Expectation Formation with Correlated Variables*

Simin He† Simas Kučinskas‡

March 2023

Abstract

We experimentally study how people form expectations about correlated variables. Subjects forecast a time-series variable A. In treatment *Baseline*, subjects only observe past values of A. In treatment *Correlated*, they additionally observe a correlated variable B; A is equally predictable and has the same univariate properties in both treatments. Subjects are significantly less accurate and underreact more in *Correlated*, inconsistent with Bayesian learning. A structural-model estimation indicates that subjects (i) underestimate the level of correlation; and (ii) are insensitive to actual correlation. Our study provides first direct evidence of correlation neglect in the domain of expectation formation.

Keywords: correlated information; overreaction; underreaction; expectation formation.

JEL Classification: C53, C93, D83, D84.

*We are very grateful to Joep Sonnemans for numerous discussions at the early stages of the project, and Zini Yang for her excellent research assistance. For comments and feedback, we also thank Kai Barron, Dirk Engelmann, Xavier Gabaix, Thomas Graeber, John Hey, Artūras Juodis, Agnė Kajackaitė, Eben Lazarus, Jaimie Lien, Yueran Ma, Bin Miao, Florian Peters, David Thesmar, Matthias Weber, Georg Weizsäcker, and Florian Zimmermann, as well as seminar audiences at the Berlin Behavioral Economics Workshop, ESA North American Meetings (Los Angeles, 2019), Experimental Social Science Conference (Tokyo, 2019), DIW, Humboldt University of Berlin, Tsinghua University, and Wake Forest University. Simin He acknowledges the National Natural Science Foundation of China (Grant 72022010) and the Program for Innovative Research Team at Shanghai University of Finance and Economics for their financial support. A previous version of the paper was circulated as "Complexity and Expectation Formation."

†School of Economics, Shanghai University of Finance and Economics (he.simin@mail.shufe.edu.cn, hesimin.com).

‡Faculty of Economics and Business Administration, Humboldt University of Berlin (simas.kucinskas@hu-berlin.de, simaskucinskas.com).

1 Introduction

Motivation. Correlated information structures are pervasive in reality. For example, most macroeconomic variables tend to move together. Firms in the same industry are subject to a common set of shocks. House prices in nearby places are highly correlated. How are expectations formed in these settings? Most previous experimental work focuses on univariate forecasting tasks. As a result, we know little about how people form expectations when faced with multiple correlated variables. Given the central role of expectations in driving economic behavior (Manski, 2018; Coibion et al., 2018; Gennaioli and Shleifer, 2018), that is an important open research question.

What we do. We study how people form expectations about correlated variables using a large-scale online experiment.[1] In the experiment, subjects forecast a target variable A. In treatment *Baseline*, subjects only observe past values of the target variable. In treatment *Correlated*, they additionally observe an informative correlated variable B. Our abstract setup maps directly to the examples mentioned above. For instance, A could represent inflation, and B may stand for oil prices, or the variables may denote house prices in different regions.

Our novel experimental design ensures that the target variable A is equally predictable across treatments. The only difference in *Correlated* is that information is split across two variables. Importantly, our design also keeps constant all univariate properties of the target variable (such as persistence), thus avoiding potential confounding effects. This feature is a key advantage of our experimental approach.

Key findings. We find that subjects in *Correlated* make less accurate forecasts, scoring 13% lower than in *Baseline*. That is a large difference: correlated information accounts for one third of the deviation from rational expectations. This finding is inconsistent with Bayesian updating, as Bayesian updating predicts no difference in accuracy.

Next, we ask how the reaction to new information varies across the two treatments. While subjects overreact to news in *Baseline*, they exhibit strong underreaction in *Correlated*. The magnitude of the effect is large. Consider a shock that increases the true conditional expectation of the target variable by one unit. In *Baseline*, subjects overshoot and increase their forecasts by 1.10 on average. In *Correlated*, in contrast, subjects undershoot and only increase their forecasts by 0.60. This finding suggests that differences in information structures may affect under- and overreaction in expectation formation.

We then zoom in on how subjects respond to new information about A versus B. We find that subjects overreact to the target variable A to a similar degree in both treatments.

[1] Our experiment is preregistered at the AEA RCT Registry as AEARCTR-0004316 (link).

However, they exhibit strong underreaction to the correlated variable B in *Correlated*, which explains the resulting overall underreaction.

In addition to our main treatment, we also investigate the effects of changing the persistence of the target variable A, and the level of correlation between A and B. Subjects become more accurate and underreact less in *Correlated* when persistence is higher, or the correlation is lower. These findings assuage concerns that our baseline findings may be driven by differences in how the data are graphically presented to subjects: one line in *Baseline*, versus two lines in *Correlated*. Focusing on differences within the *Correlated* treatment—in which the data are presented identically—we obtain the same results.

Finally, we provide suggestive evidence that subjects find the *Correlated* treatment more cognitively demanding. First, subjects in *Correlated* take around 20% more time to make their predictions.[2] Second, we estimate a simple structural model of expectation formation that allows for potential biases in the perception of persistence and correlation. We find that subjects underestimate the level of correlation, and the perceived correlation is not sensitive to the actual correlation. In contrast, the perceived persistence varies much more strongly with the actual level of persistence. This finding suggests that the task of estimating correlations is more cognitively demanding than estimating persistence.

Follow-up experiment. To test the robustness of our findings and better understand the underlying behavioral mechanisms, we also conduct a follow-up experiment with two further treatments (not pre-registered). First, we successfully replicate the findings from the original experiment with new subjects. In the first novel treatment, we provide subjects with information on how to interpret persistence and correlation, and only allow subjects to participate in the forecasting task after successfully passing a quiz designed to test their understanding. In the second treatment, we further inform subjects of the specific formula used for the data-generating process. We find that subjects continue to forecast less accurately in *Correlated* treatments. This result demonstrates that the findings in the main experiment are robust, and they are not sensitive to the information provided to the subjects. Moreover, the new treatments indicate that our results are unlikely to be driven by subject confusion or failure to pay attention to B. Instead, it appears that the subjects find it genuinely difficult to make use of correlated information.

Takeaways. Our results suggest that subjects find it cognitively demanding to use information from the correlated variable B. As a result, subjects appear to use a simplified mental model of the forecasting problem that largely ignores B. This interpretation, as well as our main findings, accords with an emerging literature on correlation neglect

[2] Rubinstein (2007) finds that more cognitively demanding choices are associated with longer response times.

in belief formation (Enke and Zimmermann, 2019; Hossain and Okui, 2020). We are the first to provide direct evidence of correlation neglect in the domain of expectation formation. Since correlated information structures are pervasive in reality, our findings suggest that correlation neglect may be important for understanding expectation formation.

Related Literature

Our paper contributes to the growing experimental literature on expectation formation (Assenza, Bao, Hommes and Massaro, 2014; Frydman and Nave, 2017). The two papers most closely related to our work are Afrouzi, Kwon, Landier, Ma and Thesmar (forthcoming) and Enke and Zimmermann (2019).

Our baseline treatment is based on the experimental paradigm of Afrouzi, Kwon, Landier, Ma and Thesmar (forthcoming) which in turn builds on Hey (1994). Afrouzi et al study how people form expectations when predicting future values of an AR(1) process; they also develop a novel theoretical model of expectation formation that is consistent with their empirical findings. The authors find that subjects exhibit persistent overreaction over a wide range of treatments and parameter values. We introduce correlated information and document that it can induce underreaction.

Enke and Zimmermann (2019) document correlation neglect in a belief-formation experiment.[3] In contrast, we study correlated information in an expectation-formation experiment. In the experiments of Enke and Zimmermann, failure to notice that signals are correlated leads to *overreaction*. In our experiment, failure to notice the correlation leads to *underreaction*. Our findings are also related to Graeber (2018) who documents that subjects tend to ignore multiple causes in a belief-formation experiment. Overall, we find the two lines of research to be highly complimentary, and it is encouraging to see similar findings emerge from very different experimental paradigms.

Our paper is related to a large literature on learning-to-forecast experiments; see Hommes (2011) and Assenza, Bao, Hommes and Massaro (2014, Section 3). In contrast to this literature, our experiment does not have any *feedback effects* from expectations to realizations. We made this design choice to ensure that the experimental setting is as simple and transparent to subjects as possible. In addition, the learning-to-forecast paradigm requires the experimenter to take a stand on whether feedback effects are positive or negative (Heemeijer, Hommes, Sonnemans and Tuinstra, 2009; Bao, Hommes, Sonnemans and Tuinstra, 2012). Our results are qualitatively consistent with the learn-

[3] Benjamin (2019) surveys the belief-formation literature. For other recent studies providing evidence of correlation neglect in various decision domains, see Eyster, Rabin and Weizsacker (2015); Eyster and Weizsacker (2016); Hossain and Okui (2020); Rees-Jones, Shorrer and Tergiman (2020).

ing model of Hommes and Zhu (2014) in which agents use a simplified univariate cognitive model, although that model also features feedback effects which are not present in our experiment.

Our study is also related to work on complexity and cognitive constraints more generally (e.g., Kalaycı and Serra-Garcia, 2016). In finance, complex information is more likely to be ignored by investors (You and Zhang, 2009; Umar, 2019).[4] Matthies (2018) provides evidence that beliefs about covariance are compressed towards moderate values, consistent with our own findings on correlation perceptions, while Ambuehl and Li (2018) find that people underreact to the informativeness of signals. Finally, our work is related to a growing theoretical literature on complexity and cognitive constraints, including Gabaix (2019), Angeletos and Huo (2021), Lian (2021), and Ilut and Valchev (2023).

2 Experimental Design

Subjects make one-step-ahead predictions of the *target variable* A_t. Subjects are rewarded for the accuracy of their predictions. Following Dwyer, Williams, Battalio and Mason (1993) and Afrouzi, Kwon, Landier, Ma and Thesmar (forthcoming), we use a bounded linear scoring rule. For each prediction, subjects receive a score S calculated by

$$S = 100 \cdot \max\left\{0, 1 - |e|/\sigma_\varepsilon\right\},$$

where e is the forecast error (realized value of A_t – forecast of A_t), and σ_ε is the standard deviation of shocks to A_t. We cumulate the scores received in each round and convert the total score at the end of the experiment to dollars using a conversion rate of 500 points = 1 dollar. Subjects also receive a base payment of 1 dollar for participating in the experiment. In Appendix B, we show that if the agents are risk neutral and perceive the distribution of A_t to be differentiable, symmetric and unimodal, then the scoring rule induces the subjects to truthfully reveal their subjective expectation of A_t.

Throughout the paper, we refer to forecasts that maximize the expected score for each period as *optimal forecasts*. Given the results in Appendix B and our assumption below that shocks are normally distributed, the optimal forecast of A_{t+1} is the conditional expectation $\mathbb{E}_t[A_{t+1}]$. These optimal forecasts are also equal to full-information

[4] Our paper is also consistent with the recent work of D'Acunto, Hoang, Paloviita and Weber (2019a,b) who document that IQ is a key determinant of expectations, with high-IQ men exhibiting more accurate inflation expectations. In an experiment based on Afrouzi, Kwon, Landier, Ma and Thesmar (forthcoming), they also find that subjects exhibiting higher cognitive ability perform better at predicting AR(1) processes.

rational expectations (Muth, 1961).

2.1 Data-Generating Processes

2.1.1 Baseline Treatment

In the *Baseline* treatment, the target variable follows an AR(1):

$$A_t = \mu(1-\phi) + \phi A_{t-1} + \varepsilon_t, \varepsilon_t \sim \text{i.i.d. } \mathcal{N}(0, \sigma_\varepsilon^2). \tag{1}$$

Here, $\phi \in (-1, 1)$ denotes the *persistence* of A_t, and μ is the unconditional mean, $\mathbb{E}[A_t]$. The variance of the shocks is given by $\sigma_\varepsilon^2 > 0$. For optimal one-step-ahead forecasts, σ_ε^2 is also the mean-squared forecast error. The initial value for the recursion is drawn from the unconditional distribution of A_t.

2.1.2 Correlated Treatment

In the *Correlated* treatment, subjects also predict the target variable A_t. However, they additionally observe a second variable B_t, referred to as the *correlated variable*. The two variables are generated by a bivariate VAR(1):

$$\begin{aligned} A_t &= \mu(1-\phi_1-\phi_2) + \phi_1 A_{t-1} + \phi_2 B_{t-1} + \varepsilon_t \\ B_t &= \mu(1-\phi) \qquad\qquad\qquad\quad\;\; + \phi B_{t-1} \;\; + \eta_t \end{aligned} \tag{2}$$

The shocks follow $(\varepsilon_t, \eta_t)^\top \sim$ i.i.d. $\mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$,[5] and we parametrize

$$\begin{aligned} \phi_1 &= \frac{\phi(1-\rho^2)}{1-\phi^2\rho^2} \\ \phi_2 &= \frac{\rho(1-\phi^2)}{1-\phi^2\rho^2} \\ \mathbf{\Sigma} &= \begin{pmatrix} \sigma_\varepsilon^2 & 0 \\ 0 & \frac{\sigma_\varepsilon^2(1-\phi^2\rho^2)}{1-\rho^2} \end{pmatrix} \end{aligned} \tag{3}$$

Here, $\phi \in (-1, 1)$ measures the *persistence* of both A_t and B_t, while $\rho \in (-1, 1)$ is the *correlation* between A_t and B_{t-1}. Hence, ρ captures the extent to which A_t is predictable from past values of B_t. We initialize the process using the stationary distribution of A_t and B_t. Note that the resulting parametrization is parsimonious: There is only one new parameter (ρ) relative to *Baseline*.

[5] Note that while we use the same symbol (ε_t) to denote the shocks to A_t in both treatments, each subjects is confronted with a unique series of shocks in the experiment.

Why this process? As we demonstrate in Appendix C, the data-generating process in Eqs. (2) and (3) is the unique bivariate VAR(1) process that satisfies the following desired experimental properties:

- P1 (Constant predictability.) Predictability of A_t is the same as in *Baseline*;
- P2 (Constant univariate properties of A_t.) The univariate autocorrelation function of A_t is the same as in *Baseline*;
- P3 (Symmetry.) A_t and B_t have the same univariate autocovariance function.

By "constant predictability," we mean that the mean-squared error of optimal forecasts is the same across treatments. This requirement ensures that the optimal score is the same across treatments. Hence, P1 pins down the variance of shocks to A_t. Since we wish to understand how expectations respond to particular shocks, the covariance matrix of shocks must be diagonal. Otherwise, shocks would not be uniquely defined (see, e.g., Hamilton, 1994, Chapter 11), and the question of whether subjects underreact to shocks would be ill defined. Hence, shocks to A_t and B_t must be uncorrelated. Finally, the variance of shocks to B_t is pinned down by P3.

Satisfying P2 is more difficult. The reason is that for a generic bivariate VAR(1), A_t has the univariate autocovariance function of an ARMA(2,1) process (Zellner and Palm, 1974; Hamilton, 1994, p. 349). Appendix C provides the details on how we construct the stochastic process. Intuitively, while the univariate autocovariance function of A_t is that of an ARMA(2, 1) for a generic VAR(1), it is possible to obtain an AR(1) autocovariance function if parameters are chosen judiciously. The requirement for an AR(1) representation yields a particular non-linear equation. Solving that equation yields the parametrization in Eq. (3).[6]

The construction of the data-generating process ensures an apples-to-apples comparison across the *Baseline* and *Correlated* treatments. That is, if we observe differences in the way subjects form expectations, the differences must stem from changes in the information structure, *not* changes in the univariate properties of A_t. In contrast, if we were to compare how subjects predict A_t generated by a generic AR(1) to A_t generated by a generic VAR(1), differences in the information structure would be confounded with changes in the univariate properties of A_t (such as persistence).

[6] The alternative to our approach would be to choose an ARMA(2,1) process for *Baseline* and use a generic VAR(1) for *Correlated*. We do not find this alternative appealing, for three reasons. First, the resulting univariate process is no longer simple and transparent. For example, optimal predictions, even with full knowledge of parameters, generally require the full history of past values of A_t, or using the Kalman filter to estimate the unobserved shocks (see, e.g., Hamilton, 1994, pp. 83–84). Second, we could no longer easily compare our results to those in the existing literature (Hey, 1994; Afrouzi, Kwon, Landier, Ma and Thesmar, forthcoming). Finally, the approach would substantially increase the number of free parameters that need to be chosen by the experimenter.

Statistical properties. The key properties of the stochastic process in Eqs. (2) and (3) are collected in the proposition below. To state the proposition, let the *univariate autocovariance function* of a scalar stochastic process z_t be denoted by $\gamma_k^z \equiv \mathbb{C}\text{ov}(z_t, z_{t-k})$.

Proposition 1. *The stochastic process for the Correlated treatment satisfies:*

1. *The univariate autocovariance functions of A_t and B_t are identical and equal to*

$$\gamma_k^A = \gamma_k^B = \frac{\phi^k \sigma_\varepsilon^2 (1-\phi^2\rho^2)}{(1-\phi^2)(1-\rho^2)}; \tag{4}$$

2. $\mathbb{C}\text{orr}(A_t, B_{t-1}) = \rho$ *and* $\mathbb{C}\text{orr}(A_t, B_t) = \phi\rho$*;*
3. *The MA(∞) representation of A_t is given by*

$$A_t = \mu + \sum_{\ell=0}^{\infty} \left[\frac{\phi\,(1-\rho^2)}{1-\phi^2\rho^2}\right]^\ell \varepsilon_{t-\ell} + \sum_{\ell=0}^{\infty} \frac{1}{\phi\rho} \left\{\phi^\ell - \left[\frac{\phi\,(1-\rho^2)}{1-\phi^2\rho^2}\right]^\ell\right\} \eta_{t-\ell}.$$

4. *A_t has the univariate representation*

$$A_t = \mu(1-\phi) + \phi A_{t-1} + v_t$$

where the innovations v_t have mean zero, are serially uncorrelated, and have variance $\mathbb{V}\text{ar}(v_t) = \sigma_\varepsilon^2(1-\phi^2\rho^2)/(1-\rho^2)$. *Moreover,* $\mathbb{C}\text{ov}(v_t, A_{t-k}) = 0$ *for* $k > 0$.

Proof. In Appendix E. □

The key takeaway from Proposition 1 is that the target variable A_t has the univariate autocorrelation function of an AR(1) process with persistence ϕ. Hence, the univariate autocorrelation function of the target variable is the same in *Baseline* and *Correlated*. We emphasize that the unconditional variance of the target variable is *not* constant across treatments. The reason is that we keep the predictability of the target variable fixed across treatments. Intuitively, if a subject ignores the correlated variable B_t and only uses past values of the target variable A_t when making forecasts, the target variable must seem less predictable in *Correlated*. For that to be the case, the target variable must be unconditionally more volatile in *Correlated*.

The response of A_t to a shock in A_t itself (ε_t) is always geometrically decaying, as shown in the proposition. However, the response of A_t to a shock in B_t (η_t) can be hump shaped and hence more complicated.

Proposition 1 also provides a simple way to compare the relative usefulness of A_t and B_t in predicting A_{t+1}. The optimal forecast of A_{t+1} is given by $\mathbb{E}[A_{t+1}|A_t, B_t]$.

Now suppose that instead of looking at both A_t and B_t, a subject only considers A_t. In that case, the optimal forecast given the restricted information set is

$$\mathbb{E}[A_{t+1}|A_t] = \mu(1-\phi) + \phi A_t,$$

and the forecast errors are given by v_t with mean zero and variance $\sigma_\varepsilon^2(1-\phi^2\rho^2)/(1-\rho^2)$, as characterized in the proposition. Next, imagine that the subject only considers B_t. The optimal forecast, given the restricted information set, is now given by

$$\mathbb{E}[A_{t+1}|B_t] = \mu(1-\rho) + \rho B_t.$$

The resulting forecast errors have mean zero and variance $\sigma_\varepsilon^2(1-\phi^2\rho^2)/(1-\phi^2)$. As a result, only looking at A_t yields more accurate forecasts if and only if $|\phi| > |\rho|$. Since we only consider positive values for the parameters in our experiment, we will say that A_t is *more important* than B_t in predicting A_{t+1} if $\phi > \rho$, and vice versa when $\phi < \rho$.

2.2 Experimental Procedures

2.2.1 Treatments

In a between-subjects design, subjects are randomly assigned to either the *Baseline* or *Correlated* treatment. In *Baseline*, subjects observe past values of the target variable A_t, and are asked to submit one-step-ahead forecasts of A_{t+1}. In the experimental instructions, subjects are told that *"Past values of "Variable A" are related to future values of "Variable A", and the relationship is stable."* In *Correlated*, subjects additionally observe the correlated variable B_t, and are asked to submit one-step-ahead forecasts of the target variable A_{t+1}. Subjects are informed that *"Past values of "Variable A" and "Variable B" are related to future values of "Variable A", and the relationship is stable."*

In addition to the information structure, we also vary the persistence of A_t and B_t (ϕ) as well as the degree to which A_t and B_{t-1} are correlated (ρ). We have in total six treatment arms, and the treatment summary can be seen in Table 1. In the *Baseline* treatments, the persistence ϕ is either low (0.25) or high (0.75). In the *Correlated* treatments, the persistence ϕ is either low (0.25) or high (0.75), and the correlation ρ is either low (0.25) or high (0.75). In all treatments, we set the expected value of the variables to $\mu = 100$ and use a standard deviation of $\sigma_\varepsilon = 10$ for the shocks to the target variable A_t.

The data series for each treatment are pre-generated, using the parameters for that

Table 1

Treatments Summary

Notes: The table summarizes the treatments that we run experimentally. We vary the information structure (*Baseline* vs. *Correlated*), persistence, and correlation, as shown in the table, with six treatment arms in total. In all treatments, we set $\mu = 100$ and $\sigma_\varepsilon = 10$.

Treatments:	Baseline	Correlated	
		Low correlation ($\rho = 0.25$)	High correlation ($\rho = 0.75$)
Low persistence ($\phi = 0.25$)	Baseline-1	Correlated-1	Correlated-3
High persistence ($\phi = 0.75$)	Baseline-2	Correlated-2	Correlated-4

treatment. Each subject observes a unique pre-generated data series from that treatment. Variables A and B mostly fall in the range of 50–150.

2.2.2 Subjects Recruitment

We used Amazon's Mechanical Turk (AMT) to recruit subjects and conduct our experiment. AMT is an online labor market that is commonly used in social sciences (e.g., Horton, Rand and Zeckhauser, 2011; Cavallo, Cruces and Perez-Truglia, 2017; Kuziemko, Norton, Saez and Stantcheva, 2015; DellaVigna and Pope, 2018).

On July 1st, 2019, we created a Human Intelligence Task (HIT) titled "Prediction task – forecast future values of given variable" with a recruiting quota of 1,000 subjects. When recruiting, we used block randomization (in blocks of six) to randomly allocate each worker into one of the six treatment arms. For participating in the experiment, subjects received a base payment of 1 dollar. In addition, subjects were also paid a bonus according to their forecasting accuracy, as discussed in Section 2.

In order to increase the quality of data on AMT, we imposed the following requirements when recruiting workers: (1) we restricted all workers to be from the USA, (2) we restricted the workers to have a success rate of 95% or higher, and (3) the workers needed to have completed at least 500 HITs.

2.2.3 Experimental Protocol

Our experiment was embedded on Amazon Turk through its External Question API. An AMT worker who saw our posted HIT could preview our experiment on the online platform.[7] In the preview, the worker learned the expected length and expected earnings

[7] The full experimental instructions are provided in Appendix A.

of the experiment. After the preview, a worker who met our selection criteria could accept the task to participate. Once the worker accepted, a unique user ID was assigned to this worker. This user ID determined the treatment the worker was assigned to. After accepting the experiment, the worker had 60 minutes to finish the entire experiment.

During the experiment, the workers first read the instructions about the experiment. Then, they proceeded to the prediction page. In the prediction page, they first saw the initial 40 values of the target variable A_t (and the correlated variable B_t in *Correlated*). They were asked to make predictions of the target variable for the next 40 rounds. After they made each prediction, they were shown the actual realization of the target variable in that round (as well as the realization of the correlated variable in *Correlated*)), and their score for that round was added to their total score. The experimental screen is shown in Figure 1. After finishing the 40 predictions, they filled out a short questionnaire, and were then shown their total score and total payment.

The worker was paid only if the experiment was fully completed within 60 minutes. Otherwise, AMT recruited another worker instead. The HIT was available until 1,000 different workers completed the task.

2.2.4 Pilot

Before conducting our experiment, we ran a small pilot study on AMT on May 22, 2019; we submitted our preregistration after the pilot study. In the pilot, we created an HIT with the same title and recruited 25 workers. We conducted this pilot to estimate (i) time necessary to finish the task; and (ii) average earnings. According to the pilot, we adjusted our base payment and the exchange rate. In addition, we changed the text in the instructions to state that the average expected time to complete the experiment is around 12 minutes.

3 Results

We now turn to the experimental results. We first discuss the sample selection criteria that we impose. Then, we describe the results of our empirical analysis. Our empirical analysis follows the analysis plan laid out in the preregistration closely, except for a small number of deviations summarized in Appendix F.

Figure 1

Screenshot of Prediction Page: *Correlated* Treatment

Notes: Screenshot of the experimental screen in *Correlated*. The only difference in *Baseline* is that there is no variable B that is shown. Subjects submit their scores by either clicking on the blue area or typing their prediction in the text box. The experimental screen in *Baseline* is provided in Figure A.1.

3.1 Data Filtering

In total, 995 subjects successfully completed the experiment, among whom 17 encountered technical problems.[8] After excluding these subjects, we are left with 978 subjects and 39,120 individual forecasts in the initial sample.[9]

Data quality on AMT can be problematic (Ahler, Roush and Sood, 2019). In addition to the requirements on subjects at the recruiting stage, we also exclude potentially low-quality data using the following filters:

1. Exclude the subject if his or her total response time is less than 3 minutes;
2. Exclude the subject if his or her total score is less than 200 points.
3. Exclude an observation if its absolute forecast error is larger than 500.

After imposing the above filters, we are left with a final sample of 945 subjects and 37,788 individual forecasts. Summary statistics for the final sample are shown in Table 2, and the sample selection process is summarized in Appendix G (Table A.2). The average bonus payment is \$2.15. Since the subjects also received a \$1.00 base payment for participating, the average earnings are \$3.15. On average, subjects finish the experiment in around 13 minutes, spending around 10 seconds per each forecasting round. In the final sample, 159 subjects are in *Baseline*-1, 148 in *Baseline*-2, 166 in *Correlated*-1, 156 in *Correlated*-2, 154 in *Correlated*-3, and 162 in *Correlated*-4. To check our randomization and test for differential dropout rates, we perform balance tests between the subjects in the *Baseline* and *Correlated* treatments. We find no significant differences in their dropout rates and other characteristics.[10]

3.2 Forecast Accuracy

We first consider the effect of correlated information on forecast accuracy. As seen in Figure 2, subjects earn a substantially lower score in *Correlated*. The average score is 29.59 per period in *Baseline*, and 25.65 in *Correlated* (MW test, $p < 0.001$). Therefore, including correlated information hurts forecast accuracy significantly. Quite strikingly, average scores in all *Correlated* treatments are lower than in the *Baseline* treatments. Figure 2 also shows that the average scores in all treatments deviate substantially from those of the optimal forecasts (or full-information rational expectations). Compared to

[8] Our quota on AMT was 1,000 subjects. However, 5 subjects had mismatched IDs in the database or missing records. 6 subjects made duplicated forecasts in some periods, and 11 subjects completed the experiment in more than 60 minutes, both of which should not have been possible unless there were technical problems.

[9] The summary statistics of the initial sample can be seen in Appendix G (Table A.1).

[10] Detailed results of the balance tests can be found in Appendix G (Table A.3).

Table 2
Summary Statistics

Notes: Summary statistics for the final sample used for the analysis. Forecast errors are defined as prediction minus the realization. Prediction times per round are shown for all rounds excluding the first round.

	N	Mean	Med.	Std. Dev.	Min.	Max.
Total Time (min)	945	13.20	9.48	10.30	3.02	57.25
Time / Round (sec)	36,843	10.27	7.00	24.51	0.00	1790.00
Score / Round	37,788	26.93	3.28	33.27	0.00	99.99
Forecast Error	37,788	0.99	1.10	17.18	-189.01	299.35
Abs. Forecast Error	37,788	12.31	9.67	12.02	0.00	299.35
Bonus ($)	945	2.15	2.17	0.58	0.44	4.02

the difference between the actual scores and the optimal forecasts, the effect of correlated information is quantitatively large. Roughly a third of the difference between the average score in *Correlated* and the optimal score can be attributed to correlated information. Hence, we have our first main result:

Result 1. *Correlated information has a negative effect on forecast accuracy.*

Next, we investigate whether the treatment effect of correlated information depends on the parameters of the stochastic process (persistence and correlation). We run OLS regressions in which we regress average scores per round on a treatment dummy as well as an interaction term of the treatment dummy with a dummy indicating high persistence or correlation.

The regression results for the effect of persistence are given below:

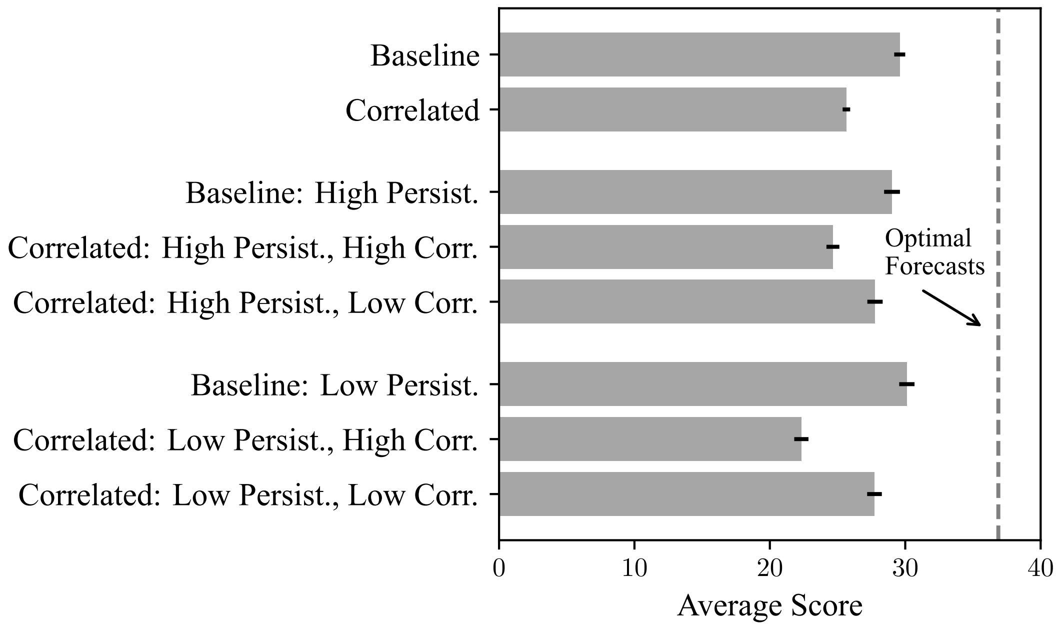
$$\text{avg. score} = \underset{(0.57)}{30.12} - \underset{(0.70)}{4.99}\ \text{correlated} - \underset{(0.81)}{1.10}\ \text{high pers.} + \underset{(0.98)}{2.15}\ \text{correlated} \times \text{high pers.}$$

Here, "high pers." is a dummy variable that equals one when persistence is equal to 0.75, and "correlated" is a dummy variable equal to one for subjects in *Correlated*. Heteroskedasticity-robust standard errors are shown in the parentheses. The regression has one observation for each subject. We observe that subjects' forecast accuracy in *Correlated* is higher when persistence is high. Specifically, the average score is higher by 2.15 in the *Correlated* treatments in which persistence is high. Note, however, that even in treatments with high persistence, the overall treatment effect is still negative ($-4.99 + 2.15 = -2.84$).

Figure 2

Average Scores by Treatment

Notes: Average scores by treatment. Plus/minus one standard error bars shown, with standard errors robust to heteroskedasticity. The average score obtained by the optimal forecasting rule (full-information rational expectations) is given by the dashed vertical line.



We next turn to the effects of correlation:

$$\text{avg. score} = \underset{(0.41)}{29.59} - \underset{(0.56)}{1.85}\ \text{correlated} - \underset{(0.53)}{4.23}\ \text{correlated} \times \text{high corr.}$$

Here, "high corr." is a dummy variable that equals one when correlation is equal to 0.75. From the second regression, we see that the negative effect of correlated information on accuracy is especially high when correlation between A and B is high, reducing the average score by 4.23. In the *Correlated* treatments with low correlation, the effect is still negative and statistically significant but smaller at –1.85.

In conclusion, we find that the effect of correlated information depends on correlation and persistence in the following manner:

Result 2. *The negative effect of correlated information on forecast accuracy is smaller when persistence is high and greater when correlation is high.*

A potential explanation of this result is that in *Correlated*, subjects pay more attention to the target variable A than to the correlated variable B. In treatments with high persistence, past values of A are relatively more important for prediction, while past values of B are relatively more important when persistence is low, as discussed in Section 2. Therefore, when the correlated variable is introduced along with high persistence, not paying much attention to it is not as detrimental to accuracy. In contrast, the correlated variable is relatively more important in treatments with high correlation. If subjects fail to pay enough attention to the correlated variable, their accuracy is reduced more when the correlation is high. In order to directly test whether subjects indeed pay more attention to A than to B, we compare subjects' reaction to new information in A and B in our tests below.

A possible concern with our results is that even a fully Bayesian subject may perform worse in *Correlated* since there are more parameters that need to be estimated. In particular, subjects have to estimate the correlation between A and B.

We now provide evidence that this is unlikely to be the case. First, to approximate the behavior of a Bayesian agent, we calculate least-squares forecasts for the data used in our experiment. We find that least-squares forecasts are very highly correlated with full-information rational-expectations (correlation above 0.97 in both *Baseline* and *Correlated*). Second, we conduct a placebo test to test whether the accuracy scores of least-squares forecasts are statistically different in *Baseline* and *Correlated* (see Table A.5). We find no statistically significant difference. Hence, a Bayesian agent would not exhibit differences in accuracy across treatments. Therefore, the treatment effects that we find are inconsistent with Bayesian updating.

3.3 Under- and Overreaction

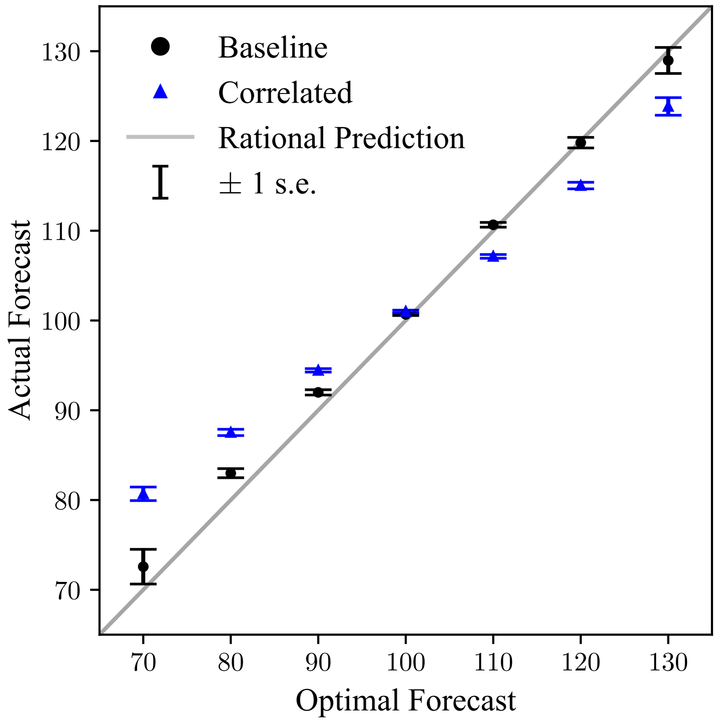
Our second test is whether subjects underreact to new information more in the *Correlated* treatments. More precisely, we ask two questions. First, we test whether subjects exhibit more *overall* underreaction in *Correlated* (i.e., underreaction to the combined information content of A_t and B_t). Second, we zoom in on the reaction to new information contained in A_t and B_t separately. Our methodology for measuring under- and overreaction follows Kucinskas and Peters (forthcoming) closely.

First, we perform a simple test based on the raw data. In particular, we study the relationship between the observed predictions and the optimal forecasts. If subjects behaved optimally, the observed predictions should line up perfectly against the optimal forecasts. In contrast, if subjects underreacted to new information, we should observe a less than one-for-one relationship. This exercise is similar to the classic tests of Mincer and Zarnowitz (1969) and De Bondt and Thaler (1990). The key difference is that we correlate actual predictions with the optimal forecasts (which are unobserved in the

Figure 3

Response of Expectations to New Information

Notes: Response to new information in *Baseline* and *Correlated* treatments. Binned scatterplot of actual vs. optimal forecasts. Plus/minus one standard error bars shown, with standard errors robust to heteroskedasticity. The bins used in the graph are $[65, 75), [75, 95), \ldots, [125, 135]$.



field).

The results of this exercise are shown in a binned scatterplot in Figure 3. As shown in the graph, forecasts in *Baseline* move almost one-for-one with the optimal forecasts. This result suggests that subjects in *Baseline* respond to new information fairly accurately. In contrast, the response to new information in *Correlated* is substantially muted, and compressed towards the simple default of predicting the unconditional mean of 100. This finding suggests that subjects in *Correlated* underreact to new information. More broadly, this exercise is consistent with the recent study of Enke and Graeber (2019) who document similar patterns for choice under risk and uncertainty, as well as belief updating.

We now perform a series of more detailed tests based on Kucinskas and Peters (forthcoming). Given the results in the existing literature, we focus on how subjects react to the most recent new information. Hence, our regressions will focus on how forecasts respond to the most recent shocks $\varepsilon_{i,t}$ (shock to A_t) and $\eta_{i,t}$ (shock to B_t), with i denoting a particular subject.

Forecast errors are defined as $\text{fe}_{i,t+1} = \mathbb{F}_{i,t}[A_{i,t+1}] - A_{i,t+1}$ with $\mathbb{F}_{i,t}[A_{i,t+1}]$ denoting the forecast of subject i at time t. A positive forecast error indicates overprediction, and a negative forecast error indicates underprediction. We define *scaled shocks* by $\tilde{\eta}_{i,t} = \phi_2 \eta_{i,t}$ and

$$\tilde{\varepsilon}_{i,t} = \begin{cases} \phi \varepsilon_{i,t} & \text{in the } \textit{Baseline} \text{ treatment} \\ \phi_1 \varepsilon_{i,t} & \text{in the } \textit{Correlated} \text{ treatment} \end{cases}$$

Economically, scaled shocks measure by how much optimal forecasts should be adjusted after a shock is realized. For example, if the realized shock to A_t is $\varepsilon_{i,t}$, the optimal (i.e., full-information rational expectations) forecast should be revised by $\tilde{\varepsilon}_{i,t}$. It is necessary to scale shocks in this manner because the response of A_t to shocks varies across treatments (as we change persistence and correlation). Hence, we use scaled shocks in all our regressions.[11]

Finally, the optimal forecast revision is given by

$$\text{rev}_t^* \equiv \mathbb{E}_t[x_{t+1}] - \mathbb{E}_{t-1}[x_{t+1}] = \begin{cases} \phi \varepsilon_t & \text{in the } \textit{Baseline} \text{ treatment} \\ \phi_1 \varepsilon_t + \phi_2 \eta_t & \text{in the } \textit{Correlated} \text{ treatment} \end{cases}$$

More details on the construction of optimal forecast revisions are provided in Appendix D. In Appendix D, we also derive theoretical predictions for under- and overreaction for a structural model of expectation formation that we estimate in Section 3.4.

To measure overall under- or overreaction, we estimate

$$\text{fe}_{i,t+1} = \alpha_i + \beta \text{rev}_{i,t}^* + \gamma(\text{rev}_{i,t}^* \times \text{correlated}_i) + u_{i,t}.$$

and test $H_0 : \gamma = 0$. The standard errors are clustered by subject. Under full-information rational expectations, forecast errors are unpredictable by variables available at the time of making the forecast, implying $\beta = \gamma = 0$. In this regression, β measures overall overreaction in *Baseline*, while $\beta + \gamma$ provides overall overreaction in *Correlated*.

Economically, β can be interpreted as follows. Consider a shock that increases the true conditional expectation of A by one unit. Then, subjects increase their forecast by $1 + \beta$ on average. Therefore, a positive estimate of β suggests that subjects react too much (overreaction), and a negative estimate of β indicates that subjects react too little (underreaction).

[11] Regressing forecast errors on the scaled shocks is equivalent to regressing forecast errors on past shocks and normalizing the estimated coefficients by the true impulse responses, as done by Coibion and Gorodnichenko (2012).

Table 3

Overreaction: Main Results

Notes: Estimated levels of overreaction in *Baseline* and *Correlated*, with sample splits by the level of persistence (ϕ). Standard errors clustered by subject in parentheses. The last column gives the difference in overreaction levels between *Baseline* and *Correlated*, and the bottom row gives the difference between the low- and high-persistence treatments. The table shows estimates of β from the regression of

$$\text{forecast error}_{i,t} = \alpha_i + \beta\, \text{rev}^*_{i,t-1} + u_{i,t}$$

for the appropriate subsample, where $\text{rev}^*_{i,t}$ is the optimal forecast revision. Differences between treatments are estimated by running regressions with appropriate interaction terms.

	Baseline	Correlated	B−C
Full Sample	0.11 (0.03)	-0.43 (0.02)	0.53 (0.04)
Low Persistence	0.74 (0.10)	-0.64 (0.03)	1.38 (0.10)
High Persistence	0.04 (0.04)	-0.18 (0.03)	0.22 (0.05)
L−H	0.70 (0.11)	-0.46 (0.04)	1.16 (0.11)

Table 3 summarizes the estimated overreaction levels in *Baseline* and *Correlated*. The first row shows the results for the full sample, followed by estimates for different persistence levels. We observe that in *Baseline* there is modest overreaction for the full sample. Subjects overshoot and exhibit an overreaction level of 0.10. By estimating overreaction separately for treatments with different persistence levels, we find that overreaction in *Baseline* is driven by strong overreaction in the low-persistence treatment. This result is similar to the existing literature (Afrouzi, Kwon, Landier, Ma and Thesmar, forthcoming), though our results indicate that such overreaction may disappear when persistence is sufficiently high. This result is consistent with the model put forward by Gabaix (2019) which predicts that overreaction is more likely to occur when the target variable is less persistent.

In contrast, subjects exhibit underreaction in *Correlated*. For one unit increase in the true conditional expectation of A, subjects only increase their forecast by 0.57. Moreover, we find that subjects underreact more when persistence is low. Hence, we have the following result:

Result 3. *Subjects overreact in Baseline but underreact in Correlated.*

Table 4 performs a similar exercise but now splitting the sample by correlation lev-

Table 4

Overreaction: Effects of Correlation

Notes: Estimated levels of overreaction in *Baseline* and *Correlated*, with sample splits by the level of correlation (ρ). Standard errors clustered by subject in parentheses. See Table 3 for further explanations.

	Baseline	Correlated	B−C
Full Sample	0.11	-0.43	0.53
	(0.03)	(0.02)	(0.04)
Low Correlation	0.11	-0.07	0.18
	(0.03)	(0.04)	(0.05)
High Correlation	0.11	-0.56	0.67
	(0.03)	(0.03)	(0.04)
L−H	–	0.49	-0.49
		(0.05)	(0.05)

els. In *Correlated*, subjects underreact strongly when correlation between A and B is high but only mildly when correlation is low. One possible explanation is that when correlation is high, B carries a substantial amount of useful information for making forecasts, yet subjects do not incorporate this information sufficiently. This behavior leads to overall underreaction. The intuition for differences with respect to persistence is similar: When persistence is lower, A becomes relatively less important for prediction than B. In the next section, we provide evidence that subjects indeed underreact more to B than to A.

Summarizing, the treatment effect of correlated information on underreaction depends on the parameters of the process as follows:

Result 4. *In Correlated, subjects underreact more when persistence decreases or correlation increases.*

To measure overreaction to information about A_t and B_t separately, we estimate

$$\text{fe}_{i,t+1} = \alpha_i + \beta\tilde{\varepsilon}_{i,t} + \gamma(\tilde{\varepsilon}_{i,t} \times \text{correlated}_i) + u_{i,t},$$

and test $H_0 : \gamma = 0$. Similarly to before, β measures overreaction to shocks to A_t in *Baseline*, and $\beta + \gamma$ gives the same measure for *Correlated*. Then, we use only data from *Correlated* and estimate

$$\text{fe}_{i,t+1} = \alpha_i + \beta\tilde{\eta}_{i,t} + u_{i,t}$$

Table 5

Overreaction to A vs. B.

Notes: Estimated levels of overreaction to shocks in A (target variable) and B (correlated variable). Standard errors clustered by subject in parentheses. The table shows estimates of β from the regression of

$$\text{forecast error}_{i,t} = \alpha_i + \beta\,\text{scaled shock}^*_{i,t-1} + u_{i,t},$$

where the scaled shocks are defined in the text and represent the revision in the optimal forecast stemming from the realization of that shock.

	Baseline	Correlated	B−C
Shock to A	0.11 (0.03)	0.10 (0.04)	0.00 (0.05)
Shock to B	–	-0.71 (0.03)	–
Shock to A−Shock to B	–	0.81 (0.05)	–

to measure overreaction to shocks to B_t.

Table 5 contains the estimated overreaction levels to A and B separately. We find that, in *Correlated*, subjects overreact to the target variable A to a similar degree as in *Baseline*. This result helps alleviate a potential concern with our findings. Specifically, one may worry that subjects are confused by multiple variables presented in *Correlated*, and are therefore unable to understand or perform the task. But this is very unlikely given that the amount of overreaction to A is similar across treatments.

Subjects, however, exhibit strong underreaction to B. This result indicates that the overall underreaction found in *Correlated* is driven by subjects underreacting to B. Therefore, we conclude that:

Result 5. *The overall underreaction in Correlated is entirely driven by underreaction to the correlated variable B.*

The level of underreaction to B is quantitatively large, implying that only around 30% of the informational content in B is incorporated by the subjects. This finding appears consistent with sparsity-based models of limited attention (Gabaix, 2014, 2017).

Finally, we note that our estimates are robust to the choice of benchmark used to measure overreaction. In the estimates above, we implicitly measure how subjects respond to shocks, and then compare that response to the true reaction of the target variable. However, the true reaction of the process is unknown to the subjects. Hence, a potential concern is that even a fully rational agent may exhibit under- or overreaction.

Table 6

Differences in Response Times

Notes: Average response times across *Baseline* and *Correlated*. Response times per round, excluding the first ten rounds, are averaged for each subject and then compared across the two treatments. Heteroskedasticity-robust standard errors shown; Mann-Whitney p-value is provided in the bottom row.

	Baseline	Correlated	B−C
Time / Round (sec)	7.82	9.35	-1.53
s.e.	0.36	0.29	MW p-value: < 0.001

That is unlikely to be the case, for two reasons. First, as discussed previously, full-information rational expectations are very highly correlated with least-squares forecasts (correlation above 0.97). Second, we obtain essentially identical numbers if we use least-squares forecasts as our benchmark. To do that, we repeat the analysis above. However, instead of using forecast errors as our left-hand side variable, we use the difference $\mathbb{F}_{i,t}[A_{i,t+1}] - \mathbb{F}^{\text{OLS}}_{i,t}[A_{i,t+1}]$, where $\mathbb{F}^{\text{OLS}}_{i,t}[A_{i,t+1}]$ is the ordinary least-squares forecast. This alternative approach yields very similar results, and the detailed tables can be found in Appendix G.

3.4 Mechanism Questions

Our final tests investigate the channels through which forecasts are less accurate and underreaction occurs in *Correlated*.

Is the *Correlated* treatment more cognitively demanding? We first ask whether the *Correlated* treatment is more cognitively demanding by studying (a) prediction times; and (b) slope of the learning curve.

For (a), we compare the average prediction times for the last 30 rounds of the experiment in *Baseline* and *Correlated* using the Mann-Whitney test.[12] Table 6 presents the average response times by treatment condition. We observe that subjects take longer to make a forecast in *Correlated*, and the difference is statistically significant. This result indicates that *Correlated* is indeed more cognitively demanding for the subjects. This finding can potentially explain why subjects perform worse in *Correlated*.

For (b), we calculate the average score in the first and the last twenty rounds for each subject. We then approximate the slope of the learning curve by the difference between the two scores. Finally, we compare the slopes across treatments using the

[12] Subjects in our experiment may not start the experiment right away after accepting the task on MTurk, or submit a few predictions to estimate how long it may take to finish the experiment and then continue working on other tasks. As a result, prediction times in the first few rounds are not very accurate measures of prediction times.

Table 7

Testing for Learning Effects

Notes: Average learning rates across *Baseline* and *Correlated*. The learning rate is defined as the difference between the average score per round in the second and first halves of the experiment. Heteroskedasticity-robust standard errors shown; Mann-Whitney p-value is provided in the bottom row.

	Baseline	Correlated	B−C
Learning	-0.00	0.92	-0.92
s.e.	0.63	0.44	MW p-value: 0.16

Mann-Whitney test. Table 7 shows the slope of the learning curves in *Baseline* and *Correlated*, respectively.[13] We can see that subjects exhibit no learning at all in *Baseline*. In *Correlated*, there is at most a modest role for learning, and the learning rates are not statistically significant between the two treatments.[14] This result indicates that learning cannot explain the differences in accuracy between *Baseline* and *Correlated*.

Result 6. *Including correlated information yields longer response times, but it does not seem to induce more learning.*

How do subjects perceive the data-generating process? Next we study how the subjects perceive the parameters of the data-generating process. For that, we estimate the parameters of the simple structural model of expectation formation presented in Appendix D.

For subjects in *Baseline*, we run the regression

$$\text{forecast}_{i,t} = \alpha_i + \phi_p A_{i,t} + u_{i,t}$$

for each subject i separately to estimate the perceived correlation level ϕ_p. Figure 4 shows the estimated perceived persistence levels in each of the *Baseline* treatments. We see that the perceived persistence levels are fairly close to the actual ones, with some overestimation when the actual persistence is low. This result suggests that subjects are able to perceive the underlying persistence to some extent.

For subjects in *Correlated*, we first run the regression

$$\mathbb{F}_{i,t}[A_{i,t+1}] = \alpha_i + \phi_{1,p} A_{i,t} + \phi_{2,p} B_{i,t} + u_{i,t}$$

to estimate the perceived law of motion for A_t. We then invert the true formulas in

[13] Figure A.2 in Appendix G shows the average scores over time in *Baseline* and *Correlated*.

[14] We obtain the same result (not reported here) by regressing individual scores on a linear time trend and an interaction term of the time trend with the treatment dummy.

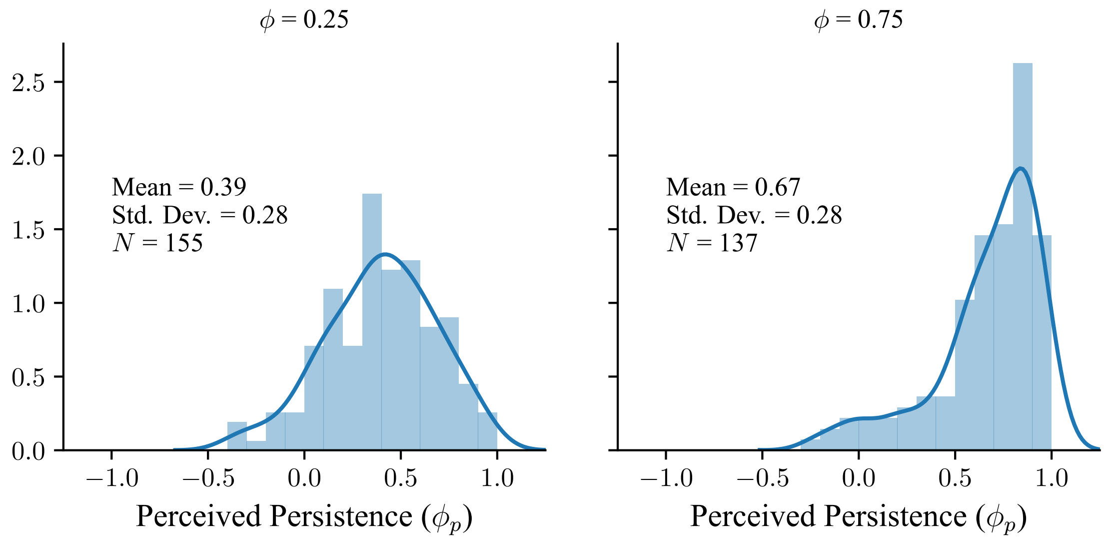
Figure 4

Perceived Persistence in *Baseline* Treatments

Notes: Subject-level estimates of perceived persistence; only for subjects in *Baseline*. The estimated persistence levels are given by $\hat{\phi}_{i,p}$ in the regression

$$\text{forecast}_{i,t} = \alpha_i + \phi_{i,p} A_{i,t} + u_{i,t}.$$

The regression is estimated for each subject i separately. Each panel shows estimates for subjects from a particular experimental condition (specific true persistence ϕ). Kernel density estimates with a Gaussian kernel and histograms plotted.



Eq. (3) to obtain the perceived persistence ϕ_p and correlation level ρ_p by

$$\begin{aligned} \phi_p &= \frac{[1 + (\phi_{1,p})^2 - (\phi_{2,p})^2] - \sqrt{[1 + (\phi_{1,p})^2 - (\phi_{2,p})^2]^2 - 4(\phi_{1,p})^2}}{2\phi_{1,p}} \\ \rho_p &= \frac{[1 - (\phi_{1,p})^2 + (\phi_{2,p})^2] - \sqrt{[1 + (\phi_{1,p})^2 - (\phi_{2,p})^2]^2 - 4(\phi_{1,p})^2}}{2\phi_{2,p}} \end{aligned} \tag{5}$$

We perform this exercise separately for each *Correlated* treatment (i.e., for each level of ϕ and ρ).

Figures 5 and 6 show the perceived correlation and persistence levels in each of the *Correlated* treatments. From Figure 5, we see that the perceived correlation tends to be lower than the actual correlation, and it is not sensitive to changes in the actual correlation. From Figure 6, in contrast, we observe that the perceived persistence increases with actual persistence, and the perceived levels are fairly close to the actual ones. All distributions are unimodal and fairly normally distributed. This observation suggests

Figure 5

Perceived Correlation

Notes: Subject-level estimates of perceived correlation; only for subjects in *Correlated*. The estimates are obtained by regressing

$$\text{forecast}_{i,t} = \alpha_i + \phi_{1,i} A_{i,t} + \phi_{2,i} B_{i,t} + u_{i,t}$$

for each subject i separately and then inverting the estimated slope coefficients $\phi_{1,i}$ and $\phi_{2,i}$ by Eq. (5) to obtain the perceived correlation level. Each panel shows estimates for subjects from a particular experimental condition (specific true persistence ϕ and correlation ρ). Kernel density estimates with a Gaussian kernel and histograms plotted.

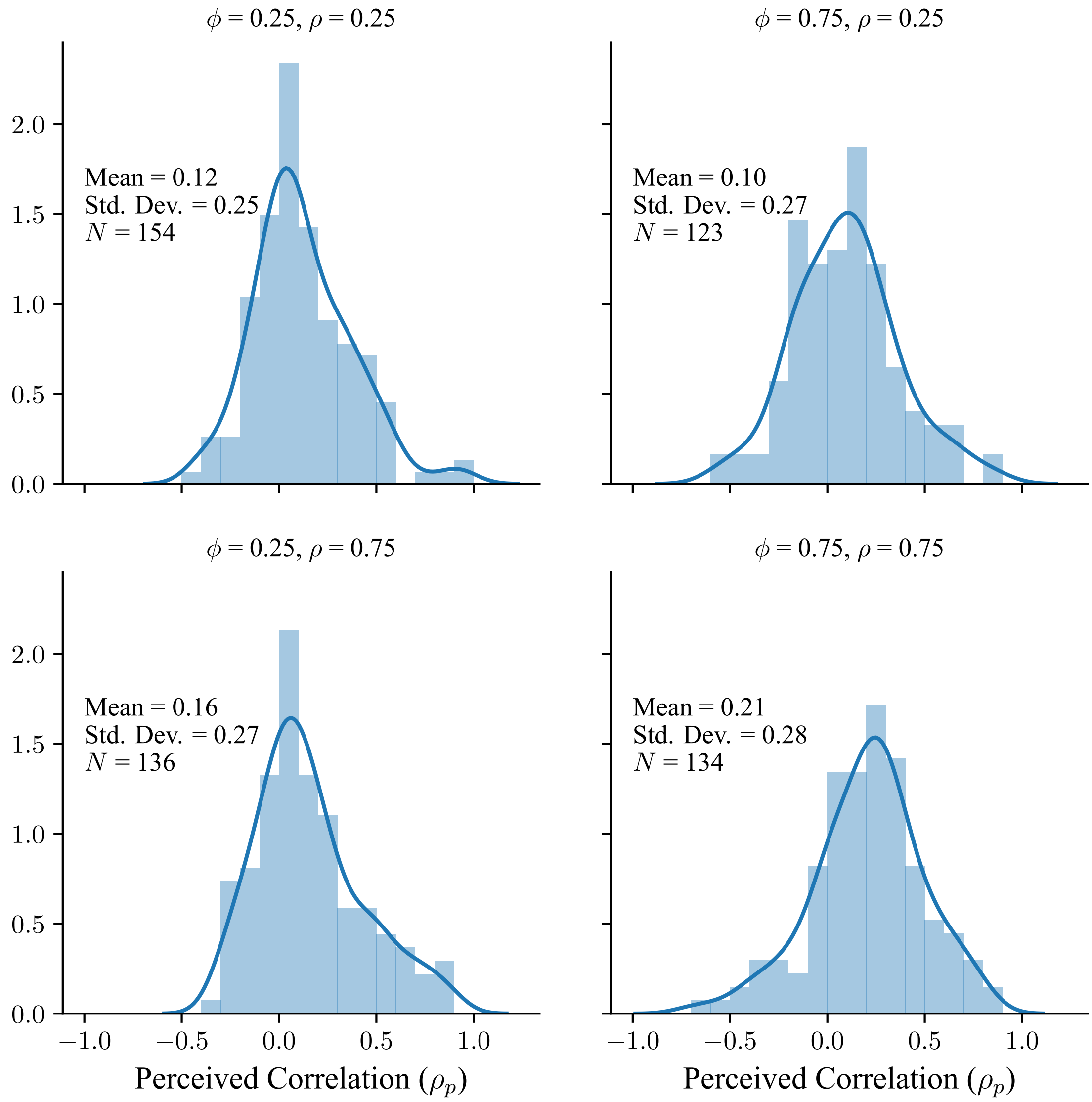
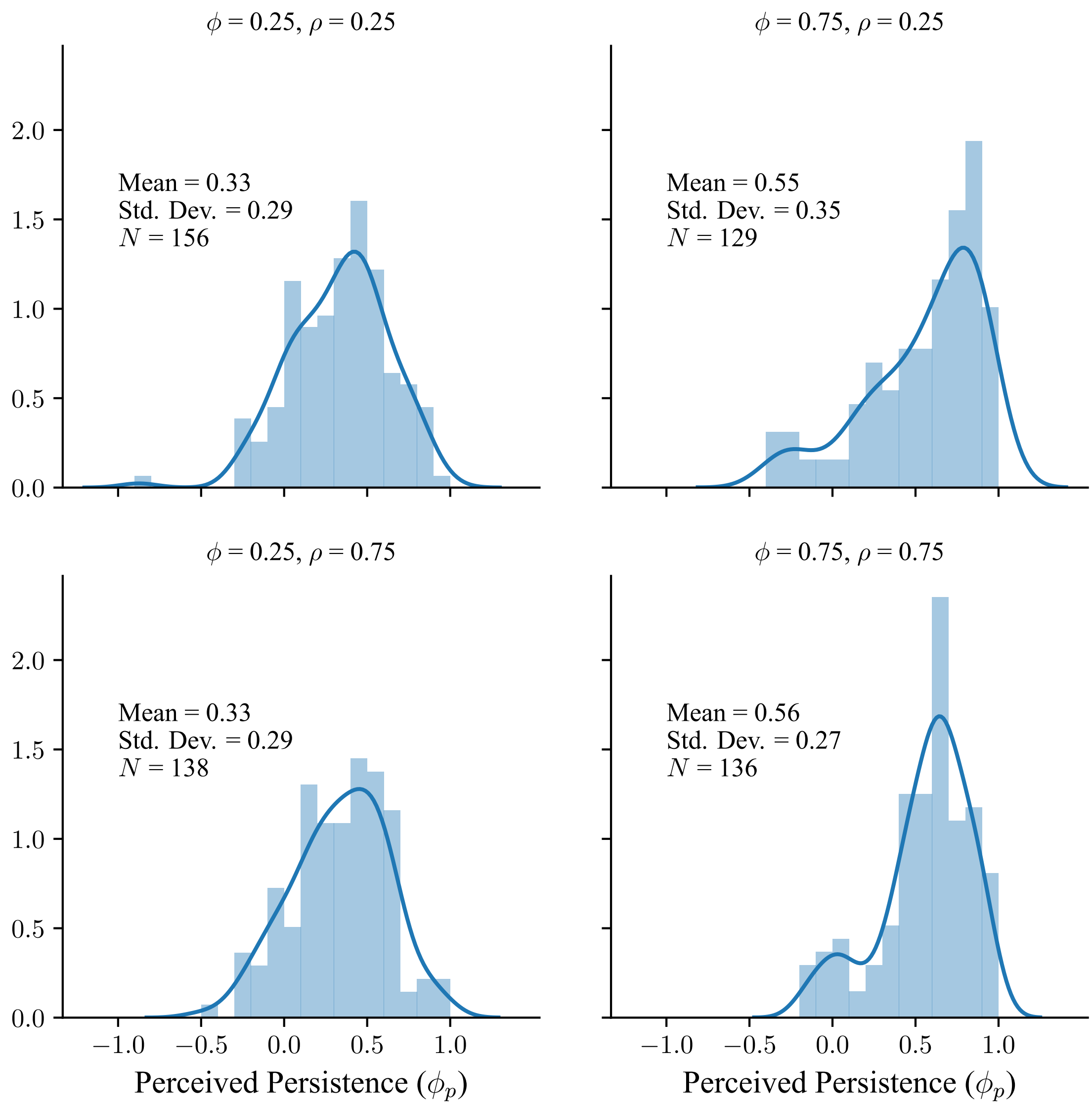


Figure 6

Perceived Persistence in *Correlated* Treatments

Notes: Subject-level estimates of perceived persistence; only for subjects in *Correlated*. Each panel shows estimates for subjects from a particular experimental condition (specific true persistence ϕ and correlation ρ). Kernel density estimates with a Gaussian kernel and histograms plotted. See Table 5 for details.



that most subjects struggle with correlated information.

Result 7. *Subjects' perceptions of persistence are fairly accurate and responsive to changes in persistence. However, subjects tend to underestimate correlation and are not sensitive to changes in correlation.*

Do individual characteristics affect correlation neglect? In our questionnaire, we elicit demographic information such as gender, whether subjects have taken a course in statistics, and educational attainment. Furthermore, we ask whether subjects find variable B useful in the *Correlated* treatment.[15] We now perform an OLS regression in which we regress the average score per round on a treatment dummy and four individual characteristics ("female", "taken statistical course", "having a university degree or above", "find B useful"), as well as the interaction terms of the treatment dummy with the four dummy variables of the individual characteristics. The coefficients of the interaction terms are displayed in Table A.9 in Appendix G. As seen in this table, none of these characteristics have a significant impact on correlation neglect.

Result 8. *The level of correlation neglect appears to be uncorrelated with demographic characteristics.*

4 Follow-Up Experiment

4.1 Motivation

Our main experimental design has two potential shortcomings. First, it is possible that subjects ignore B simply because they do not understand that B is important for predicting A. This leaves open the possibility that our key findings are driven by subject confusion or failure to pay attention. Second, since we do not provide subjects with an explicit formula for the underlying data-generating process, we cannot tightly control for prior beliefs.

In a follow-up experiment (not pre-registered), we conduct two additional treatments specifically designed to address these concerns. In the first treatment, we provide an intuitive explanation to help subjects understand the correlation between past and future values of A, and the correlation between A and B; in addition, we only allow subjects to proceed to the forecasting task after successfully passing a quiz designed to test their understanding. In the second treatment, we additionally inform subjects of the specific

[15] We also include an attention check question, which is passed by 99% of subjects in the final sample. Because there is no sufficient variation for this question, we do not include it in the regression.

formula used in the data-generating process. As we document below, even with this extra information, all of the earlier findings remain essentially unchanged.

4.2 Experimental Design and Procedures

To limit the number of new treatments, we set persistence and correlation to 0.75 throughout (i.e., $\phi = \rho = 0.75$). We begin by replicating two treatments from the main experiments (specifically, *Baseline*-2 and *Correlated*-4). We replicate these treatments to ensure that results from the follow-up experiment are not driven by changes in the subject pool. In addition, this replication serves to check the robustness of our original findings. To the extent that the replication yields similar results to our original experiment, we may be more confident that our key findings are reliable.

In the first new treatment (called *Info*), we include the following statement in the experimental instructions to improve the subjects' understanding of the correlation structure:

> *Here you can see how the prediction page will look like: [Insert A.1 for Baseline / Figure 1 for Correlated].*
>
> *Future values of "Variable A" are positively correlated with past values of both "Variable A" and "Variable B". That means that when "Variable A" or "Variable B" is larger in the current period, "Variable A" tends to be larger in the next period. When making predictions, you should therefore try to use information contained in past values of both "Variable A" and "Variable B".*[16]

In addition, we only allow subjects to proceed to the forecasting task upon successfully completing a quiz; the quiz is designed to test the subjects' understanding of the correlation structure in the experiment. The questions we used in the quiz can be found in the Appendix (Section A.3).

The second new treatment (called *Info-DGP*) builds on *Info* by additionally informing the subjects of the formula for the data-generating process (without specifying the exact parameters) Specifically, we add the following information to the experimental instructions:

> *"Variable A" in the next period = c + a * "Variable A" in current period + b * "Variable B" in current period" + "shock".*[17]

[16] We do not mention "Variable B" in the *Baseline* treatments.
[17] We do not mention "Variable B" in the *Baseline* treatments.

Table 8

Follow-Up Treatments: Summary

Notes: The table summarizes the follow-up treatments that we run. We vary the information structure (*Baseline* vs. *Correlated*), and the information provided, as shown in the table, with six treatment arms in total. In all treatments, we set $\mu = 100$, $\sigma_\varepsilon = 10$, and $\phi = 0.75$.

	Baseline ($\phi = 0.75$)	Correlated ($\phi = 0.75$, $\rho = 0.75$)
Replication	Baseline-Replication	Correlated-Replication
Info	Baseline-Info	Correlated-Info
Info & DGP	Baseline-Info-DGP	Correlated-Info-DGP

Finally, we provide a short explanation of the formula to help subjects better understand it. In total, we run six treatments using a between-subjects design, and the treatment names are presented in Table 8.

Following our main experiment, we used Amazon's Mechanical Turk (AMT) to recruit subjects and conduct our experiment. On June 23rd, 2022, we created a Human Intelligence Task (HIT) titled "Prediction task – forecast future values of given variable" with a recruiting quota of 750 subjects. We imposed the same recruiting requirements as the main experiment. To adjust for inflation, we increased both the base payment and the exchange rate of the bonus payment by 25%.[18]

4.3 Experimental Results

4.3.1 Data Filtering

In total, 747 subjects successfully completed the experiment, among whom 26 encountered technical problems.[19] After excluding these subjects, we are left with 721 subjects and 28,840 individual forecasts in the initial sample. We impose the same filters as in our main experiment, and we are left with a final sample of 608 subjects and 24,301 individual forecasts. Summary statistics for the final sample are shown in Table 9, and the sample selection process is summarized in Appendix G (Table A.10). The average bonus payment is \$2.42. Since the subjects also received a \$1.25 base payment for participating, the average earnings are \$3.67. On average, subjects finish the experiment in

[18] We changed the base payment from \$1 to \$1.25, and the conversion rate of the bonus payment from 500 points = \$1 to 400 points = \$1.

[19] Our quota on AMT was 750 subjects, but only 747 subjects successfully completed the experiment (3 subjects failed to complete the experiment, they had mismatched IDs or missing records). Among the 747 subjects, 26 completed the experiment in more than 60 minutes, which should not have been possible unless there were technical problems.

Table 9

Summary Statistics: Follow-Up Experiment

Notes: Summary statistics for the final sample used for the analysis. Forecast errors are defined as prediction minus the realization. Prediction times per round are shown for all rounds excluding the first round. Quiz trials equal the number of attempts made by subjects to pass the quiz in *Info* and *Info-DGP* treatments.

	N	Mean	Med.	Std. Dev.	Min.	Max.
Total Time (min)	608	15.42	11.03	12.12	3.03	59.10
Time / Round (sec)	23,693	10.12	6.00	35.81	0.00	2642.00
Score / Round	24,301	24.17	0.00	32.46	0.00	99.99
Forecast Error	24,301	0.91	2.11	23.05	-259.58	414.12
Abs. Forecast Error	24,301	15.12	10.85	17.42	0.00	414.12
Bonus ($)	608	2.42	2.43	0.83	0.55	4.72
Quiz Trials	412	2.30	2.00	1.76	1.00	15.00

around 15.4 minutes, spending around 10 seconds per each forecasting round. In the final sample, 95 subjects are in *Baseline-Replication*, 99 in *Correlated-Replication*, 107 in *Baseline-Info*, 104 in *Correlated-Info*, 104 in *Baseline-Info-DGP*, and 99 in *Correlated-Info-DGP*. To check our randomization and test for differential dropout rates, we perform balance tests between the subjects in *Baseline* and *Correlated* treatments.[20] We find no significant differences in their dropout rates and other characteristics, except for the attention check.[21]

4.3.2 Forecast Accuracy

As seen in Figure 7, subjects in the *Replication* treatments earn a significantly lower score than in the main experiment, in both *Baseline* (MW test, $p = 0.018$) and *Correlated* (MW test, $p = 0.005$). This suggests that data quality on AMT may have declined over the past few years.[22] Nevertheless, we replicate our most critical finding that subjects earn a substantially lower score in *Correlated-Replication* than in *Baseline-Replication* (MW test, $p < 0.001$). In both our new treatments, subjects earn a lower score when Variable B is present: Subjects obtain a lower a score in *Correlated-Info* than in *Baseline-Info* (MW test, $p = 0.045$), and a lower score in *Correlated-Info-DGP*

[20] Detailed results of the balance tests can be found in Appendix G (Table A.11).

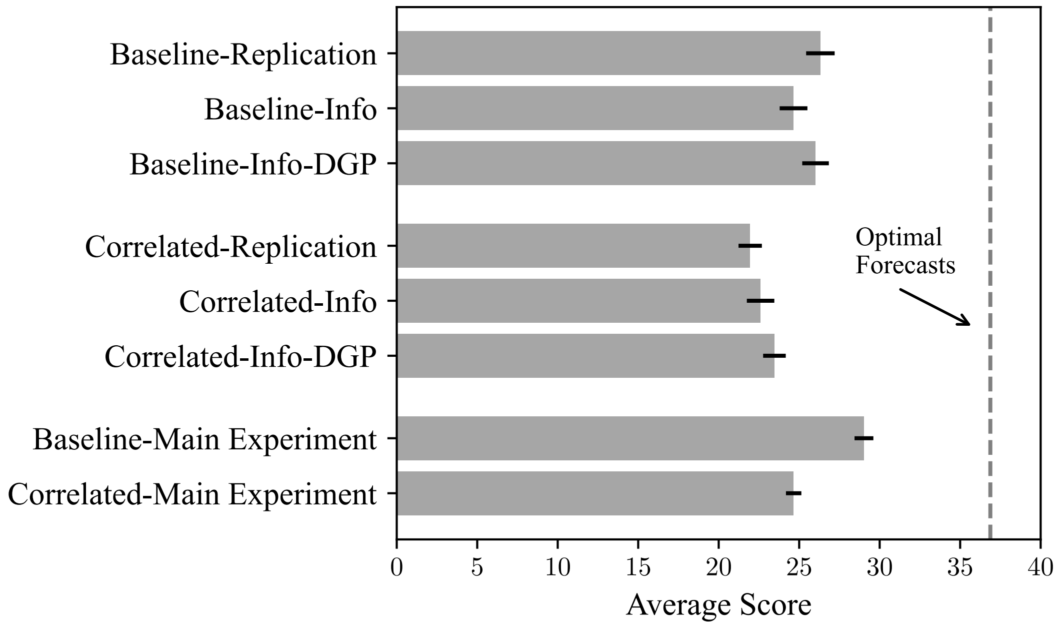
[21] Subjects in the *Correlated* treatments pass the attention check significantly better than subjects in the *Baseline* treatments, suggesting that subjects in the *Correlated* pay more attention in the experiment. This difference can only work against our finding that subjects in the *Correlated* treatments earn a lower accuracy score than subjects in the *Baseline* treatments.

[22] This interpretation is consistent with a higher fraction of excluded sample under the same filtering criterion, as can be seen in Table A.2 and Table A.10 in Appendix G.

Figure 7

Average Scores by Treatment: Follow-Up Experiment

Notes: Average scores by treatment. Plus/minus one standard error bars shown, with standard errors robust to heteroskedasticity. The average score obtained by the optimal forecasting rule (full-information rational expectations) is given by the dashed vertical line. *Baseline-Main Experiment* and *Correlated-Main Experiment* give the average scores of treatments from Section 3 with the same persistence and correlation parameters for comparison purposes (i.e., *Baseline*-2 and *Correlated*-4).



than in *Baseline-Info-DGP* (MW test, $p = 0.016$).

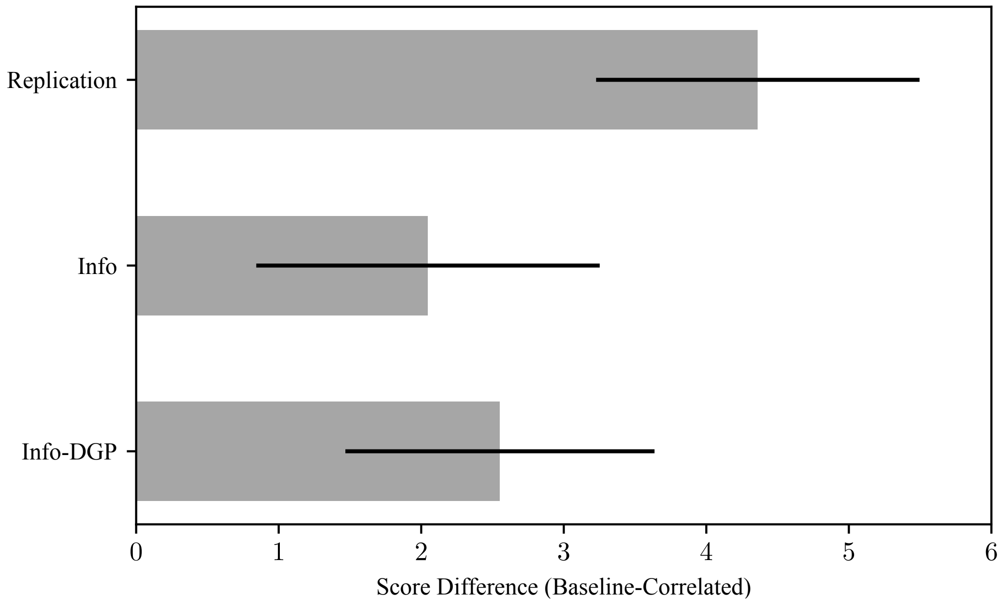
Result 9. *The negative effect of correlated information persists when additional information on the correlation structure is provided, with or without the formula for the data-generating process.*

We then investigate the effect of providing additional information on persistence and correlation. As seen in Figure 7, subjects in *Info* do not earn a significantly different score compared to *Replication*, both in *Baseline* (MW test, $p = 0.220$) and *Correlated* (MW test, $p = 0.933$). This result suggests that the additional information on the correlation structure fails to improve forecasting accuracy.

Next, we ask whether providing subjects with the formula for the data-generating process (in combination with the extra information in *Info*) has an effect. As seen in Figure 7, the average scores are not significantly different between *Baseline-Info-DGP* and *Baseline-Replication* (MW test, $p = 0.857$), nor between *Correlated-Info-DGP* and *Correlated-Replication* (MW, $p = 0.216$). This result indicates that providing subjects

Figure 8
Difference in Difference: Follow-Up Experiment

Notes: Differences of average scores (*Baseline* − *Correlated*) in each treatment condition. Plus/minus one standard error bars shown, with standard errors robust to heteroskedasticity.



with the formula for the data-generating process fails to improve their forecast accuracy.

Result 10. *Providing information on persistence and correlation, whether in combination with the data-generating process formula or not, does not lead to an improvement in forecast accuracy.*

Finally, we study whether the additional information has an impact on reducing the accuracy gap between *Baseline* and *Correlated* (i.e., whether the treatment effect of correlated information is affected by the new treatments). Figure 8 shows the differences of average scores between *Baseline* and *Correlated* under each condition. The score difference is 4.36 in *Replication*, 2.04 in *Info*, and 2.55 in *Info-DGP*. Although the difference in *Replication* is more than twice as large as that in *Info*, the differences are not significantly different (t test, $p = 0.154$). Similarly, no differences between any other two conditions are statistically significant (*Replication* vs. *Info-DGP*, t test, $p = 0.270$; *Info* vs. *Info-DGP*, t test, $p = 0.753$).

Result 11. *Providing information on persistence and correlation, whether in combination with the data-generating process formnula or not, does not reduce the accuracy gap caused by correlated information.*

Table 10

Overreaction in Each Treatment: Follow-Up Experiment

Notes: Estimated levels of overreaction in *Baseline* and *Correlated*, under each treatment condition. Standard errors clustered by subject in parentheses. *Main Experiment* gives the estimates for treatments from Section 3 with the same persistence and correlation parameters for comparison purposes (i.e., *Baseline*-2 and *Correlated*-4). See Table 3 for further explanations.

	Baseline	Correlated	B−C
Replication	-0.19 (0.06)	-0.36 (0.06)	0.17 (0.08)
Info	-0.23 (0.07)	-0.33 (0.06)	0.10 (0.09)
Info-DGP	-0.13 (0.06)	-0.36 (0.05)	0.24 (0.08)
Main Experiment	0.04 (0.04)	-0.26 (0.03)	0.30 (0.05)

4.3.3 Under- and Overreaction

We finally test whether subjects underreact to new information more in the *Correlated* treatments. Following the approaches in section 3.3, we test whether subjects exhibit more overall underreaction in *Correlated*. As shown in Figure 9, compared to *Baseline*, the response to new information in *Correlated* is more muted, and more compressed towards the simple default of predicting the unconditional mean of 100. This figure, therefore, replicates the results in Figure 3.

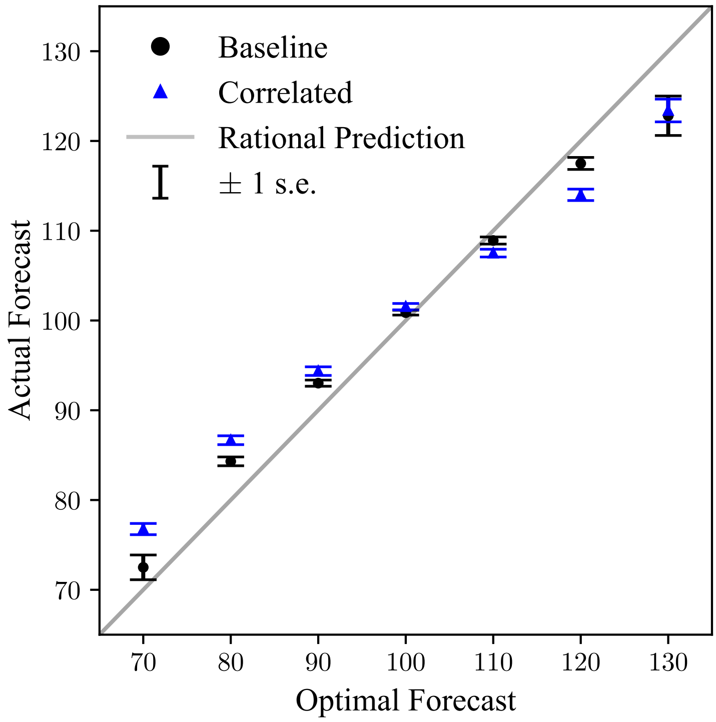
Table 10 summarizes the estimated overreaction levels for each treatment of the follow-up experiment, as well as estimates for the same persistence and correlation levels from the main experiment. We observe that in *Baseline*, there is underreaction in the three new treatments, whereas we observed a very modest overreaction in the main experiment. In *Correlated*, underreaction in the three new treatments is more severe compared to the original results from the main experiment. The difference in overreaction level (*B-C*) is always positive and significant in all four conditions. These results replicate our previous finding that subjects underreact more in *Correlated* than in *Baseline*.

Result 12. *When provided with the information on persistence and correlation, with or without the formula for the data-generating process, subjects always underreact more in Correlated than in Baseline.*

Figure 9

Response to New Information: Follow-Up Experiment

Notes: Response to new information in *Baseline* and *Correlated* treatments. Binned scatterplot of actual vs. optimal forecasts. Plus/minus one standard error bars shown, with standard errors robust to heteroskedasticity. The bins used in the graph are $[65, 75), [75, 95), \ldots, [125, 135]$.



4.4 Mechanism Questions

The results of the follow-up experiment allow us to better understand the underlying behavioral mechanisms for why subjects in *Correlated* tend to underreact to the correlated variable.

First, providing the formula for the data generating process allows us to have a better control for prior beliefs. In a placebo test (see Table A.13 in Appendix G), we find that least-squares learning—a natural approximation to Bayesian updating—leads to forecasts that are very close to full-information rational expectations in *Info-DGP*. This result suggests that Bayesian subjects should not have a lower score in *Correlated*. Therefore, a lower score that we observe in *Correlated* strongly suggests that subjects are not Bayesian.

Second, by providing additional information on the correlation structure as well as the data-generating process, we are able to rule out the possibility that subjects hold an incorrect prior that the correlated variable is unimportant. It appears that subjects understand that the correlated variable is potentially useful, but they find it too cognitively demanding to incorporate it into their forecasts. Therefore, subjects behave as if they largely ignore the correlated variable. Mathematically, persistence is just a special type of correlation. However, people seem to view persistence and correlation very differently. In particular, making use of information contained in the correlated variable appears to be much more cognitively demanding.

5 Conclusions and Limitations

Correlated information structures are ubiquitous in practice. However, little is known about how people form expectations when faced with correlated variables. In this paper, we studied expectation formation with correlated variables in a large-scale preregistered experiment.

We found a strong effect of correlated information, and the effect persisted even when subjects were provided with increasingly detailed intuitive explanations of the data-generating process. Even though actual predictability was kept constant, subjects performed substantially worse when faced with multiple correlated variables. Subjects also exhibited underreaction to new information, even though overreaction was observed in the baseline treatment. While subjects' perceptions of persistence were fairly sensitive to actual persistence, correlation perceptions were significantly less sensitive. This result suggests that people find estimating correlations more cognitively demanding (Gabaix, 2019; Enke and Graeber, 2019; Woodford, 2020).

Overall, we provide evidence that when forming expectations, people have limited ability to incorporate information contained in correlated variables. This finding may have important policy implications. Rather than simply providing more information, our results suggest that it may be more effective to focus on helping individuals better use the available information. Consistent with this idea, recent research on central bank communication suggests that simpler messages have a stronger impact on expectations (Haldane and McMahon, 2018; Bholat, Broughton, Meer and Walczak, 2019; Kryvtsov and Petersen, 2019).

Due to correlation neglect, expectations can exhibit significant context dependence. In other words, when correlated variables provide a lot of useful information, we may see more underreaction. Correlation neglect can therefore help explain why both underreaction and overreaction coexist in the field, in addition to the mechanisms already proposed in the literature (Enke et al., 2020; Bordalo et al., 2020; Kohlhas and Walther, 2021; Bordalo et al., 2023). Future work using field data may help answer whether different empirical correlation structures indeed lead to differences in under- vs overreaction, and whether correlation neglect is consistent with the amount of under- and overreaction in the data.

We conclude by listing the *key limitations* of the present study:

- Our results may be specific to the subject pool used in the study (MTurk workers);
- We do not have a strong theory as to why subjects find incorporating information in B more cognitively demanding;
- Subjects may have not fully recognized the importance of variable B or may have disregarded it during the experiment. While our follow-up treatments (*Info* and *Info-DGP*) attempt to address this issue, we cannot completely eliminate this possibility as a factor in our results.

Addressing these limitations may be an interesting research avenue for future work.

References

Afrouzi, H., Kwon, S.Y., Landier, A., Ma, Y., Thesmar, D., forthcoming. Overreaction in expectations: Evidence and theory. The Quarterly Journal of Economics.

Ahler, D.J., Roush, C.E., Sood, G., 2019. The micro-task market for lemons: Data quality on Amazon's Mechanical Turk. Working paper.

Ambuehl, S., Li, S., 2018. Belief updating and the demand for information. Games and Economic Behavior 109, 21–39.

Angeletos, G.M., Huo, Z., 2021. Myopia and anchoring. American Economic Review 111, 1166–1200.

Assenza, T., Bao, T., Hommes, C., Massaro, D., 2014. Experiments on expectations in macroeconomics and finance, in: Duffy, J. (Ed.), Experiments in Macroeconomics (Research in Experimental Economics). Emerald Group Publishing Limited, Bingley. volume 17, pp. 11–70.

Bao, T., Hommes, C., Sonnemans, J., Tuinstra, J., 2012. Individual expectations, limited rationality and aggregate outcomes. Journal of Economic Dynamics and Control 36, 1101–1120.

Benjamin, D.J., 2019. Errors in probabilistic reasoning and judgment biases, Elsevier, Amsterdam. volume 2. chapter 2, pp. 69–186.

Bholat, D., Broughton, N., Meer, J.T., Walczak, E., 2019. Enhancing central bank communications using simple and relatable information. Journal of Monetary Economics 108, 1–15.

Bordalo, P., Conlon, J.J., Gennaioli, N., Kwon, S.Y., Shleifer, A., 2023. Memory and probability. The Quarterly Journal of Economics 138, 265–311.

Bordalo, P., Gennaioli, N., Ma, Y., Shleifer, A., 2020. Overreaction in macroeconomic expectations. American Economic Review 110, 2748–82.

Bordalo, P., Gennaioli, N., Shleifer, A., 2018. Diagnostic expectations and credit cycles. The Journal of Finance 73, 199–227.

Cavallo, A., Cruces, G., Perez-Truglia, R., 2017. Inflation expectations, learning and supermarket prices: Evidence from survey experiments. American Economic Journal: Macroeconomics 9, 1–35.

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

Coibion, O., Gorodnichenko, Y., Kamdar, R., 2018. The formation of expectations, inflation and the Phillips Curve. Journal of Economic Literature 56, 1447–1491.

D'Acunto, F., Hoang, D., Paloviita, M., Weber, M., 2019a. Cognitive abilities and inflation expectations. American Economic Review: Papers & Proceedings 109, 562–566.

D'Acunto, F.D., Hoang, D., Paloviita, M., Weber, M., 2019b. IQ, expectations, and choice. Working paper.

De Bondt, W.F.M., Thaler, R.H., 1990. Do security analysts overreact? The American Economic Review: Papers & Proceedings 80, 52–57.

DellaVigna, S., Pope, D., 2018. What motivates effort? Evidence and expert forecasts. Review of Economic Studies 85, 1029–1069.

Dwyer, G.P., Williams, A.W., Battalio, R.C., Mason, T.I., 1993. Tests of rational expectations in a stark setting. The Economic Journal 103, 586–601.

Enke, B., Graeber, T., 2019. Cognitive uncertainty. Working paper.

Enke, B., Schwerter, F., Zimmermann, F., 2020. Associative memory and belief formation. Working paper.

Enke, B., Zimmermann, F., 2019. Correlation neglect in belief formation. Review of Economic Studies 86, 313–332.

Eyster, E., Rabin, M., Weizsacker, G., 2015. An experiment on social mislearning. Available at SSRN 2704746 .

Eyster, E., Weizsacker, G., 2016. Correlation neglect in portfolio choice: Lab evidence. Available at SSRN 2914526 .

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

Gabaix, X., 2014. A sparsity-based model of bounded rationality. The Quarterly Journal of Economics , 1661–1710.

Gabaix, X., 2017. Behavioral macroeconomics via sparse dynamic programming. Working paper.

Gabaix, X., 2019. Behavioral inattention, Elsevier, Amsterdam. volume 2. chapter 4, pp. 261–343.

Gennaioli, N., Shleifer, A., 2018. A Crisis of Beliefs: Investor Psychology and Financial Fragility. Princeton University Press, Princeton.

Graeber, T., 2018. Inattentive inference. Working paper.

Haldane, A., McMahon, M., 2018. Central bank communications and the general public. American Economic Review: Papers & Proceedings 108, 578–83.

Hamilton, J.D., 1994. Time Series Analysis. Princeton University Press, Princeton.

Heemeijer, P., Hommes, C., Sonnemans, J., Tuinstra, J., 2009. Price stability and volatility in markets with positive and negative expectations feedback: An experimental investigation. Journal of Economic Dynamics & Control 33, 1052–1072.

Hey, J.D., 1994. Expectations formation: Rational or adaptive or ...? Journal of Economic Behavior and Organization 25, 329–349.

Hommes, C., 2011. The heterogeneous expectations hypothesis: Some evidence from the lab. Journal of Economic Dynamics and Control 35, 1–24.

Hommes, C., Zhu, M., 2014. Behavioral learning equilibria. Journal of Economic Theory 150, 778–814.

Horton, J.J., Rand, D.G., Zeckhauser, R.J., 2011. The online laboratory: Conducting experiments in a real labor market. Experimental Economics 14, 399–425.

Hossain, T., Okui, R., 2020. Belief formation under signal correlation. Available at SSRN 3218152 .

Ilut, C., Valchev, R., 2023. Economic agents as imperfect problem solvers. The Quarterly Journal of Economics 138, 313–362.

Kalaycı, K., Serra-Garcia, M., 2016. Complexity and biases. Experimental Economics 19, 31–50.

Kohlhas, A., Walther, A., 2021. Asymmetric attention. American Economic Review, forthcoming.

Kryvtsov, O., Petersen, L., 2019. Central bank communication that works: Lessons from lab experiments. Working paper.

Kucinskas, S., Peters, F.S., forthcoming. Measuring under- and overreaction in expectation formation. The Review of Economics and Statistics.

Kuziemko, I., Norton, M.I., Saez, E., Stantcheva, S., 2015. How elastic are preferences for redistribution? Evidence from randomized survey experiments. American Economic Review 105, 1478–1508.

Lian, C., 2021. A theory of narrow thinking. The Review of Economic Studies 88, 2344–2374.

Lütkepohl, H., 2005. New Introduction to Multiple Time Series Analysis. volume 39. Springer-Verlag, Berlin.

Mankiw, G.N., Reis, R., 2002. Sticky information versus sticky prices: A proposal to replace the New Keynesian Phillips Curve. The Quarterly Journal of Economics 117, 1295–1328.

Manski, C.F., 2018. Survey measurement of probabilistic macroeconomic expectations: Progress and promise. NBER Macroeconomics Annual 32, 411–471.

Matthies, B., 2018. Biases in the perception of covariance. Working paper.

Mincer, J.A., Zarnowitz, V., 1969. The evaluation of economic forecasts, in: Mincer, J.A. (Ed.), Economic Forecasts and Expectations: Analysis of Forecasting Behavior and Performance. National Bureau of Economic Research. chapter 1, pp. 3–46.

Muth, J.F., 1961. Rational expectations and the theory of price movements. Econometrica 29, 315–335.

Rees-Jones, A., Shorrer, R., Tergiman, C.J., 2020. Correlation neglect in student-to-school matching, in: Proceedings of the 21st ACM Conference on Economics and Computation, pp. 467–468.

Rubinstein, A., 2007. Instinctive and cognitive reasoning: A study of response times. Economic Journal 117, 1243–1259.

Schlag, K.H., Tremewan, J., van der Weele, J.J., 2015. A penny for your thoughts: A survey of methods for eliciting beliefs. Experimental Economics 18, 457–490.

Umar, T., 2019. Complexity aversion when seeking alpha. Working paper.

Woodford, M., 2003. Imperfect common knowledge and the effects of monetary policy, in: Aghion, P., Frydman, R., Stiglitz, J., Woodford, M. (Eds.), Knowledge, Information, and Expectations in Modern Macroeconomics: In Honor of Edmund S. Phelps. Princeton University Press, Princeton, pp. 25–58.

Woodford, M., 2020. Modeling imprecision in perception, valuation and choice. Annual Review of Economics 12, forthcoming.

You, H., Zhang, X.j., 2009. Financial reporting complexity and investor underreaction to 10-K information. Review of Accounting Studies 14, 559–586.

Zellner, A., Palm, F., 1974. Time series analysis and simultaneous equation econometric models. Journal of Econometrics 2, 17–54.

Appendix A Experimental Instructions

A.1 Preview

Welcome to this experiment on forecasting!

The experiment is about predicting future values of a given variable. For participating in the experiment, you will receive a **base payment** of **$1.00**. You can also receive an additional **bonus payment**, and the amount of the bonus payment depends on your performance. We estimate that the average bonus payment will be around **$2.00**. The task will take around **12 minutes** to complete. To receive the payment (both base and bonus), you are required to finish the full experiment.

A.2 Instructions

Welcome to this experiment! Please read the following instructions carefully.

For participating in the experiment, you will receive a **base payment** of **$1.00**. You can also receive an additional bonus payment, and the amount of the **bonus payment** depends on your performance. We estimate that the average bonus payment will be around **$2.00**. The task will take around **12 minutes** to complete. To receive the payment (both base and bonus), you are required to finish the full experiment.

TASK DESCRIPTION

In this experiment, you are going to **predict** future values of **"Variable A"**.

(Only in all the *Baseline* treatments.) The experiment lasts for 40 rounds. In each round, you will see past values of "Variable A". Past values of "Variable A" are related to future values of "Variable A", and the relationship is stable. At the beginning of the experiment, you will see data from 40 previous rounds. Then, you will have to make predictions for 40 rounds.

(Only in the *Baseline-Info* and *Baseline-Info-DGP* treatments.) Here you can see how the prediction page will look like: (Figure A.1 here).

Past and future values of "Variable A" are positively correlated. That means that when "Variable A" is larger in the current period, "Variable A" tends to be larger in the next period. When making predictions, you should therefore try to use information contained in past values of "Variable A".

(Only in the *Baseline-Info-DGP* treatment.) We use the following formula to generate values of “Variable A”:

“Variable A” in the next period = c + a * (“Variable A” in current period) + shock.

In the formula above, a and c are both fixed positive numbers; you are not informed about the exact values of these numbers. The “shock” is an unpredictable random number which is independent across periods and zero on average.

(Only in all the *Correlated* treatments.) The experiment lasts for 40 rounds. In each round, you will see past values of “Variable A” and “Variable B”. Past values of “Variable A” and “Variable B” are related to future values of “Variable A”, and the relationship is stable. At the beginning of the experiment, you will see data from 40 previous rounds. Then, you will have to make predictions for 40 rounds.

(Only in the *Correlated-Info* and *Correlated-Info-DGP* treatments.) Here you can see how the prediction page will look like: (Figure 1 here).

Future values of “Variable A” are positively correlated with past values of both “Variable A” and “Variable B”. That means that when “Variable A” or “Variable B” is larger in the current period, “Variable A” tends to be larger in the next period. When making predictions, you should therefore try to use information contained in past values of both “Variable A” and “Variable B”.

(Only in the *Correlated-Info-DGP* treatment.) We use the following formula to generate values of “Variable A”:

“Variable A” in the next period = c+ a * (“Variable A” in current period) + b * (“Variable B” in current period”) + “shock”.

In the formula above, a, b and c are all fixed positive numbers; you are not informed about the exact values of these numbers. The “shock” is an unpredictable random number which is independent across periods and zero on average.

BONUS PAYMENT

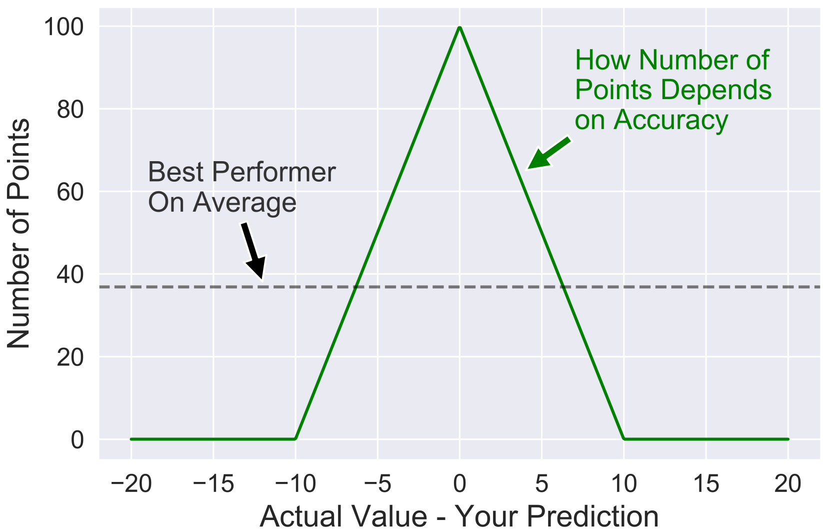
You receive points for accurate predictions. At the end of the experiment, your points will be converted to dollars at the rate of **500 points = $1**. Your total earnings in dollars will therefore be **$1.00 + your total points/500**.

How many points you receive depends on the accuracy of your prediction:

- The closer your prediction is to the actual value, the more points you receive;
- If your prediction is more than 10 units away from the actual value, you receive no points.

We estimate that the best performer will on average receive 37 points per round.

Graphically, the number of points you receive depends on accuracy as follows:



The exact formula for the number of points is $100 * \max\{0, 1 - |D|/10\}$ where D is the difference between the actual value and your prediction.

A.3 Quiz

Now you are going to answer a few questions to test your understanding of the experiment. After you fill all your answers, you can click "Confirm". You will be notified if any of your answers is wrong.

Note: You will not be able to participate in the experiment until you answer all questions correctly. If you quit at this stage, you will not receive any payment.

Please select one correct answer for each of the following question.

(Only in the *Baseline-Info*, *Baseline-Info-DGP*, *Correlated-Info*, and *Correlated-Info-DGP* treatments.)

1. To obtain a high score in this experiment, you should:
   a. Make predictions as quickly as possible. b. Make predictions that are as accurate as possible.
2. Choose the correct statement:
   a. Future values of "Variable A" are completely random.
   b. Future values of "Variable A" are positively correlated with past values of "Variable A".
   c. Future values of "Variable A" are negatively correlated with past values of "Variable A".
3. If the current value of "Variable A" is high, then:
   a. The value of "Variable A" in the next period will always be high.
   b. The value of "Variable A" in the next period tends to be high but could sometimes be low.
   c. Nothing can be said about the value of "Variable A" in the next period.

   (Only in the *Correlated-Info* and *Correlated-Info-DGP* treatments.)
4. Choose the correct statement:
   a. Future values of "Variable A" are completely random.
   b. Future values of "Variable A" are positively correlated with past values of "Variable B".
   c. Future values of "Variable A" are negatively correlated with past values of "Variable B".
5. If the current value of "Variable B" is high, then:
   a. The value of "Variable A" in the next period will always be high.
   b. The value of "Variable A" in the next period tends to be high but could sometimes be low.
   c. Nothing can be said about the value of "Variable A" in the next period.

6. In order to make accurate predictions, you should:
   a. Only consider past values of “Variable A”.
   b. Only consider past values of “Variable B”.
   c. Consider past values of both “Variable A” and Variable B”.
   d. Make random predictions.

A.4 Questionnaire

Questions marked with an asterisk are compulsory. The possible answers we provided in closed-ended questions are given in the parentheses.

1. *Age
2. *Gender (Male, female)
3. *Have you ever taken a class on statistics or forecasting? (Yes, no)
4. *What is the highest level of educational degree that you hold? (Below high school, high school, college, graduate school, other)
5. *I am someone who finds it easy to concentrate and can work on tasks for a long time. (Completely agree, agree, neutral, disagree, strongly disagree)
6. *I have a good eye for detail and often notice things that others miss. (Completely agree, agree, neutral, disagree, strongly disagree)
7. *Please select “disagree” among the following options to show that you are paying attention. (Completely agree, agree, neutral, disagree, strongly disagree)
8. *In the experiment, I found past values of “Variable A” to be useful when making predictions. (Completely agree, agree, neutral, disagree, strongly disagree)
9. *(Only in all the *Correlated* treatments) In the experiment, I found past values of “Variable B” to be useful when making predictions. (Completely agree, agree, neutral, disagree, strongly disagree)
10. Do you have any additional comments about the experiment?

Figure A.1

Screenshot of Prediction Page: *Baseline* Treatment

Notes: Screenshot of the experimental screen in *Baseline*. Subjects submit their scores by either clicking on the blue area or typing their prediction in the text box.

Appendix B Incentive Payments

In this appendix, we provide conditions under which the incentive scheme that we use in the experiment induces the subjects to report their subjective expected value.

The incentive structure used in the experiment is

$$S = \max\{0, S_0 - \beta|e|\},$$

where S_0 and β are strictly positive parameters, $e = A - F$ is the forecast error, A denotes the actual realized value, and F is the forecast. If $\beta = S_0/(k\sigma)$, then the subjects receive 0 whenever their forecasts are more than k standard deviations away from the realized value.

Assuming that the subjects are *risk neutral*, the optimal forecast F_i^* of subject i is chosen as

$$F_i^* = \arg\max_F \mathbb{E}_i[S] = \arg\min_F \mathbb{E}_i[\min\{K, |e|\}] \text{ where } K \equiv S_0/\beta. \tag{6}$$

Here, the expectation $\mathbb{E}_i[\cdot]$ is taken with respect to the *subjective* distribution of A, with its cdf given by $G_i(A)$. In light of Eq. (6), F_i^* can be thought of as arising from minimizing a loss function that depends on the absolute forecast error, $|e|$. Differently from the standard absolute error loss function, however, this loss function is *clipped* at K.

The following proposition demonstrates that when the subjective distribution function is unimodal and symmetric, the optimal forecast is given by the subjective expected value. The result implies, in particular, that if the subjects in the experiment correctly perceive variables to be normally distributed, they should optimally report their expected value.

Proposition 2. *Suppose that the subjective distribution function of a risk-neutral subject i is (i) symmetric around point μ_i; (ii) unimodal; and (iii) has a differentiable density function. Then, the optimal forecast is*

$$F_i^* = \mu_i = \mathbb{E}_i[A] = \textit{median}_i(A) = \textit{mode}_i(A).$$

Proof. In Appendix E. □

We emphasize the two key conditions (which are both well known) that are necessary for the result to obtain. First, the subjects must be risk neutral. If they are not, their forecasts will no longer be equal to the subjective expected value (for more discussion, see Schlag, Tremewan and van der Weele, 2015). Second, the subjective distribution function must be symmetric and unimodal. If the distribution function were not symmetric, then even without any clipping in the loss function, the median and the expected value would differ from each other, and the scoring rule would induce the subjects to reveal their subjective median. With clipping, they would reveal something approximating their subjective median. If the subjects correctly perceive the variables to be normally distributed in our experiment, then the second condition is satisfied. While we have not been able to obtain a formal proof, we conjecture that the result remains valid under risk aversion if all variables are perceived to be jointly normally distributed (as they are in our experiment).

Appendix C Construction of Bivariate Process

In this appendix we provide a detailed discussion of how the bivariate process used in the *Correlated* treatment (Eqs. (2) and (3)) is constructed.

Consider a general bivariate VAR(1) process

$$\underbrace{\begin{pmatrix} x_{1t} \\ x_{2t} \end{pmatrix}}_{\equiv \mathbf{x}_t} = \underbrace{\begin{pmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{pmatrix}}_{\equiv \mathbf{A}} \begin{pmatrix} x_{1,t-1} \\ x_{2,t-2} \end{pmatrix} + \underbrace{\begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}}_{\equiv \boldsymbol{\varepsilon}_t} \text{ with } \boldsymbol{\Sigma} \equiv \mathbb{V}\text{ar}(\boldsymbol{\varepsilon}_t) = \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}$$

where $\sigma_1, \sigma_2 > 0$. We can rewrite this equation as

$$\underbrace{\begin{pmatrix} 1-\phi_{11}L & -\phi_{12}L \\ -\phi_{21}L & 1-\phi_{22}L \end{pmatrix}}_{\equiv \mathbf{A}(L)} \begin{pmatrix} x_{1t} \\ x_{2t} \end{pmatrix} = \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} \Leftrightarrow \mathbf{A}(L)\mathbf{x}_t = \boldsymbol{\varepsilon}_t.$$

Let $\mathbf{A}^*(L)$ denote the adjugate matrix of $\mathbf{A}(L)$ so that $\mathbf{A}^*(L)\mathbf{A}(L) = \mathbf{A}(L)\mathbf{A}^*(L) = |\mathbf{A}(L)|\mathbf{I}$. Pre-multiplying the above equation by $\mathbf{A}^*(L)$, we obtain that

$$|\mathbf{A}(L)|\mathbf{x}_t = \mathbf{A}^*(L)\boldsymbol{\varepsilon}_t.$$

In the (2×2) case, explicit expressions for $\mathbf{A}^*(L)$ and $|\mathbf{A}(L)|$ are

$$|\mathbf{A}(L)| = (1-\phi_{11}L)(1-\phi_{22}L) + \phi_{12}\phi_{21}L^2$$
$$\mathbf{A}^*(L) = \begin{pmatrix} 1-\phi_{22}L & \phi_{12}L \\ \phi_{21}L & 1-\phi_{11}L \end{pmatrix}$$

Hence, x_{1t} has the univariate representation as

$$[(1-\phi_{11}L)(1-\phi_{22}L) + \phi_{12}\phi_{21}L^2]x_{1t} = (1-\phi_{22}L)\varepsilon_{1t} + \phi_{12}\varepsilon_{2,t-1}. \tag{7}$$

The right-hand side is an MA(1) process, by virtue of being the sum of an MA(1) process and independent noise. Hence, x_{1t} follows an ARMA(2,1) process (Zellner and Palm, 1974; Hamilton, 1994, p. 349).

We wish to impose restrictions on the parameters of the VAR(1) process so that (i) x_{1t} has a univariate AR(1) representation; (ii) x_{1t} is predictable by past values of x_{2t}, i.e., $\phi_{12} \neq 0$; (iii) both x_{1t} and x_{2t} are serially correlated; (iv) x_{1t} and x_{2t} have the same univariate autocovariance function (symmetry). It turns out that these conditions impose very strong restrictions on the coefficient matrix $\mathbf{A}$.

First, consider the right-hand side (RHS) of Eq. (7). The invertible representation of the RHS is given by $(1-\theta L)v_t$ where

$$\theta = \frac{[(1+\phi_{22}^2)+(\phi_{12}^2\sigma_2^2/\sigma_1^2)] - \sqrt{[(1+\phi_{22}^2)+(\phi_{12}^2\sigma_2^2/\sigma_1^2)]^2 - 4\phi_{22}^2}}{2\phi_{22}},$$

see, e.g., Hamilton (1994, pp. 102–103). To get an AR(1) representation for x_{1t}, we need to get rid of any MA term. We first check for what parameter values $\theta = 0$. Clearly, if $\phi_{22} = 0$, then $\theta = 0$. On the other hand, suppose that $\theta = 0$ but, for contradiction, that $\phi_{22} \neq 0$. Then, we see from the expression above that $\phi_{22} = 0$, a contradiction. Hence, $\theta = 0 \Leftrightarrow \phi_{22} = 0$. But if $\phi_{22} = 0$, for the univariate representation of x_{1t} to be an AR(1) and not an AR(2), we need either $\phi_{12} = 0$ or $\phi_{21} = 0$. If $\phi_{12} = 0$, x_{1t} is not predictable from past values of x_{2t}; if $\phi_{21} = 0$, then x_{2t} is white noise and serially uncorrelated. Hence, we discard this case and impose $\phi_{22} \neq 0$.

Given the above, to obtain an AR(1) process with the desired properties, we need some of the factors of the AR lag polynomial on the left-hand side (LHS) to cancel against the MA lag polynomial on the RHS. At the same time, to get an AR(1) process and not an AR(2), we need to make sure that there are no L^2 terms on the LHS. For the latter, there are two possibilities.

The first possibility is that $\phi_{11}\phi_{22} + \phi_{12}\phi_{21} = 0$. In that case, the lag polynomial on the LHS becomes $[1-(\phi_{11}+\phi_{22})L]$. However, for the MA term on the RHS to cancel out, we need $\phi_{11}+\phi_{22} = \theta$. (Otherwise, we would get an ARMA(1,1) process.) But in that case, x_{1t} is white noise and serially uncorrelated, and hence we impose that $\phi_{11}\phi_{22} + \phi_{12}\phi_{21} \neq 0$.

As a result, to get rid of the L^2 term on the LHS, we need $\phi_{21} = 0$. Then, for the roots to cancel, we need either $\phi_{11} = \theta$ or $\phi_{22} = \theta$. First, we check $\phi_{22} = \theta$. Some algebra, however, shows that

$$\phi_{22} = \theta \quad \Rightarrow \quad 4\phi_{12}^2\phi_{22}^2\frac{\sigma_2^2}{\sigma_1^2} = 0,$$

which implies that $\phi_{12} = 0$, violating the requirement that x_{1t} be predictable by x_{2t}.

Therefore, we set $\phi_{11} = \theta$ and solve the equation for ϕ_{12} to finally obtain

$$\phi_{12} = \pm\frac{\sigma_1}{\sigma_2}\sqrt{\frac{\phi_{22}}{\phi_{11}} + \phi_{11}\phi_{22} - (1+\phi_{22}^2)},$$

where we impose $\phi_{11} \neq 0$ and $[\phi_{22}/\phi_{11} + \phi_{11}\phi_{22} - (1+\phi_{22}^2)] \geq 0$.

The only thing that remains to get to Eq. (3) is to parametrize the process appropriately. Letting $\mathbf{\Gamma}_k = \mathbb{E}[\mathbf{x}_t \mathbf{x}_{t-k}^\top]$ denote the autocovariance matrix of $\mathbf{x}_t$, it is well known (e.g., Lütkepohl, 2005, p. 27) that

$$\begin{aligned} \text{vec}(\mathbf{\Gamma}_0) &= (\mathbf{I} - \mathbf{A} \otimes \mathbf{A})^{-1} \text{vec}(\mathbf{\Sigma}) \\ \mathbf{\Gamma}_k &= \mathbf{A} \mathbf{\Gamma}_{k-1} \end{aligned}$$

We use this result to calculate that $\mathbb{V}\text{ar}(x_{2t}) = \sigma_2^2/(1-\phi_{22}^2)$ and $\mathbb{V}\text{ar}(x_{1t}) = \phi_{22}\sigma_1^2/[\phi_{11}(1-\phi_{22}^2)]$.[23] Hence, to ensure symmetry and $\mathbb{V}\text{ar}(x_{1t}) = \mathbb{V}\text{ar}(x_{2t})$, we set

$$\sigma_2 = \sigma_1 \sqrt{\frac{\phi_{22}}{\phi_{11}}}. \tag{8}$$

Next, we calculate that

$$\rho \equiv \mathbb{C}\text{orr}(x_{1t}, x_{2,t-1}) = \frac{\sqrt{\frac{\phi_{22}}{\phi_{11}} + \phi_{11}\phi_{22} - (1 + \phi_{22}^2)}}{(1 - \phi_{11}\phi_{22})\sqrt{\frac{\phi_{22}}{\phi_{11}}}}. \tag{9}$$

Setting $\phi \equiv \phi_{22}$ yields a quadratic equation for ϕ_{11}. Solving the resulting equation for $\phi_{11} \neq \phi^{-1}$ (which would both imply $\phi_{12} = 0$ and violate stationarity for x_{1t}), we obtain

$$\phi_{11} = \frac{\phi\,(1 - \rho^2)}{1 - \phi^2 \rho^2}. \tag{10}$$

Finally, note that Eqs. (8) and (9) imply that

$$\rho = \frac{\phi_{12}}{(1 - \phi_{11}\phi_{22})}.$$

Setting $\phi_{22} = \phi$ and using Eq. (10) yields

$$\phi_{12} = \frac{\rho(1 - \phi^2)}{1 - \phi^2 \rho^2},$$

concluding the calculation.

[23] Alternatively, we can calculate the necessary variances from first principles by using the methods in the proof of Proposition 1, exploiting the fact that $\phi_{21} = 0$ yields an upper-triangular structure for $\mathbf{A}$.

Appendix D Theoretical Predictions

We now derive the theoretical predictions for over- and underreaction to new information for a simple model of expectations. This model nests full-information rational expectations—as well as various deviations from rational expectations—as special cases.

Suppose that while the true parameters governing the data-generating process are given by ϕ and ρ, the subjects perceive these parameters to be equal to ϕ_p and ρ_p; all parameters are assumed to lie in $(0, 1)$. Otherwise, the subjects correctly perceive the structure of the process.[24] Hence, their forecasts are given by

$$\mathbb{F}_t[A_{t+1}] = \begin{cases} \mu(1-\phi_p) + \phi_p A_t & \text{in the } \textit{Baseline} \text{ treatment} \\ \mu(1-\phi_{1,p}-\phi_{2,p}) + \phi_{1,p}A_t + \phi_{2,p}B_t & \text{in the } \textit{Correlated} \text{ treatment} \end{cases}$$

Here, $\phi_{1,p}$ and $\phi_{2,p}$ are given by the expressions in Eq. (3) with ϕ and ρ replaced by their perceived counterparts.

A variety of existing models of expectation formation is captured by this simple formulation. In the special case $\phi_p = \phi$ and $\rho_p = \rho$, we recover full-information rational expectations. When $\phi_p > \phi$, subjects perceive the variables to be more persistent than they actually are, whereas with $\phi_p < \phi$ subjects think that they are less persistent. In *Baseline*, underreaction to new information is captured by $\phi_p < \phi$, while $\phi_p > \phi$ yields overreaction. [25] In *Correlated*, under- and overreaction is driven by both ϕ_p and ρ_p, as characterized below. Partially ignoring the informational content of B_t would be captured by $\rho_p < \rho$, while mistakenly thinking that B_t is more important for predicting A_t than is actually the case is given by $\rho_p > \rho$.

We now derive the predictions of the model for under- and overreaction to new information. To quantify under- and overreaction, we follow the methodology of Kucinskas and Peters (forthcoming) and calculate the theoretically predicted *bias coefficients*. Bias coefficients are direct measures of under- and overreaction. These coefficients are equal to the difference between the perceived response of A_t to some shock to the actual response of A_t to that shock. A positive bias coefficient indicates overreaction, while a negative coefficient means underreaction.

[24] Given our focus on under- and overreaction, it is without loss of generality to assume that subjects correctly perceive μ.

[25] For example, the sticky-information model of Mankiw and Reis (2002) and the noisy-information model of Woodford (2003) both predict underreaction to new information in the AR(1) case. While the pattern of underreaction is not exactly the same as in the model with a misperceived level of persistence, it is qualitatively similar (see Kucinskas and Peters, forthcoming). By the same logic, extrapolative or diagnostic expectations (Bordalo, Gennaioli and Shleifer, 2018) can be approximated by $\phi_p > \phi$.

We calculate bias coefficients for the response of expectations to the current shocks, ε_t and η_t. Similarly to Coibion and Gorodnichenko (2012), we normalize the coefficients by the true response of A_t. In *Baseline*, the bias coefficient with respect to ε_t (shock to A_t) is given by

$$b_A^{\text{baseline}} = \frac{\phi_p - \phi}{\phi}.$$

In *Correlated*, bias coefficients with respect to ε_t (shock to A_t) and η_t (shock to B_t) are equal to

$$b_A^{\text{correlated}} = \frac{\phi_{1,p} - \phi_1}{\phi_1}$$
$$b_B^{\text{correlated}} = \frac{\phi_{2,p} - \phi_2}{\phi_2}$$

Finally, we calculate a measure of *overall overreaction*. This measure combines the reaction to all shocks that are present in a given treatment. We obtain this measure from the population regression of forecast errors on the *optimal forecast revision*. The optimal forecast revision is given by

$$\text{rev}_t^* \equiv \mathbb{E}_t[x_{t+1}] - \mathbb{E}_{t-1}[x_{t+1}] = \begin{cases} \phi\varepsilon_t & \text{in the } \textit{Baseline} \text{ treatment} \\ \phi_1\varepsilon_t + \phi_2\eta_t & \text{in the } \textit{Correlated} \text{ treatment} \end{cases}$$

We obtain our measure of overall overreaction by the slope coefficient β in

$$\text{fe}_{t+1} = \alpha + \beta \text{rev}_t^* + u_{t+1},$$

where forecast errors fe_{t+1} are defined as $\text{fe}_{t+1} = \mathbb{F}_t[A_{t+1}] - A_{t+1}$. With full-information rational expectations, $\beta = 0$. If subjects overreact to new information (as proxied by rev_t^*) and adjust the forecast more than is optimal, $\beta > 0$, while underreaction yields $\beta < 0$. Note that in *Baseline*, β is equal to b_A^{baseline}, as it should be given that there is only one shock in *Baseline*. In *Correlated*, β is given by a weighted average of the two shock-specific bias coefficients:

$$\left(\frac{\phi_1^2\sigma_\varepsilon^2}{\phi_1^2\sigma_\varepsilon^2 + \phi_2^2\sigma_\eta^2}\right) b_A^{\text{correlated}} + \left(\frac{\phi_2^2\sigma_\eta^2}{\phi_1^2\sigma_\varepsilon^2 + \phi_2^2\sigma_\eta^2}\right) b_B^{\text{correlated}}.$$

The weights are given by the fraction of variance in optimal forecast revisions due to each shock.[26] For the model of *diagnostic expectations* (Bordalo, Gennaioli and Shleifer, 2018), the coefficient β is equal to the representativeness parameter (θ in the notation of Bordalo, Gennaoili, and Shleifer; c.f. their Proposition 1). This observation provides an additional justification for using β as a measure of overall overreaction.

The following result is immediate from the definitions above.

Proposition 3. *All else equal, overreaction depends on the perceived parameters as follows:*

1. *In Baseline, overreaction to ε_t (and hence also overall overreaction) is increasing in the perceived persistence ϕ_p.*
2. *In Correlated, overreaction to ε_t is increasing in the perceived persistence ϕ_p and decreasing in the perceived correlation ρ_p. Overreaction to η_t is decreasing in the perceived persistence ϕ_p and increasing in the perceived correlation ρ_p. The effect of changes in ϕ_p and ρ_p on overall overreaction is ambiguous.*

Intuitively, overreaction to the shock to A_t (i.e., ε_t) is more likely when subjects perceive persistence of A_t to be higher or the predictive content of B_t lower. The effects on overreaction to the shock to B_t (i.e., η_t) go in the opposite direction. If subjects underestimate the predictive content of B_t, they are more likely to underreact to η_t. Given the opposing effects, the effect of perceived parameters on overall overreaction is ambiguous. In particular, even if people underestimate the importance of B_t in predicting future values of A_t, they may nevertheless overreact to shocks to A_t as well as exhibit overall overreaction. Hence, whether or not people exhibit more underreaction in *Correlated* is an empirical question.

A key insight from the proposition is that the way people react to different sources of information in *Correlated* is interdependent. If people neglect the informational content of B_t more (ρ_p is lower), that increases underreaction to B_t. However, that simultaneously makes the agent overreact more to A_t. The total effect on overall overreaction is ambiguous and depends on the parameter values. This result, while simple, captures an important fact. If subjects have limited attention, focusing too much on one variable must necessarily lead to too little attention allocated to other variables.

[26] Our measure of overall under- and overreaction is closely related to the composite bias coefficients proposed by Kucinskas and Peters (forthcoming). The key difference is that composite bias coefficients are given by a non-linear function of the shock-specific bias coefficients, rather than a weighted average as here. The key advantage of the measure used in the present paper is its robustness to noise in expectations, as the estimation employs the true shocks.

Appendix E Proofs

Proposition 1

Property 1. First, calculate that

$$\mathbb{V}\text{ar}(B_t) = \frac{\sigma_\varepsilon^2(1-\phi^2\rho^2)}{(1-\phi^2)(1-\rho^2)}.$$

Since B_t is just an AR(1) process, its autocovariance function satisfies $\gamma_k^B = \phi\gamma_{k-1}^B$ and hence is given by Eq. (4).

To find γ_k^A, start by deducing

$$\mathbb{C}\text{ov}(A_t, B_t) = \frac{\phi_2\phi}{1-\phi_1\phi}\,\mathbb{V}\text{ar}(B_t) = \frac{\sigma_\varepsilon^2\phi\rho(1-\phi^2\rho^2)}{(1-\phi^2)(1-\rho^2)}.$$

Then, from Eq. (2), we have

$$\mathbb{V}\text{ar}(A_t) = \phi_1^2\,\mathbb{V}\text{ar}(A_t) + \phi_2^2\,\mathbb{V}\text{ar}(B_t) + 2\phi_1\phi_2\,\mathbb{C}\text{ov}(A_t, B_t) + \sigma_\varepsilon^2.$$

Substituting the expressions for ϕ_1, ϕ_2 and $\mathbb{C}\text{ov}(A_t, B_t)$ yields $\mathbb{V}\text{ar}(A_t) = \mathbb{V}\text{ar}(B_t)$.

Now for $k > 0$

$$\begin{aligned}\mathbb{C}\text{ov}(A_t, A_{t-k}) &= \mathbb{C}\text{ov}(\phi_1 A_{t-1} + \phi_2 B_{t-1} + \varepsilon_t, A_{t-k}) \\ &= \phi_1\,\mathbb{C}\text{ov}(A_t, A_{t-k+1}) + \phi_2\,\mathbb{C}\text{ov}(B_t, A_{t-k+1})\end{aligned}$$

From Eq. (2), $\mathbb{C}\text{ov}(B_t, A_{t-k}) = \phi^k\,\mathbb{C}\text{ov}(B_t, A_t)$. Therefore, γ_k^A satisfies the difference equation

$$\gamma_k^A = \phi_1\gamma_{k-1}^A + \phi_2\phi^{k-1}\frac{\sigma_\varepsilon^2\phi\rho(1-\phi^2\rho^2)}{(1-\phi^2)(1-\rho^2)},$$

with the initial condition $\gamma_0^A = \mathbb{V}\text{ar}(A_t)$. It is straightforward to check that the solution to this difference equation is given by Eq. (4).

Property 2. Note that for $k > 0$

$$\begin{aligned}\mathbb{C}\text{ov}(A_t, B_{t-k}) &= \phi_1\,\mathbb{C}\text{ov}(A_t, B_{t-k+1}) + \phi_2\,\mathbb{C}\text{ov}(B_t, B_{t-k+1}) \\ &= \left[\frac{\phi(1-\rho^2)}{1-\phi^2\rho^2}\right]\mathbb{C}\text{ov}(A_t, B_{t-k+1}) + \frac{\rho\phi^{k-1}\sigma_\varepsilon^2}{1-\rho^2}.\end{aligned}$$

This is again a difference equation, with the initial condition given by $\mathbb{C}\text{ov}(A_t, B_t)$.

Solving the difference equation yields

$$\mathbb{C}\text{ov}(A_t, B_{t-k}) = \frac{\sigma_\varepsilon^2(1-\phi^2\rho^2)\phi^{k-1}}{\rho(1-\phi^2)(1-\rho^2)}\left[1 - \frac{(1-\rho^2)^k}{(1-\phi^2\rho^2)^{k-1}}\right].$$

Hence, $\mathbb{C}\text{ov}(A_t, B_t) = \phi\rho\,\mathbb{V}\text{ar}(A_t)$ and $\mathbb{C}\text{ov}(A_t, B_{t-1}) = \rho\,\mathbb{V}\text{ar}(A_t)$, and since $\mathbb{V}\text{ar}(A_t) = \mathbb{V}\text{ar}(B_t)$, the result follows.

Property 3. For a general stationary VAR(1) process $\mathbf{x}_t = \mathbf{b} + \mathbf{A}\mathbf{x}_{t-1} + \boldsymbol{\varepsilon}_t$, we have that $\mathbf{x}_t = (\mathbf{I}-\mathbf{A})^{-1}\mathbf{b} + \sum_{\ell=0}^{\infty}\mathbf{A}^\ell\boldsymbol{\varepsilon}_{t-\ell}$. In the present case,

$$\mathbf{A} = \begin{pmatrix} \phi_1 & \phi_2 \\ 0 & \phi \end{pmatrix} \text{ and } \mathbf{b} = \begin{pmatrix} \mu(1-\phi_1-\phi_2) \\ \mu(1-\phi) \end{pmatrix}.$$

First, calculate that $(\mathbf{I}-\mathbf{A})^{-1}\mathbf{b} = (\mu,\mu)^\top$. Second, we need to find $\mathbf{A}^\ell$. We claim that

$$\mathbf{A}^\ell = \begin{pmatrix} \left[\frac{\phi(1-\rho^2)}{1-\phi^2\rho^2}\right]^\ell & \frac{1}{\phi\rho}\left\{\phi^\ell - \left[\frac{\phi(1-\rho^2)}{1-\phi^2\rho^2}\right]^\ell\right\} \\ 0 & \phi^\ell \end{pmatrix}.$$

The proof is by induction. The expression is clearly true for $\ell = 1$. Suppose it is true for an arbitrary integer $\ell \geq 1$. It is then straightforward to check that

$$\mathbf{A}\mathbf{A}^\ell = \mathbf{A}^\ell\mathbf{A} = \mathbf{A}^{\ell+1},$$

proving the induction step. The result in the text then follows.

Property 4. By definition of v_t,

$$v_t = \mu(\phi - \phi_1 - \phi_2) + (\phi_1 - \phi)A_{t-1} + \phi_2 B_{t-1} + \varepsilon_t.$$

As a result, we calculate that for $k > 0$

$$\begin{aligned}\mathbb{C}\text{ov}(v_t, v_{t-k}) = (\phi_1-\phi)^2\,\mathbb{C}\text{ov}(A_t, A_{t-k}) + (\phi_1-\phi)\phi_2\,\mathbb{C}\text{ov}(A_t, B_{t-k}) \\ + (\phi_1-\phi)\,\mathbb{C}\text{ov}(A_t, \varepsilon_{t-k+1}) \\ + (\phi_1-\phi)\phi_2\,\mathbb{C}\text{ov}(B_t, A_{t-k}) + \phi_2^2\,\mathbb{C}\text{ov}(B_t, B_{t-k}).\end{aligned}$$

Now $\mathbb{C}\text{ov}(A_t, \varepsilon_{t-k}) = \phi_1^k\sigma_\varepsilon^2$. Substituting the expressions for the covariances from above and simplifying yields $\mathbb{C}\text{ov}(v_t, v_{t-k}) = 0$.

Calculating the variance yields

$$\mathbb{V}\text{ar}(v_t) = \mathbb{V}\text{ar}(A_t)\left[(\phi_1-\phi)^2 + \phi_2^2 + 2(\phi_1-\phi)\phi_2\frac{\phi_2\phi}{1-\phi_1\phi}\right] + \sigma_\varepsilon^2.$$

Substituting the expressions and simplifying yields the result in the proposition.

Finally, for $k > 0$

$$\mathbb{C}\text{ov}(v_t, A_{t-k}) = (\phi_1 - \phi)\,\mathbb{C}\text{ov}(A_t, A_{t-k+1}) + \phi_2\,\mathbb{C}\text{ov}(B_t, A_{t-k+1}),$$

which is also seen to be 0.

Proposition 2

Write out the objective function in Eq. (6) explicitly as

$$\begin{aligned}\mathbb{E}_i[\min\{K, |e|\}] &= \int_{-\infty}^{F-K} K\,\mathrm{d}\,G_i(A) + \int_{F-K}^{F+K} |A-F|\,\mathrm{d}\,G_i(A) + \int_{F+K}^{+\infty} K\,\mathrm{d}\,G_i(A)\\ &= K\left\{1 - [G_i(F+K) - G_i(F-K)]\right\} + \int_{F-K}^{F} (F-A)\,\mathrm{d}\,G_i(A)\\ &\quad + \int_{F}^{F+K} (A-F)\,\mathrm{d}\,G_i(A)\end{aligned}$$

Differentiate w.r.t. F and set to zero to find

$$G_i(F) = \frac{1}{2}\left[G_i(F+K) + G_i(F-K)\right]. \tag{11}$$

Since the subjective distribution of A is symmetric, it satisfies $G_i(x) = 1 - G_i(2\mu_i - x)$. But then $F = \mu_i$ satisfies Eq. (11) since

$$G_i(\mu_i) = \frac{1}{2} = \frac{1}{2}[G_i(\mu_i + K) + G_i(\mu_i - K)].$$

Since the subjective distribution is unimodal, $g'(x) > 0$ for $x < \mu_i$ and $g'(x) < 0$ for $x > \mu_i$. Hence, $G_i(x)$ is strictly convex for $x < \mu_i$ and strictly concave for $x > \mu_i$. But then, for $F < \mu_i$, we have that

$$G_i(F) = G_i\left[\frac{1}{2}(F+K) + \frac{1}{2}(F-K)\right] > \frac{1}{2}[G_i(F+K) + G_i(F-K)],$$

and the inequality is reversed for $F > \mu_i$. Hence, the objective function is strictly decreasing for $F < \mu_i$ and strictly increasing for $F > \mu_i$. Hence, $F = \mu_i$ is the unique global minimizer.

Appendix F Deviations From Analysis Plan

In our empirical analysis, we follow the plan laid out in the preregistration closely. In a few instances, however, we have deviated slightly from the original plan, as we now document:

- Filter (3) in the sample selection criteria ("Exclude an observation if its absolute forecast error is larger than 500.") was not in the original analysis plan. Filter (3) was added afterwards because there were some extreme forecasts in the experiment that we had not anticipated. These extreme forecasts make most of the empirical analysis noisy or very sensitive to these extreme forecasts. The reason for the extreme forecasts is that subjects could type in any number for their prediction using a text box. As shown in Table A.2, only 12 individual forecasts (out of 39,120) are removed by filter 3.
- In the preregistration, we phrased our accuracy tests in terms of total scores (not per-round scores). However, per-round scores are easier to interpret, and hence we have chosen to present the final results in per-round terms. This difference has no effect on the statistical results as the average per period score is equal to the total score divided by 40.
- Figure 3 was not included in our original analysis plan. We included this graph after becoming aware of the work by Enke and Graeber (2019) that inspired this figure, having already collected the experimental data.
- The analysis using OLS forecasts (Tables A.5-A.8) was not included in our original analysis plan. We used OLS forecasts to: (i) test whether a Bayesian agent would be less accurate in *Correlated* (placebo test for differences in accuracy scores); and (ii) measure under- and overreaction with respect to a Bayesian benchmark rather than full-information rational expectations. These exercises were inspired by feedback from our colleagues after having collected the data.
- Figure A.2 was not included in our original analysis plan. We included this graph to depict potential learning effects in a simple graphical manner.
- In the original analysis plan, we planned to do another round of data-quality checks by further excluding subjects who failed the attention question in the Questionnaire, and subjects whose total response time was in the top or bottom 5% of the sample, and subjects with duplicated IPs. We did not perform this additional round of checks as we have already done many other robustness analyses.
- In the original analysis plan, we referred to the *Baseline* treatment as *Simple*, and the *Correlated* treatment as *Complex*. We changed the names of the treatments to better reflect the experimental manipulation.

- The new treatments in Section 4 were not pre-registered. We conducted these additional treatments after receiving constructive feedback from referees. However, the statistical analysis of data from these treatments followed exactly the same approach as in Section 3.

Appendix G Additional Tables and Figures

Table A.1

Summary Statistics: Initial Sample

Notes: Summary statistics for the initial sample (all subjects that successfully completed the experiment). Forecast errors are defined as prediction minus the realization. Prediction times per round are shown for all rounds excluding the first round.

	N	Mean	Med.	Std. Dev.	Min.	Max.
Total Time (min)	978	13.20	9.44	10.47	1.47	57.25
Time / Round (sec)	38,142	10.22	7.00	26.02	0.00	1790.00
Score / Round	39,120	26.39	1.09	33.14	0.00	99.99
Forecast Error	39,120	4.82	0.85	683.13	-8108.36	100987.94
Abs. Forecast Error	39,120	19.51	9.89	682.87	0.00	100987.94
Bonus ($)	978	2.11	2.15	0.65	0.00	4.02

Table A.2

Sample Selection Procedures

Notes: Sample size after various data filters imposed in our analysis. Total number of forecasts shown.

	Forecasts
Finished experiment:	39,803
Remove subjects with duplicated periods:	−243
Remove subjects taking more than 1 hour:	−440
Initial sample:	39,120
Remove subjects finishing in $\leq$ 3 minutes:	−560
Remove subjects with total score $\leq$ 200:	−760
Remove forecasts with absolute errors $\geq$ 500:	−12
Final sample:	37,788

Table A.3

Balance Tests

Notes: Balance tests for randomization in *Baseline* and *Correlated*. Mann-Whitney p-values are reported. *Dropout Rate* is the fraction of subjects that started the experiment on MTurk but did not finish. The number is calculated using data on all subjects in the experiment, excluding only those who had technical problems. *Extreme Forecast Error* is the fraction of subjects that made a forecast error larger than 500 in absolute value. The number is calculated using our initial sample; the remaining statistics are calculated using our final sample. $\geq$ *University* is the fraction of subjects with college or graduate education. *Attention Check* is the fraction of subjects that successfully pass our attention screen. *Eye for Detail* and *Concentration* give the fraction of subjects who agree or strongly agree with survey questions that ask for a self-reported evaluation of attention to detail and level of concentration.

	Baseline	Correlated	p-value
Dropout Rate	0.09	0.07	0.32
Extreme Forecast Error	0.02	0.01	0.20
Age	38.65	38.00	0.51
Fraction Female	0.51	0.47	0.24
$\geq$ University	0.72	0.75	0.40
Attention Check	1.00	1.00	0.23
Taken Statistics Course	0.34	0.30	0.19
Eye for Detail	0.74	0.70	0.17
Concentration	0.80	0.80	0.92

Table A.4

Summary Statistics by Treatment

Notes: Summary statistics for the various treatments. Standard errors robust to heteroskedasticity.

	Avg. Score	s.e.	N
Baseline	29.59	0.41	307
Correlated	25.65	0.28	638
Baseline: High Persist.	29.02	0.59	148
Correlated: High Persist., High Corr.	24.66	0.48	162
Correlated: High Persist., Low Corr.	27.77	0.57	156
Baseline: Low Persist.	30.12	0.57	159
Correlated: Low Persist., High Corr.	22.33	0.52	154
Correlated: Low Persist., Low Corr.	27.73	0.54	166

Table A.5

Placebo Test: OLS Forecasts

Notes: Placebo test for differences in accuracy scores of OLS forecasts in *Baseline* and *Correlated*. In *Baseline*, OLS forecasts are generated by first estimating

$$A_t = b_0 + b_1 A_{t-1} + u_t$$

and then forecasting A_{t+1} as $\hat{b}_0 + \hat{b}_1 A_t$. In *Correlated*, OLS forecasts are generated by first estimating

$$A_t = b_0 + b_1 A_{t-1} + b_2 B_{t-1} + u_t$$

and then forecasting A_{t+1} as $\hat{b}_0 + \hat{b}_1 A_t + \hat{b}_2 B_t$. When forecasting A_{t+1}, only data for periods $1, 2, \ldots, t$ is used in the OLS estimation.

We then calculate accuracy scores for these forecasts in the same way as we do for subjects in the experiment, and test whether there is a statistically significant difference in forecast accuracy. Heteroskedasticity-robust standard errors shown; Mann-Whitney p-value is provided in the bottom row.

	Baseline	Correlated	B−C
OLS Accuracy Score	36.34	36.00	0.34
s.e.	0.31	0.22	MW p-value: 0.44

Table A.6

Overreaction using OLS forecasts: Main Results

Notes: Estimated levels of overreaction in *Baseline* and *Correlated*, with sample splits by the level of persistence (ϕ). Standard errors clustered by subject in parentheses. The last column gives the difference in overreaction levels between *Baseline* and *Correlated*, and the bottom row gives the difference between the low- and high-persistence treatments. The table shows estimates of β from the regression of

$$\text{prediction}_{i,t} - \text{OLS forecast}_{i,t} = \alpha_i + \beta\, \text{rev}^*_{i,t-1} + u_{i,t}$$

for the appropriate subsample, where $\text{rev}^*_{i,t}$ is the optimal forecast revision. Differences between treatments are estimated by running regressions with appropriate interaction terms.

	Baseline	Correlated	B−C
Full Sample	0.10 (0.03)	-0.44 (0.02)	0.54 (0.04)
Low Persistence	0.76 (0.10)	-0.65 (0.03)	1.41 (0.10)
High Persistence	0.02 (0.03)	-0.21 (0.03)	0.23 (0.04)
L−H	0.73 (0.10)	-0.45 (0.04)	1.18 (0.11)

Table A.7

Overreaction using OLS forecasts: Effects of Correlation

Notes: Estimated levels of overreaction in *Baseline* and *Correlated*, with sample splits by the level of correlation (ρ). Standard errors clustered by subject in parentheses. See Table A.6 for further explanations.

	Baseline	Correlated	B−C
Full Sample	0.10 (0.03)	-0.44 (0.02)	0.54 (0.04)
Low Persistence	0.76 (0.10)	-0.65 (0.03)	1.41 (0.10)
High Persistence	0.02 (0.03)	-0.21 (0.03)	0.23 (0.04)
L−H	0.73 (0.10)	-0.45 (0.04)	1.18 (0.11)

Table A.8

Overreaction to A vs. B using OLS forecasts.

Notes: Estimated levels of overreaction to shocks in A (target variable) and B (correlated variable). Standard errors clustered by subject in parentheses. The table shows estimates of β from the regression of

$$\text{prediction}_{i,t} - \text{OLS forecast}_{i,t} = \alpha_i + \beta\,\text{scaled shock}^*_{i,t-1} + u_{i,t},$$

where the scaled shocks are defined in the text and represent the revision in the optimal forecast stemming from the realization of that shock.

	Baseline	Correlated	B−C
Shock to A	0.10 (0.03)	0.09 (0.04)	0.01 (0.05)
Shock to B	–	-0.72 (0.02)	–
Shock to $A-$Shock to B	–	0.81 (0.05)	–

Figure A.2

Average Scores During the Experiment

Notes: Average total scores by period for the *Baseline* (solid black line) and *Correlated* (dashed blue line) treatments.

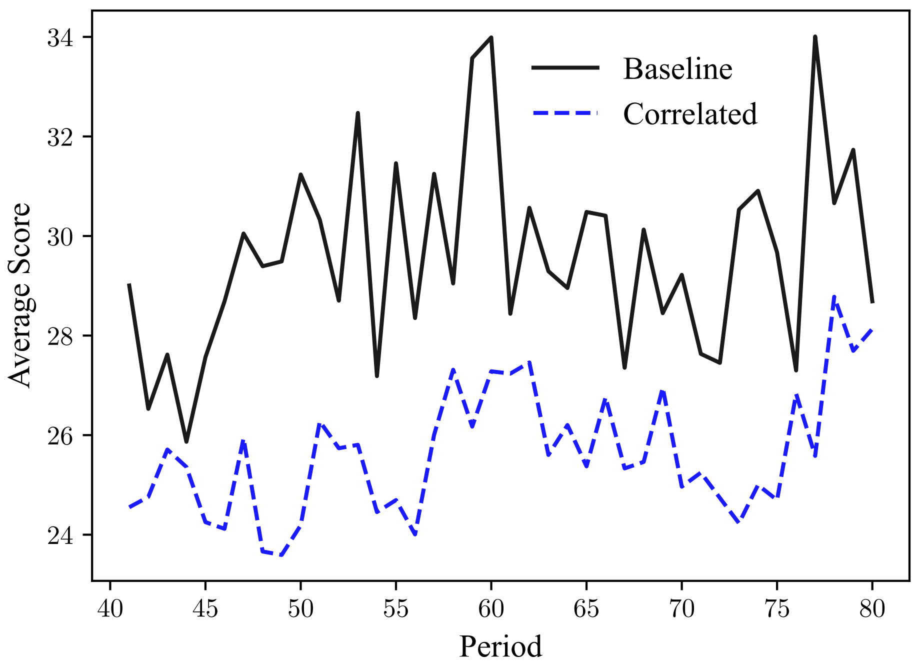


Table A.9

Regressions on Individual Characteristics

Notes: The table shows estimated of γ from the regression of

$$\text{avg. score}_i = \alpha_i + \beta\,\mathrm{X}_i + \gamma\,\text{correlated}_i\,\mathrm{X}_i + \delta\,\text{correlated}_i + u_{i,t},$$

where X_i contains the four individual characteristics “female”, “taken statistical course”, “$\geq$ University”, and “find B useful”. “$\geq$ *University*” is the fraction of subjects with college or graduate education. *** indicates statistical significance at the 0.01 level.

Variables	Avg. Score
Female × *Correlated*	0.181
	(0.981)
Taken Statistical Course × *Correlated*	0.454
	(1.096)
$\geq$ *University* × *Correlated*	-1.444
	(1.159)
Find B Useful × *Correlated*	0.624
	(0.566)
Correlated	-3.429***
	(1.072)
Observations	945
R-squared	0.075

Table A.10

Sample Selection Procedures: Follow-Up Experiment

Notes: Sample size after various data filters imposed in our analysis. Total number of forecasts shown.

	Forecasts
Finished experiment:	29,880
Remove subjects with duplicated periods:	−0
Remove subjects taking more than 1 hour:	−1,040
Initial sample:	28,840
Remove subjects finishing in $\leq$ 3 minutes:	−1,240
Remove subjects with total score $\leq$ 200:	−3,280
Remove forecasts with absolute errors $\geq$ 500:	−19
Final sample:	24,301

Table A.11

Balance Tests: Follow-Up Experiment

Notes: Balance tests for randomization in *Baseline* and *Correlated*. Mann-Whitney p-values are reported. *Dropout Rate* is the fraction of subjects that started the experiment on MTurk but did not finish. The number is calculated using data on all subjects in the experiment, excluding only those who had technical problems. *Extreme Forecast Error* is the fraction of subjects that made a forecast error larger than 500 in absolute value. The number is calculated using our initial sample; the remaining statistics are calculated using our final sample. $\geq$ *University* is the fraction of subjects with college or graduate education. *Attention Check* is the fraction of subjects that successfully pass our attention screen. *Eye for Detail* and *Concentration* give the fraction of subjects who agree or strongly agree with survey questions that ask for a self-reported evaluation of attention to detail and level of concentration.

	Baseline	Correlated	p-value
Dropout Rate	0.01	0.00	0.57
Extreme Forecast Error	0.04	0.04	0.99
Age	37.18	37.23	0.87
Fraction Female	0.43	0.43	0.98
$\geq$ University	0.89	0.86	0.29
Attention Check	0.97	1.00	0.01
Taken Statistics Course	0.53	0.55	0.62
Eye for Detail	0.75	0.79	0.29
Concentration	0.76	0.75	0.93

Table A.12

Summary Statistics by Treatment: Follow-Up Experiment

Notes: Summary statistics for the various treatments. Standard errors robust to heteroskedasticity.

	Avg. Score	s.e.	N
Baseline	25.63	0.49	306
Correlated	22.67	0.44	302
Baseline-Replication	26.32	0.89	95
Baseline-Info	24.65	0.86	107
Baseline-Info-DGP	26.02	0.83	104
Correlated-Replication	21.96	0.73	99
Correlated-Info	22.60	0.85	104
Correlated-Info-DGP	23.46	0.70	99

Table A.13

Placebo Test: OLS Forecasts in *Info-DGP*

Notes: Placebo test for differences in accuracy scores of OLS forecasts in *Baseline-Info-DGP* and *Correlated-Info-DGP*. Heteroskedasticity-robust standard errors shown; Mann-Whitney p-value is provided in the bottom row. See Table A.5 for further explanations.

	Baseline-Info-DGP	Correlated-Info-DGP	B−C
OLS Accuracy Score	36.20	36.05	0.15
s.e.	0.55	0.57	MW p-value: 0.92