i=0}^{t}\left(\frac{y_i - \frac{\mu_G+\mu_B}{2}}{\sigma}\right)\left(\frac{\mu_G - \mu_B}{\sigma}\right) \quad (35)$$

$$= \sum_{i=0}^{t} \underbrace{\left(\epsilon_i + \frac{\mu_G - \mu_B}{2\sigma}\right)}_{\text{strength of evidence}} \underbrace{\left(\frac{\mu_G - \mu_B}{\sigma}\right)}_{\text{diagnosticity}} \tag{36}$$

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 (36) 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 (36) 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:

$$\log\left(\frac{\pi_{y|G}}{\pi_{y|B}}\right) = \sum_{i=1}^{t} \left(\frac{y_i - \left(\rho y_{i-1} + (1-\rho)\frac{\mu_G + \mu_B}{2}\right)}{\sigma}\right) \left(\frac{(1-\rho)(\mu_G - \mu_B)}{\sigma}\right) \tag{37}$$

$$= \sum_{i=1}^{t} \underbrace{\left(\epsilon_i + \frac{\mu_G - \mu_B}{2\left(\frac{\sigma}{1-\rho}\right)}\right)}_{\text{strength of evidence}} \underbrace{\left(\frac{\mu_G - \mu_B}{\frac{\sigma}{1-\rho}}\right)}_{\text{diagnosticity}} \tag{38}$$

The strength of the evidence and diagnosticity now depend on the same set of features as in the iid case in (36), 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:

$$\log\left(\frac{\pi_{y|G}}{\pi_{y|B}}\right) = \left(\epsilon_T + \frac{1-\rho}{2}\right)(1-\rho) \tag{39}$$

B.2.2 Forecasting

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

$$\mathbb{E}[y_{t+h}|\{y_i\}_{i=0}^t] = \rho^h y_t + (1-\rho^h)\mathbb{E}[\mu|\{y_i\}_{i=0}^t] \tag{40}$$

and the revision from observing an additional signal is:

$$\mathbb{E}[y_{t+h}|\{y_i\}_{i=0}^t] - \mathbb{E}[y_{t+h}|\{y_i\}_{i=0}^{t-1}] = \rho^h\sigma\epsilon_t + (1-\rho^h)\left(\mathbb{E}\left[\mu|\{y_i\}_{i=0}^t\right] - \mathbb{E}[\mu|\{y_i\}_{i=0}^{t-1}]\right) \tag{41}$$

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 (41) 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.3 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).

B.3 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.

$$\frac{\pi_{G|y}}{\pi_{B|y}} = \frac{\Phi\left(\frac{\hat{\mu}_{MLE}-M}{\frac{1}{\sqrt{T}}\frac{\sigma}{1-\rho}}\right)}{1-\Phi\left(\frac{\hat{\mu}_{MLE}-M}{\frac{1}{\sqrt{T}}\frac{\sigma}{1-\rho}}\right)} = \frac{\Phi\left(\frac{\mu-M}{\frac{1}{\sqrt{T}}\frac{\sigma}{1-\rho}} + \frac{\bar{\epsilon}_t}{\frac{1}{\sqrt{T}}}\right)}{1-\Phi\left(\frac{\mu-M}{\frac{1}{\sqrt{T}}\frac{\sigma}{1-\rho}} + \frac{\bar{\epsilon}_t}{\frac{1}{\sqrt{T}}}\right)} \tag{42}$$

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

$$\hat{\mu}_{MLE} = \frac{1}{T}\sum_{i=1}^{t}\left(\frac{y_i - \rho y_{i-1}}{1-\rho}\right) = \mu + \frac{\sigma\bar{\epsilon}_t}{1-\rho} \tag{43}$$

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:

$$\mathbb{E}_t[y_{t+h}] = \rho^h y_t + (1-\rho^h)\mathbb{E}_t[\mu] = \rho^h y_t + (1-\rho^h)\left(\mu + \frac{\sigma\bar{\epsilon}_t}{1-\rho}\right) \tag{44}$$

Forecast revision in the shock condition is given by:

$$\mathbb{E}_t[y_{t+h}] - \mathbb{E}_{t-1}[y_{t+h}] = \rho^h\sigma\epsilon_t + (1-\rho^h)\left(\mathbb{E}_t[\mu] - \mathbb{E}_{t-1}[\mu]\right) \tag{45}$$

$$= \rho^h\sigma\epsilon_t + \left(\frac{1-\rho^h}{1-\rho}\right)\left(\frac{\sigma\epsilon_t}{t} - \frac{\sigma\bar{\epsilon}_{t-1}}{t}\right) \tag{46}$$

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

B.4 Allocation of Attention Across Multiple Features

Agents decide how to allocate attention across features by minimizing mean squared errors subject to their budget constraint on attention. To simplify notation, we omit task and time subscripts in the derivation.

$$\max_{\{\tau_{sk}\}_{k=1}^{K}} \quad \mathbb{E}[(S(\mathbf{x}) - \tilde{S}(\tilde{\mathbf{x}}))^2] \qquad \text{s.t} \qquad \sum_{k=1}^{K} c_k\tau_{sK} \leq C \tag{47}$$

where $\tilde{S}(\tilde{\mathbf{x}}) \equiv \prod_{k=1}^{K}(\tilde{x}_k)$, and $\tilde{x}_k \equiv \mathbb{E}\left[x_k|\hat{x}_k, \mathcal{D}_k\right] = \frac{\tau_{sk}}{\tau_{sk}+\tilde{\tau}_{0k}}\hat{x}_k + \frac{\tilde{\tau}_{0k}}{\tau_{sk}+\tilde{\tau}_{0k}}\tilde{\mu}_k$ and $\mathbb{V}[x_k|\hat{x}_k, \mathcal{D}_k] = (\tau_{sk}+\tilde{\tau}_{0k})^{-1}$, and where the cognitive default $\mathcal{D}_k$ is defined in the text. Taking a first-order delta-method (Taylor) expansion of $\tilde{S}$ around the realized state $\mathbf{x}$, we obtain:[33]

$$\mathbb{E}[(S(\mathbf{x}) - \tilde{S}(\tilde{\mathbf{x}}))^2] \approx \sum_{k=1}^{K} \frac{\zeta_k^2}{\tau_{sk} + \tilde{\tau}_{0k}} \tag{48}$$

where $\zeta_k^2 \equiv \mathbb{E}\left[(S/x_k)^2\right]$ and the expectation is taken under the subjective prior. Setting up the lagrangian:

$$\mathcal{L} = \sum_{k=1}^{K} \frac{\zeta_k^2}{(\tau_{sk} + \tilde{\tau}_{0k})} + \lambda \left(\sum_{k=1}^{K} c_k \tau_{sk} - C \right) \tag{49}$$

Taking the first order condition for any feature with $\tau_{sk} > 0$:

$$-\frac{\zeta_k^2}{(\tau_{sk} + \tilde{\tau}_{0k})^2} + \lambda c_k = 0 \implies \tau_{sk}^* = \frac{\zeta_k}{\sqrt{\lambda c_k}} - \tilde{\tau}_{0k} \tag{50}$$

Imposing the non-negativity constraint:

$$\tau_{sk}^* = \max\left\{0, \frac{\zeta_k}{\sqrt{\lambda c_k}} - \tilde{\tau}_{0k}\right\} \tag{51}$$

and $\lambda > 0$ is chosen to satisfy the budget constraint $\sum_{k=1}^{K} c_k \tau_{sk}^* = C$. As is standard in these models, the expression in (51) shows that ceteris paribus attention to variable k is higher when the variable is more relevant (higher ζ_k), cheaper to acquire information about (lower c_k), more volatile (lower $\tilde{\tau}_{0k}$) or if the attention budget is larger (larger C).

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. (2024a), 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

[33] Formally, letting $\mathbf{e} := \mathbf{x} - \tilde{\mathbf{x}}$, a first-order expansion gives $\tilde{S}(\mathbf{x} - \mathbf{e}) \approx S(\mathbf{x}) - \nabla S(\mathbf{x}) \cdot \mathbf{e}$. Re-arranging yields: $S(\mathbf{x}) - \tilde{S}(\mathbf{x} - \mathbf{e}) \approx \nabla S(\mathbf{x}) \cdot \mathbf{e}$. Squaring, taking expectations under the subjective prior, and using $\mathbb{E}[e_k e_j] = 0$ for $k \neq j$ yields the expression in (48).

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. (2024a).

Choice Features. Choice features $i \in M_O$ capture both features that relate to the data generating process, as well as relevant realizations:[34]

$$y(\omega) = (e_\theta, e_{\epsilon_T|\theta}, \rho, \epsilon_T) \tag{52}$$

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.

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

$$y(\alpha_O) = (e_\theta(\alpha_O), e_{\epsilon_T|\theta}(\alpha_O), \rho(\alpha_O), \epsilon_T(\alpha_O)) \tag{53}$$

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:

$$\rho(\alpha_O) = \alpha_\rho \rho + (1-\alpha_\rho)\bar{\rho} \tag{54}$$

$$\epsilon_T(\alpha_O) = \begin{cases} \alpha_\epsilon \epsilon_T + (1-\alpha_\epsilon)\bar{\epsilon}_T & \text{if } \pi_{\theta|\epsilon_T} \geq \pi_\theta \\ \alpha_\epsilon \epsilon_T - (1-\alpha_\epsilon)\bar{\epsilon}_T & \text{if } \pi_{\theta|\epsilon_T} < \pi_\theta \end{cases} \tag{55}$$

[34]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.

so that when evaluating the shock, the DM is correctly specified about its direction, but misperceives its magnitude, as in Augenblick et al. (2025). Moreover, the cognitive defaults are determined by

$$\bar{\rho} = \sum_{j=1}^{J} \left(\frac{S(n_j, N) \times Pr(n_j|Q)}{\sum_k S(n_k, N) \times Pr(n_k|Q)} \right) \rho_j \tag{56}$$

$$\bar{\epsilon}_T = \sum_{j=1}^{J} \left(\frac{S(n_j, N) \times Pr(n_j|Q)}{\sum_k S(n_k, N) \times Pr(n_k|Q)} \right) \epsilon_{T,j} \tag{57}$$

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:

$$k = (y, q, n) \tag{58}$$

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:

$$k = (\pi_\theta, \pi_{\epsilon_T|\theta}, \rho, \epsilon_T, q_{focus}, q_{unit}, q_\Theta, n) \tag{59}$$

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 are encountered, $F_c \in \mathbb{R}_+$. We consider five categories, which have been documented empirically in prior work (Fan et al., 2025, 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:

$$k_{ai} = (\pi_\theta, \varphi, \varphi, \varphi, q_u \in \{\pi, y\}, q_f = \theta, q_\Theta = \{G, B\}, n) \tag{60}$$

where φ indicates that the corresponding feature in (67) 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:

$$k_{ss} = (\varphi, \pi_{\epsilon_T|\theta}, \rho, \epsilon_T, q_u \in \{\pi, y\}, q_f = y, q_\Theta = \{\theta\}, n) \tag{61}$$

where φ indicates that the corresponding feature in (67) 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:

$$k_{ss} = (\pi_\theta, \pi_{\epsilon_T|\theta}, \rho, \epsilon_T, q_u \in \{\pi, y\}, q_f = y, q_\Theta = \{\theta\}, n) \tag{62}$$

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

$$k_{er} = (\pi_\theta, \pi_{\epsilon_T|\theta}, \rho, \epsilon_T, \varphi, q_u = y, q_f = \theta, q_\Theta \subseteq \mathbb{R}, n) \tag{63}$$

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:

$$k_{ne} = (\varphi, \varphi, \rho = 1, \epsilon_T, \varphi, q_u = y, q_f = y, \varphi, n) \tag{64}$$

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:

$$\pi_G|y \propto \left(\frac{\pi_G}{\pi_B}\right)\left(\frac{\pi_{y|G}}{\pi_{y|B}}\right) \tag{65}$$

Assuming without loss of generality that $\pi_G|y > \pi_{B|y}$, we can write the likelihood as:[35]

$$\frac{\pi_{y|G}}{\pi_{y|B}} = \exp\left(\left(\epsilon_T + \frac{1-\rho}{2}\right)(1-\rho)\right) \tag{66}$$

[35]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.

The relevant context vector for this problem is given by:

$$k_t = \left(\underbrace{\pi_\theta, \pi_{\epsilon_T|\theta}, \rho, \epsilon_T}_{y\text{: event features}}, \quad \underbrace{q_{\text{unit}} = \pi, q_{\text{focus}} = \theta, q_\Theta = \{G, B\}}_{q\text{: question features}}, \quad \underbrace{n}_{n\text{: non payoff-relevant features}} \right) \tag{67}$$

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:

$$\pi_{G|\theta} = \pi_G \tag{68}$$

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

$$\pi_G|y \propto \frac{\pi_{y|G}}{\pi_{y|B}} \tag{69}$$

where:

$$\frac{\pi_{y|G}}{\pi_{y|B}} = \exp\left(\left(\epsilon_T(\alpha_O) + \frac{1-\rho(\alpha_O)}{2}\right)(1-\rho(\alpha_O))\right) \tag{70}$$

and:

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

$$\alpha_\epsilon \geq 0 \qquad \bar{\epsilon}_T = \sum_{j=1}^{J} \left(\frac{S(n_j, N) \times Pr(n_j|Q)}{\sum_k S(n_k, N) \times Pr(n_k|Q)} \right) \epsilon_{T,j} \tag{72}$$

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:

$$E[y_{t+1}|y] = \rho(\alpha_O)^{h(\alpha_O)} y_T + \left(1 - \rho(\alpha_O)^{h(\alpha_O)}\right) \mathbb{E}[\mu|y] \tag{73}$$

The relevant context vector for this problem is given by:

$$k_t = \left(\pi_\theta, \pi_{\epsilon_T|\theta}, \rho, \epsilon_T, h, q_u = y, q_f = y, q_\Theta = \{G, B\}, n\right) \tag{74}$$

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

$$E[y_{t+1}|y] = \pi_G \times \mu_G \tag{75}$$

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

$$\mathbb{E}_t[y_{t+1}|y] = \rho(\alpha_O)^{h(\alpha_O)}(\rho y_{T-1} + (1-\rho)\mu_\theta + \epsilon_T(\alpha_O)) + \left(1 - \rho(\alpha_O)^{h(\alpha_O)}\right) \left(\pi_{G|\theta} \times \mu_G\right) \tag{76}$$

where:

$$\pi_G|y \propto \frac{\pi_{y|G}}{\pi_{y|B}} \tag{77}$$

with:

$$\frac{\pi_{y|G}}{\pi_{y|B}} = \exp\left(\left(\epsilon_T(\alpha_O) + \frac{1-\rho(\alpha_O)}{2}\right)(1-\rho(\alpha_O))\right) \tag{78}$$

and:

$$\alpha_\rho \geq 0 \qquad \bar{\rho} = \sum_{j=1}^{J}\left(\frac{S(n_j, N) \times Pr(n_j|Q)}{\sum_k S(n_k, N) \times Pr(n_k|Q)}\right)\rho_j \tag{79}$$

$$\alpha_\epsilon \geq 0 \qquad \bar{\epsilon}_T = \sum_{j=1}^{J}\left(\frac{S(n_j, N) \times Pr(n_j|Q)}{\sum_k S(n_k, N) \times Pr(n_k|Q)}\right)\epsilon_{T,j} \tag{80}$$

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:

$$\mathbb{E}[y_{t+1}|y] = \mathbb{1}\left(\int (\pi_\theta \pi_{y|\theta}) \times \mu_\theta \, d\theta > 50\right)\mu_G \tag{81}$$

Naive Extrapolation. When the DM fits the problem to this category, the resulting heuristic is given by:

$$\mathbb{E}[y_{t+1}|y] = y_T \tag{82}$$

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?"

$$E[y_{t+1}|y] = \rho(\alpha_O)^{h(\alpha_O)} y_T + \left(1 - \rho(\alpha_O)^{h(\alpha_O)}\right) \mathbb{E}[\mu|y] \tag{83}$$

Context features given by:[36]

$$k_t = (\pi_\theta, \pi_{y_T|\theta}, \pi_{\theta|y_T}, \rho, \epsilon_T, h, q_u = y, q_f \in \{y, \theta\}, q_\Theta \in \{G, B\}, n) \tag{84}$$

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.

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%

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

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

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 $\rho \in \{0, 0.2, 0.4, 0.6, 0.8, 1\}$[37]

[37] These are the parameters used in AKLMT.

(b) unconditional variance $\frac{\sigma^2}{1-\rho^2} \in \{20, \dots, 50, 60, 70, 80, 90, 100\}$[38]

(c*) if we relax normality of ϵ_t, then there are more features (among others):

i. number of possible values ϵ_t can take $\in \{2, \dots, \text{continuous}\}$[39]

ii. whether ϵ_t is numerical or non-numerical (e.g. color).

2. Prior knowledge over possible DGPs:

(a) number of possible states/bags $\in \{1, 2, 3, 5, 11, \text{continuous}\}$[40]

(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 $\in \{1, \dots, T\}$

(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):

1. Inference:

(a) probability distribution over states/means

(b) expected value of underlying process

2. Forecast – horizon, $h \in \{1, 2, 5, 10\}$[41]:

(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.

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

[39] 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?

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

[41] These are the horizons used in AKLMT.

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.

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		$\mu_A - \mu_B = 100$	$\frac{p(signal=1 \mid value=100)}{p(signal=1 \mid value=0)}$	$\frac{p(up \mid Good)}{p(up \mid Bad)}$	$\frac{p(G \mid \text{Bag i})}{p(R \mid \text{Bag j})}$
	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	$\times$**	$\times$****	✓	$\times$	$\times$
	probability		✓	✓	✓	✓
Forecast	horizon	$h \in \{1, 2, 5, 10\}$	$h = 1$	$h = 1$	$h = 1$	$\times$
	value	✓	✓	✓	$\times$	
	probability	$\times$	$\times$****	✓	✓	

* 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.[42]

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

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

[42]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.[43]

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).

[43] 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.[44]

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.[45]

[44] 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.

[45] 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

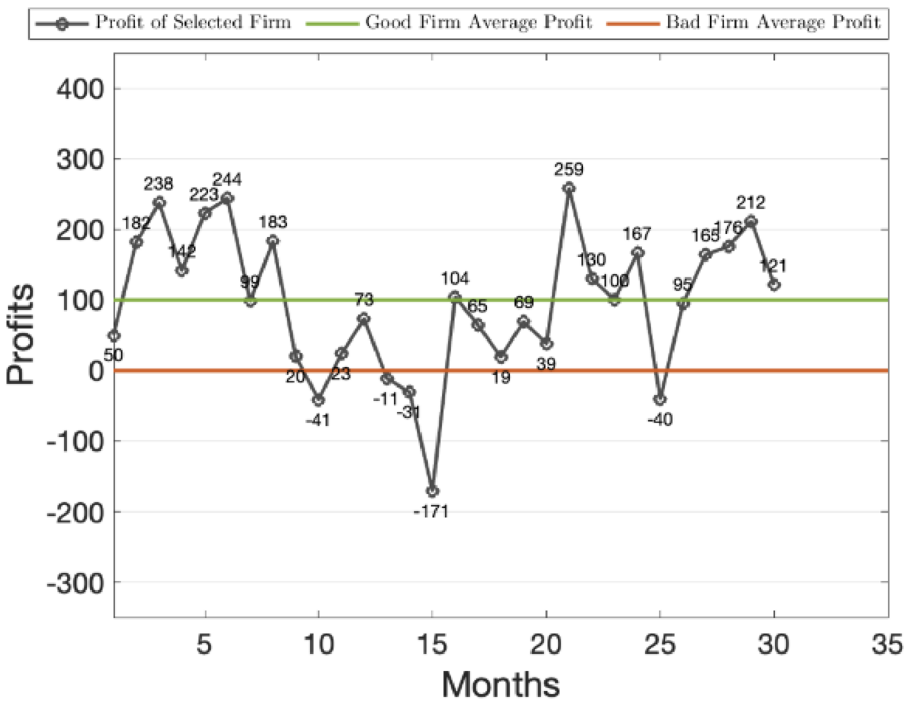
- The **firm's average profit** is **100** if the firm is **good**. This means that a good firm's profits fluctuate around 100.
- The **firm's average profit** is **0** if the firm is **bad**. This means that a bad firm's profits fluctuate around 0.
- The **unpredictable fluctuations** in each period 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.

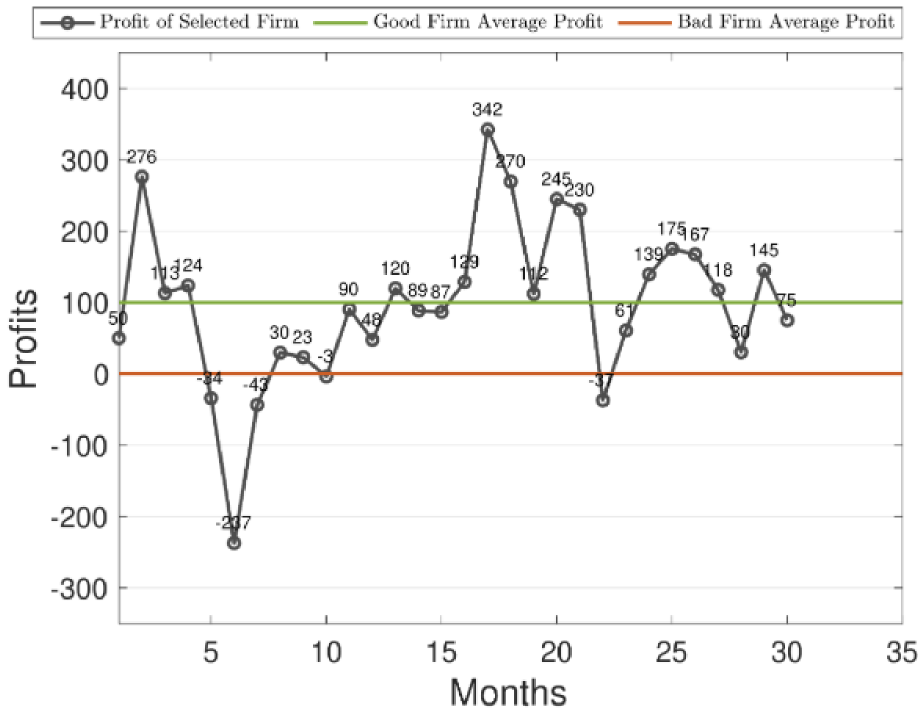
CHECK YOUR UNDERSTANDING on the difference between good and bad firms.

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.

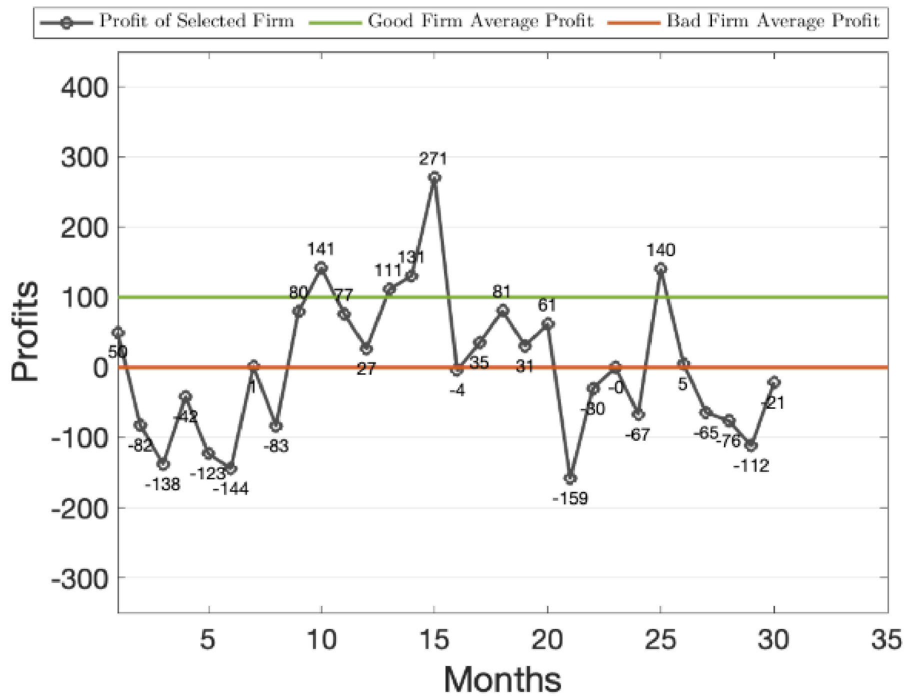


CHECK YOUR UNDERSTANDING on the difference between good and bad firms.

profit in current month =

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

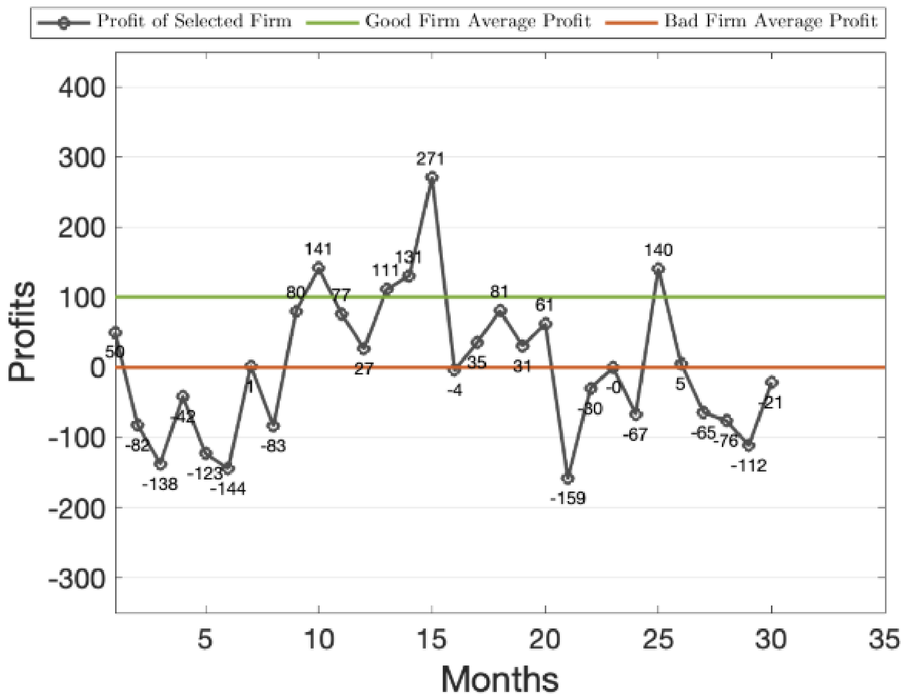
+ unpredictable fluctuations



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:

10 Bad Firms B B B B B B B B B B G G G G G G G G G G 10 Good Firms

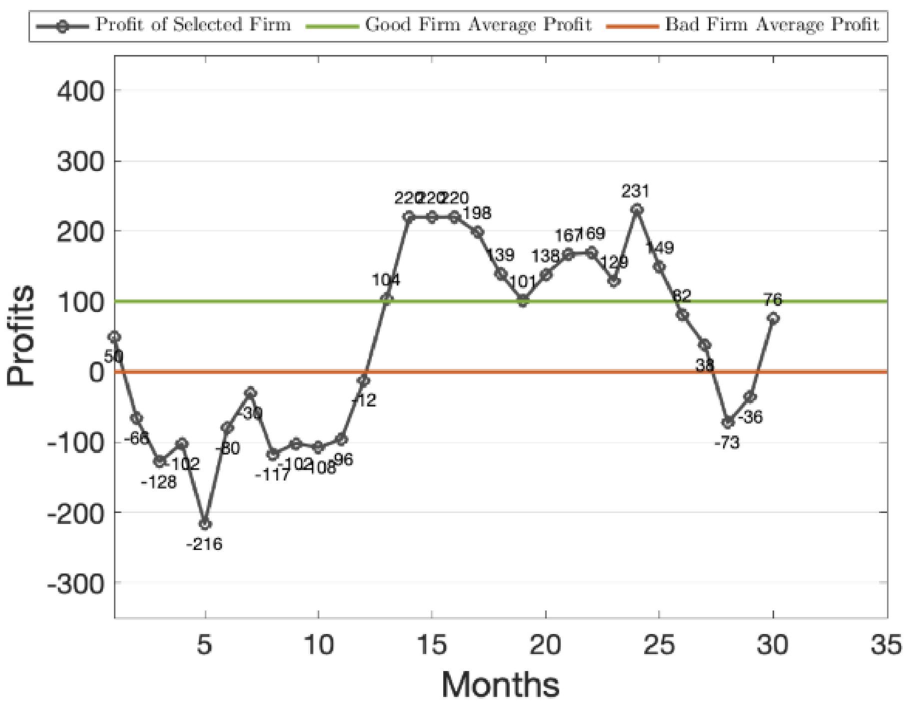
Each of the 20 firms is equally likely to be picked. **In each round, one firm is randomly picked from this pool.** You will not know whether a firm is good or bad, but we will give you information about the profits of the firm over time. You will then make a guess as to whether the firm is good or bad.

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?

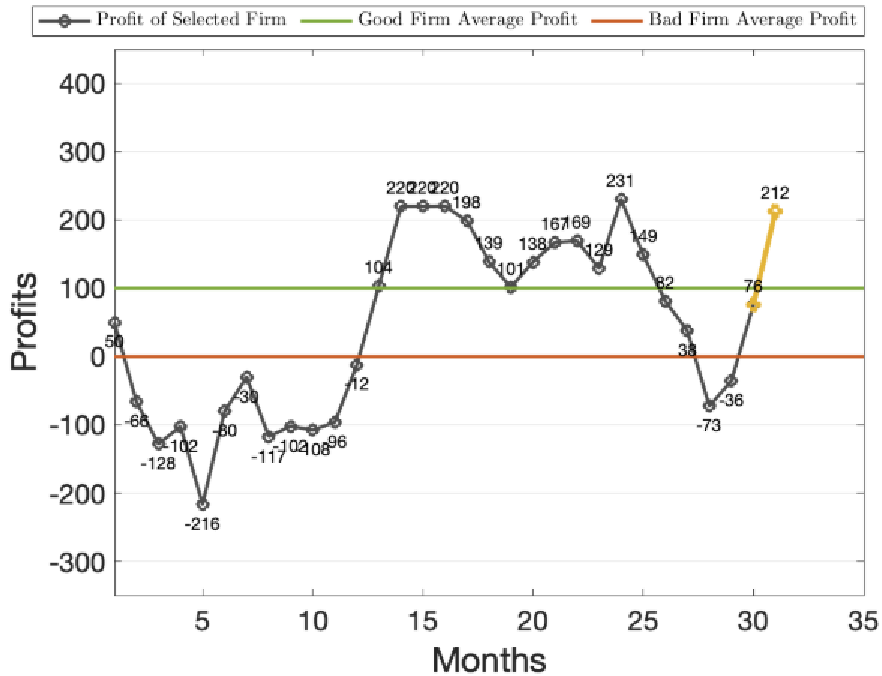
Month 31 profits INCREASE the chances that the selected firm is good

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.