Can Agents Add and Subtract When Forming Beliefs? Evidence from the Lab and Field

Pascal Kieren, Jan Müller-Dethard, and Martin Weber[1]

March, 2021

Abstract

Bayes' Theorem has an implicit, fundamental rule of how subjects should incorporate informationally equivalent signals of opposite direction: two opposite-directional signals should cancel out such that prior beliefs remain constant. In this study, we test whether agents always follow this simple counting heuristic. We find that this is not the case. Whenever a sequence of signals that go in the same direction is interrupted by a signal of opposite direction, agents violate the simple counting heuristic and strongly overreact to the signal of opposite direction. In contrast to that, subjects correctly follow the counting heuristic whenever opposite-directional signals alternate. Building on our experimental findings, we empirically analyze announcement and post-announcement stock return reactions. In line with our experimental evidence, we find that initial stock reactions are significantly stronger and subsequent price drifts weaker for opposite-directed earnings surprises than for same-directed earnings surprises.

Keywords: *Belief Formation, Bayes Theorem, Information Processing, Overreaction*

JEL Classification: *D81, D83, D84, G41*

[1] Pascal Kieren is at the University of Mannheim, Finance Department, L9, 1-2, 68161 Mannheim, pascal.kieren@gess.uni-mannheim.de, Jan Müller-Dethard is at the University of Mannheim, Finance Department, L9, 1-2, 68161 Mannheim, mueller-dethard@bank.bwl.uni-mannheim.de, Martin Weber is at the University of Mannheim, Finance Department, L9, 1-2, 68161 Mannheim, and CEPR, London, weber@bank.bwl.uni-mannheim.de

Introduction

Probabilistic beliefs are essential to decision-making under risk in various economic problems including investments in financial markets, purchasing insurance, attaining education, or when searching for employment. Standard models assume that individuals update their beliefs according to Bayes' Theorem. Besides the prescription of how to calculate posterior probabilities, Bayes' Theorem has an implicit, fundamental rule of how subjects should incorporate information *signals of opposite direction*. In the usual case of updating about two states of the world from independent binomial signals, two unequal signals should cancel out. Thus, taken together they should not affect prior beliefs.

To illustrate this idea, imagine a physician who seeks to design a treatment program to improve a patient's condition. To refine her certainty that a patient is suffering from a particular disease, the physician has access to a diagnostic test, which can yield positive and negative results. Suppose, the first two tests yield negative results, making the physician rather optimistic about the patient's condition. Yet, to further reduce uncertainty, the doctor decides to administer two more diagnostic tests. Assume, the first test yields a positive result, whereas the second yields again a negative result. Would the doctor be just as optimistic as she was after the first two tests? In other words, assuming all tests are equally informative, are *two* negative test results just as good as *three* negative results and *one* positive, as prescribed by Bayes' Theorem?

In this article, we ask whether individuals follow this simple, counting-based rule when updating their beliefs. To test this, we create an environment in which subjects repeatedly observe binary signals to learn about one of two states of the world. While such a binary decision-making problem appears to presents a specific, commonly used and simplified setting in experimental research, it applies to many every-day decision problems (e.g. are we in a good or bad stock market regime, should I take an umbrella for the walk or not). Throughout this paper, we refer to signals that are in line with the true underlying state of the world as *confirming* signals and otherwise as *disconfirming* signals. We exogenously manipulate the number of subsequent confirming signals that gets interrupted by a single disconfirming signal. This setup allows us to test (i) how subjects update their priors after a *disconfirming* signal conditional on the number of previously observed confirming signals; and (ii) the extent to which they revise their priors after the disconfirming signal is followed by another confirming signal (i.e. *corrected*). In both cases, Bayes' Rule makes a simple, yet important prediction: An agent should reduce (increase) his prior after a disconfirming (confirming) signal by the same magnitude than he increased (reduced) it after the previous confirming (disconfirming) signal.

To implement this framework, we conduct three bookbag-and-poker-chip experiments in the spirit of Grether (1980) with 1800 participants. All experiments follow the same basic design. Over the course of six periods, we provide subjects with information about a risky asset which can either draw from a "good distribution" or from a "bad distribution". Both distributions are binary with a high outcome of +5 and a low outcome of –5. In the good distribution, the higher payoff occurs with 70 % probability while the lower payoff occurs with 30 % probability. In the bad distribution, the probabilities are reversed, i.e. the lower payoff occurs with 70 % probability while the higher payoff occurs with 30 % probability. To create situations which are consistent with our framework, we use a stratified sample of price paths. More precisely, we examine six price paths for the good distribution and six price paths for the bad distribution. In each of the six periods of a price path, subjects subsequently observe payoffs of the risky asset. After each payoff, we ask them to provide a probability estimate that the risky asset draws from the good distribution and how confident they are about their estimate. In Experiment 2 and 3 we run variations of our baseline experiment to test the robustness and underlying drivers of our findings. In Experiment 2, we change the informational content of the positive signal (i.e. the diagnosticity). In Experiment 3, we reduce the uncertainty about the underlying distribution by providing subjects with the full outcome history in advance. For comparability, the price paths we use in both variations remain identical to the baseline experiment.

Our experimental findings can be summarized as follows. First, we consistently find that subjects strongly overreact whenever a sequence of confirming signals is interrupted by a single disconfirming signal. Across all experiments, subjects update their prior beliefs on average by 3.54 % immediately before observing the disconfirming signal, whereas they update their prior beliefs on average by 15.38 % after the subsequent disconfirming signal. In relative terms, subjects update their priors by 334 % too much after a disconfirming signal, thereby acting as if one single disconfirming signal would carry the weight of up to three confirming signals.

Second, we find that this overreaction is almost entirely corrected once subjects observe another confirming signal following the disconfirming signal. More precisely, after observing a confirming signal following the disconfirming signal, they update their prior beliefs again by 13.65 %, compared to their initial overreaction of 15.38 %. In other words, subjects almost completely correct their initial overreaction if the disconfirming signal gets revered.

Third, we find that both the overreaction and the subsequent correction do not critically depend on subjects having extreme priors. Even with a diagnosticity of only 60 %, two subsequent confirming signals are sufficient to observe a pronounced overreaction after a disconfirming signal. In such a setting not only the experimentally observed subjective priors, but also the objective Bayesian probabilities are low with on average 72 % and 69 %, respectively.

Fourth, the observed overreaction after a disconfirming signal becomes stronger the more confirming signals individuals previously encountered. Even though – in absolute terms – the observed overreaction should become smaller as subjective priors converge to one, we find that a single disconfirming signal can completely revert up to five confirming signals the later it occurs. This implies that – in contrast to the Bayesian prediction – signals are not invariant to the order in which they occur. In other words, observing one single disconfirming signal followed by five confirming signals is different compared to observing five confirming signals that are followed by a single disconfirming signal. Whereas subjects mostly correct their strong overreaction if they can, the violation of the counting heuristic is most severe when subjects have no opportunity to collect further information.

In a next step, we apply our experimental insights to a real-case application in financial markets, that is stock return reactions to earnings announcements, and test whether the sequence of earnings surprises affects asset prices. In particular, we examine how the direction of the prior earnings surprise – same- or opposite-directed – relative to the current surprise affects announcement and post-announcement stock returns. We argue that the presence of a large subset of investors who display similar over- and underreaction patterns as observed in our well-identified experimental results can generate stock price over- and underreactions to news, respectively, and, in turn, return predictability and post-announcement price drift. If the earnings surprise is opposite-directed to the previous earnings surprise, we predict a stronger announcement stock return reaction than if it is same-directed. This follows from the experimentally observed updating behavior that shows a stronger absolute reaction after an opposite-directed signal relative to a prior same-directed signal. Moreover, and given the stronger initial reaction after opposite-directed earnings surprises, we assume a weaker post-announcement drift following opposite-directed earnings surprises.

To test this hypothesis, we examine stock returns and earnings announcement of more than 1400 US firms over the period from 2009 to 2020. Consistent with our experimental data, we find that initial stock reactions are significantly stronger for opposite-directed earnings surprises than for same-directed earnings surprises. That is, after positive earnings surprises, we observe on average stronger return reactions if the surprise is following a prior negative surprise, compared to when it is following a prior positive surprise. This larger initial reaction – however – reverts in the long-run with a weaker post-earnings announcement drift for opposite-directed positive surprises and a stronger post-earnings announcement drift for same-directed positive surprises. For negative surprises, we observe a similar pattern. In particular, we observe on average a stronger negative return reaction if the negative earnings surprise is following a prior positive surprise, compared to when it is following a prior negative surprise. In the long run, prices of stocks which experienced an opposite-directed negative earnings surprise drift even upwards, suggesting an initial overreaction, while stocks which experienced a same-directed negative earnings surprise show only a marginal or even no drift. This results in a convergence of prices. Remarkably, however, the initial difference carries forward and is not fully corrected even 60 days after the initial earnings announcement.

Our paper contributes to several strands of literature. First, we contribute to the various studies that document biases and heuristics in probabilistic reasoning (for an overview see Camerer, 1987, 1995, Benjamin, 2019). A common finding, by and large is that people update too little, with three exceptions as noted by Benjamin (2019): (i) People overinfer from signals if the diagnosticity is low, (ii) people may overinfer when signals go in the same direction of the priors (i.e. prior-biased updating), and (iii) people may overinfer when priors are extreme and signals go in the opposite direction of the priors (due to base-rate neglect). Especially, (ii) and (iii) push in opposite directions which makes it important to understand when one or the other dominates. Our study suggests that whenever subjects violate the simple counting heuristic implied by Bayes, individuals generally overreact to signals of opposite direction of their priors. A violation occurs whenever a sequence of signals that go in the same direction is interrupted by a signal of opposite direction. Importantly, we find that this overreaction is independent of subjects having extreme priors and requires only a sequence of two signals that go in the same direction. Conversely, we find that subjects generally underinfer in situations in which they cannot or do not violate the counting heuristic. This is either because there are (i) only signals of same direction, or (ii) positive and negative signals alternate.

Second, we also contribute to the literature on the co-existence of both over- and underreaction of prices to information in financial markets (De Bond & Thaler, 1985). In order to generate both over- and underreaction, the most prominent models rely on one (or an interaction) of the following building blocks: multiple behavioral biases, information asymmetry and a long-lasting exogenous belief distortion (i.e. there is limited learning even after a long history of signals). Barberis et al. (1998) propose a model in which agents suffer from two types of biases – the representativeness heuristic and the conservatism bias to generate both over- and underreaction. Hong and Stein (1999) build a model with two types of investors – news watchers and momentum traders – and with asymmetric information. Daniel et al. (1998) also rely on the interaction of two types of behavioral biases: overconfidence and the self-attribution bias. Our results indicate that both over- and underreaction to information can be generated by a single belief distortion. In both our lab and field data, we show that individuals generally underinfer when they observe information signals that confirm a history of prior signals (e.g. observing a positive earnings surprise when prior earnings announcements of the same company were also associated with positive surprises). However, whenever individuals observe extreme signals that are at odds with the history of prior signals, agents generally overinfer. One potential mechanism underlying this belief distortion is that agents must selectively allocate their resources (with extreme events being associated with higher mental costs) despite having access to the same information.

Finally, we contribute to the literature on stock return reactions to earnings announcements. One of the most puzzling anomalies in financial economics is the underreaction to earnings announcements resulting in the so-called post-earnings announcement drift (Bernard & Thomas, 1989, 1990). While an extensive body of literature consensually reports that stock prices appear to drift after major news announcements, the reasons for its occurrence are controversial and challenge market efficiency. Even though recent studies find an attenuated drift for large-cap US stocks (Martineau, 2019; Richardson et al., 2010), it remains puzzling why the effect has remained or even still remains so robust across many countries and over time (Hung et al., 2015). Motivated by the updating behavior observed in various experiments, we propose an interesting new dimension which affects the initial price sensitivity towards an announcement and consequently the intensity of a price drift. This is how the direction of an earnings announcement surprise relates to the direction of the prior earnings announcement surprise. Per se, this information should not have any impact on prices given that in a well-functioning efficient market all relevant information should have already been priced in. However, we find that indeed investors tend to incorporate new positive information in a more comprehensive manner when it is contrary to the prior information, resulting in less underreaction and a smaller post-announcement drift. As such, we show that underreactions to earnings news and thus price drifts are most prevalent in situations in which new information points in the same direction than prior information.

Our paper is structured as follows. In Section 1, we present an empirical framework, briefly review the existing literature and state our hypotheses. In Section 2, we describe the experimental design and present our experimental results. In Section 3, we turn to an application in financial markets, describe our data, empirical methodology and present results. Finally, in Section 4 we conclude.

1. Empirical Framework and Hypotheses

In this section, we describe the framework which serves as a basis for our hypotheses as well as the later empirical analyses and then relate the existing literature to our established framework. Suppose there is an agent who wants to learn about the quality of a risky asset. The risky asset can either be in a good or bad state. Over a number of periods, the agent may receive good (+) or bad (-) signals from which he can learn about the quality of the risky asset. This framework of how the agent's beliefs about the asset being in the good state should evolve can best be illustrated using the following graph.

[INSERT FIGURE 1 ABOUT HERE]

Figure 1 illustrates three phases of how Bayesian beliefs evolve over a sequence of four outcomes. The first phase ("confirming signals") resembles a sequence of same-directed signals. A signal which (i) confirms the underlying distribution and (ii) follows another same-directed signal will be referred to as a *confirming signal.* Thus, if a signal is to be referred as a confirming signal, an agent must have observed at least two signals. The second phase ("disruptive signal") defines the situation when a sequence of confirming signals (phase 1) is disrupted by a signal of opposite direction than the previously observed signal. A signal which disrupts a sequence of same-directed signals will be referred to as a *disconfirming signal*. The third phase ("correction") resembles the case when a previously observed disconfirming signal is reverted. A signal which follows on a disconfirming signal and has the opposite direction than the previously-observed disconfirming signal is referred to as a *correction*.

In our framework with binary information signals, an agent should update his prior beliefs according to the following formula:

$$P_t^{Bayes} = P(G|\delta_t)^{Bayes} = \frac{\theta^{\delta_t}}{\theta^{\delta_t} + (1-\theta)^{\delta_t}}, \quad \delta_t = g_t - b_t \qquad (1)$$

where P_t^{Bayes} is the posterior probability that the risky asset pays from the good distribution (G) and θ refers to the diagnosticity of the good signal. The number of good signals observed until period t is referred to as g_t, while the number of bad signals observed until period t is referred to as b_t.

Applying the formula to our described framework from Figure 1 provides several implications on how agents should update their beliefs. Overall, note that the Bayesian agent in our setting is indifferent regarding the order of the signals, since only the difference δ_t is relevant. This feature of the described framework has implications which are especially relevant for the second and the third phase in Figure 1. For the second phase this implies that an agent should reduce the probability estimate after a disconfirming signal by the same magnitude than he increased it after the previous confirming signal. In other words, a Bayesian agent would report the same probability estimate than he did two signals ago. As such he simply cancels the previously observed confirming signal. Referring to the framework in Figure 1, the Bayesian agent would state the same probability estimate as he did after observing the first positive signal. For the third phase, a similar logic applies. In particular, after observing a correction (i.e. the reversion of the disconfirming signal) agents should also only cancel the previously observed disconfirming signal and should again, end up with the same probability estimate as they did two signals ago. In both scenarios (disruption and correction), a Bayesian agent would follow a counting heuristic which means that one positive and one negative signal simply cancel out.

In contrast, agents in the first phase cannot rely on a simple counting heuristic in determining the precise probability estimate. That means after observing two same-directional signals, the counting heuristic does not provide any insight by how much they need to adjust the prior estimate. In other words, to state the correct magnitude of the change in probability estimate, the agent needs to know Bayes' Rule.

Based on the established framework, we formulate the following hypotheses[2]:

[2] The hypotheses are formulated for the good distribution. In the bad distribution, subjects should adjust their priors in the opposite direction.

Hypothesis H1: Disruption (Phase 2)

After observing a disconfirming signal, an agent should reduce his prior probability estimate by the same magnitude than he increased it after the previous confirming signal.

Hypothesis H2: Correction (Phase 3)

After a previous disconfirming signal got reverted, an agent should cancel the previously observed disconfirming signal and end up with the same probability estimate as he did two signals ago.

It is important to stress that our framework and the later experimental design do not crucially depend on agents being Bayesian. Instead, it is sufficient for agents to know that two directionally inconsistent signals cancel each other out. In other words, for the basic updating rule we are testing, it is not essential that agents state the correct *absolute* Bayes estimate. We are rather interested in the *changes* in probability estimates after subjects incorporate new signals into their prior beliefs.

As discussed, Bayes Theorem provides clear and testable predictions on how individuals should revise their beliefs after a sequence of confirming signals is interrupted by a single disconfirming signal as well as after its subsequent reversal (i.e. correction). While this is perfect normative advice, the literature on probabilistic reasoning has identified various situations in which individuals systematically deviate from Bayes and either over- or underinfer. Using bookbag-and-poker-chip experiments, some studies find underinference when a new signal confirms the prior hypothesis and no or only very little revision of beliefs when a new signal disconfirms the prior hypothesis, consistent with prior-biased inference (Pitz et al., 1976; Geller & Pitz, 1968; Pitz, 1969). In contrast to this, Du Charme and Peterson (1968) observe in experiments with normally distributed signals overinference in response to a disconfirming signal. However, Eil and Rao (2011) as well as Möbius et al. (2014) find no evidence for prior-biased inference at all. Recently, Charness and Dave (2017) establish a conceptual framework which combines both under- and overinference and test it experimentally. They find prior-biased inference. In particular, they observe overinference after a confirming signal in updating problems with equal prior probabilities of the states and high diagnosticity of 70%. However, and opposing to Charness and Dave (2017), Pitz et al. (1967) find for the identical level of diagnosticity underinference after a confirming signal. In brief, while there are several studies showing that individuals deviate from Bayes, the evidence *in which way* and *when* they deviate is mixed and apparently inconsistent.

2. Experimental Evidence

2.1 Experimental Design

One-thousand-eight-hundred-and-seven individuals (1159 males, 648 females, mean age 34 years, 10 years standard deviation) were recruited from Amazon Mechanical Turk (MTurk) to participate in three online experiments. MTurk advanced to a widely used and accepted recruiting platform for economic experiments. Not only does it offer a large and diverse subject pool compared to lab studies (which frequently rely on students), but it also provides a response quality similar to that of other subject pools (Buhrmester et al., 2011; Goodman et al., 2013).

An environment to study the role of disconfirming information signals requires (i) a sequential set-up with room for subjective belief formation, (ii) control over Bayesian beliefs, (iii) variation in the number of confirming signals prior to a disconfirming signal, and (iv) an incentive-compatible belief elicitation. Our design accommodates all of these features.

A. Baseline Design

To study the role of disconfirming information signals, we provide subjects with information about a risky asset (similar to Grether, 1980). In all of our experiments, the risky asset has an initial value of 50 which either increases or decreases over the course of six periods depending on the asset's payoffs. The payoffs are either drawn from a "good distribution" or from a "bad distribution". Both distributions are binary with a high outcome of +5 and a low outcome of –5. In the good distribution, the higher payoff occurs with 70 % probability while the lower payoff occurs with 30 % probability. In the bad distribution, the probabilities are reversed, i.e. the lower payoff occurs with 70 % probability while the higher payoff occurs with 30 % probability.

Since we only focus on a single disconfirming signal within six periods, we differentiate between six possible price paths per distribution. These price paths resemble our treatments. The first treatment dimension depicts the underlying distribution and therefore the domain (good or bad), while the second treatment dimension depicts the period in which the disconfirming signal occurs (from period one to period six). Table 1 provides an overview of all twelve treatments.

[INSERT TABLE 1 ABOUT HERE]

For example, in treatment G-3, the risky asset pays from the good distribution and the disconfirming signal appears in period three after two confirming signals (i.e. the sequence would be: positive, positive, negative, positive, … signal). A key feature of our design is that we shift the single disconfirming signal between a sequence of six signals. That allows us to test how subjects update their beliefs after observing a single disruptive, disconfirming signal conditional on the number of previously observed confirming signals. Additionally, the design makes it possible to investigate how subjects update their beliefs after the disconfirming signal is reverted.

Across all experiments, subjects make forecasting decisions in six consecutive periods. At the beginning of the experiment, the computer randomly determines the distribution of the risky asset (which can be good or bad) and the period in which the disconfirming signal will occur (which can be from one to six). In each of the six rounds, subjects observe a payoff of the risky asset. After each round, we ask them to provide a probability estimate that the risky asset draws from the good distribution and how confident they are about their estimate. To keep the focus on the forecasting task and to not test their memory performance, we display the prior outcomes in a price-line-chart next to the questions. To ensure that subjects have a sufficient understanding of the forecasting task, they had to correctly answer four comprehension questions before they could continue (see Appendix A).

The experiment concluded with a brief survey about subjects' socio-economic background, self-assessed statistic skills, as well as a measure of risk preferences and financial literacy adopted from Kuhnen (2015). Subjects' belief elicitation was incentivized. Participants were paid a participation fee and a variable fee based on the accuracy of the probability estimate provided. Specifically, they received 25 cents for each probability estimate within 10 % (+/-5%) of the objective Bayesian value. Across all studies, it took participants approximately 7 minutes to complete the experiment and participants earned \$1.50 on average.

B. Experimental Variations

We conducted two variations of our baseline experiment, referred to as *Reduced Diagnosticity* and *Reduced Uncertainty*. The two additional experiments are designed to identify whether the belief updating after a disconfirming signal depends on (i) the diagnosticity of the signal (i.e. its informational content), (ii) subjects' uncertainty about the distribution (i.e. whether the asset turns out to be good or bad), and (iii) whether subjects do not anticipate the disconfirming signal (i.e. are surprised about the disruption of the sequence of confirming signals).

Experiment Reduced Diagnosticity: In the experiment *Reduced Diagnosticity* we change the informational content that subjects can infer from a positive signal. This means, we change the probability of the higher outcome in the good distribution from 70 % to 60 % and of the lower outcome from 30 % to 40 %, respectively. In the bad distribution, we change the probability of the lower outcome from 70% to 60% and of the higher outcome from 30 % to 40 %, respectively. On the one hand, we expected to observe – as Bayes' Theorem implies – lower (higher) absolute levels of probability estimates in the good (bad) distribution given the reduced diagnosticity of signals. On the other hand, we expect to observe no impact of diagnosticity on the fundamental counting rule we are testing. Within our empirical framework, the increase (decrease) in posterior probability after a confirming signal in the good (bad) distribution should remain exactly as much as the decrease (increase) in posterior probability after a subsequent disconfirming signal, irrespective of how informative the signal is.

Experiment Reduced Uncertainty: In the experiment *Reduced Uncertainty* we combine aspects (ii) and (iii) from above. To do so, we change the previously framed forward-looking updating task to a backward-looking updating task. In detail, subjects in the baseline experiment are asked to make a forecasting decision without knowing the future outcome history. In the *Reduced Uncertainty* experiment, we show subjects the full outcome history beforehand. Then, we ask them to provide probability estimates period by period as in the baseline experiment for exactly the same outcome history they have seen in advance. Importantly, subjects were still incentivized to provide probability forecasts which only incorporate the information subjects had in each period. In other words, the objective Bayesian probabilities are identical to the baseline experiment. By showing subjects the entire outcome history beforehand, we already eliminate most of the uncertainty regarding the underlying distribution and any of the potential surprise related to the period in which the disruptive signal occurs. Additionally, before the first period, we directly ask subjects two questions: (i) we ask them to count the number of positive and the number of negative payoffs in the outcome history and (ii) we ask them to state the period in which the disconfirming signal occurs.

C. Demographics

Table 2 presents summary statistics for all our three experiments. Overall 1807 subjects participated in our studies, with an average age of 33.79 years in Experiment 1 (33.59 years in Experiment 2, and 35.01 years in Experiment 3). Thirty-five percent (forty-one percent, thirty-two percent) were female. Subjects reported average statistic skills of 4.46 out of 7 (4.42, 4.42) and their level of risk aversion, measured by how much of an endowment of 10,000 they are willing to invest risky in a broad equity index, is as follows. Subjects invest on average 4,470 (4,420, 5,000) in the risky asset. Across all experiments subjects report medium financial literacy. In particular, they make 1.73 (1.70, 1.70) out of three possible basic errors.

2.2 Experimental Results

2.2.1 Main Results

In this section, we first present results of our baseline experiment of how individuals update their beliefs after disconfirming signals as well as of how they revise their probability estimates after a correction. Then, we test the robustness of our findings with respect to the diagnosticity of the information signals and finally examine how the reduction of uncertainty with respect to the underlying distribution affects subjects' updating behavior.

A. Baseline Experiment

Figure 2 presents subjects' average updating tendency over all periods for each treatment of our baseline experiment individually. Panel A shows the results of those treatments in which the underlying distribution is *good* and Panel B shows the results of those treatments in which the underlying distribution is *bad*.

[INSERT FIGURE 2 ABOUT HERE]

To be consistent with our framework in Section 2, we focus our analysis on the treatments in which subjects observe at least two subsequent same-directional signals before a disconfirming signal occurs. This is the case for our treatments G-3, G-4, G-5, and G-6 (B-3, B-4, B-5, and B-6). We will analyze the results of treatments G-1 and G-2 (B-1 and B-2) in a separate section at the end of this chapter. From Figure 2, we observe that subjects in the good distribution increase their prior beliefs by 6.44 % on average after a confirming signal, whereas they decrease their prior beliefs by 18.63 % on average after observing a disconfirming signal. In the bad distribution, the findings look similar. Subjects decrease their prior beliefs by 5.38 % on average after a confirming signal, while they increase their prior beliefs by 16.94 % on average after observing a disconfirming signal. In relative terms, this means that subjects in the good distribution update their prior beliefs after a disconfirming signal with a magnitude that is approximately three times as large as if they update after a confirming signal. This ratio is more or less independent of the distribution, albeit a little bit stronger in the bad distribution. Given the difference in updating behavior, Figure 2 suggest that subjects strongly overreact after a disconfirming signal. In particular, subjects update their beliefs after a disconfirming signal as if they failed to incorporate up to three previously observed confirming signals.

Next, we investigate how individuals update their prior beliefs after a disconfirming signal gets reverted. In particular, we examine whether and to what extent subjects correct the observed overreaction after a disconfirming signal. We find that subjects in the good distribution increase their probability estimate on average by 17.11 %. Similarly, in the bad distribution, subjects decrease their probability estimates on average by 14.16 %. In essence, the previously observed overreaction after a disconfirming signal is almost entirely corrected. This finding holds independent of the distribution.

From these descriptive statistics alone, it becomes already evident that subjects fail to follow a simple counting heuristic when they incorporate inconsistent signals in their beliefs. In other words, they do not adhere to the simple updating rule in which they count the difference between positive and negative signals. Instead, they strongly overreact after a disconfirming signal. Interestingly however, this is not the case, if an inconsistent (i.e. disconfirming) signal is reverted. Then, subjects appear to follow the counting heuristic implied by Bayes' Rule and fully correct their prior overreaction.

Besides the descriptive analysis, we also run regressions, in which we can control for the objective posterior probability. To investigate how individuals update their prior beliefs both in response to disconfirming signals and subsequent confirming signals (i.e. the correction of the disconfirming signal), we estimate the following model[3]:

$$\Delta p_{i,t} = \beta_1 \Delta Objective\ Prior_{i,t} + \beta_2 Disconfirm_{i,t} + \beta_3 Correction_{i,t} + \varepsilon_{i,t},$$

where $\Delta p_{i,t}$ is the difference in subjects' probability estimates between two subsequent periods and $\Delta Objective\ Prior_{i,t}$ is the difference in the objective Bayesian probability between two subsequent periods. Finally, $Disconfirm_{i,t}$ and $Correction_{i,t}$ are two indicator variables which equal one if subject i observes a disconfirming signal or a correction in period t, respectively. In the above specification we can test both for Bayesian behavior and in which way individuals depart from it. If subjects were perfect Bayesian, we would expect that $\widehat{\beta_1} = 1$, and $\widehat{\beta_2} = \widehat{\beta_3} = 0$. In other words, subjects always update their prior beliefs according to Bayes Rule, while neither a disconfirming signal (which disrupts a sequence of confirming signals) nor a subsequent correction would explain any additional variation. Conversely, $\widehat{\beta_1} < (>)1$, $\widehat{\beta_2} < (>)0$, and $\widehat{\beta_3} < (>)0$ would signal underinference (overinference), to subsequent confirming signals, to disconfirming signals, and to corrections, respectively. Results are reported in Table 3.

[INSERT TABLE 3 ABOUT HERE]

The findings support our previously drawn conclusions. Even after controlling for the objective posterior, we find an economically strong and highly statistically significant overreaction after a disconfirming signal. Additionally, we find that the initial overreaction is almost entirely corrected if the disconfirming signal is reverted. While in the bad distribution, both effects are of similar magnitude and thus cancel out, we find a slightly asymmetric effect in the good distribution. Whereas the correction is of similar strength as in the bad distribution, the overreaction is stronger. As such the overreaction in the good distribution is not entirely corrected.

Next, we examine how our model in which we explicitly control for a disconfirming signal and a subsequent correction performs compared to the standard Bayes model. When

[3] Since we investigate changes in subjective probability estimates, we estimate the model without constant to be consistent with the theoretical benchmark. However, results are qualitatively similar if we estimate the model on levels or with constant. For the ease of interpretation, we report the specification without constant.

comparing the explanatory power of the two models, we find that the standard Bayesian model explains roughly 14% (10%) in the good (bad) distribution, while our model explains roughly 22% (14%). Irrespective of the distribution, our model explains roughly 50% more of the variation of subjects' probability estimates than the standard Bayesian model.

Moreover, Table 3 implies that subjects generally underinfer which is consistent with several studies on Bayesian updating (see Benjamin, 2019). Interestingly, our results suggest that the observed underinference is mostly driven by subsequent confirming signals. When differentiating between the good and the bad distribution, we find that the observed underinference is stronger when subjects update their beliefs from a sequence of confirming bad signals than when updating their beliefs from a sequence of confirming good signals. This finding is consistent with the recently identified good news-bad news effect reported by Eil and Rao (2011) as well as Möbius et al. (2014). However, for our main finding, it remains to stress that we do not find such an asymmetric effect across domains.

B. Reducing the Diagnosticity of Information Signals

In this section, we report results of our second experiment in which we vary the informational content of the signals. Like in our baseline experiment, Figure 3 presents subjects' general updating behavior over all periods for each treatment individually. Panel A shows the results of those treatments in which the underlying distribution is *good* and Panel B shows the results of those treatments in which the underlying distribution is *bad*.

[INSERT FIGURE 3 ABOUT HERE]

Overall, the findings look very similar to our baseline experiment. In particular, we find that subjects in the good distribution increase their prior beliefs by 7.15 % on average after a confirming signal, whereas they decrease their prior beliefs by 14.81 % on average after observing a disconfirming signal. In the bad distribution, subjects decrease their prior beliefs by 3.65 % on average after a confirming signal, while they increase their prior beliefs by 7.15 % on average after observing a disconfirming signal. Like in our baseline experiment, subjects update their beliefs after a disconfirming signal as if they failed to incorporate up to three previously observed confirming signals. Despite the lower diagnosticity in the second experiment, the observed overreaction after a disconfirming signal persists.

This finding even holds after controlling for the objective Bayesian probability as to be seen in Table 4. The observed overreaction after a disconfirming signal remains economically large and statistically significant. In comparison to the results from our baseline experiment, the magnitude with which subjects update their prior after a disconfirming signal is smaller. However, this is to be expected since the updating magnitude strongly correlates with the diagnosticity. Consistent with our previous findings, we find that subjects correct their priors after a disconfirming signal is reverted. Interestingly, we find that in contrast to the baseline experiment, subjects seem to not sufficiently correct their previous overreaction which can especially be seen in the bad distribution. Overall, even in a setting with lower diagnosticity subjects still do not follow the simple counting heuristic when observing a disconfirming signal. Instead, they show a strong overreaction which they partly correct subsequently.

[INSERT TABLE 4 ABOUT HERE]

C. Reducing the Uncertainty About the Underlying Distribution

In the following, we discuss the results of our third experiment in which we reduce subjects' uncertainty about the underlying distribution. This variation of the design allows us to exclude the possibility that subjects falsely infer trends or price reversal. Additionally, we control for the possibility that subjects do not anticipate (i.e. are surprised by) the disconfirming signal as they observe the full outcome history in advance. Results on individuals' updating behavior are reported in Figure 4. Again, Panel A shows the results of those treatments in which the underlying distribution is *good* and Panel B shows the results of those treatments in which the underlying distribution is *bad*.

[INSERT FIGURE 4 ABOUT HERE]

We find that both, overreaction after a disconfirming signal and subsequent correction even persist in a setting in which the uncertainty about the underlying distribution is dramatically reduced. In particular, the Bayesian probability of the asset being in the good distribution is 96.74 %. As such after subjects observe the full outcome history there should be barely any uncertainty left about the distribution. Besides almost no uncertainty about the underlying distribution, there is also no uncertainty about the period in which the disconfirming signal will occur. First, the graphical representation of the full outcome history in the form of a price chart is known to subjects and makes the period in which the disconfirming signal occurs easily identifiable. Second, we also explicitly ask participants to state the period in which the disconfirming signal occurs prior to the forecasting task. As such our design should eliminate any potential surprise subjects may experience when observing a disconfirming signal. In the light of the still persistent overreaction, we can confidentially rule out that surprise effects or uncertainty about the underlying distribution drive the results. Moreover, we can also exclude that subjects overreact after a disconfirming signal because they potentially anticipate a new trend, given that they know that a disconfirming signal will subsequently be reverted.

We run the same regression as previously to control for the objective Bayesian posterior probability, while also investigating potential differences to the baseline experiment. The results are reported in Table 5.

[INSERT TABLE 5 ABOUT HERE]

A direct comparability is given as Bayes probabilities are identical across treatments in the baseline and the reduced uncertainty experiment. First, we can confirm all prior findings. Subjects strongly overreact after a disconfirming signal and subsequently correct the overreaction. Second, when comparing the effect sizes between the two experiments, we find that the overreaction as well as the subsequent correction are slightly more pronounced in the baseline treatment. Even though the reduced uncertainty experiment was designed to significantly decrease the overreaction resulting from disconfirming signals, the effect is still economically strong and statistically significant.

D. Additional Treatments G-1 and G-2 (B-1 and B-2)

Finally, we analyze the results of treatments G-1 and G-2 (B-1 and B-2) for which – per definition – our empirical framework does not apply. In these treatments, the single opposite-directional signal occurs either directly in the first period or in the second period. As such these treatments describe price paths for which the pre-requisite for phase 1 of our framework (i.e. at least two confirming signals prior to the disconfirming signal) is not fulfilled. Nevertheless, they allow us to analyze how subjects update their beliefs (i) in situations without prior outcome history (G-1 and B-1) and (ii) in situations with exclusively alternating signals (G-2 and B-2).

Figure 5 reports the results split by distribution. Across all experiments, we find that subjects do not significantly update their beliefs downwards if the first signal is bad.[4] In contrast to that, subjects significantly update their beliefs upwards if the first signal is good. Their first probability estimate is almost identical to the objective Bayesian probability and this finding holds for both, the two experiments with high diagnosticity (70 %) and the experiment with low diagnosticity (60 %). In period 2, when the bad signal of period 1 is reverted, subjects state probability estimates significantly above the objective probability of 50 %, while when the good signal of period 1 is reverted, subjects are almost perfect Bayesian. In other words, subjects in the B-1 treatment almost perfectly adhere to the investigated counting rule implied by Bayes' Theorem, while subjects in the G-1 treatment clearly violate this rule. In particular, they seem to violate this rule because they ignored or were averse to adjust their beliefs downwards following the first bad signal.

[INSERT FIGURE 5 ABOUT HERE]

This pattern is mirrored when looking at the treatments G-2 and B-2. In these treatments, the signals alternate up until period 3. Subjects, who observe first a good, second a bad, and then again a good signal, are almost perfect Bayesian. Across all experiments, they follow the counting rule and increase their probability estimate after the good signal in period 3 as much as they decreased it after the bad signal in period 2 which in turn they previously increased exactly as much as after the good signal in period 1. In contrast to that, subjects who first

[4] We follow the terminology used in the empirical framework section and also refer to a bad signal in the first period drawn from an asset with a good distribution as a disconfirming signal, even though subjects cannot know at this point in time that the signal disconfirms the true underlying distribution. The same logic applies to a good signal in the first period drawn from the good distribution which we refer to as a confirming signal.

observe a bad, second a good, and then again a bad signal do only partly follow the counting rule. Like subjects in the G-1 treatment, they do not significantly adjust the probability estimate downwards if the first signal is bad, but correctly – as implied by the counting rule – decrease their probability estimate in period 3 by the amount by which they previously increased it in period 2. This robust pattern can be found across all experiments.

Taken together, we can complement our findings from treatments G-3 to G-6 (B-3 to B-6) as follows: We find that subjects adhere to the counting rule implied by Bayes' Theorem in situations with no prior sequence of same-directional signals and in situations with exclusively alternating signals. Interestingly however, subjects seem to have problems following this rule right at the beginning of the updating task, when the first signal is bad. In these cases, they act as if they ignore the bad signal and consequently update too much after the subsequent good signal.

2.2.2 Signal Ordering

One aspect of the counting heuristic we have not discussed so far is that Equation 1 of the established framework also implies that a Bayesian is indifferent regarding the order in which outcomes occur. In other words, observing a disconfirming signal followed by five subsequent confirming signals should lead to the same posterior probability as first observing five subsequent confirming signals followed by a disconfirming signal. Since our experimental design explicitly varies the round in which the single disconfirming signal occurs, we can directly test this relation. To do so, we estimate the following model:

$$P_{i,6} = \beta_0 + \beta_1 D_{i\,|\,Round=2} + \beta_2 D_{i\,|\,Round=3} + \beta_3 D_{i\,|\,Round=4} + \beta_4 D_{i\,|\,Round=5} + \beta_5 D_{i\,|\,Round=6} + \varepsilon_{i,t},$$

where $P_{i,6}$ is the subjective posterior in round 6, and $D_{i\,|\,Round=t}$ are indicator variables denoting the round in which participants encountered the disconfirming signal (with round 1 being the baseline category). Note that the Bayesian posterior in our setting is the same for each treatment and only depends on the underlying distribution (good or bad) and the underlying diagnosticity. To accommodate this feature, we estimate the model separately for each distribution and split by diagnosticity of the signal. Results are reported in Table 6.

[INSERT TABLE 6 ABOUT HERE]

We find that the round in which the disconfirming signal occurs plays an important role in how individuals form their posterior beliefs. In particular, the later the disconfirming signal occurs, the stronger the overreaction which ultimately leads to a lower final posterior after round 6. This result holds independent of the underlying distribution and is of similar magnitude across different diagnosticities. One potential driver of this further inconsistency is that individuals generally overreact after disconfirming signals, which is mostly corrected after subsequently observing another confirming signal. However, if subjects observe the disconfirming signal in the final period (where the objective prior in the good distribution is as high as 96.74 %!) subjects can no longer correct their strong overreaction, causing them to be substantially more pessimistic (or optimistic if the underlying distribution is the bad one) about the underlying distribution than they should be. This relation can be especially seen by the considerably higher coefficients of the disconfirming dummy for round 6.

Overall, this result highlights once more the fact that individuals consistently violate the counting heuristic after they encounter disconfirming signals. However, whereas they mostly correct their strong overreaction if they can, the violation is most severe when subjects have no opportunity to collect further information.

2.2.3 Robustness Checks

In this section we will replicate our main analyses on different subsamples to validate its robustness against extreme outliers or individuals who are inattentive and as such more likely to suffer from a bias in probabilistic reasoning. Besides validating the robustness of our main finding, such an analysis might also provide valuable insights into which subgroup is most likely to violate the counting heuristic.

In particular, we conduct splits regarding (i) extreme outliers; (ii) ”speeders”; and (iii) below median forecasters. Extreme outliers are individuals whose subjective priors largely deviate from the Bayesian benchmark. Following the classification of Enke and Graeber (2019), we define extreme outliers as individuals who report a subjective posterior $p_S < 25\ \%$ ($> 75\ \%$) when the Bayesian posterior is $p_B > 75\ \%$ ($< 25\ \%$). Speeders are defined as subjects who are in the bottom decile of the response time distribution. Finally, we also investigate whether the here documented effect is only driven by individuals who lack the statistical skills to correctly perform the forecasting task, or whether even individuals who are closer to

Bayesian behavior exhibit a pronounced bias. To examine this relation, we define the squared deviation of subjects' probability estimate in each period from the objective posterior probability as a measure of forecasting quality and conduct median splits. Results are reported in Table 7. Panel A reproduces the analysis split by extreme outliers, Panel B splits the sample by speeders, and Panel C reports results split by forecasting ability.

[INSERT TABLE 7 ABOUT HERE]

Overall, results are very similar, with two sets of results warrant a brief discussion. First, throughout each subsample, we consistently find an economically strong and statistically significant overreaction following a disconfirming signal, which is mostly corrected after observing a subsequent confirming signal. While the overreaction is even more pronounced for outliers and individuals with below-median forecasting ability, it is mostly unaffected by individuals' response time. This suggest that systematic violations of the counting heuristic appear to be a general phenomenon even though they correlate with participants' statistical skills. Yet, given that response time does not play a major role, attention does not appear to be a major driver. Second, when splitting the sample by extreme outliers, it becomes apparent that outliers are mostly clustered in the bad distribution. This confirms our previous finding, that a greater fraction of individuals struggles to forecast the bad distribution, even though both tasks should be – at least from a Bayesian perspective – equivalent.

3. From the Laboratory to the Field: Confirming and Contrary Earnings Surprises and Stock Return

Results from the laboratory show a substantial discrepancy between how people should incorporate an opposite-directed signal following a sequence of same-directed signals and how they actually update their beliefs in these settings. Based on this insight, we turn to a real-case application in financial markets, that is stock return reactions to earnings announcements, and test whether the sequence of earnings surprises affects asset prices. In particular, we examine how the direction of the prior earnings surprise – same- or opposite-directed – relative to the current surprise affects announcement and post-announcement stock returns. Applying our experimental findings to financial markets, we expect to observe a stronger announcement stock return reaction if the earnings surprise is opposite-directed to the previous earnings surprise than if it is same-directed. Moreover, and given the stronger initial reaction, we assume a weaker post-announcement drift following opposite-directed earnings surprises.

3.1 Data and Methodology

We obtain data for daily stock returns, market value-weighted index returns, daily stock prices (adjusted for splits and dividends), and daily trading volumes from the CRSP-Compustat merged database from June 2009 to May 2020. Additionally, we use quarterly earnings announcement data and analyst earnings forecasts from the union of the CRSP-Compustat merged database and I/B/E/S for the same period. To clearly identify the earnings announcement dates, we compare the earnings announcements available from the CRSP-Compustat merged database with the dates available at I/B/E/S. When the announcement dates differ, we follow DellaVigna and Pollet (2009) and take the earlier date. Our primary sample of 53,168 firm-quarter observations consists of all firm quarters from June 2009 to May 2020 and 1,456 unique firms with sufficient data to estimate the required variables (i.e. current earnings surprise and future abnormal returns) as described below.

The signals we are investigating are firms' quarterly earnings surprises. Following Hirshleifer et al. (2009), we measure an earnings surprise as the difference between announced earnings as reported by I/B/E/S and the consensus earnings forecast, defined as the median of the most recent forecasts from individual analysts using the I/B/E/S detail tape. We normalize the difference by the stock price at the end of the corresponding quarter.

For our main analysis, we classify earnings surprises in either contrary or confirming signals depending on how the sign of an earnings surprise proceeds the sign of the previous earnings surprise. More precisely, an earnings surprise is defined as a contrary signal if its sign is opposite to the sign of the previous earnings surprise (e.g. a positive earnings surprise is followed by a negative earnings surprise). We define an earnings surprise as a confirming signal if its sign is the same as the sign of the previous earnings surprise (e.g. a positive earnings surprise is followed by a positive earnings surprise).

Our dependent variable is the cumulative abnormal return over two windows, the announcement window [0,1] and the post-announcement window [2,61] measured in trading days relative to the announcement date. We define the cumulative abnormal return as the difference between the buy-and-hold return of the announcing firm in quarter q and the market return in quarter q,

$$CAR[0,1]_{iq} = \prod_{k=t}^{t+1}(1 + r_{ik}) - \prod_{k=t}^{t+1}(1 + r_m)$$

$$CAR[2,61]_{iq} = \prod_{k=t+2}^{t+61}(1 + r_{ik}) - \prod_{k=t+2}^{t+61}(1 + r_m),$$

where r_{ik} is the return of the firm i and r_m is the return of the value-weighted market portfolio on day k, where t is the announcement date of quarter q's earnings.

3.2 Empirical Results

We first perform univariate analyses to examine the effect of contrary versus confirming earnings surprises on (i) announcement date returns and (ii) post-earnings announcement date returns. In each quarter, we perform a two-way independent sort of quarterly earnings surprises in that quarter into 2 x 5 = 10 groups based on the sign of the earnings surprise relative to the previous surprise (i.e. contrary versus confirming) and the magnitude of the earnings surprise (Quintile 1: most negative earnings surprise, Quintile 5: most positive earnings surprise).[5] For each group of confirming and contrary earnings surprises, we calculate the average cumulative abnormal return over the announcement and the post-announcement window for the most negative earnings surprise quintile and the most positive earnings surprise quintile and the difference in cumulative abnormal returns between the two extreme earnings surprise quintiles.

The difference in abnormal announcement day returns between earnings surprise quintile 5 and 1 measures the stock price reaction to earnings news, whereby a larger difference indicates that investors react more strongly to earnings news on the announcement day. The difference in post-announcement abnormal returns between earnings surprise quintile 5 and 1 measures underreaction to earnings news as reflected in a subsequent drift. Our experimental results predict a stronger CAR[0,1] difference (stronger announcement-day reaction) and a weaker CAR[2,61] difference (weaker post-earnings announcement drift) for contrary earnings surprises compared to confirming earnings surprises.

[5] We also perform sorts based on earnings surprise deciles instead of quintiles. The results are similar and reported in the Appendix.

[INSERT FIGURE 6 ABOUT HERE]

Figure 6 illustrates the development of the average cumulative abnormal returns over 61 days after the earnings announcement for the top and the bottom earnings surprise quintile conditional on whether the earnings surprise was confirming (i.e. same-directed relative to the previous earnings surprise of the firm) or contrary (i.e. opposite-directed relative to the previous earnings surprise of the firm). In line with the literature, we observe that stock prices adjust strongly in the direction of the earnings surprise immediately after the announcement. That is, after a positive earnings surprise stock prices increase, whereas after a negative earnings surprise they decrease. However, Figure 6 shows that this initial announcement reaction depends not only on the magnitude of the announcement-day earnings surprise, but also on its sign relative to the previous earnings surprise. Consistent with our experimental findings, we find that initial announcement stock return reactions are significantly stronger for opposite-directed earnings surprises than for same-directed earnings surprises. That is, after *positive* earnings surprises, we observe on average stronger return reactions if the surprise follows a prior *negative* earnings surprise, compared to when it follows a prior *positive* earnings surprise. The same holds true for negative earnings surprises that follow a prior positive earnings surprise. In that case, the initial stock return reaction is almost twice the size compared to the case if the same negative earnings surprise follows a prior same-directed, i.e. negative, earnings surprise.

Going forward, Figure 6 demonstrates that this larger initial reaction of opposite-directed earnings surprises reverts in the long-run. In particular, we observe for the top earnings surprise quintile that the cumulative abnormal return following the contrary earnings surprise converges to the cumulative abnormal return following the confirming earnings surprise. Approximately 42 days after the announcement, they become statistically insignificant from one another. While stocks in both, the contrary and confirming top surprise quintile clearly drift upwards over the 60 days following the announcement date, the post-earnings announcement drift is stronger for those stocks that belong to the confirming earnings surprise group. In particular, there seems to be a slight reversal between day 50 and 60 following the announcement date for stocks that have experienced a contrary positive earnings surprise. For negative surprises, we observe a similar pattern, albeit with a slower convergence. Remarkably, however, the initial difference in stock return reactions between contrary and confirming surprises in the bottom surprise quintile carries forward and is not fully corrected even 60 days after the initial earnings announcement.

[INSERT TABLE 8 ABOUT HERE]

In Table 8, we provide a more detailed overview of the announcement and post-announcement abnormal returns following confirming and disconfirming earnings surprises. Consistent with our prior conjecture, we show that investors' 2-day announcement reactions to earnings surprises are more sensitive following announcements that are of opposite direction. While on average cumulative abnormal returns for the top surprise quintile are around 3.15% if they follow a previous same-directed earnings surprise, they are around 4.20% if they follow a previous opposite-directed earnings surprise, resulting in a difference in abnormal returns of 1.05% ($p<0.01$). For the bottom surprise quintile, the difference is even larger with 1.77% ($p<0.01$). Announcement abnormal returns for earnings surprises that are confirming the previous earnings surprise are around -2.56%, whereas they are around -4.33% if they are contrary to the previous earnings surprise. This indicates that the stock price reactions to earnings announcements are stronger when they are of opposite direction to the prior announcement than when they are of same direction.

For post-announcement abnormal returns, we observe a weak or even absent post-earnings announcement drift, measured as the difference between Quintile 5 and 1, if the earnings surprise is contrary to the previous quarter earnings surprise. However, for confirming earnings surprises, we find a positive post-earnings announcement drift. In this case, the difference in post-announcement abnormal returns between the top and bottom surprise quintile is 1.07% ($p<0.01$). Comparing the post-announcement returns within quintiles between contrary and confirming surprises, we show that there is a stronger same-directional drift for confirming positive surprises than for contrary positive surprises and a weaker opposite-directional drift for confirming negative surprises than for contrary negative surprises.

To back up these findings, we run multivariate regression analyses on the sample containing all observations in the top and bottom earnings surprise quintile. In separate analyses, we regress the CAR[0,1] and the CAR[2,61] on a disconfirming dummy which is one if the earnings surprise is of opposite direction than the prior surprise, a FE 5 dummy that is one if the observation belongs to the top earnings surprise quintile, the interaction between the two variables, and controls. Table 9 reports the regression results. Consistent with our results in Table 8, we confirm that the initial stock return reaction is significantly stronger for contrary surprises than for confirming surprises. This holds after controlling for conventional variables in the announcement drift literature such as for example size, book-to-market, reporting lag, earnings persistence, and share turnover. For the post-announcement return, we find a significant positive drift for the top surprise quintile, but no negative drift for the bottom surprise quintile. Furthermore, we find the tendency of stock price reversion for stocks in the top surprise quintile after contrary earnings surprises.

Overall, our findings are in line with the under- and overreaction patterns observed in our experimental data. On average, investors initially underreact after positive earnings surprises. This underreaction is particularly pronounced for confirming positive surprises, that is if a positive surprise follows a prior positive surprise. In contrast to that, after negative earnings surprises, investors tend to initially overreact. This overreaction is especially pronounced for contrary negative surprises, that is if a negative surprise follows a prior positive surprise. Over the subsequent 60 days, abnormal returns of confirming and contrary surprise stocks converge. This convergence is asymmetric in the sense that in the top surprise quintile it is the *confirming* surprise stocks that drift stronger in the *same* direction of surprise (i.e. correcting the initial stronger *underreaction*), while in the bottom surprise quintile it is the *contrary* surprise stocks that drift stronger in the *opposite* direction of surprise (i.e. correcting the initial stronger *overreaction*).

4 Discussion and Conclusion

The goal of this study is to test whether subjects follow a simple counting heuristic in belief updating as implied by Bayes Rule: two informationally equivalent signals of opposite direction should always cancel out. However, our study suggests that this is not the case. Whenever a sequence of signals that go in the same direction is interrupted by a signal of opposite direction, subjects violate the counting heuristic and strongly overreact to the signal of opposite direction. In contrast to that, subjects correctly follow the counting heuristic whenever opposite-directional signals alternate. Our results show a clear and robust pattern of over- and underreaction following violations of a simple counting heuristic. This pattern does not depend on the diagnosticity of the signals, on individuals' limited memory capacity, on signals not being anticipated, and the uncertainty of the underlying state.

Our findings have relevant implications for various fields of research, among others investors' belief formation and trading behavior in financial markets. One of the most robust paradoxes in the financial literature is the co-existence of both over- and underreaction of prices to new information (Hong & Stein, 1999; Barberis & Shleifer, 2003; Greenwood & Shleifer, 2014; Barberis, Greenwood, Jin & Shleifer, 2015, 2018). This finding has important applications for excess stock market volatility, bubbles, and cross-sectional phenomena of stock returns such as for example momentum and long-term reversal and in particular the post-earnings-announcement drift. The underlying mechanism of when one may observe overreaction and underreaction are – however – still incomprehensively understood. Models which generate both over- and underreaction to stock prices rely on a number of different mechanisms including multiple behavioral biases, information asymmetry and a long-lasting exogenous belief distortion. Our findings from both the lab and field demonstrate that over- and underreaction to information can be generated by a single belief distortion without assuming asymmetric information.

References

Barberis, N., Greenwood, R., Jin, L., & Shleifer, A. (2015). X-CAPM: An extrapolative capital asset pricing model. *Journal of financial economics*, 115(1), 1-24.

Barberis, N., Greenwood, R., Jin, L., & Shleifer, A. (2018). Extrapolation and bubbles. *Journal of Financial Economics*, 129(2), 203-227.

Barberis, N., & Shleifer, A. (2003). Style investing. *Journal of financial Economics*, 68(2), 161-199.

Barberis, N., Shleifer, A., & Vishny, R. (1998). A model of investor sentiment. *Journal of Financial Economics*, 49(3), 307-343.

Benjamin, D. J. (2019). Errors in probabilistic reasoning and judgment biases. *In Handbook of Behavioral Economics: Applications and Foundations 1* (Vol. 2, pp. 69-186). North-Holland.

Bernard, V. L., & Thomas, J. K. (1989). Post-earnings-announcement drift: delayed price response or risk premium?. *Journal of Accounting Research*, 27, 1-36.

Bernard, V. L., & Thomas, J. K. (1990). Evidence that stock prices do not fully reflect the implications of current earnings for future earnings. *Journal of Accounting and Economics*, 13(4), 305-340.

Camerer, C. F. (1987). Do biases in probability judgment matter in markets? Experimental evidence. *The American Economic Review*, 77(5), 981-997.

Camerer, C. (1995). Individual decision making. *Handbook of Experimental Economics.*

Cassella, S., & Gulen, H. (2018). Extrapolation bias and the predictability of stock returns by price-scaled variables. *The Review of Financial Studies*, 31(11), 4345-4397.

Charness, G., & Dave, C. (2017). Confirmation bias with motivated beliefs. *Games and Economic Behavior*, 104, 1-23.

Daniel, K., Hirshleifer, D., & Subrahmanyam, A. (1998). Investor psychology and security market under-and overreactions. *The Journal of Finance*, 53(6), 1839-1885.

De Bondt, W. F., & Thaler, R. (1985). Does the stock market overreact? *The Journal of Finance*, 40(3), 793-805.

DellaVigna, S., & Pollet, J. M. (2009). Investor inattention and Friday earnings announcements. *The Journal of Finance*, 64(2), 709-749.

DuCharme, W. M., & Peterson, C. R. (1968). Intuitive inference about normally distributed populations. *Journal of Experimental Psychology*, 78(2p1), 269.

Enke, B., & Graeber, T. (2019). *Cognitive Uncertainty* (No. w26518). National Bureau of Economic Research.

Eil, D., & Rao, J. M. (2011). The good news-bad news effect: asymmetric processing of objective information about yourself. *American Economic Journal: Microeconomics*, 3(2), 114-38.

Geller, E. S., & Pitz, G. F. (1968). Confidence and decision speed in the revision of opinion. *Organizational Behavior and Human Performance*, 3(2), 190-201.

Goodman, J. K., Cryder, C. E., & Cheema, A. (2013). Data collection in a flat world: The strengths and weaknesses of Mechanical Turk samples. *Journal of Behavioral Decision Making*, 26(3), 213-224.

Gómez, A., Brooks, M. L., Buhrmester, M. D., Vázquez, A., Jetten, J., & Swann Jr, W. B. (2011). On the nature of identity fusion: Insights into the construct and a new measure. *Journal of Personality and Social Psychology*, 100(5), 918.

Greenwood, R., & Shleifer, A. (2014). Expectations of returns and expected returns. *The Review of Financial Studies*, 27(3), 714-746.

Grether, D. M. (1980). Bayes rule as a descriptive model: The representativeness heuristic. *The Quarterly Journal of Economics*, 95(3), 537-557

Hirshleifer, D., Lim, S. S., & Teoh, S. H. (2009). Driven to distraction: Extraneous events and underreaction to earnings news. *The Journal of Finance*, 64(5), 2289-2325.

Hong, H., & Stein, J. C. (1999). A unified theory of underreaction, momentum trading, and overreaction in asset markets. *The Journal of Finance*, 54(6), 2143-2184.

Hung, M., Li, X., & Wang, S. (2015). Post-earnings-announcement drift in global markets: Evidence from an information shock. *The Review of Financial Studies*, 28(4), 1242-1283.

Kuhnen, C. M. (2015). Asymmetric learning from financial information. *The Journal of Finance*, 70(5), 2029-2062.

Martineau, C. (2019). The evolution of financial market efficiency: Evidence from earnings announcements. *Available at SSRN 3111607*.

Möbius, M. M., Niederle, M., Niehaus, P., & Rosenblat, T. S. (2014). Managing self-confidence. *NBER Working paper*.

Pitz, G. F. (1969). The influence of prior probabilities on information seeking and decision-making. *Organizational Behavior and Human Performance*, 4(3), 213-226.

Pitz, G. F., Downing, L., & Reinhold, H. (1967). Sequential effects in the revision of subjective probabilities. *Canadian Journal of Psychology/Revue canadienne de psychologie*, 21(5), 381.

Richardson, S., Tuna, I., & Wysocki, P. (2010). Accounting anomalies and fundamental analysis: A review of recent research advances. *Journal of Accounting and Economics*, 50(2-3), 410-454.

Figure 1: Empirical Framework

The figure illustrates the empirical framework of this study. We examine subjects' belief updating behavior over three phases: Phase 1 describes a sequence of signals that go in the same direction (i.e. confirming an underlying distribution). Phase 2 describes a situation in which a sequence of previously observed same-directional signals is interrupted by a single signal of opposite direction (i.e. disconfirming signal). Finally, Phase 3 defines the situation when a disconfirming signal is immediately reverted (i.e. correction). The blue dots present the objective probabilities (i.e. the beliefs according to Bayes' Theorem) that the asset pays from the good distribution given the sequence of signals.

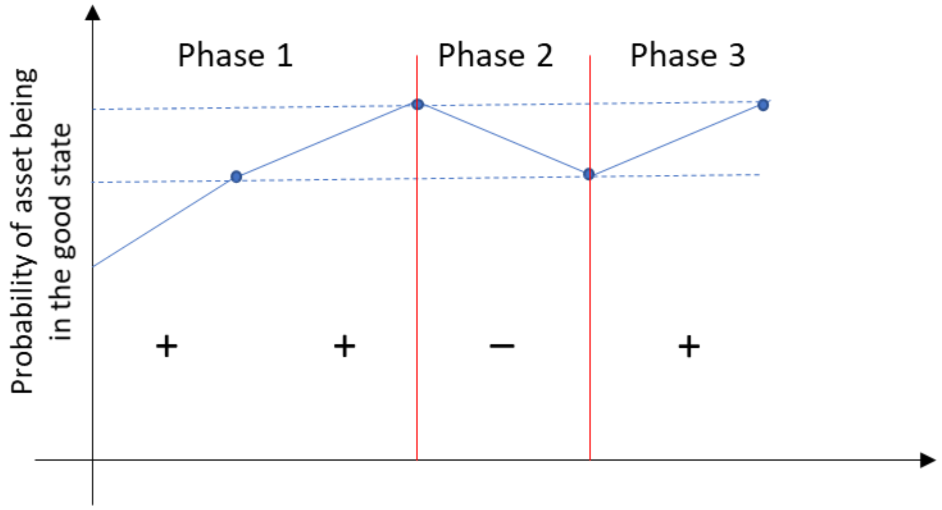
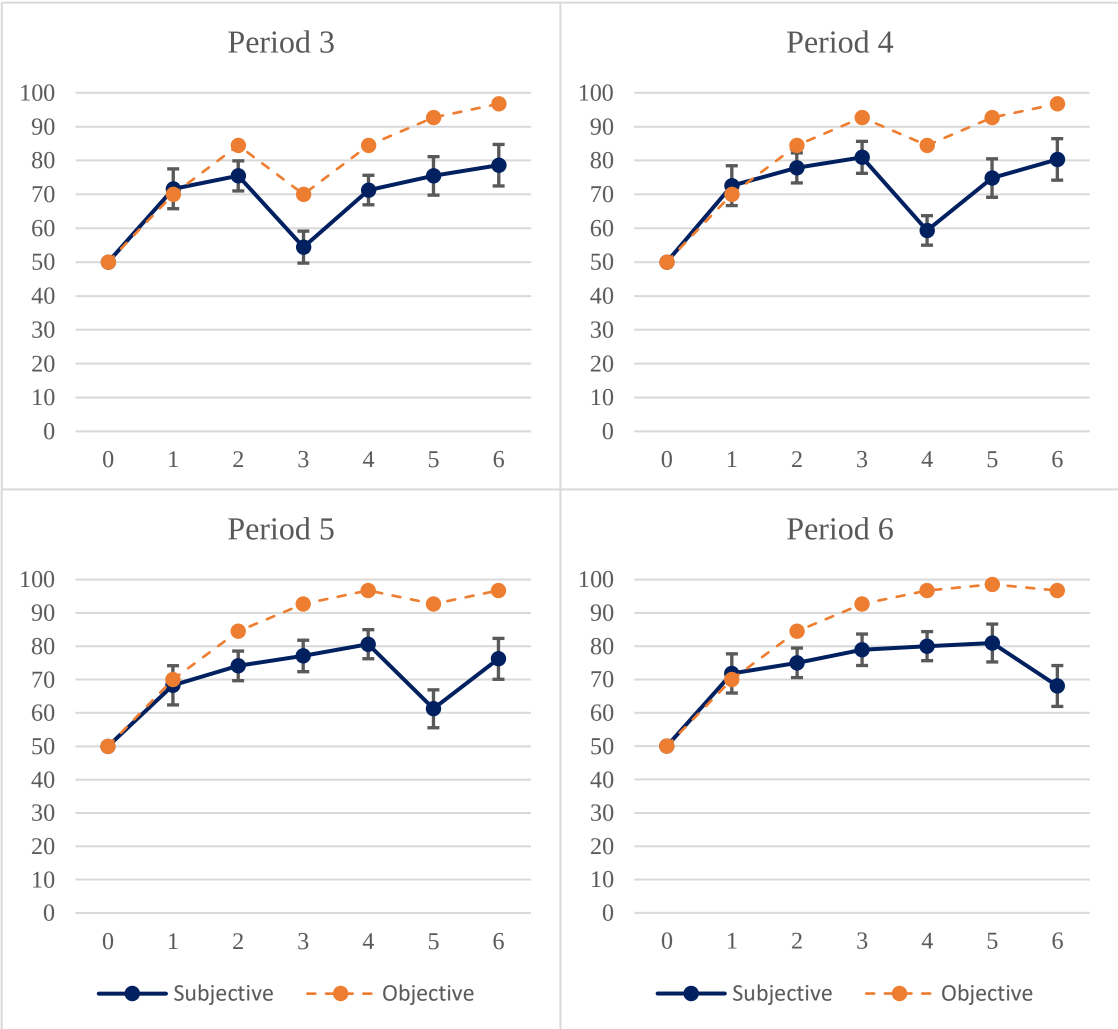


Figure 2: Subjects' Average Updating Behavior – Experiment 1

Panel A displays subjects' average probability estimates over six consecutive periods in the good distribution for treatments G-3 to G-6 individually. Panel B displays subjects' average probability estimates over six consecutive periods in the bad distribution for treatments B-3 to B-6 individually. The dashed line shows the objective Bayesian posterior probabilities and the solid line shows subjects' average probability estimates. Displayed are 95% confidence intervals.

Panel A: Good Distribution



Panel B: Bad Distribution

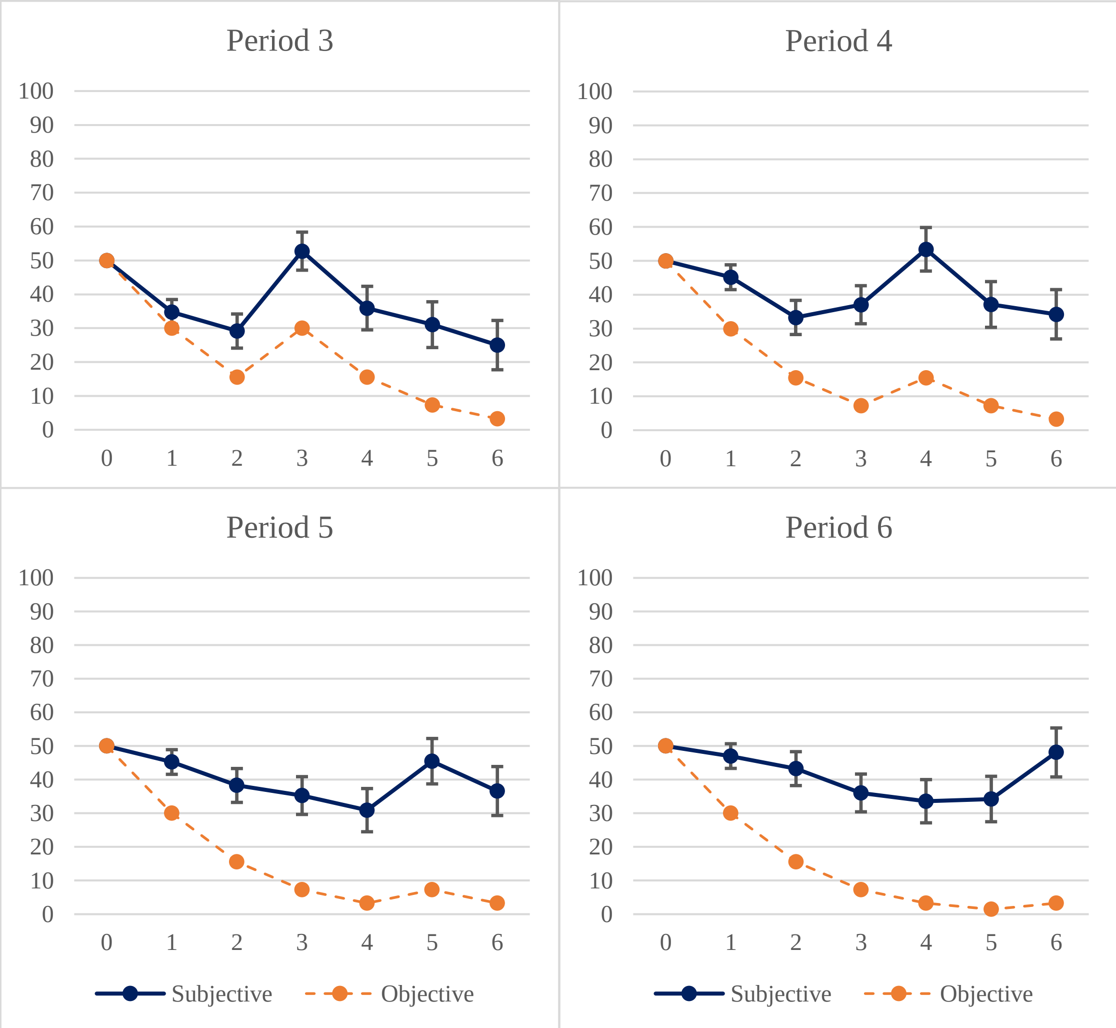
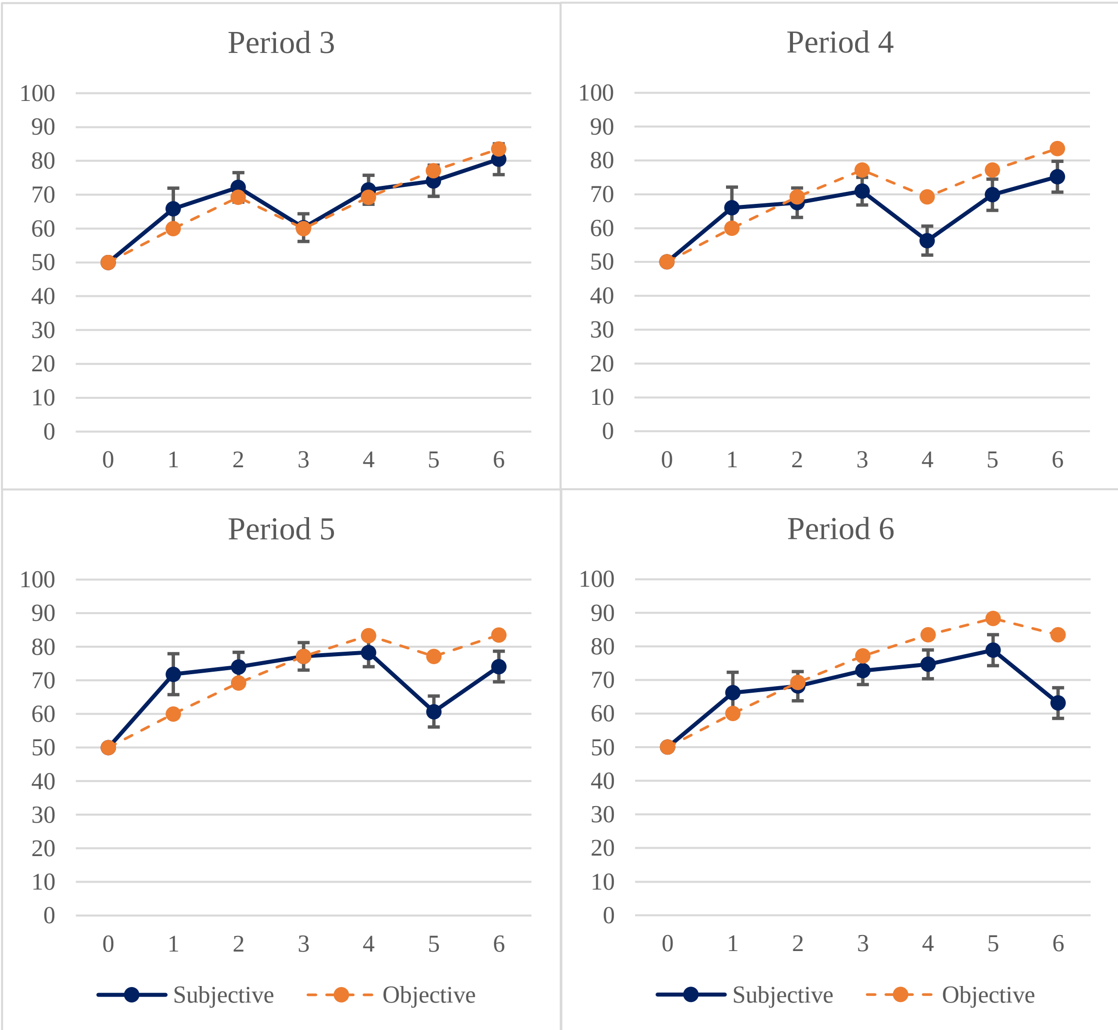


Figure 3: Subjects' Average Updating Behavior – Experiment 2

Panel A displays subjects' average probability estimates over six consecutive periods in the good distribution for treatments G-3 to G-6 individually. Panel B displays subjects' average probability estimates over six consecutive periods in the bad distribution for treatments B-3 to B-6 individually. The dashed line shows the objective Bayesian posterior probabilities and the solid line shows subjects' average probability estimates. Displayed are 95% confidence intervals.

Panel A: Good Distribution



Panel B: Bad Distribution

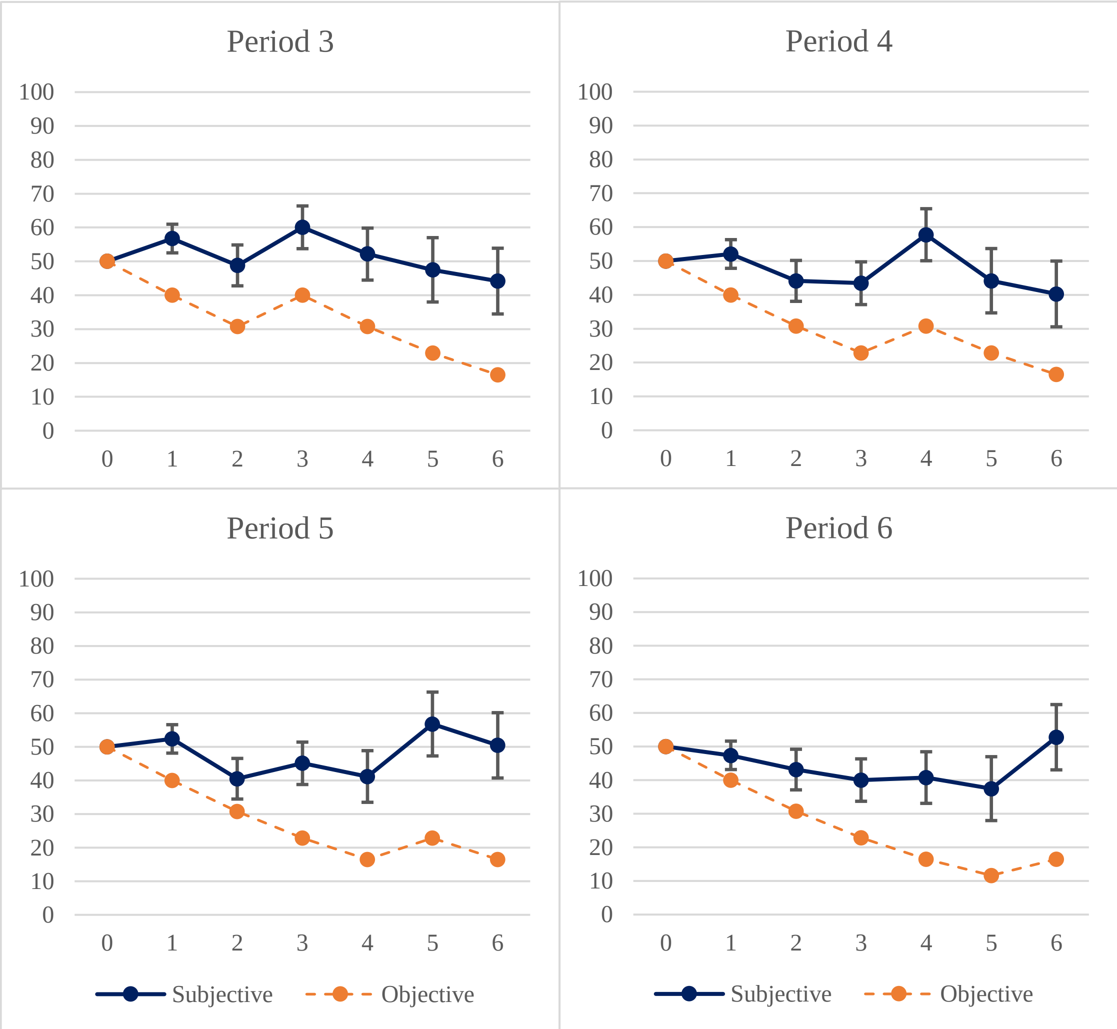
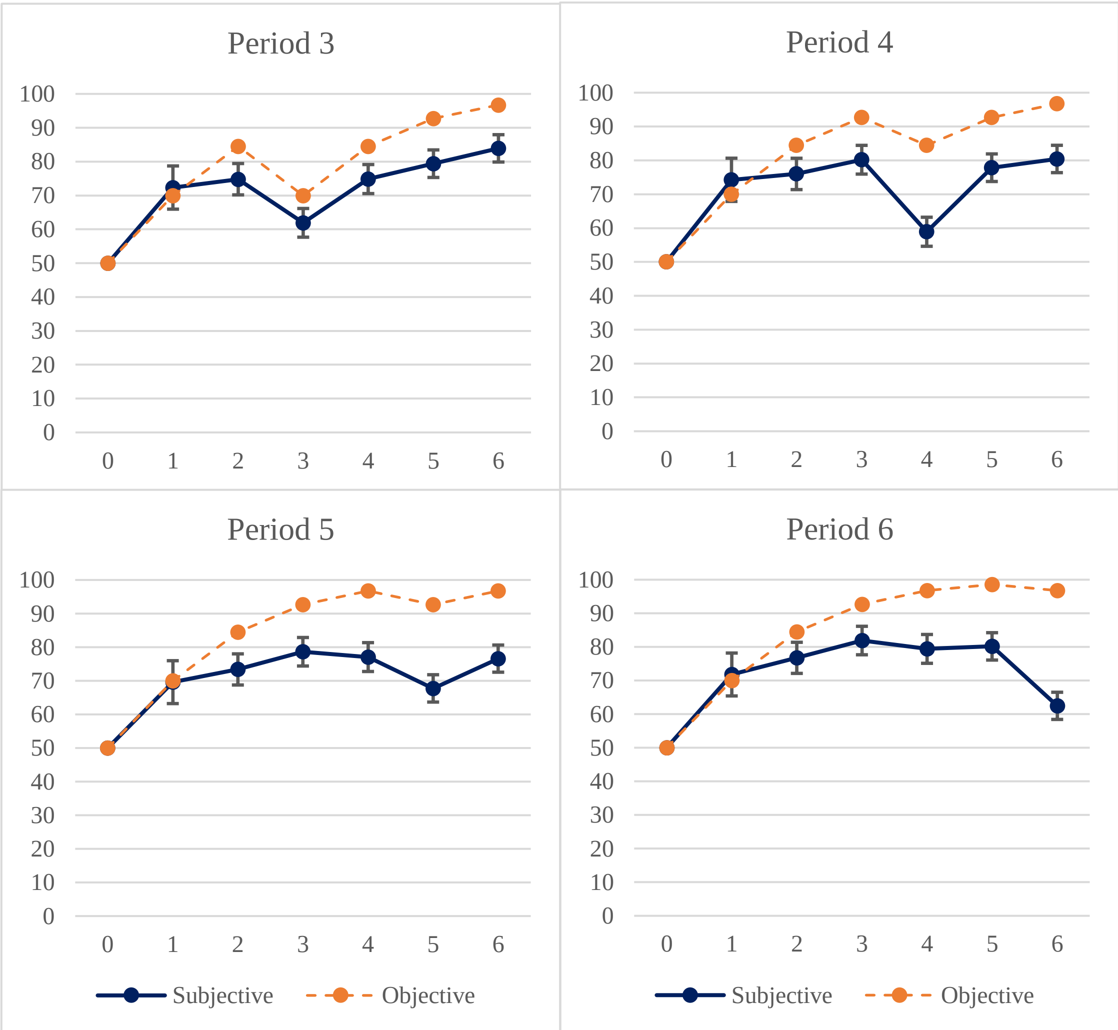


Figure 4: Subjects' Average Updating Behavior – Experiment 3

Panel A displays subjects' average probability estimates over six consecutive periods in the good distribution for treatments G-3 to G-6 individually. Panel B displays subjects' average probability estimates over six consecutive periods in the bad distribution for treatments B-3 to B-6 individually. The dashed line shows the objective Bayesian posterior probabilities and the solid line shows subjects' average probability estimates. Displayed are 95% confidence intervals.

Panel A: Good Distribution



Panel B: Bad Distribution

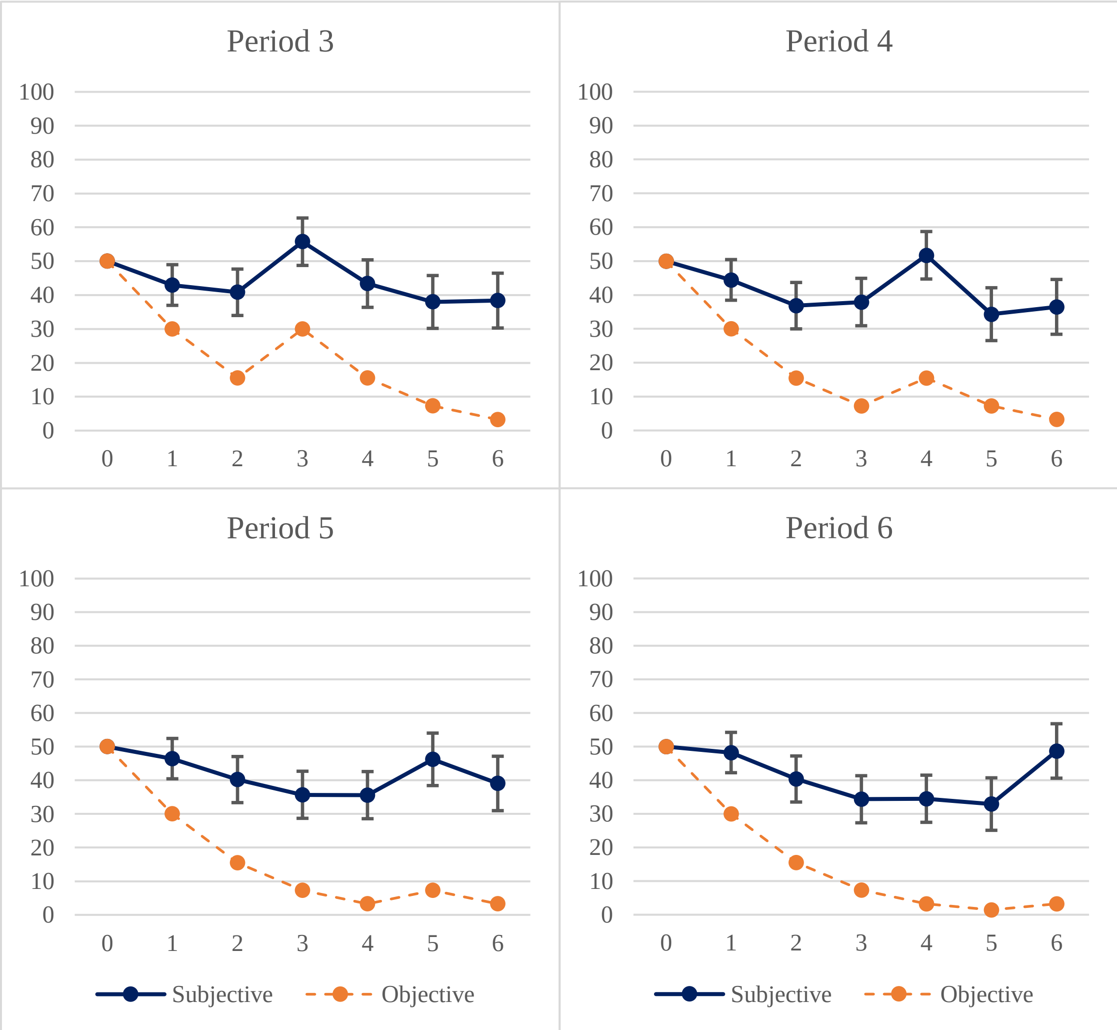
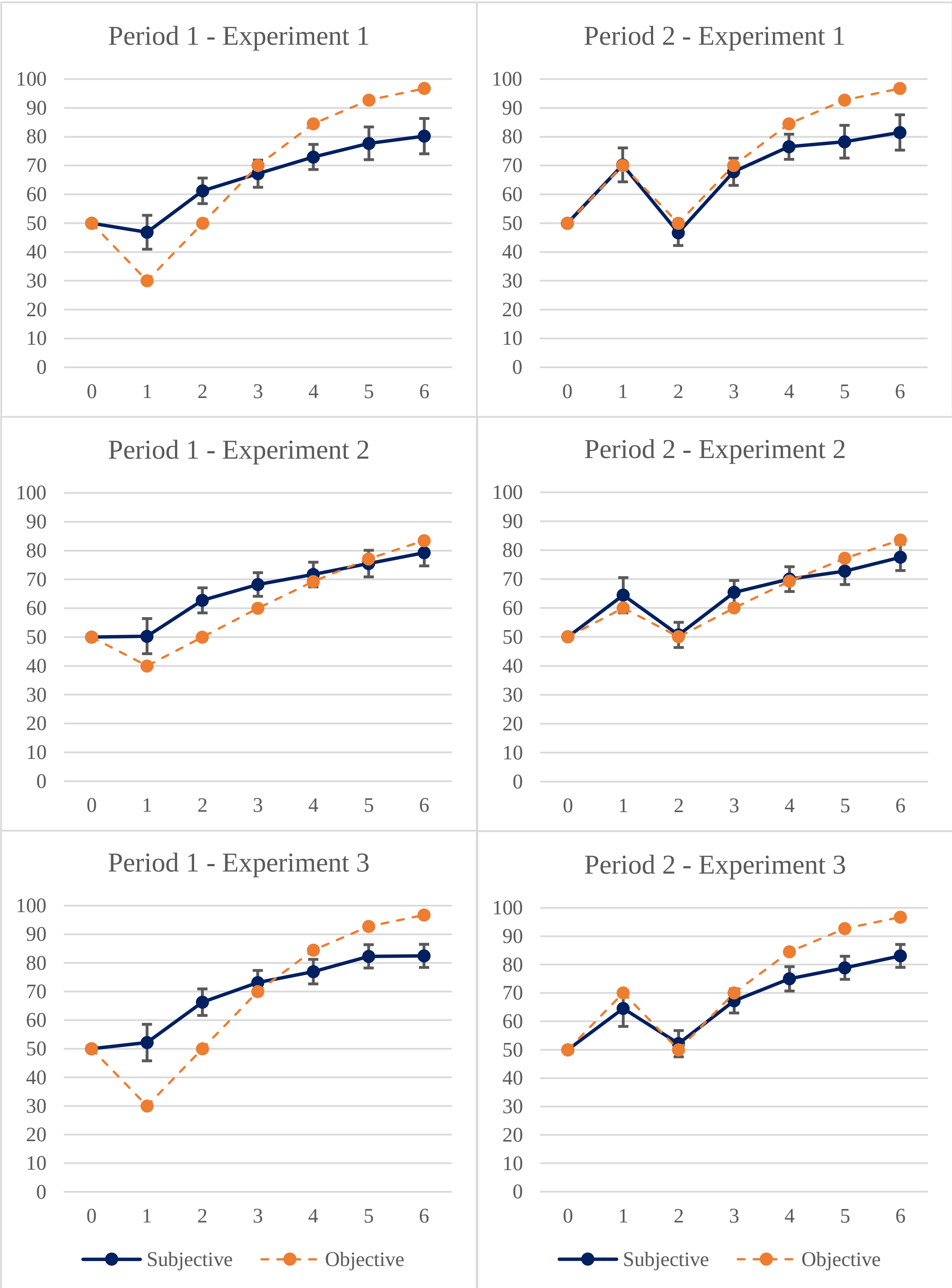


Figure 5: Subjects' Average Updating Behavior – Additional Treatments

Panel A and Panel B display subjects' average probability estimates over six consecutive periods in the good and the bad distribution for treatments G-1 (B-1) and G-2 (B-2) individually by experiment. The dashed line shows the objective Bayesian posterior probabilities and the solid line shows subjects' average probability estimates. Displayed are 95% confidence intervals.

Panel A: Good Distribution



Panel B: Bad Distribution

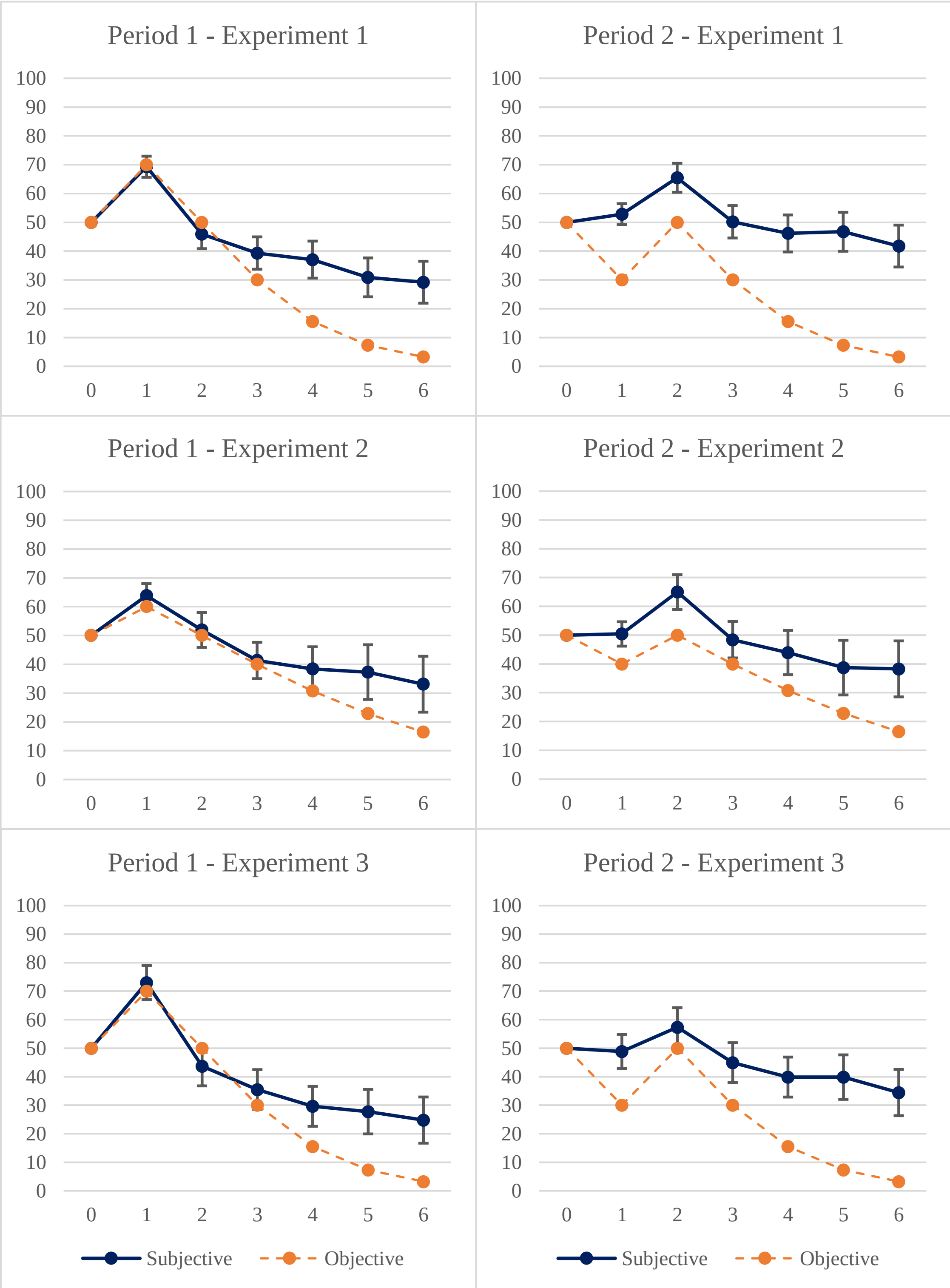


Figure 6: Cumulative Abnormal Returns of Extreme Earnings Surprise Quintiles by Confirming and Disconfirming Surprises after the Earnings Announcement

Using quarterly earnings announcements from June 2009 to May 2020, we calculate the average cumulative abnormal returns (CAR[0,t]) for extreme earnings surprise quintiles by whether the announcement was following a same-directed announcement surprise (confirming) or an opposite-directed announcement surprise (contrary). Earnings surprise quintiles are formed based on independent sorts of quarterly earnings announcements by the corresponding forecast error.

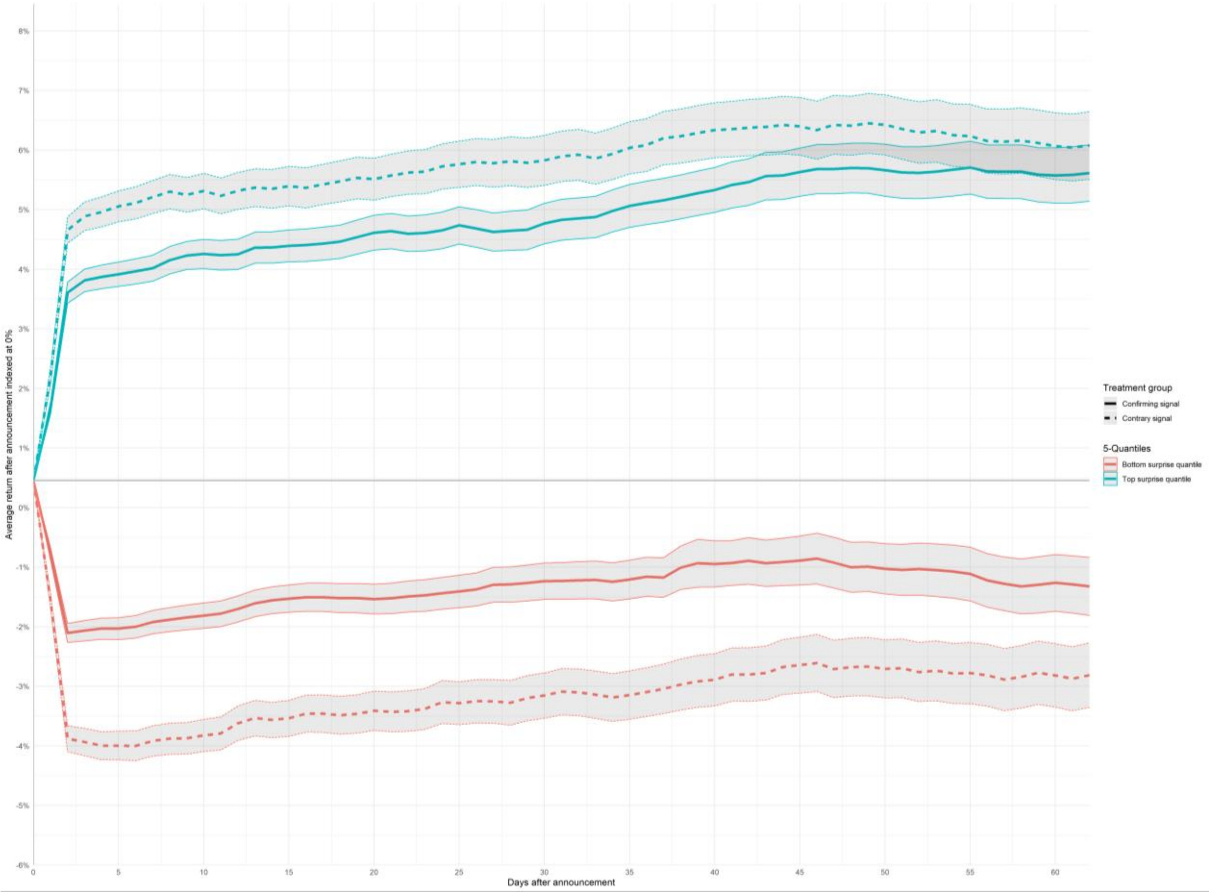


Table 1: Overview of Treatments

This table provides an overview of all treatments in our experiments. Overall, there are twelve treatments, six in the good distribution and six in the bad distribution, defined by the period in which the disruptive signal occurs. The "–" sign represents a negative (bad) signal and the "+" sign a positive (good) signal.

			Good Distribution			
Treatment	1	2	3	4	5	6
G-1	–	+	+	+	+	+
G-2	+	–	+	+	+	+
G-3	+	+	–	+	+	+
G-4	+	+	+	–	+	+
G-5	+	+	+	+	–	+
G-6	+	+	+	+	+	–
			Bad Distribution			
Treatment	1	2	3	4	5	6
B-1	+	–	–	–	–	–
B-2	–	+	–	–	–	–
B-3	–	–	+	–	–	–
B-4	–	–	–	+	–	–
B-5	–	–	–	–	+	–
B-6	–	–	–	–	–	+

Table 2: Summary Statistics

This table shows summary statistics for our experimental data. Reported are the mean and the standard deviation (in parentheses) for each experiment individually. *Female* is an indicator variable that equals 1 if a participant is female. *Statistic skills* denotes participants' self-assed statistical skills on a 7-point Likert scale. *Risk preferences* are elicited by asking subjects to split an endowment between a risky and a risk-free asset (reported is the fraction invested risky). *Financial literacy* was assessed by asking subjects to identify the correct formula for calculating the expected value of the portfolio they selected. Through multiple choice answers, participants could make three basic errors (reported is the number of basic errors).

Variable	**Experiment 1** Baseline (N = 601)	**Experiment 2** Reduced Diagnosticity (N = 602)	**Experiment 3** Reduced Uncertainty (N = 604)
Age	33.79	33.59	35.01
	(9.89)	(9.17)	(9.83)
Female	0.35	0.41	0.32
	(0.48)	(0.49)	(0.47)
Statistic Skills (1-7)	4.46	4.42	4.42
	(1.64)	(1.64)	(1.68)
Risk Preferences	44.7%	44.2%	50.0%
	(2.94)	(2.89)	(2.98)
Financial Literacy (1-3)	1.73	1.70	1.70
	(0.93)	(0.91)	(0.93)

Table 3: Updating Behavior After Disconfirming Signal and Correction – Experiment 1

This table reports the results of four OLS regressions on how subjects update their posterior beliefs after a disconfirming signal and a correction in the baseline experiment. We report the results of OLS regressions for each distribution individually (good and bad distribution). The dependent variable in the regression model, *Change in Posterior Probability Estimate*, is the change in subjective posterior beliefs that the asset is paying from the good distribution between period *t* and period *t-1*. Independent variables include the *Disconfirm* dummy, an indicator variable that equals 1 if participants observe a disconfirming signal and zero otherwise, the *Correction* dummy, an indicator variable that equals 1 if a disconfirming signal is subsequently reverted, as well as *Change in Bayes*, which is the change in the correct Bayesian probability that the stock is good between period *t* and period *t-1*. Reported are coefficients and t-statistics (in parentheses) using robust standard errors. *, **, *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Dependent Variable	*Change in Posterior Probability Estimate*			
	Good Distribution		**Bad Distribution**	
Change in Bayes	0.770***	0.377***	0.718***	0.384***
	(14.64)	(8.02)	(13.03)	(7.51)
Disconfirm		-15.94***		12.38***
		(-9.15)		(7.37)
Correction		11.57***		-11.05***
		(7.36)		(-6.76)
Observations	1782	1782	1824	1824
R^2	0.138	0.218	0.097	0.142

Table 4: Updating Behavior After Disconfirming Signal and Correction – Experiment 2

This table reports the results of four OLS regressions on how subjects update their posterior beliefs after a disconfirming signal and a correction in Experiment 2 with lower diagnosticity than in the baseline experiment. We report the results of OLS regressions for each distribution individually (good and bad distribution). The dependent variable in the regression model, *Change in Posterior Probability Estimate*, is the change in subjective posterior beliefs that the asset is paying from the good distribution between period *t* and period *t-1*. Independent variables include the *Disconfirm* dummy, an indicator variable that equals 1 if participants observe a disconfirming signal and zero otherwise, the *Correction* dummy, an indicator variable that equals 1 if a disconfirming signal is subsequently reverted, as well as *Change in Bayes*, which is the change in the correct Bayesian probability that the stock is good between period *t* and period *t-1*. Reported are coefficients and t-statistics (in parentheses) using robust standard errors. *, **, *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Dependent Variable	*Change in Posterior Probability Estimate*			
	Good Distribution		**Bad Distribution**	
Change in Bayes	0.860***	0.430***	0.877***	0.524***
	(16.47)	(9.08)	(13.98)	(8.84)
Disconfirm		-11.53***		10.06***
		(-8.82)		(6.74)
Correction		9.355***		-6.649***
		(6.57)		(-4.18)
Observations	1872	1872	1740	1740
R^2	0.112	0.169	0.087	0.116

Table 5: Updating Behavior After Disconfirming Signal and Correction – Experiment 3

This table reports the results of four OLS regressions on how subjects update their posterior beliefs after a disconfirming signal and a correction in Experiment 3. We report the results of OLS regressions for each distribution individually (good and bad distribution). The dependent variable in the regression model, *Change in Posterior Probability Estimate*, is the change in subjective posterior beliefs that the asset is paying from the good distribution between period *t* and period *t-1*. Independent variables include the *Disconfirm* dummy, an indicator variable that equals 1 if participants observe a disconfirming signal and zero otherwise, the *Correction* dummy, an indicator variable that equals 1 if a disconfirming signal is subsequently reverted, as well as *Change in Bayes*, which is the change in the correct Bayesian probability that the stock is good between period *t* and period *t-1*. Reported are coefficients and t-statistics (in parentheses) using robust standard errors. *, **, *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Dependent Variable	*Change in Posterior Probability Estimate*			
	Good Distribution		**Bad Distribution**	
Change in Bayes	0.603***	0.294***	0.666***	0.362***
	(12.89)	(6.87)	(11.68)	(7.03)
Disconfirm		-11.77***		9.559***
		(-6.41)		(6.10)
Correction		9.978***		-11.03***
		(7.53)		(-6.77)
Observations	1884	1884	1740	1740
R^2	0.088	0.135	0.086	0.122

Table 6: Outcome Ordering

This table reports the results of OLS regressions on how subjects updating behavior after a disconfirming signal and correction depends on their prior beliefs. We report the results of OLS regressions for each experiment (Experiment 1 and 3 pooled) and distribution (good and bad distribution) individually. The dependent variable in the regression model, *Posterior Probability Estimate in Period 6*, is the absolute subjective posterior belief that the asset is paying from the good distribution in period 6. Independent variables include *Condition t* dummies which are indicator variables for each period *t*. Reported are coefficients and t-statistics (in parentheses) using robust standard errors. *, **, *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Dependent Variable	*Posterior Probability Estimate in Period 6*			
	Experiment 1 & 3		**Experiment 2**	
	Good Distribution	Bad Distribution	Good Distribution	Bad Distribution
Disconfirm Round 2	0.912	11.45**	-1.755	5.194
	(0.39)	(2.52)	(-0.53)	(0.77)
Disconfirm Round 3	-0.374	5.306	1.224	11.09*
	(-0.16)	(1.36)	(0.41)	(1.74)
Disconfirm Round 4	-1.070	8.198**	-4.059	7.177
	(-0.46)	(2.17)	(-1.21)	(1.21)
Disconfirm Round 5	-5.043**	10.63***	-5.145	17.34***
	(-2.00)	(2.68)	(-1.61)	(2.76)
Disconfirm Round 6	-16.09***	21.24***	-16.09***	19.67***
	(-5.15)	(5.10)	(-4.33)	(3.05)
Constant	80.45***	26.16***	78.25***	32.10***
	(45.94)	(9.72)	(34.38)	(6.69)
Observations	611	594	312	290
R^2	0.094	0.046	0.101	0.049

Table 7: Forecasting Ability and Extreme Outliers

This table reports the results of OLS regressions on how subjects update their posterior beliefs after a disconfirming signal and a correction across all experiments split by extreme outliers (Panel A), the time it takes subjects to finish the experiment (Panel B), and subjects' forecasting ability (Panel C). We report the results of OLS regressions for each subsample of individuals (with above-median versus below-median updating ability, no outlier versus outlier, and speeders versus non-speeders) and for each distribution (good and bad distribution) individually. Speeders are defined as the fastest 10% of the subjects. Non-speeders are defined as the remaining 90% of the subjects. The dependent variable in the regression model, *Change in Posterior Probability Estimate*, is the change in subjective posterior beliefs that the asset is paying from the good distribution between period *t* and period *t-1*. Independent variables include the *Disconfirm* dummy, an indicator variable that equals 1 if participants observe a disconfirming signal and zero otherwise, the *Correction* dummy, an indicator variable that equals 1 if a disconfirming signal is subsequently reverted, as well as *Change in Bayes*, which is the change in the correct Bayesian probability that the stock is good between period *t* and period *t-1*. Reported are coefficients and t-statistics (in parentheses) using robust standard errors. *, **, *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Panel A: Extreme Outliers

Dependent Variable	*Change in Posterior Probability Estimate*			
	Good Distribution		**Bad Distribution**	
	No Outlier	Outlier	No Outlier	Outlier
Change in Bayes	0.397***	-0.0964	0.583***	-0.128*
	(15.00)	(-0.51)	(18.76)	(-1.72)
Disconfirm	-11.41***	-35.85***	10.45***	12.19***
	(-14.21)	(-4.45)	(12.32)	(5.22)
Correction	8.757***	36.26***	-9.312***	-10.33***
	(11.80)	(5.28)	(-9.87)	(-4.41)
Observations	5238	300	3882	1422
R^2	0.181	0.222	0.242	0.031

Panel B: Speeders versus Non-Speeders

Dependent Variable	*Change in Posterior Probability Estimate*			
	Good Distribution		**Bad Distribution**	
	Non-Speeders	Speeders	Non-Speeders	Speeders
Change in Bayes	0.370***	0.149	0.415***	0.299***
	(12.86)	(1.63)	(12.20)	(3.43)
Disconfirm	-13.75***	-7.236**	11.25***	7.325***
	(-13.89)	(-2.36)	(11.46)	(3.15)
Correction	10.81***	5.825**	-10.07***	-5.991*
	(12.57)	(2.08)	(-10.37)	(-1.90)
Observations	5028	510	4734	570
R^2	0.190	0.040	0.143	0.039

Panel C: Forecasting Ability

Dependent Variable	*Change in Posterior Probability Estimate*			
	Good Distribution		**Bad Distribution**	
	Above Median	Below Median	Above Median	Below Median
Change in Bayes	0.625***	0.0137	0.823***	0.111**
	(23.44)	(0.30)	(26.97)	(2.48)
Disconfirm	-6.218***	-21.97***	5.924***	13.95***
	(-8.90)	(-11.48)	(7.29)	(10.35)
Correction	5.420***	16.59***	-5.780***	-12.13***
	(8.36)	(9.71)	(-6.62)	(-8.42)
Observations	3270	2268	2154	3150
R^2	0.267	0.154	0.388	0.079

Table 8: Cumulative Abnormal Returns of Extreme Earnings Surprise Quintiles by Confirming and Disconfirming Surprises

Using quarterly earnings announcements from June 2009 to May 2020, we calculate the average 2-day announcement cumulative abnormal returns (CAR[0,1]) and 60-day post-announcement cumulative abnormal return (CAR[2,61]) for extreme earnings surprise quintiles by whether the announcement was following a same-directed announcement surprise (confirming) or a opposite-directed announcement surprise (disconfirming). Earnings surprise quintiles are formed based on independent sorts of quarterly earnings announcements by the corresponding forecast error.

	Average CAR[0,1] for Earnings Surprise Quintile 1 and 5			**Average CAR[2,61] for Earnings Surprise Quintile 1 and 5**		
	Quintile 1	Quintile 5	Difference	Quintile 1	Quintile 5	Difference
Confirming	-2.56%	3.15%	5.71%***	0.81%	1.89%	1.07%***
Disconfirming	-4.33%	4.20%	8.53%***	1.15%	1.39%	0.25%
Difference	1.77%***	-1.05%***	-2.82%***	-0.33%	0.49%	0.83%

Table 9: Market Reactions to Earnings News: Multivariate Tests

This table reports the multivariate tests of the effects of the sign of the announcement relative to the prior announcement on the relation between announcement or post-announcement returns and earnings surprises. The dependent variable is indicated under each column heading. FE is the earnings surprise quintile (FE=1: lowest,5: highest) and Disconfirming is a dummy that equals 1 if the sign of the current earning surprise is inconsistent with the sign of the prior earning surprise, and 0 otherwise. FE5 is an indicator variable for the top earnings deciles (FE=5). Control variables include size and book-to-market deciles, log (1+# Analysts), Reporting Lag, Reporting Lag squared and cubed, institutional ownership (IO), Earnings Volatility, Earnings Persistence, Share Turnover, and indicator variables for year, month, day of week, and Fama-French 10 industry classification. Standard errors adjusted for heteroskedasticity and clustering by the day of announcement are in parentheses. $*, **, ***$indicate significance at the 10%, 5%, and 1% level, respectively.

	(1) CAR[0,1]	(2) CAR[2,61]
Disconfirming (0/1)	-0.01774***	0.00334
	(0.00163)	(0.00427)
FE5	0.05711***	0.01074***
	(0.00145)	(0.00389)
Disconfirming x	0.02823***	-0.00826
DC5	(0.00237)	(0.00588)
Controls	X	X
Observations	22,681	22,681
R^2	0.139	0.0003

Appendix

A. Experimental Instructions and Screenshots

Instructions Bayesian Updating (Exemplary for Experiment 1)

In this part we would like to test your forecasting abilities.

You will make forecasting decisions in one block consisting of 6 rounds.

Suppose you find yourself in an environment, in which a risky asset with an initial value of 50 can either increase by 5 or decrease by 5. The probability of either outcome (5 or -5) depends on the state in which the asset is (**good** state or **bad** state). If the risky asset is in the **good** state, then the probability that the risky asset increases in value by 5 is 70% and the probability that it decreases in value by 5 is 30%. If the risky asset is in the **bad** state, then the probability that the risky asset increases in value by 5 is 30% and the probability that it decreases in value by 5 is 70%.

The computer determines the state at the beginning of the block (consisting of 6 rounds). Within a block, the state does not change and remains fixed.

At the beginning of the block, you do not know which state the risky asset is in. The risky asset may be in the good state or in the bad state with equal probability.

At the beginning of each round, you will observe the payoff of the risky asset (5 or -5). After that, we will ask you to provide a probability estimate that the risky asset is in the good state and ask you how sure you are about your probability estimate. While answering these questions, you can observe the price development in a chart next to the question.

There is always an objective correct probability that the risky asset is in the good state. This probability depends on the history of payoffs of the risky asset already. As you observe the payoffs of the risky asset, you will update your beliefs whether or not the risky asset is in the good state.

Objective Bayesian Posterior Probabilities

This table provides all possible values for the objectively correct probability that the asset is in the good state for every possible combination of trials and outcomes. The initial prior for good and bad distribution is set to 50%. The objective Bayesian posterior probability that the asset is in the good state, after observing t high outcomes in n trials so far is given by: $\frac{1}{1+\frac{1-p}{p}\cdot(\frac{q}{1-q})^{n-2t}}$, where *p* is the initial prior before any outcome is observed that the stock is in the good state (50% here), and *q* is the probability that the value increase of the asset is the higher one (70% in Experiment 1 & 3, and 60% in Experiment 2).

		Experiment 1 & 3 (q = 70%)	**Experiment 2 (q = 60%)**
n (number of trials so far)	**t (number of high outcomes so far)**	**Probability [stock is good t high outcomes in n trials]**	**Probability [stock is good t high outcomes in n trials]**
0	0	50.00%	50.00%
1	0	30.00%	40.00%
1	1	70.00%	60.00%
2	0	15.52%	30.77%
2	1	50.00%	50.00%
2	2	84.48%	69.23%
3	0	7.30%	22.86%
3	1	30.00%	40.00%
3	2	70.00%	60.00%
3	3	92.70%	77.14%
4	0	3.26%	16.49%
4	1	15.52%	30.77%
4	2	50.00%	50.00%
4	3	84.48%	69.23%
4	4	96.74%	83.51%
5	0	1.43%	11.64%
5	1	7.30%	22.86%
5	2	30.00%	40.00%
5	3	70.00%	60.00%
5	4	92.70%	77.14%
5	5	98.57%	88.36%
6	0	0.62%	8.7%
6	1	3.26%	16.49%
6	2	15.52%	30.77%
6	3	50.00%	50.00%
6	4	84.48%	69.23%
6	5	96.74%	83.51%
6	6	99.38%	91.93%

Screenshots of Experiment 1

Figures B1 to B3 present the screens of the forecasting task as seen by subjects in the experiment (example round 4). One round consists of three sequential screens. First, subjects saw the payoff of the risky asset in the respective round. Second, the cumulated payoffs of the risky asset are shown in a price-line-chart and subjects are asked to provide a probability estimate that the risky asset pays from the good distribution. Finally, subjects are asked on a 9-point Likert scale how confident they are in their probability estimate.

Figure A1: Payoff screen

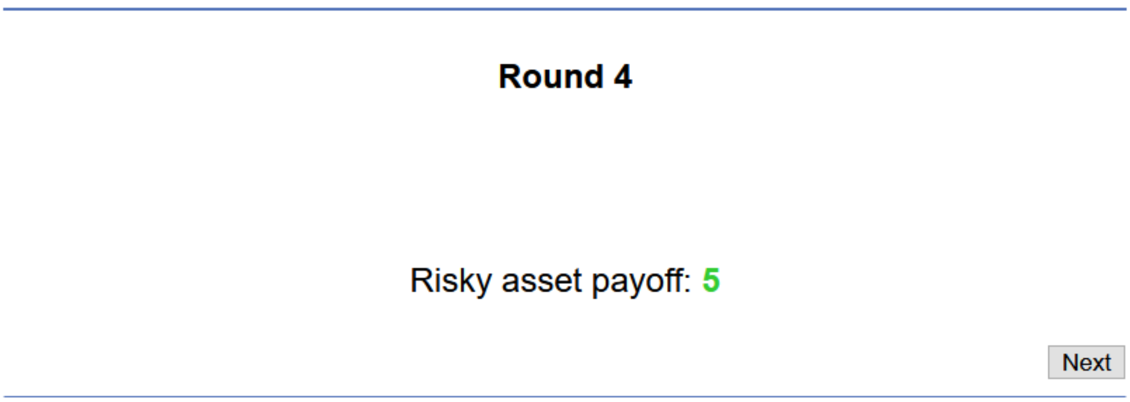
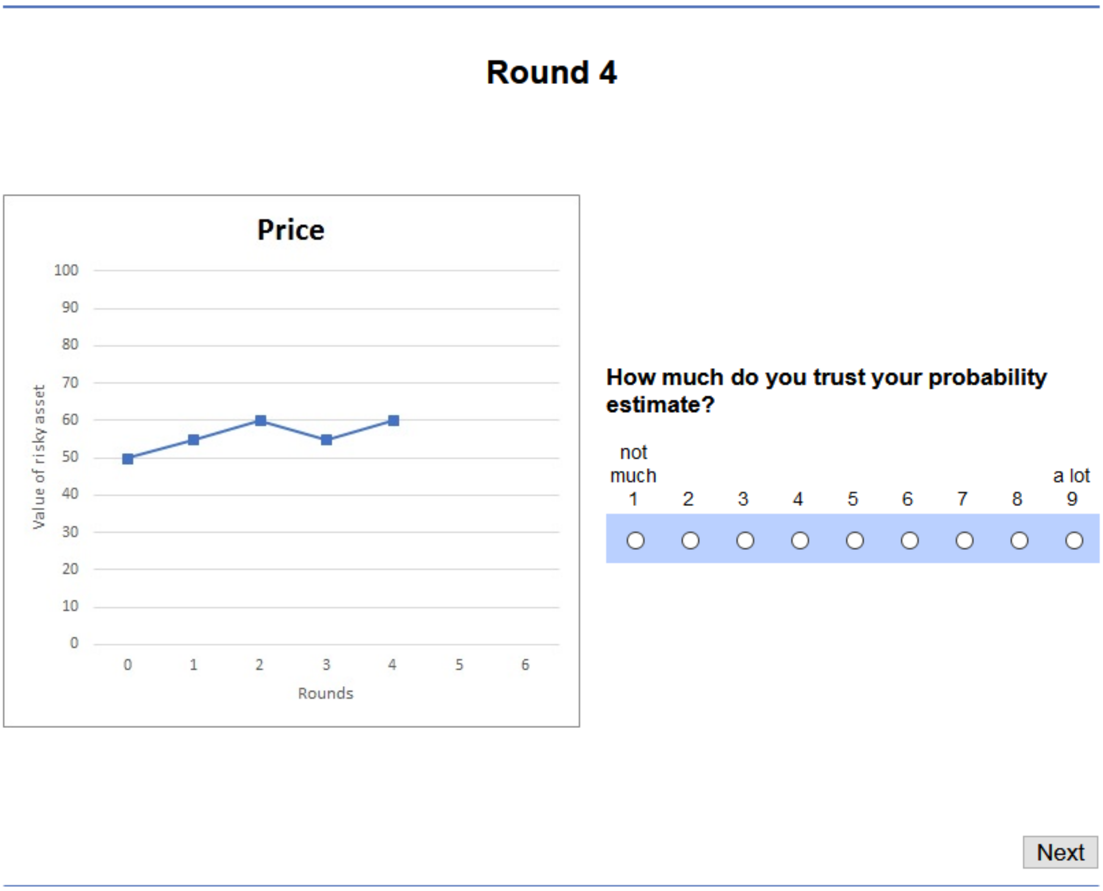


Figure A2: Probability estimate screen

Figure A3: Confidence level screen



Comprehension Question for Bayesian Updating Task

Below we report the comprehension questions that participants had to answer correctly after reading the instructions to proceed to the Bayesian Updating task. Correct responses are displayed in *italic*.

1. If you see a series of +5, what is more likely?
    a. *The risky asset is in the good state.*
    b. The risky asset is in the bad state.

2. You observe a -5, how do you have to update your probability estimate that the asset draws from the good distribution?
    a. *I reduce the probability estimate that the asset is in the good distribution.*
    b. I increase the probability estimate that the asset is in the good distribution.

3. The correct probability estimate is let's say 0.70. Which probability estimate(s) would be in the range such that you earn 25 cents? [Note: You can check multiple boxes.]
    a. 0.55
    b. *0.67*
    c. *0.75*
    d. 0.85
    e. 0.87

4. At the beginning of the first period, the probability that the risky asset is in the good state is 50%.
    a. *True*
    b. False