

Memory Overload and Asymmetric Underreaction*

Lorenz Goette[†], Hua-Jing Han[‡], Benson Tsz Kin Leung[§]

July 4, 2026

Abstract

We show that greater memory overload results in a disproportionately greater underreaction to belief-challenging information compared to belief-confirming information. In an experiment, we keep constant the information subjects receive but vary the difficulty of memorizing the signals. In the treatment condition, the preceding signal disappears as the next signal appears; in the control condition, the preceding signal remains visible. Memory load is therefore greater in the treatment than in the control condition. We find that in the treatment condition, subjects underreact more to belief-challenging information than in the control condition, but updating to belief-confirming information is not significantly affected. Our results caution against information campaigns that ignore the roles of memory load and attention.

Keywords: memory overload, belief formation, underreaction

JEL codes: D83, D91

1 Introduction

Memory is central to belief formation. Recent experimental and empirical work shows that features of memory, including the associative structure, the high retrievability of narrative content, and experience-based encoding, shape beliefs (Enke et al. (2024); Graeber et al. (2024); Bordalo et al. (2025)). A growing theoretical literature also demonstrates that memory constraints can generate confirmation bias even under optimal information processing (Wilson (2014); Jehiel and Steiner (2019); Leung (2020)). We build on this work to study, theoretically and experimentally, how higher memory load asymmetrically affects belief updating in response to belief-confirming versus belief-challenging information.

Our theoretical framework extends Leung (2020) and provides the foundation for our main hypothesis: under greater memory load, subjects underreact more to belief-challenging signals, while their updating to belief-confirming signals is largely unaffected. Consider a decision maker (DM) who seeks to match an action to the

*Funding by the German Research Foundation (DFG) through CRC TR 224 (Project B07) is gratefully acknowledged. We thank Rafael Suchy and Zhi Hao Lim for their excellent service as research assistants. Benson Tsz Kin Leung would like to acknowledge the financial support from the Digital Chair Jean-Jacques Laffont.

[†]National University of Singapore. Email: ecslfg@nus.edu.sg

[‡]Comma Soft AG.

[§]Hong Kong Baptist University. Email: btkleung@hkbu.edu.hk

true state, which is either “high” or “low” with equal probability. The DM receives at most two signals, s_1 and s_2 , but has only one unit of memory capacity to memorize a signal and update her belief. Upon observing s_1 , the DM must decide whether to memorize s_1 and update her belief immediately or reserve capacity for s_2 . Reserving capacity entails forgetting s_1 ; memorizing s_1 precludes memorizing and updating her belief with s_2 later. The choice trades off the benefit of using the current signal against the opportunity cost of missing a future signal.

Now suppose the DM’s prior favors the “high” state. If s_1 is belief-confirming and leads to a “high” guess, the opportunity cost of memorizing s_1 is the chance that s_2 will be belief-challenging and overturn the guess to “low”. Because both the prior and s_1 favor the “high” state, that chance is small, so the opportunity cost of memorizing s_1 is low and the DM has strong incentive to memorize s_1 . In contrast, if s_1 is belief-challenging and leads to a “low” guess, the DM’s belief is still relatively close to (50%, 50%), and the probability that s_2 will support a “high” guess is comparatively large. The opportunity cost of memorizing s_1 is therefore higher, increasing the DM’s incentive to forgo memorizing a belief-challenging s_1 . This asymmetry in opportunity costs gives the DM greater incentive to memorize belief-confirming signals and to skip belief-challenging signals, producing asymmetric underreaction when the DM faces a memory constraint.

We test this prediction in an experiment that varies memory load while holding constant the information set across treatment and control, thereby isolating the effect of memory load on updating. Subjects observe integer signals (1 to 8) about an underlying binary state (“low” or “high”) drawn with equal prior probability. The low distribution places greater weight on small numbers; the high distribution places greater weight on large numbers. Subjects are randomly paired each round, with one in the treatment condition and the other in the control condition. Information arrives in two phases. In Phase 1, subjects in each pair view the same sequence of five numbers under an identical display; this induces subject-specific elicited posteriors away from 50-50, which we use to classify subsequent Phase-2 signals as belief-confirming or belief-challenging. The Phase-2 signal sequence is belief-confirming if the aggregate likelihood ratio of the seven Phase-2 signals favors the same state as the Phase-1 posterior, and is belief-challenging if it favors the opposite state. In Phase 2, subjects in both treatment and control groups view the same sequence of (at most) seven numbers over 30 seconds and then report their posterior belief about the state. The key manipulation is as follows: In the control condition, previously observed numbers remain visible, allowing multiple signals to be seen simultaneously. In the treatment condition, each new signal replaces the previous one, so only the current signal is visible and earlier signals must be memorized. Moreover, the timing of signal presentation is identical across conditions, and subjects advance through the sequence at the same pace. Because the control

display permits at-a-glance aggregation while the treatment requires memorization, the treatment induces higher memory load. Consistent with this, subjects in the treatment make larger updating errors than in the control.

Our results show that higher memory load produces asymmetric underreaction, supporting our main hypothesis. In the treatment condition, subjects underreact to belief-challenging information relative to the control, while responses to belief-confirming information are statistically indistinguishable across conditions. Holding the signal sequence constant, greater memory load leads to more biased processing, driven by reduced responsiveness to belief-challenging signals rather than changes in responses to belief-confirming signals.

These findings indicate that, beyond preferences, the information environment itself can drive biased belief updating behavior. This echoes a recent related study by [Bohren et al. \(2025\)](#), which shows that *“representational distortions can generate decision anomalies that have previously been attributed to preferences”*.¹ The policy implications differ from those implied by purely utility-based mechanisms. One way to weaken biased belief updating is to make information easier to process and memorize. Conversely, information campaigns that convey complex and hard-to-memorize information may disproportionately dampen the impact of belief-challenging messages and thus backfire. Similarly, work environments that impose higher memory load can degrade decision quality.

Our paper contributes to the experimental literature on how individuals update their beliefs. [Eil and Rao \(2011\)](#), [Ertac \(2011\)](#), [Grossman and Owens \(2012\)](#) and [Möbius et al. \(2022\)](#) study how subjects’ beliefs about their task performance, IQ, or beauty evolve with information, and find evidence of overconfidence. [Liang \(2025\)](#) shows that subjects substantially discount signals when the quality of the information source is ambiguous. Similar to this paper, recent experimental literature finds that the information environment, or how information is presented, affects the prevalence of biases and how individuals update their beliefs. [Enke and Zimmermann \(2019\)](#) finds that many participants neglect the correlation between signals when forming their beliefs, and that reducing complexity mitigates correlation neglect. [Enke \(2020\)](#) finds that computational complexity affects whether subjects take into account unobserved signals when they update their beliefs. [Gonçalves et al. \(2021\)](#) finds that the complexity of information retraction induces underreaction. [Ba et al. \(2025\)](#) study a model in which subjects have distorted representations of the information environment, and show — theoretically and experimentally — that the resulting bias (overreaction or underreaction) depends on environmental features such as signal precision and prior structure.

The remainder of the paper is organized as follows. The next section presents

¹[Bohren et al. \(2025\)](#) show that presenting information sequentially versus simultaneously—which taxes bounded memory and limited attention, respectively—leads to different representational distortions of the choice environment and hence different decisions.

a stylized theoretical model that motivates our hypothesis and clarifies the mechanism underlying the treatment effect. Section 3 outlines the experimental design. Section 4 presents descriptive statistics for our sample and our main variables. Section 5 reports the results. Section 6 concludes with implications for information campaigns, advertising, and managerial decisions.

2 Model

In this section, we present the theoretical analysis underpinning our main hypothesis and the intuition behind it.

The model closely follows Leung (2020). While the example does not perfectly mirror our experimental design, it captures its key features and motivates our hypothesis. Consider a subject who must guess whether the “high” or “low” distribution was randomly chosen in a given round to generate the signals she observes. If her guess is correct, she receives one unit of utility; otherwise, she receives zero. Her prior belief is denoted by $(p_H, 1 - p_H)$ where p_H is the prior probability assigned to the “high” distribution. We assume $p_H > 0.5$.

Before the subject makes a guess, she receives at most two signals, denoted by s_1 and s_2 . Each signal is a number drawn from the interval $(0, \infty)$. The “high” distribution is more likely to generate a higher number, whereas the “low” distribution is more likely to generate a lower number. Formally, let F_H and F_L denote the CDFs in the high and low states with densities f_H and f_L . The signals satisfy the monotone likelihood-ratio property, and, without loss of generality, we normalize the signal so that

$$s = \frac{f_H(s)}{f_L(s)}.$$

We also assume that the two signal distributions are symmetric around 1:

$$1 - F_H(s) = F_L(s^{-1}) \text{ for all } s, \quad (1)$$

i.e., the probability of receiving a signal greater than s under the high distribution equals the probability of receiving a signal smaller than s^{-1} under the low distribution.

We consider a memory constraint that creates a trade-off in memory capacity allocation. The subject updates beliefs using information stored in memory but can memorize only one of the two signals. More precisely, she engages in the following sequential game: After seeing the first number s_1 , she decides whether to memorize it. If she chooses to memorize s_1 , she forgoes s_2 ; on the other hand, if she chooses to ignore s_1 , she proceeds to receive the next signal s_2 and memorizes it. Finally, she makes a guess based on her memorized signal. We analyze the dual-self perfect Bayesian equilibrium, such that when self-1 ignores s_1 , self-2 forms a rational belief

about s_1 that is consistent in equilibrium.² We assume that if the equilibrium prescribes always memorizing s_1 — so self-2 never receives s_2 on path — but self-1 deviates by forgetting some signal, then self-2's belief is the prior p_H . For ease of exposition, we assume that the subject memorizes the number if she is indifferent.

Proposition 1. *In any equilibrium, there exist thresholds $\bar{s} \geq \frac{1-p_H}{p_H} > \underline{s} > 0$ such that the subject memorizes s_1 if and only if $s_1 \geq \bar{s}$ or $s_1 \leq \underline{s}$. Moreover, if the following inequality holds,*

$$\frac{\left(1 - F_H\left(\frac{1-p_H}{p_H}\right)\right) / \left(1 - F_L\left(\frac{1-p_H}{p_H}\right)\right)}{F_L\left(\frac{1-p_H}{p_H}\right) / F_H\left(\frac{1-p_H}{p_H}\right)} > \left(\frac{1-p_H}{p_H}\right)^2, \quad (2)$$

then there exists an equilibrium with asymmetric thresholds satisfying $\bar{s} < \frac{1}{\underline{s}}$. On the other hand, denote $\tilde{\theta} = \frac{1-p_H}{p_H} \cdot \frac{1-F_L\left(\frac{1-p_H}{p_H}\right)}{1-F_H\left(\frac{1-p_H}{p_H}\right)}$. If the following inequality holds,

$$\frac{\left(1 - F_H(\tilde{\theta})\right) / \left(1 - F_L(\tilde{\theta})\right)}{F_L(\tilde{\theta}) / F_H(\tilde{\theta})} > \left(\frac{1-p_H}{p_H}\right)^2, \quad (3)$$

for example when p_H is large enough, no equilibrium exists with $\bar{s} \geq \frac{1}{\underline{s}}$.

By symmetry of the signal distributions around 1, asymmetric thresholds with $\bar{s} < 1/\underline{s}$ imply two immediate consequences. First, the probability of memorizing s_1 in state “low” is strictly lower than in state “high”. Second, the subject memorizes $s_1 > 1$ if she memorizes $(s_1)^{-1}$ but the reverse is not true. Hence, under $\bar{s} < 1/\underline{s}$, the subject is more likely to disregard belief-challenging signals than belief-confirming signals. The proposition extends [Leung \(2020\)](#) by providing tractable sufficient conditions in symmetric environments that guarantee the existence of an equilibrium with disproportionate underreaction to belief-challenging signals, and the non-existence of equilibria with disproportionate underreaction to belief-confirming signals.

The proof of the Proposition is presented in [Appendix A.1](#). To build intuition, we first parse the left- and right-hand sides of Equation (2). The subject trades off between allocating her updating capacity to the current signal s_1 and the future signal s_2 . We distinguish two cases: 1) $s_1 \geq \frac{1-p_H}{p_H}$; 2) $s_1 < \frac{1-p_H}{p_H}$.

1. $s_1 \geq \frac{1-p_H}{p_H}$ (the current signal induces a “high” guess). The relevant counterfactual set for s_2 is $\{s_2 < \frac{1-p_H}{p_H}\}$, i.e., signals that would reverse the guess to “low”. On this set, the average likelihood ratio in favor of the low state equals $F_L\left(\frac{1-p_H}{p_H}\right) / F_H\left(\frac{1-p_H}{p_H}\right)$, which corresponds to the denominator of the left-hand

²More precisely, if self-1's equilibrium strategy is to forgo $s_1 \in S_F \subseteq (0, \infty)$, self-2's rational belief on the high distribution (before observing s_2) equals $\frac{p_H \int_{S_F} f_H(s) ds}{p_H \int_{S_F} f_H(s) ds + (1-p_H) \int_{S_F} f_L(s) ds}$.

side of Equation (2). A larger value raises the opportunity cost of not memorizing s_2 , making the subject more inclined to save her memory capacity for s_2 .

2. $s_1 < \frac{1-p_H}{p_H}$ (the current signal induces a “low” guess). The relevant counterfactual set is $\{s_2 \geq \frac{1-p_H}{p_H}\}$, i.e., signals that would reverse the guess to “high”. On this set, the average likelihood ratio in favor of the high state equals $[1 - F_H(\frac{1-p_H}{p_H})]/[1 - F_L(\frac{1-p_H}{p_H})]$, which corresponds to the numerator of the left-hand side of Equation (2). A larger value raises the opportunity cost of memorizing s_1 and missing s_2 .

The left-hand side of Equation (2) therefore compares the informativeness of reversal-inducing signals across the two cases. When reversal signals favoring “high” guess are, in this average-likelihood sense, more informative than those favoring “low” guess, the DM has stronger incentives to memorize belief-confirming signals than belief-challenging signals.

More importantly, the right-hand side of Equation (2), $((1-p_H)/p_H)^2$, generates an asymmetric tendency to memorize belief-confirming rather than belief-challenging signals even when signals inducing “high” and “low” guesses are equally informative. Intuitively, when s_1 confirms a prior that favors the high state, the posterior shifts further toward “high”, reducing the probability that a subsequent s_2 would induce a “low” guess and thus lowering the opportunity cost of memorizing s_1 . In contrast, when s_1 challenges the prior, the posterior moves closer to $1/2$, so the probability that s_2 would induce a “high” guess—and the opportunity cost of memorizing s_1 —remains relatively high. This asymmetry in opportunity costs produces a biased updating pattern under memory constraints, consistent with related insights in [Wilson \(2014\)](#) and [Jehiel and Steiner \(2019\)](#).

Equation (3) has a structure and interpretation similar to Equation (2). Both conditions compare the informativeness of reversal-inducing signals in the same average-likelihood sense, and differ only in where the comparison is evaluated: Equation (2) at the prior threshold $\frac{1-p_H}{p_H}$, and Equation (3) at the smaller threshold $\tilde{\theta} < \frac{1-p_H}{p_H}$, which makes the latter more demanding. Intuitively, the opposite bias ($\bar{s} \geq \frac{1}{\underline{s}}$) must be ruled out in all equilibria, not just one, and $\tilde{\theta}$ captures the most demanding case. The two conditions become closer when p_H increases or when the information structure becomes noisier.³

Based on Proposition 1, the subject’s posterior after the first signal s depends on whether it is memorized. The posterior $\hat{p}_H(s)$ after receiving the first signal s

³For the signal distributions used in our experiment, numerical computation shows that an equilibrium with $\bar{s} < 1/\underline{s}$ exists for any $p_H > 0.5$, while no equilibrium with $\bar{s} \geq 1/\underline{s}$ exists for $p_H > 0.59$. For $p_H < 0.59$, equilibria with $\bar{s} < 1/\underline{s}$ and $\bar{s} \geq 1/\underline{s}$ may coexist, but the equilibrium with $\bar{s} \geq 1/\underline{s}$ is only marginal: $\bar{s}$ is barely larger than $1/\underline{s}$ ($\bar{s} \cdot \underline{s} \approx 1.014$).

can be written compactly as

$$\hat{p}_H(s) = (1 - \mathbb{1}_{s > \bar{s} \text{ or } s < \underline{s}}) \frac{p_H c}{p_H c + 1 - p_H} + \mathbb{1}_{s > \bar{s} \text{ or } s < \underline{s}} \left(\frac{p_H s}{p_H s + 1 - p_H} \right),$$

or equivalently as the log-odds form,

$$\log \frac{\hat{p}_H(s)}{1 - \hat{p}_H(s)} = \log \frac{p_H}{1 - p_H} + \mathbb{1}_{s > \bar{s} \text{ or } s < \underline{s}} \log s + (1 - \mathbb{1}_{s > \bar{s} \text{ or } s < \underline{s}}) \log c,$$

for $c = \frac{\int_{\underline{s}}^{\bar{s}} f_H(s') ds'}{\int_{\underline{s}}^{\bar{s}} f_L(s') ds'}$, i.e., the likelihood ratio of not memorizing s in the high versus the low state.

Equation (4) shows how forgetting produces underreaction at the signal level. We define underreaction as the subject's posterior being less responsive to the signal value s than the Bayesian posterior. When the subject memorizes s (i.e., $s \geq \bar{s}$ or $s \leq \underline{s}$), her posterior log-odds shift by $\log s$ —perfectly aligned with Bayes' rule. When she forgets s (i.e., $\underline{s} < s < \bar{s}$), her posterior log-odds shift only by $\log c$, which is independent of the actual signal value. Forgetting therefore eliminates the posterior's responsiveness to s : for any $s \in (\underline{s}, \bar{s})$, the subject's posterior is constant, while the Bayesian's posterior varies with $\log s$. Moreover, when the equilibrium thresholds satisfy $\bar{s} < 1/\underline{s}$, the forgotten range covers a wider span of belief-challenging signal values ($s < 1$) than of belief-confirming ones ($s > 1$); hence forgetting occurs disproportionately for belief-challenging signals, and the loss of responsiveness is concentrated on those belief-challenging signals. We summarize these implications in the following corollary.

Corollary 1. *In any equilibrium with $\bar{s} < 1/\underline{s}$:*

1. *the subject underreacts to signals relative to the Bayesian benchmark, strictly so for any $s \in (\underline{s}, \bar{s})$; and*
2. *underreaction is stronger for belief-challenging signals ($s < 1$) than for belief-confirming signals ($s > 1$).*

The corollary follows directly from Equation (4) together with the symmetry of the signal distributions in Equation (1) and the threshold structure obtained in Proposition 1. We test the resulting prediction in our experiment. The next section describes the experimental design, and we state the formal hypothesis at the end of the section.

3 Experimental Design

In the experiment, subjects complete 12 rounds of a guessing task involving belief updating with multiple signals.

3.1 States and Information in the Guessing Task

In each round of the guessing task, subjects observe two sequences of integers from 1 to 8. Each integer is drawn independently from one of two distributions, denoted “high” and “low”. The probability mass functions for these distributions are shown in Figure 1.

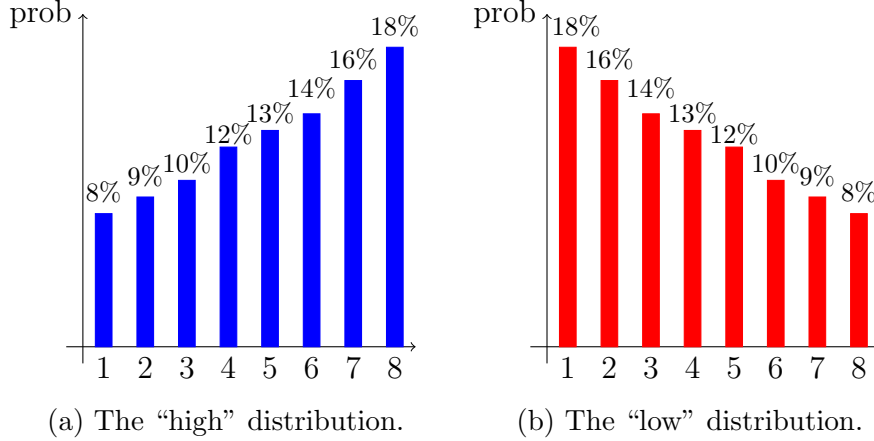


Figure 1: The two distributions shown in bar charts.

As shown in Figure 1, the “high” distribution is more likely to generate larger numbers, whereas the “low” distribution is more likely to generate smaller numbers. Therefore, subjects can infer which distribution generated the numbers they observe in a given round. We explain the rationale for the parameters of the two distributions in Online Appendix A.2. Briefly speaking, the distributions are sufficiently informative that subjects can easily make inferences, but not so informative that belief-challenging signals are rare; moreover, signal informativeness increases steadily as values move toward the extremes (1 and 8).

To make the task more accessible, we refer to the generating distributions as “computers”: The high computer is more likely to produce higher numbers, and the low computer is more likely to produce lower numbers.

3.2 Procedure of the Guessing Task

Pairing and Assignment to Treatment and Control. Subjects play 12 rounds of the guessing task and are alternately assigned to the treatment and control conditions across rounds. At the beginning of each round, each treatment subject is randomly matched with a control subject to form a pair. During that round, the two matched subjects observe the same sequence of numbers drawn from the same underlying distribution (either the “high” or “low” distribution with equal probability). This pairing and synchronized information allow us to single out the effect of memory overload by keeping the two subjects’ information constant.

Figure 2 illustrates the timeline of a round; we detail each step below.

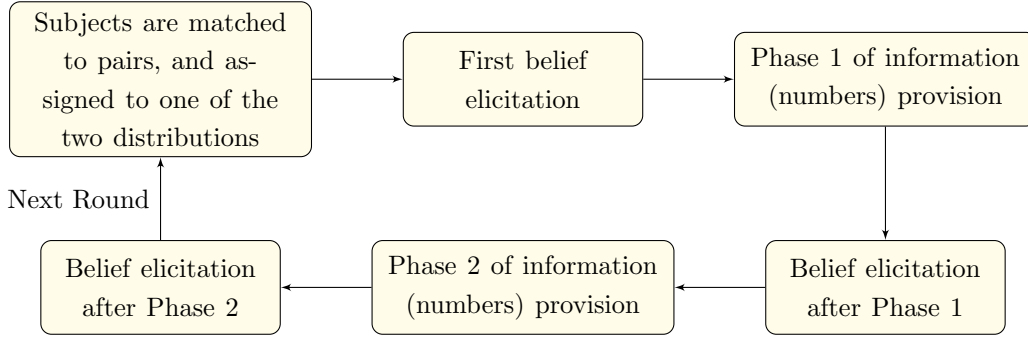


Figure 2: Sequence of a round.

First Belief Elicitation and Belief Elicitation Mechanism. Before seeing any numbers, subjects complete the first belief elicitation at the start of each round. We use a variant of the incentive-compatible binarized scoring rule (Hossain and Okui (2013)); a screenshot is provided in Figure A2 in the Appendix. During the belief elicitation, first, subjects indicate whether the “high” or “low” distribution is more likely to have been selected. Second, they choose between the following two options: earn 8€ if their chosen distribution is the true distribution, or earn 8€ with probability $x\%$, where x starts at 50% and increases in 5% increments per row until 95%. As shown in Hossain and Okui (2013), the mechanism is incentive-compatible.

This first elicitation checks whether subjects hold a 50-50 prior before seeing any numbers and ensures they recognize that a new round has begun. There is a soft 30-second time limit for this (and subsequent) belief elicitation.⁴ Afterwards, subjects see two sequences of numbers drawn by the selected distribution in two phases of information provision, with a second belief elicitation between the two phases and a third elicitation after the second phase.

Phase 1 of Information Provision. In Phase 1, all subjects see five numbers displayed on the screen for 30 seconds (Figure 3). The two matched subjects in a pair (one in the treatment and another in the control condition) see the same five numbers. After 30 seconds, subjects are redirected to the second belief-elicitation page.

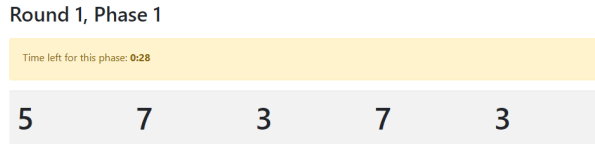


Figure 3: A screenshot of Phase 1.

⁴If a subject spends more than 30 seconds, a warning appears indicating that time is up. However, subjects are not automatically redirected to the next page.

Belief Elicitation After Phase 1. After Phase 1, we elicit subjects’ beliefs using the same table as in the first elicitation. Subjects’ choices from the first elicitation are shown as the default. Phase 1 naturally induces heterogeneous posteriors across subjects away from 50-50. This heterogeneity allows us to define belief-confirming and belief-challenging information and to study whether the treatment effect on belief updating is asymmetric across these two types of information.

Phase 2 of Information Provision. In Phase 2, subjects can see up to seven numbers within a strict time limit of 30 seconds. Paired subjects in the treatment and control conditions see the same numbers but under different screen layouts.

The Treatment Condition The treatment layout is illustrated in Figure 4. Subjects see one number at a time and decide when to advance by clicking the blue “Next” button. When they click “Next”, the next number is revealed and the preceding number disappears. Moreover, subjects cannot return to previous numbers. Thus, treatment subjects face a trade-off between spending more time on the current number and saving time for subsequent numbers.

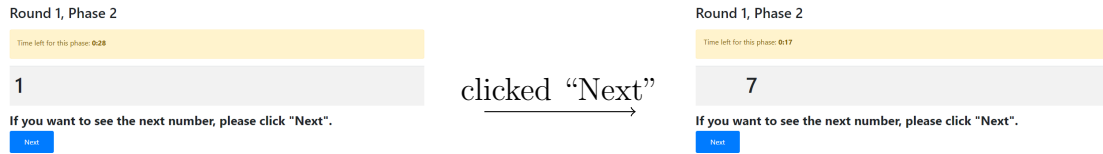


Figure 4: A screenshot of Phase 2 in the treatment condition.

The Control Condition In the control condition, subjects cannot influence when the next number appears, and preceding numbers remain visible as additional numbers are displayed (Figure 5). They start with one number on the screen and, as they advance through the sequence, see two, three, four, etc., numbers on the screen simultaneously. To ensure identical information and timing within each pair of subjects, control subjects advance through the sequence at the same moments that their matched treatment counterparts click “Next”. However, subjects are unaware that their clicks determine their counterpart’s progression, and control subjects are unaware that a counterpart controls their progression.



Figure 5: A screenshot of Phase 2 in the control condition.

The main difference between the treatment and control conditions is that subjects face no (or minimal) memory constraints in the control condition, unlike their counterparts in the treatment condition. The control subjects do not have to memorize each signal, and it is easier for them to aggregate the information conveyed by the Phase 2 numbers. In other words, subjects in the **control** condition are exposed to **lower** memory overload than those in the treatment condition, even though they receive the same information.

Belief Elicitation After Phase 2. By comparing how subjects update their beliefs across treatment and control for belief-confirming versus belief-challenging information, we draw insights into whether memory overload has an asymmetric impact on belief updating with respect to the two types of information.

New Round. Following the belief elicitation after Phase 2, a new round begins. Subjects are randomly re-matched and assigned to either the “high” or “low” distribution with equal probability. They are redirected to a screen reminding them that a new round has started.

3.3 Procedural Details

We conducted 12 sessions of the experiment involving 260 subjects at a university laboratory. The participants were university students and recruited via h-root (Bock et al. (2014)). The experiment was programmed and run in oTree (Chen et al. (2016)). Each session lasted approximately 2 hours and 15 minutes. For payment, three rounds were selected at random; from each of those rounds, one decision (from the first, second, and third belief elicitation, respectively) was paid. Average earnings were 18.12€ per subject, plus a 7€ participation fee.

3.4 Hypothesis

Our experimental design is constructed to isolate the causal effect of memory load on belief updating, holding the information set constant across treatment and control. As noted above, the treatment condition imposes a higher memory load than the control. Proposition 1 characterizes the consequence: in any equilibrium with $\bar{s} < 1/\underline{s}$, memory load induces forgetting—and hence reduced responsiveness—concentrated on belief-challenging signals (the forgotten range $(\underline{s}, \bar{s})$ covers a wider span of $s < 1$ than $s > 1$). The control condition, by contrast, imposes minimal memory load: preceding signals remain visible and need not be memorized, so there is no asymmetric forgetting in the control. Comparing the two conditions, we therefore expect the treatment effect—the additional underreaction induced by memory load—to be concentrated on belief-challenging signals while leaving updating to belief-confirming signals largely unchanged. Formally:

Hypothesis 1. *Subjects in the treatment and control conditions underreact similarly to belief-confirming information, but subjects in the treatment condition underreact more to belief-challenging information than subjects in the control condition.*

4 Descriptive Statistics and Main Variables

In this section, we present descriptive statistics for our sample before introducing the main variables of interest.

4.1 Data

The average age of the subjects was 23, and the maximum age was 30. A total of 109 subjects were male, and 151 subjects were female. 50% of the subjects reported that they had taken economics or statistics courses.

4.1.1 Observations

Each subject in each round provides one observation for the guessing task. The 260 subjects played 12 rounds, yielding a total of 3,120 observations. Unfortunately, a technical glitch in the first four sessions sometimes under-recorded the number of Phase-2 signals subjects have seen by one. For example, the system might record 6 even when the subject saw 7 numbers in Phase 2. Thus, when the recorded number of Phase-2 signals is less than 7, we cannot distinguish whether the subject truly saw fewer than 7 numbers or whether the count was misrecorded. Therefore, we dropped observations from the first four sessions if the recorded number of Phase-2 signals seen by the subject was less than 7, amounting to 264 out of 1,008 observations.⁵ The technical glitch was fixed in later sessions. After dropping these observations, we have 2,856 observations. Online Appendix A.4 reports robustness checks in which we exclude all observations from the first four sessions; the results are unchanged.

Furthermore, in the analysis presented in the main text, we further restrict to observations where the first elicited belief is compatible with a 50-50 prior over the high and low distributions.⁶ The first elicitation checks that subjects understood the mechanism, recognized that they were at the beginning of a new round, and knew a new distribution had been drawn with equal probability. In 615 of the 2,856

⁵As shown in Table 1, in almost all (97%) of the observations, subjects saw all 7 numbers in Phase 2. Therefore, we are confident that dropping these data does not systematically affect our results. Note that among the 264 of 1,008 observations, some subjects did see all 7 numbers but were not recorded as such, which is why the fraction 264/1008 is much higher than the 3% rate we observed in sessions 5 to 12.

⁶That is, the subject chose option 2 (earning 8€ with a fixed probability) in all rows in the table shown in Figure A2, except possibly in the first row.

Table 1: Frequencies and proportions of observations where different numbers of signals have been seen in Phase 2.

Number of signals seen in Phase 2	All sessions, restricted to observations with first elicited belief (50%, 50%)		Sessions 5 - 12, restricted to observations with first elicited belief (50%, 50%)	
	Frequency	Percentage	Frequency	Percentage
4	2	0.09%	2	0.12%
5	9	0.40%	9	0.55%
6	34	1.52%	34	2.07%
7	2196	97.99%	1691	97.27%

Notes. Note that in this table, we do not include observations in sessions 1 to 4 where the number of signals seen in Phase 2 is less than 7. Throughout all sessions, no subjects saw fewer than four signals.

observations (21.53%), subjects’ choices in the first belief elicitation fail the check.^{7,8} Including these 615 observations does not change our results, and the analysis is presented in Online Appendix A.3. After excluding these observations, the final analysis sample contains 2,241 observations. Unless otherwise noted, all results in the main text use this sample.

4.1.2 Numbers of Signals Seen in Phase 2

Table 1 shows descriptive statistics for the number of signals subjects have seen in Phase 2. Only a small fraction of subjects saw fewer than 7 numbers. In sessions 5–12, where the technical glitch was absent, fewer than 3% of subjects saw fewer than 7 numbers. Therefore, our results are unlikely to be driven by dropping first-four-session observations in which the recorded number of Phase-2 signals was below seven.

4.1.3 Randomization of High and Low States

Among all 2,241 final observations, 48.77% were assigned to the “high” distribution, whereas 51.23% were assigned to the “low” distribution. The composition is not exactly half-and-half because we dropped some observations as noted above.

⁷Among subjects who have never taken statistics or economics courses, 26.70% of observations fail the check, whereas the proportion is 16.27% among subjects who have taken such courses. We control for whether subjects have taken statistics/economics courses in all analyses.

⁸For example, in 84 observations, subjects’ choices in the first belief elicitation indicate at least 95% confidence in one distribution (subjects chose option 1 in all rows in the table shown in Figure A2). This might be due to a misunderstanding of the belief elicitation mechanism. However, most subjects corrected such mistakes in later rounds; only 6 subjects indicated at least 95% confidence in more than six rounds.

4.1.4 Treatment and Control Condition

Each subject was alternately assigned to the treatment and control conditions across the 12 rounds of the experiment. Among the 2,241 final observations, 1,124 are from the control condition and 1,117 are from the treatment condition. Again, the composition is not exactly half-and-half because of dropped observations.

4.2 Variables of Interest

4.2.1 Elicited Beliefs

We denote the first elicited belief by p_0 and the elicited beliefs after Phase 1 and Phase 2 by p_1 and p_2 , respectively. All elicited beliefs are normalized so that p_0 , p_1 , and p_2 represent the subject's assessed probability that the "high" distribution was chosen. By construction, $p_0 = 0.5$ in all observations used in the analysis, as we exclude observations with $p_0 \neq 0.5$. In contrast, p_1 and p_2 are pinned down by the point at which subjects switch from option 1 to option 2 in the belief elicitation table. For example, if after Phase 1 a subject judged the "high" distribution more likely and switched from option 1 to option 2 when the random-lottery winning probability reached 70%, then $p_1 = (0.70 + 0.65)/2 = 0.675$, the midpoint of the compatible belief interval (0.65 to 0.70). In the analysis, we treat p_1 as the subject's prior for Phase 2 and study how beliefs update from p_1 to p_2 across control and treatment conditions, conditioning on whether Phase-2 signals are belief-confirming or belief-challenging.

4.2.2 Bayesian Beliefs

The Bayesian counterpart of the first elicited belief is denoted p_0^B ; correspondingly, the Bayesian counterparts of the elicited beliefs after Phase 1 and Phase 2 are p_1^B and p_2^B , respectively. Throughout the paper, to ensure comparability with the elicited beliefs, p_1^B and p_2^B are discretized into bins of size 0.05 and take the values equal to the midpoint of their bins (e.g., values in $[0.65, 0.70)$ are mapped to 0.675).

Because the state is drawn with equal probability, p_0^B is always equal to 0.5. The Bayesian belief after Phase 1, p_1^B , is constructed from the first elicited belief p_0 using the Phase-1 signals:

$$\frac{p_1^B}{1 - p_1^B} = \prod_{s_i \in \mathcal{S}_1} \frac{f_H(s_i)}{f_L(s_i)} \times \frac{p_0}{1 - p_0},$$

where $\mathcal{S}_1$ is the set of five numbers observed in Phase 1, $f_H(s_i)$ is the probability that the "high" distribution generates s_i , and $f_L(s_i)$ is the probability under the "low" distribution. Since $p_0 = 0.5$ in our analysis sample, this equals

$$\frac{p_1^B}{1 - p_1^B} = \prod_{s_i \in \mathcal{S}_1} \frac{f_H(s_i)}{f_L(s_i)} \times \underbrace{\frac{0.5}{1 - 0.5}}_{=1}.$$

Similarly, the Bayesian belief after Phase 2, p_2^B , is constructed by taking the elicited p_1 as the prior and updating with the Phase-2 signals:

$$\frac{p_2^B}{1 - p_2^B} = \prod_{s_i \in \mathcal{S}_2} \frac{f_H(s_i)}{f_L(s_i)} \times \frac{p_1}{1 - p_1},$$

where $\mathcal{S}_2$ is the set of numbers observed in Phase 2. In other words, p_2^B is the belief a Bayesian decision-maker would hold if they took their elicited prior p_1 as given and updated on $\mathcal{S}_2$ in a statistically optimal way.⁹

4.2.3 Treatment and Control Condition

For each observation i , the subject’s condition is denoted by the dummy variable T_i , which takes the value 1 if the subject is assigned to the treatment condition and 0 if the subject is assigned to the control condition.

5 Results

This section presents our experimental results. Section 5.1 verifies sensible updating in Phase 1 and observational equivalence between treatment and control before the experimental manipulation. Section 5.2 tests Hypothesis 1. Section 5.3 explores the underlying mechanism by running more granular tests of the treatment effect.

5.1 Preliminaries: Updating in Phase 1

Before presenting the main analysis on the treatment effect of memory load (in Phase 2), we analyze the relationship between elicited and Bayesian beliefs in Phase 1. While they are not the main focus, these results indicate whether subjects understand the experiment and whether our belief elicitation mechanism recovers “sensible” belief-updating behavior.

Specifically, we estimate the following two regression equations:

$$\begin{aligned} p_{1,i} &= \beta_0 + \beta_1 p_{1,i}^B + X_i + \epsilon_i; \\ p_{1,i} &= \beta_0 + \beta_1 p_{1,i}^B + \beta_2 T_i + \beta_3 T_i \times p_{1,i}^B + X_i + \epsilon_i, \end{aligned}$$

where X_i includes age, gender and coursework controls. Throughout our analyses, standard errors are clustered at the subject level. The results are presented in Table 2.

⁹This approach isolates Phase-2 updating from Phase-1 updating. We view it as unlikely that subjects revised their interpretation of Phase-1 signals during Phase 2 given the short time limit and memory load; moreover, the previously elicited belief is displayed as the default in the next elicitation, providing an anchor that separates belief formation across phases.

	(1) Elicited belief	(2) Elicited belief
Bayesian belief after Phase 1	0.764*** (0.017)	0.768*** (0.020)
Treatment		0.001 (0.013)
Treatment $\times$ Bayesian belief after Phase 1		-0.008 (0.022)
R^2	0.702	0.702
Observations	2241	2241
Subjects	235	235
Clustered standard errors at the subject level in parentheses		
[†] $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$		

Table 2: Regression analysis of elicited beliefs after Phase 1. Both Bayesian and elicited beliefs take the values of the midpoint of their respective 5% bins. Standard errors are clustered at the subject level; we control for age, gender, and statistics/economics coursework. Constants are not shown in the Table because they are subsumed by the binary fixed effects. Subjects update their beliefs in the same direction as the Bayesian benchmark, but underreact to information.

In all columns of Table 2, the regression coefficients of Bayesian belief β_1 are significantly positive but smaller than 1. Taken together, both findings suggest that, on average, subjects correctly infer the direction of the signal, i.e., they believe more in the “high” (“low”) distribution when they receive higher (lower) numbers, but underreact to signals compared to the Bayesian benchmark.¹⁰ This result coheres with the findings presented in Eil and Rao (2011), Liang (2025) and many other belief-updating experiments that involve binary states (Benjamin (2019); Ba et al. (2025)). Moreover, as a sanity check, column 2 shows that the treatment setting has no effect on the elicited belief nor the relationship between the elicited and Bayesian belief. Specifically, the coefficient on Treatment and the coefficient on the interaction Treatment $\times$ Bayesian belief are both statistically insignificant, consistent with the fact that subjects in the two conditions face identical displays during Phase 1.

5.2 Main Results: Asymmetric Updating in Phase 2

We now analyze subjects’ belief updating in Phase 2 and test our main hypothesis: higher memory load in the treatment condition leads to stronger underreaction to

¹⁰Note that $\beta_1 < 1$ unambiguously captures underreaction here because the average Bayesian posterior and the average elicited belief are both close to 0.5 (reminiscent of the martingale property of Bayesian beliefs): average $p_1^B = 0.498$ and average $p_1 = 0.527$ in Phase 1; average $p_2^B = 0.515$ and average $p_2 = 0.530$ in Phase 2. The fitted regression line therefore passes close to (0.5, 0.5), so $\beta_1 < 1$ makes it flatter than the 45° Bayesian benchmark, capturing underreaction.

belief-challenging signals than in the control condition, while underreaction to belief-confirming signals is similar in the two conditions. Specifically, we run the following regressions:

$$\begin{aligned} p_{2,i} &= \beta_0 + \beta_1 p_{2,i}^B + X_i + \epsilon_i; \\ p_{2,i} &= \beta_0 + \beta_1 p_{2,i}^B + \beta_2 T_i + \beta_3 T_i \times p_{2,i}^B + X_i + \epsilon_i, \end{aligned} \tag{5}$$

Table 3 reports the results. Note that column 3 includes only observations where Phase-2 signals are, in aggregate, belief-confirming; column 4 includes only observations where Phase-2 signals are, in aggregate, belief-challenging. Comparing columns 3 and 4 tests our hypothesis: whether the treatment asymmetrically affects belief updating with belief-confirming versus belief-challenging information.

First, as in Phase 1, the coefficient on Bayesian belief is significantly positive but smaller than 1 in all four columns, reaffirming that subjects understand what state the signals support but underreact relative to the Bayesian benchmark.

Second, comparing columns 3 and 4, the coefficient on Bayesian belief is smaller with belief-challenging signals (0.521) than with belief-confirming signals (0.762), indicating stronger underreaction to belief-challenging signals than to belief-confirming signals in Phase 2, across both treatment and control conditions.

Third, and most importantly, treatment has an asymmetric impact on the belief updating behavior with belief-confirming and belief-challenging information. In the confirmatory subsample (column 3), neither the treatment main effect nor its interaction with the Bayesian belief is significant, implying that memory load has no significant impact on how subjects update their beliefs with belief-confirming signals. In contrast, in the challenging subsample (column 4), the interaction between treatment and the Bayesian belief is significantly negative, indicating that memory load further flattens the slope and thereby increases underreaction to belief-challenging signals.¹¹ Moreover, the interaction coefficient in column 4 is significantly more negative than in column 3 (the p-value of a two-tailed t-test is 0.024). These patterns support Hypothesis 1: Higher memory load leads to greater underreaction to belief-challenging signals, but has little impact on the underreaction to belief-confirming signals.

Alternative specification based on log-odds deviation. Next, we present a complementary regression analysis based on the log-odds form of Bayes' rule. This alternative specification not only serves as a robustness check of our main result, but as it additively separates the treatment effect, it allows us to easily extend our

¹¹The positive and significant standalone treatment effect (0.057) is a by-product of the martingale property of beliefs: Within the treatment condition, a flatter slope rotates the regression line around (0.5, 0.5), increasing the y-intercept.

	(1)	(2)	(3)	(4)
	Elicited belief	Elicited belief	Elicited belief (with confirmatory signals)	Elicited belief (with challenging signals)
Bayesian belief after Phase 2	0.697*** (0.014)	0.708*** (0.017)	0.762*** (0.016)	0.521*** (0.033)
Treatment		0.017 (0.010)	0.005 (0.011)	0.057* (0.024)
Treatment $\times$ Bayesian belief after Phase 2		-0.023 (0.015)	-0.002 (0.014)	-0.101* (0.041)
R^2	0.728	0.729	0.864	0.346
Observations	2241	2241	1327	914
Subjects	235	235	230	225

Clustered standard errors at the subject level in parentheses

$^\dagger p < 0.1$, $^* p < 0.05$, $^{**} p < 0.01$, $^{***} p < 0.001$

Table 3: Regression analysis of elicited beliefs after Phase 2. Both Bayesian and elicited beliefs take values equal to the midpoint of their respective 5% bins. Standard errors are clustered at the subject level; we control for age, gender, and statistics/economics coursework. Constants are not shown in the Table because they are subsumed by the binary fixed effects. As in Phase 1, subjects underreact to information, and they underreact more to challenging than to confirming signals. The treatment amplifies underreaction only for belief-challenging signals.

analysis in Section 5.3.¹²

We estimate the following regression on the log-odds deviation $\log \frac{p_2}{1-p_2} - \log \frac{p_2^B}{1-p_2^B}$, i.e., the difference between the log-odds of the subject's elicited posterior and the Bayesian posterior:

$$\log \frac{p_2}{1-p_2} - \log \frac{p_2^B}{1-p_2^B} = \beta_0 + \beta_1 T_i + \beta_2 \text{Challenging}_i + \beta_3 T_i \times \text{Challenging}_i + X_i + \epsilon_i, \quad (6)$$

where Challenging_i is a dummy variable that equals 1 if and only if Phase-2 signals are, in aggregate, belief-challenging. β_1 measures the treatment effect on updating to belief-confirming signals, while $\beta_1 + \beta_3$ measures the treatment effect on updating to belief-challenging signals.

Table 4 presents the results. Our main focus is on column 1, 2 and 3, where we pool the negative of the log-odds deviation $\log \frac{p_2}{1-p_2} - \log \frac{p_2^B}{1-p_2^B}$ from low prior observations with $\log \frac{p_2}{1-p_2} - \log \frac{p_2^B}{1-p_2^B}$ from high prior observations. Such pooling leverages the symmetry of the experimental setup and the log-odds form of Bayes' rule, so that positive (resp. negative) values with belief-challenging (resp. belief-

¹²To see that, note that the log-odds form of Bayes' rule follows:

$$\log \frac{p_2^B}{1-p_2^B} = \log \prod_{s_i \in \mathcal{S}_2} \frac{f_H(s_i)}{f_L(s_i)} + \log \frac{p_1}{1-p_1},$$

therefore additively separating the impact of belief-updating (first term of the RHS) and the prior (second term of the RHS).

confirming) signals consistently indicate underreaction regardless of the prior. The pooling also enables us to improve the statistical power of our analysis. To easily illustrate the asymmetric treatment effect, columns 2 and 3 present the regression analysis with only belief-confirming and belief-challenging Phase-2 signals. For a robustness check, columns 4 and 5 split the data by prior ($p_1 > 0.5$ versus $p_1 < 0.5$).

In the pooled regression (column 1), treatment has a negative but insignificant standalone effect, while the interaction between the treatment and belief-challenging dummy variable is positive and significant. Thus, memory load has no effect when Phase-2 signals are (in aggregate) belief-confirming, but strengthens underreaction when Phase-2 signals in aggregate are belief-challenging, consistent with Hypothesis 1.¹³ Columns 2 and 3 show similar results: The treatment does not significantly impact the belief-updating behavior with confirmatory signals, but strengthens underreaction to belief-challenging signals. The magnitude of treatment-induced underreaction toward belief-challenging signals is equal to 0.168, whereas the corresponding estimate for belief-confirming signals is equal to $|-0.053|$.¹⁴ Similar to Table 3, these patterns support Hypothesis 1: Higher memory load leads to greater underreaction to belief-challenging signals, but has little impact on the underreaction to belief-confirming signals.

Comparing the two specifications. The log-odds specification (Table 4) reproduces the findings of the belief-level regression in Table 3: the treatment significantly increases underreaction to belief-challenging signals but does not significantly affect updating to belief-confirming signals. On the other hand, the difference between the treatment-induced underreaction to belief-challenging and to belief-confirming signals is significant in the belief-level regression ($p = 0.024$); in the log-odds specification it points in the same direction but is not significant ($p = 0.212$).¹⁵ The reason is that the two specifications differ in how they weight observations, which we discuss below.

For a given updating difference in percentage points, the log-odds deviation is larger when the Bayesian belief is more extreme.¹⁶ The log-odds specification

¹³The results in column 4 and 5 are similar, although the effects are less significant due to smaller sample sizes.

¹⁴Note that within column 1, the treatment induced underreaction toward belief-confirming signals equals 0.064 which is close to 0.053, the estimate in column 2; the treatment induced underreaction toward belief-challenging signals equals $0.224 - 0.064 = 0.16$ which is close to 0.168, the estimate in column 3. The differences are only due to the fact that we control for pair fixed effects in column 2 and 3, but not in column 1.

¹⁵To synthesize evidence across the two t-tests, we apply the Cauchy Combination test (Liu and Xie (2020)); the combined p-value is 0.022. Thus, we conclude that our analysis supports that memory overload induces stronger underreaction to belief-challenging signals than to belief-confirming signals. That said, we acknowledge the restrictions of our analysis, and an analysis with a larger sample would be fruitful for future work.

¹⁶For example, consider a prior equal to 50% with a 5 percentage point underreaction toward signals favoring the high state. When the Bayesian belief equals 57.5% (the elicited belief equals 52.5%), the log-odds deviation equals -0.088 ; when the Bayesian belief equals 97.5% (the elicited

	(1)	(2)	(3)	(4)	(5)
	Pooled	Pooled (with confirming signals)	Pooled (with challenging signals)	High prior	Low prior
Treatment	-0.064 (0.048)	-0.053 (0.054)	0.168* (0.070)	-0.048 (0.074)	0.093 (0.070)
Challenging	1.108*** (0.079)			1.149*** (0.103)	-1.068*** (0.100)
Treatment $\times$ Challenging	0.224** (0.086)			0.209† (0.124)	-0.235* (0.116)
Pair Controls	No	Yes	Yes	No	No
R^2	0.262	0.788	0.803	0.273	0.262
Observations	2241	1327	914	1219	1022
Subjects	235	230	225	229	222

Clustered standard errors at the subject level in parentheses

† $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4: Regression analysis of the log-odds deviation $\log \frac{p_2}{1-p_2} - \log \frac{p_2^B}{1-p_2^B}$. Standard errors are clustered at the subject level; age, gender, and statistics/economics coursework controls are included in all five columns. We control for pair fixed effects in columns 2 and 3, but not in columns 1, 4, and 5 as the pair controls are collinear with the challenging dummy variable. Constants are not shown in the Table because they are subsumed by the binary fixed effects. Columns 1, 2 and 3 pool the negative of the log-odds deviation from the low prior data with the log-odds deviation from the high prior data; Column 2 uses only observations where Phase-2 signals are in aggregate belief-confirming; Column 3 uses only observations where Phase-2 signals are in aggregate belief-challenging; Column 4 restricts to high prior ($p_1 \geq 0.5$) data; Column 5 restricts to low prior ($p_1 < 0.5$) data.

therefore places more weight on observations with extreme Bayesian beliefs, for example, those arising from highly informative Phase-2 signals. But by Proposition 1, when signals are very informative the subject memorizes both belief-confirming and belief-challenging signals, so there is no asymmetry among these observations. The log-odds specification thus upweights precisely the observations where the asymmetric treatment effect is weakest, attenuating the estimated effect relative to the belief-level regression in Table 3.

Relatedly, because the log-odds specification is most sensitive to extreme beliefs while our 5% elicitation bins are uniform over $[0, 1]$, measurement error at extreme beliefs inflates the variance of the log-odds estimates. For these reasons, we treat the belief-level regression (Table 3) as our primary specification and the log-odds analysis as a complementary check that yields the same qualitative conclusion.

5.3 Exploring the Underlying Mechanism

While our main analysis classifies Phase-2 evidence in aggregate as confirming or challenging, here we treat the seven Phase-2 signals individually to shed light on the mechanism behind the asymmetric treatment effect. We focus on observations where subjects saw all seven signals, and perform two regression analyses extending the log-odds specification.

Testing the fixed-capacity mechanism. We first note that our theory with limited memory capacity implies that attention spent on different signals is substitutable. Memory capacity not spent on the 1st to 6th signals would be spent on the 7th signal, as there are no future signals beyond the 7th signal. If the 1st to 6th signals are, in aggregate, belief-challenging, the treatment leads to less memory capacity spent on them, and subjects allocate their residual capacity to the 7th signal. This effectively reduces underreaction to the 7th signal. We test this prediction by extending the regression Equation (6):

$$\begin{aligned} \log \frac{p_2}{1 - p_2} - \log \frac{p_2^B}{1 - p_2^B} = & \beta_0 + \beta_1 T_i + \beta_2 \text{Challenging } 1^{\text{st}}\text{-}6^{\text{th}}_i \\ & + \beta_3 \text{Challenging } 7^{\text{th}}_i + \beta_4 \text{Challenging } 1^{\text{st}}\text{-}6^{\text{th}}_i \times \text{Challenging } 7^{\text{th}}_i \\ & + \beta_5 T_i \times \text{Challenging } 1^{\text{st}}\text{-}6^{\text{th}}_i + \beta_6 T_i \times \text{Challenging } 7^{\text{th}}_i \\ & + \beta_7 T_i \times \text{Challenging } 1^{\text{st}}\text{-}6^{\text{th}}_i \times \text{Challenging } 7^{\text{th}}_i + X_i + \epsilon_i. \end{aligned} \quad (7)$$

“Challenging $1^{\text{st}}\text{-}6^{\text{th}}$ ” equals 1 if the first six signals are, in aggregate, belief-challenging; “Challenging 7^{th} ” equals 1 if the 7th signal is belief-challenging.

Table 5 presents the results. We focus on the triple-interaction coefficient β_7 , belief equals 92.5%), the log-odds deviation equals -0.50 (a 5.7 times increase).

i.e., the interaction effect between the treatment, “Challenging 1st-6th” and “Challenging 7th” dummy variables. The fixed-capacity mechanism predicts a negative β_7 : when the first six signals are in aggregate belief-challenging, the treatment-induced underreaction to a belief-challenging 7th signal is reduced. In column 1, we present the result where we pool low and high prior data; in column 2 and 3, we focus on high and low prior data respectively. Across all three columns the coefficients β_7 are not significant ($p = 0.393, 0.849, 0.428$ respectively), and their signs are positive and thus opposite of what the capacity mechanism predicts. Moreover, the treatment effects on the 7th signal are not significantly different from those on the 1st to 6th signals. These patterns are therefore inconsistent with the fixed-capacity mechanism.

The number of challenging signals. Next, we analyze how the number of challenging signals subject see affects their underreaction by estimating the following variant of Equation (6),

$$\log \frac{p_2}{1-p_2} - \log \frac{p_2^B}{1-p_2^B} = \beta_0 + \beta_1 T_i + \beta_2 \# \text{ Challenging}_i + \beta_3 T_i \times \# \text{ Challenging}_i + X_i + \epsilon_i, \quad (8)$$

in which we replace the challenging dummy variable with the number of challenging signals that subjects see in Phase 2 ($\# \text{ Challenging}_i$). Table 6 shows that, in the pooled sample (column 1), the interaction between treatment and the number of challenging signals is positive (0.068) and statistically significant ($p = 0.038$). Quantitatively, the treatment-induced underreaction with seven challenging signals ($0.068 \times 7 = 0.476$) exceeds the marginal underreaction from one additional challenging signal in the control condition (0.370). Column 2 and 3 show similar but less significant results, due to smaller sample sizes.

Discussion In summary, Tables 5 and 6 suggest that attention across the seven signals is not reallocated in the manner predicted by a fixed-capacity mechanism: we find no evidence that residual attention shifts to the final signal when earlier signals are challenging.¹⁷ Instead, the treatment-induced underreaction strengthens as subjects receive more belief-challenging signals. A natural alternative explanation is a per-signal cost mechanism in which the cognitive cost of incorporating belief-challenging evidence exceeds that for belief-confirming evidence; as memory load increases, these costs rise and disproportionately dampen updating with belief-challenging signals. In this mechanism, attention is effectively independent across signals. We view identifying the precise cognitive process by which memory

¹⁷This test assumes that capacity not spent on the first six signals is allocated to the seventh signal within the same round. If instead residual capacity is carried over to subsequent rounds or to other cognitively demanding tasks, reallocation need not manifest as reduced underreaction to the seventh signal, in which case our test would not detect it.

	(1) Pooled	(2) High prior	(3) Low prior
Treatment	-0.047 (0.067)	-0.076 (0.099)	0.025 (0.092)
Challenging 1 st -6 th	1.167*** (0.090)	1.312*** (0.121)	-0.976*** (0.109)
Challenging 7 th	0.342*** (0.087)	0.383** (0.116)	-0.290* (0.125)
Challenging 1 st -6 th × Challenging 7 th	-0.129 (0.132)	-0.380 [†] (0.198)	-0.207 (0.176)
Treatment × Challenging 1 st -6 th	0.061 (0.117)	0.106 (0.157)	-0.008 (0.170)
Treatment × Challenging 7 th	0.080 (0.113)	0.122 (0.155)	-0.040 (0.157)
Treatment × Challenging 1 st -6 th × Challenging 7 th	0.132 (0.166)	0.199 (0.232)	-0.049 (0.258)
R^2	0.273	0.293	0.271
Observations	2196	1193	1003
Subjects	235	228	222

Clustered standard errors at the subject level in parentheses

[†] $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 5: Regression analysis of the log-odds deviation $\log \frac{p_2}{1-p_2} - \log \frac{p_2^B}{1-p_2^B}$ that tests the prediction of the capacity mechanism. Standard errors are clustered at the subject level; we control for age, gender, and statistics/economics coursework. Constants are not shown in the Table because they are subsumed by the binary fixed effects. Column 1 pools high and low priors; column 2 restricts to high priors ($p_1 \geq 0.5$); column 3 restricts to low priors ($p_1 < 0.5$). “Challenging 1st-6th” equals 1 if signals 1–6 are, in aggregate, belief-challenging; “Challenging 7th” equals 1 if the 7th signal is belief-challenging.

	(1) Pooled	(2) High prior	(3) Low prior
Treatment	-0.187 (0.103)	-0.205 (0.147)	0.189 (0.146)
# Challenging	0.370*** (0.029)	0.392*** (0.034)	-0.337*** (0.039)
Treatment $\times$ # Challenging	0.068* (0.033)	0.075 [†] (0.044)	-0.065 (0.045)
R^2	0.265	0.295	0.240
Observations	2196	1193	1003
Subjects	235	228	222
Clustered standard errors at the subject level in parentheses			
[†] $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$			

Table 6: Regression analysis of the log-odds deviation $\log \frac{p_2}{1-p_2} - \log \frac{p_2^B}{1-p_2^B}$ that analyzes how the number of Phase-2 challenging signals affects the treatment effect. Standard errors are clustered at the subject level; we control for age, gender, and statistics/economics coursework. Constants are not shown in the Table because they are subsumed by the binary fixed effects. Column 1 pools high and low priors; column 2 restricts to high priors ($p_1 \geq 0.5$); column 3 restricts to low priors ($p_1 < 0.5$). # Challenging_{*i*} is the number of Phase-2 challenging signals.

load affects belief updating as an important question for future work, with potentially distinct policy implications.

6 Concluding Remarks and Implications

In this study, we investigate the role of memory overload in asymmetric underreaction to belief-challenging versus belief-confirming information. We show that when subjects are exposed to stronger memory load, they underreact more to belief-challenging information. In contrast, our data do not find that stronger memory overload leads to greater underreaction to belief-confirming information. Moreover, the evidence is consistent with a mechanism in which incorporating belief-challenging signals is more memory-costly than incorporating belief-confirming signals.

In addition to the popular view that confirmation bias, or asymmetric updating behavior, arises from intrinsic preferences for belief-confirming information, our findings demonstrate that asymmetry in updating also depends on the informational environment. This lends credence to the growing theoretical literature showing how limited attention and ability could generate a number of behavioral anomalies. In the following, we discuss implications for economic theory, information campaigns, and managerial practice.

Modeling Limited Cognitive Resources Recently, there has been increased attention to how imperfect cognition influences economic decisions (see the discussion and summary in [Enke \(2024\)](#)). Our results speak to an understudied question: whether memory resources are substitutable or effectively independent across multiple signals. We find no evidence of substitution (i.e., reallocation to a final signal when earlier signals are belief-challenging), but we do find that the treatment effect grows with the number of belief-challenging signals, consistent with differential per-signal memory costs. We view our study as a first step and encourage further theoretical and experimental work to refine these characterizations.

Information Campaigns & Advertising Information campaigns have become ever so popular because of technological advancements that facilitate information dissemination. Yet, research on information campaigns on different topics across immigration, health, and energy conservation reports mixed results on their effectiveness ([Snyder \(2007\)](#); [Espinosa and Stoop \(2021\)](#); [Andor et al. \(2022\)](#); [Facchini et al. \(2022\)](#)). Our results suggest that one reason is memory overload: simply sending more information, without regard to recipients’ memory capacity, can disproportionately strengthen underreaction to belief-challenging content and undermine or even backfire on the campaign’s objectives. A more effective approach is to design messages that lower memory demands—e.g., by simplifying formats, aggregating key points, and reducing dispersion across sources—so that challenging information is easier to incorporate.

Similarly, memory overload could be utilized by incumbent firms as a means to deter entry. By using excessive advertising or other forms of information provision, the incumbent can strengthen memory overload, thereby causing loyal consumers to underreact to information favorable to entrants. We encourage further research on this potential deterrence channel.

Managerial Implications Memory overload in the workplace can worsen decision quality by disproportionately dampening responsiveness to belief-challenging information. Although some managerial literature emphasizes “cognitive overload” and its effects on stress and productivity ([da Silva Cezar and Maçada \(2023\)](#); [Biondi et al. \(2021\)](#)), our findings highlight memory overload as a distinct factor. The contrast between our treatment and control settings suggests how to design information environments to reduce memory load—for example, by organizing multiple pieces of information within a unified framework, reducing fragmentation and redundancy, and highlighting the most diagnostic items. Such practices can improve employees’ ability to integrate belief-challenging evidence and, in turn, decision quality.

References

- Andor, Mark A, Andreas Gerster, and Jörg Peters (2022) “Information campaigns for residential energy conservation,” *European Economic Review*, Vol. 144, p. 104094.
- Ba, Cuimin, J. Aislinn Bohren, and Alex Imas (2025) “Over- and Underreaction to Information: Belief Updating with Cognitive Constraints,” Working paper, University of Pennsylvania.
- Benjamin, Daniel J. (2019) “Errors in probabilistic reasoning and judgment biases,” in Bernheim, B. Douglas, Stefano DellaVigna, and David Laibson eds. *Handbook of Behavioral Economics: Applications and Foundations 1*, Vol. 2: Elsevier, pp. 69–186.
- Biondi, Francesco N, Angela Cacanindin, Caitlyn Douglas, and Joel Cort (2021) “Overloaded and at work: investigating the effect of cognitive workload on assembly task performance,” *Human factors*, Vol. 63, No. 5, pp. 813–820.
- Bock, Olaf, Ingmar Baetge, and Andreas Nicklisch (2014) “hroot: Hamburg Registration and Organization Online Tool,” *European Economic Review*, Vol. 71, pp. 117–120.
- Bohren, J Aislinn, Josh Hascher, Alex Imas, Michael Ungeheuer, and Martin Weber (2025) “Mental Models of Information: The Role of Attention and Memory,” *University of Chicago, Becker Friedman Institute for Economics Working Paper*, No. 2024-19, p. 93.
- Bordalo, Pedro, Giovanni Burro, Katherine Coffman, Nicola Gennaioli, and Andrei Shleifer (2025) “Imagining the future: memory, simulation, and beliefs,” *Review of Economic Studies*, Vol. 92, No. 3, pp. 1532–1563.
- Chen, Daniel L., Martn Schonger, and Chris Wickens (2016) “oTree-an open-source platform for laboratory, online and field experiments,” *Journal of Behavioral and Experimental Finance*, Vol. 9, pp. 88–97.
- Eil, David and Justin Rao (2011) “The good news-bad news effect: asymmetric processing of objective information about yourself,” *American Economic Journal: Microeconomics*, Vol. 3, No. 2, pp. 114–138.
- Enke, Benjamin (2020) “What you see is all there is,” *The Quarterly Journal of Economics*, Vol. 135, No. 3, pp. 1363–1398.
- (2024) “The cognitive turn in behavioral economics,” Technical report, Working Paper.

- Enke, Benjamin, Frederik Schwerter, and Florian Zimmermann (2024) “Associative memory, beliefs and market interactions,” *Journal of Financial Economics*, Vol. 157, p. 103853.
- Enke, Benjamin and Florian Zimmermann (2019) “Correlation neglect in belief formation,” *The review of economic studies*, Vol. 86, No. 1, pp. 313–332.
- Ertac, Seda (2011) “Does self-relevance affect information processing? Experimental evidence on the response to performance and non-performance feedback,” *Journal of Economic Behavior & Organization*, Vol. 80, No. 3, pp. 532–545.
- Espinoza, Romain and Jan Stoop (2021) “Do people really want to be informed? Ex-ante evaluations of information-campaign effectiveness,” *Experimental Economics*, Vol. 24, No. 4, pp. 1131–1155.
- Facchini, Giovanni, Yotam Margalit, and Hiroyuki Nakata (2022) “Countering public opposition to immigration: The impact of information campaigns,” *European Economic Review*, Vol. 141, p. 103959.
- Gonçalves, Duarte, Jonathan Libgober, and Jack Willis (2021) “Learning versus unlearning: An experiment on retractions,” Technical report, National Bureau of Economic Research.
- Graeber, Thomas, Christopher Roth, and Florian Zimmermann (2024) “Stories, statistics, and memory,” *The Quarterly Journal of Economics*, Vol. 139, No. 4, pp. 2181–2225.
- Grossman, Zachary and David Owens (2012) “An unlucky feeling: Overconfidence and noisy feedback,” *Journal of Economic Behavior & Organization*, Vol. 84, No. 2, pp. 510–524.
- Hossain, Tanjim and Ryo Okui (2013) “The binarized scoring rule,” *Review of Economic Studies*, Vol. 80, No. 3, pp. 984–1001.
- Jehiel, Philippe and Jakub Steiner (2019) “Selective Sampling with Information-Storage Constraints, and Resulting Behavioral Biases,” *The Economic Journal*, URL: <https://doi.org/10.1093/ej/uez068>, DOI: 10.1093/ej/uez068.
- Leung, Benson Tsz Kin (2020) “Limited cognitive ability and selective information processing,” *Games and Economic Behavior*, Vol. 120, pp. 345–369.
- Liang, Yucheng (2025) “Learning from unknown information sources,” *Management Science*, Vol. 71, No. 5, pp. 3873–3890.
- Liu, Yaowu and Jun Xie (2020) “Cauchy combination test: a powerful test with analytic p-value calculation under arbitrary dependency structures,” *Journal of the American Statistical Association*, Vol. 115, No. 529, pp. 393–402.

- Möbius, Markus M, Muriel Niederle, Paul Niehaus, and Tanya S Rosenblat (2022) “Managing self-confidence: Theory and experimental evidence,” *Management Science*, Vol. 68, No. 11, pp. 7793–7817.
- da Silva Cezar, Bibiana Giudice and Antônio Carlos Gastaud Maçada (2023) “Cognitive overload, anxiety, cognitive fatigue, avoidance behavior and data literacy in big data environments,” *Information Processing & Management*, Vol. 60, No. 6, p. 103482.
- Snyder, Leslie B (2007) “Health communication campaigns and their impact on behavior,” *Journal of nutrition education and behavior*, Vol. 39, No. 2, pp. S32–S40.
- Wilson, Andrea (2014) “Bounded Memory and Biases in Information Processing,” *Econometrica*, Vol. 82, No. 6, pp. 2257–2294.

A.1 Proofs

A.1.1 Proof of Proposition 1

Proof. We first prove that there exist two thresholds $\bar{s}, \underline{s}$ satisfying $\bar{s} > \underline{s}$ such that the subject memorizes s_1 iff $s_1 \geq \bar{s}$ or $s_1 \leq \underline{s}$. We look at two cases: 1) $s_1 \geq \frac{1-p_H}{p_H}$; 2) $s_1 < \frac{1-p_H}{p_H}$.

First, suppose the subject receives $s_1 \geq \frac{1-p_H}{p_H}$, we show that her utility of memorizing minus not memorizing increases in s_1 . Her utility of memorizing is equal to:

$$\frac{p_H f_H(s_1)}{p_H f_H(s_1) + p_L f_L(s_1)} = \frac{p_H s_1}{p_H s_1 + p_L}$$

where $p_L = 1 - p_H$. Given the equilibrium probability of memorizing in state H and L , denoted as m_H and m_L , the subject's utility of not memorizing is equal to:

$$\frac{p_H s_1}{p_H s_1 + p_L} \left(\int_{\frac{p_L(1-m_L)}{p_H(1-m_H)}}^{\infty} f_H(s) ds \right) + \frac{p_L}{p_H s_1 + p_L} \left(\int_0^{\frac{p_L(1-m_L)}{p_H(1-m_H)}} f_L(s) ds \right).$$

The utility of memorizing s_1 minus not memorizing is equal to:

$$\begin{aligned} & \frac{p_H s_1}{p_H s_1 + p_L} \left(1 - \int_{\frac{p_L(1-m_L)}{p_H(1-m_H)}}^{\infty} f_H(s) ds \right) - \frac{p_L}{p_H s_1 + p_L} \left(\int_0^{\frac{p_L(1-m_L)}{p_H(1-m_H)}} f_L(s) ds \right) \\ &= \frac{p_H s_1}{p_H s_1 + p_L} \left(\int_0^{\frac{p_L(1-m_L)}{p_H(1-m_H)}} f_H(s) ds \right) - \frac{p_L}{p_H s_1 + p_L} \left(\int_0^{\frac{p_L(1-m_L)}{p_H(1-m_H)}} f_L(s) ds \right) \\ &= - \left(\int_0^{\frac{p_L(1-m_L)}{p_H(1-m_H)}} f_L(s) ds \right) + \frac{p_H s_1}{p_H s_1 + p_L} \left(\int_0^{\frac{p_L(1-m_L)}{p_H(1-m_H)}} (f_H(s) + f_L(s)) ds \right) \end{aligned}$$

which increases in s_1 as $\frac{p_H s_1}{p_H s_1 + p_L}$ increases in s_1 .

Second, suppose the subject receives $s_1 < \frac{1-p_H}{p_H}$. Her utility of memorizing is equal to:

$$\frac{p_L f_L(s_1)}{p_H f_H(s_1) + p_L f_L(s_1)} = \frac{p_L}{p_H s_1 + p_L}.$$

The subject's utility of not memorizing is equal to:

$$\frac{p_H s_1}{p_H s_1 + p_L} \left(\int_{\frac{p_L(1-m_L)}{p_H(1-m_H)}}^{\infty} f_H(s) ds \right) + \frac{p_L}{p_H s_1 + p_L} \left(\int_0^{\frac{p_L(1-m_L)}{p_H(1-m_H)}} f_L(s) ds \right)$$

The utility of memorizing s_1 minus not memorizing is equal to:

$$\begin{aligned}
& -\frac{p_H s_1}{p_H s_1 + p_L} \left(\int_{\frac{p_L(1-m_L)}{p_H(1-m_H)}}^{\infty} f_H(s) ds \right) + \frac{p_L}{p_H s_1 + p_L} \left(1 - \int_0^{\frac{p_L(1-m_L)}{p_H(1-m_H)}} f_L(s) ds \right) \\
& = -\frac{p_H s_1}{p_H s_1 + p_L} \left(\int_{\frac{p_L(1-m_L)}{p_H(1-m_H)}}^{\infty} f_H(s) ds \right) + \frac{p_L}{p_H s_1 + p_L} \left(\int_{\frac{p_L(1-m_L)}{p_H(1-m_H)}}^{\infty} f_L(s) ds \right) \\
& = -\left(\int_{\frac{p_L(1-m_L)}{p_H(1-m_H)}}^{\infty} f_H(s) ds \right) + \frac{p_L}{p_H s_1 + p_L} \left(\int_{\frac{p_L(1-m_L)}{p_H(1-m_H)}}^{\infty} (f_H(s) + f_L(s)) ds \right)
\end{aligned}$$

which decreases in s_1 as $\frac{p_L}{p_H s_1 + p_L}$ increases in s_1 .

Thus, all equilibria are characterized as two thresholds $\bar{s} \geq \frac{1-p_H}{p_H} > \underline{s}$ which satisfy the following Equations:

$$\begin{aligned}
\bar{s} &= \frac{1-p_H}{p_H} \frac{\int_0^{\frac{p_L(1-m_L)}{p_H(1-m_H)}} f_L(s) ds}{\int_0^{\frac{p_L(1-m_L)}{p_H(1-m_H)}} f_H(s) ds} \equiv A\left(\frac{1-m_L}{1-m_H}\right) \\
\underline{s} &= \frac{p_H}{1-p_H} \frac{\int_{\frac{p_L(1-m_L)}{p_H(1-m_H)}}^{\infty} f_H(s) ds}{\int_{\frac{p_L(1-m_L)}{p_H(1-m_H)}}^{\infty} f_L(s) ds} \equiv B\left(\frac{1-m_L}{1-m_H}\right) \\
1-m_H &= \int_{\underline{s}}^{\bar{s}} f_H(s) ds \\
1-m_L &= \int_{\underline{s}}^{\bar{s}} f_L(s) ds
\end{aligned} \tag{9}$$

We now show that there always exists an equilibrium where $\bar{s} > \underline{s}$ and $\bar{s} < \frac{1}{\underline{s}}$ when Equation (2) holds. We prove the statement in 4 steps. 1) when $\bar{s}$ or $\underline{s}$ decreases, $\frac{1-m_L}{1-m_H}$ increases. Note that

$$\frac{1-m_L}{1-m_H} = \frac{\int_{\underline{s}}^{\bar{s}} f_L(s) ds}{\int_{\underline{s}}^{\bar{s}} f_H(s) ds}.$$

Note that when $\bar{s}$ decreases, as $\frac{f_L(s)}{f_H(s)}$ decreases in s , $\frac{\int_{\underline{s}}^{\bar{s}} f_L(s) ds}{\int_{\underline{s}}^{\bar{s}} f_H(s) ds}$ and $\frac{1-m_L}{1-m_H}$ increases. Similarly, when $\underline{s}$ decreases, $\frac{1-m_L}{1-m_H}$ increases.

2) When $\frac{1-m_L}{1-m_H}$ increases, $A(\cdot)$ decreases and $B(\cdot)$ increases. Recall that

$$A\left(\frac{1-m_L}{1-m_H}\right) = \frac{1-p_H}{p_H} \frac{\int_0^{\frac{p_L(1-m_L)}{p_H(1-m_H)}} f_L(s) ds}{\int_0^{\frac{p_L(1-m_L)}{p_H(1-m_H)}} f_H(s) ds}.$$

Similar to our argument in 1), as $\frac{f_L(s)}{f_H(s)}$ decreases in s , $A\left(\frac{1-m_L}{1-m_H}\right)$ decreases as $\frac{1-m_L}{1-m_H}$

increases. Now, recall that

$$B\left(\frac{1-m_L}{1-m_H}\right) = \frac{p_H}{1-p_H} \frac{\int_0^\infty \frac{p_L(1-m_L)}{p_H(1-m_H)} f_H(s) ds}{\int_0^\infty \frac{p_L(1-m_L)}{p_H(1-m_H)} f_L(s) ds}$$

As $\frac{f_H(s)}{f_L(s)}$ increases in s , $B\left(\frac{1-m_L}{1-m_H}\right)$ increases when $\frac{1-m_L}{1-m_H}$ increases.

3) Denote Equation (9) as

$$\begin{pmatrix} \bar{s} \\ \underline{s} \end{pmatrix} = \begin{pmatrix} \frac{1-p_H}{p_H} \frac{\int_0^\infty \frac{p_L(1-m_L)}{p_H(1-m_H)} f_L(s) ds}{\int_0^\infty \frac{p_L(1-m_L)}{p_H(1-m_H)} f_H(s) ds} \\ \frac{1-p_H}{p_H} \frac{\int_0^\infty \frac{p_L(1-m_L)}{p_H(1-m_H)} f_L(s) ds}{\int_0^\infty \frac{p_L(1-m_L)}{p_H(1-m_H)} f_H(s) ds} \end{pmatrix} = \tilde{A} \begin{pmatrix} \bar{s} \\ \underline{s} \end{pmatrix}$$

and consider the space where $\bar{s} \in \left[\frac{1-p_H}{p_H}, \frac{1-p_H}{p_H} \frac{\int_0^\infty \frac{p_L}{p_H} f_L(s) ds}{\int_0^\infty \frac{p_L}{p_H} f_H(s) ds} \right]$ and $\underline{s} \in \left[0, \frac{1-p_H}{p_H} \frac{\int_0^\infty \frac{p_L}{p_H} f_L(s) ds}{\int_0^\infty \frac{p_L}{p_H} f_H(s) ds} \right]$.

When $\bar{s} = \frac{1-p_H}{p_H} \frac{\int_0^\infty \frac{p_L}{p_H} f_L(s) ds}{\int_0^\infty \frac{p_L}{p_H} f_H(s) ds}$ and $\underline{s} = \frac{1-p_H}{p_H} \frac{\int_0^\infty \frac{p_L}{p_H} f_L(s) ds}{\int_0^\infty \frac{p_L}{p_H} f_H(s) ds}$, if Equation (2) holds, i.e.,

$$\frac{1-p_H}{p_H} \frac{\int_0^\infty \frac{p_L}{p_H} f_L(s) ds}{\int_0^\infty \frac{p_L}{p_H} f_H(s) ds} < \frac{1}{\frac{1-p_H}{p_H} \frac{\int_0^\infty \frac{p_L}{p_H} f_L(s) ds}{\int_0^\infty \frac{p_L}{p_H} f_H(s) ds}} \quad (10)$$

we have $1-m_L > 1-m_H$ (by the symmetry of f_H and f_L), and by 1) and 2), $\tilde{A}(\cdot) \leq \left(\frac{1-p_H}{p_H} \frac{\int_0^\infty \frac{p_L}{p_H} f_L(s) ds}{\int_0^\infty \frac{p_L}{p_H} f_H(s) ds}, \frac{1-p_H}{p_H} \frac{\int_0^\infty \frac{p_L}{p_H} f_L(s) ds}{\int_0^\infty \frac{p_L}{p_H} f_H(s) ds} \right)$. Moreover by 1) and 2), when $\bar{s}$ and/or $\underline{s}$ decreases, $\tilde{A}(\cdot)$ decreases in both dimensions. Hence $\tilde{A}$ is a one-to-one (continuous) mapping from $\left[\frac{1-p_H}{p_H}, \frac{1-p_H}{p_H} \frac{\int_0^\infty \frac{p_L}{p_H} f_L(s) ds}{\int_0^\infty \frac{p_L}{p_H} f_H(s) ds} \right] \times \left[0, \frac{1-p_H}{p_H} \frac{\int_0^\infty \frac{p_L}{p_H} f_L(s) ds}{\int_0^\infty \frac{p_L}{p_H} f_H(s) ds} \right]$ which is a non-empty, compact and convex space. By Brouwer's fixed-point theorem, there exists a fixed point $\left[\frac{1-p_H}{p_H}, \frac{1-p_H}{p_H} \frac{\int_0^\infty \frac{p_L}{p_H} f_L(s) ds}{\int_0^\infty \frac{p_L}{p_H} f_H(s) ds} \right] \times \left[0, \frac{1-p_H}{p_H} \frac{\int_0^\infty \frac{p_L}{p_H} f_L(s) ds}{\int_0^\infty \frac{p_L}{p_H} f_H(s) ds} \right]$, such that $\bar{s} < 1/\underline{s}$ (by Equation (10)).

Next, we prove that if Equation (3) holds, then no equilibrium with $\bar{s} \geq \frac{1}{\underline{s}}$ exists. Suppose toward a contradiction that some equilibrium has $\bar{s} \geq 1/\underline{s}$, i.e., $\bar{s} \underline{s} \geq 1$. First, we establish bounds for $\frac{1-m_L}{1-m_H}$. Note that $\underline{s} < \frac{1-p_H}{p_H}$ combined with $\bar{s} \geq 1/\underline{s}$ implies that $\bar{s} > \frac{p_H}{1-p_H}$. Since

$$\frac{1-m_L}{1-m_H} = \frac{\int_{\underline{s}}^{\bar{s}} f_L(s) ds}{\int_{\underline{s}}^{\bar{s}} f_H(s) ds},$$

the monotone likelihood-ratio property implies that increasing $\underline{s}$ or $\bar{s}$ decreases $\frac{1-m_L}{1-m_H}$.

Hence $\frac{1-m_L}{1-m_H} \in \left(\frac{1-F_L(\frac{1-p_H}{p_H})}{1-F_H(\frac{1-p_H}{p_H})}, 1 \right]$, where the lower bound is approached as $\underline{s} \rightarrow \frac{1-p_H}{p_H}$ and $\bar{s} \rightarrow \infty$. Next, by Equation (9), $A(\cdot)$ (and hence $\bar{s}$) decreases in $\frac{1-m_L}{1-m_H}$, while $B(\cdot)$ (and hence $1/\underline{s}$) increases in $\frac{1-m_L}{1-m_H}$; thus $\bar{s}\underline{s}$ decreases in $\frac{1-m_L}{1-m_H}$. Combining with the lower bound of $\frac{1-m_L}{1-m_H}$, it follows that, for any equilibrium with $\bar{s} \geq 1/\underline{s}$,

$$\begin{aligned} \bar{s}\underline{s} &= \frac{1-p_H}{p_H} \frac{\int_0^{\frac{p_L(1-m_L)}{p_H(1-m_H)}} f_L(s) ds}{\int_0^{\frac{p_L(1-m_L)}{p_H(1-m_H)}} f_H(s) ds} \times \frac{1-p_H}{p_H} \frac{\int_{\frac{p_L(1-m_L)}{p_H(1-m_H)}}^{\infty} f_L(s) ds}{\int_{\frac{p_L(1-m_L)}{p_H(1-m_H)}}^{\infty} f_H(s) ds} \\ &\leq \frac{1-p_H}{p_H} \frac{\int_0^{\tilde{\theta}} f_L(s) ds}{\int_0^{\tilde{\theta}} f_H(s) ds} \times \frac{1-p_H}{p_H} \frac{\int_{\tilde{\theta}}^{\infty} f_L(s) ds}{\int_{\tilde{\theta}}^{\infty} f_H(s) ds} \\ &< 1 \end{aligned}$$

where the equality is Equation (9), the first inequality holds because $\bar{s}\underline{s}$ is decreasing in $\frac{1-m_L}{1-m_H}$ (so it is largest at the lower bound, at which $\frac{p_L(1-m_L)}{p_H(1-m_H)} = \tilde{\theta}$), and the last inequality follows from Equation (3). This contradicts $\bar{s}\underline{s} \geq 1$.

Finally, we prove that Equation (3) holds when $p_H \rightarrow 1$. Note that Equation (3) can be rearranged to:

$$\frac{1-p_H}{p_H} \frac{1-F_L(\tilde{\theta})}{1-F_H(\tilde{\theta})} \times \frac{1-p_H}{p_H} \frac{F_L(\tilde{\theta})}{F_H(\tilde{\theta})} < 1.$$

We show that $\lim_{p_H \rightarrow 1} \frac{1-p_H}{p_H} \frac{1-F_L(\tilde{\theta})}{1-F_H(\tilde{\theta})} \times \frac{1-p_H}{p_H} \frac{F_L(\tilde{\theta})}{F_H(\tilde{\theta})} < 1$. First, since the ratio $\frac{\int_{\tilde{\theta}}^{\infty} f_L(s) ds}{\int_{\tilde{\theta}}^{\infty} f_H(s) ds} \in (0, 1)$ is bounded and $\frac{1-p_H}{p_H} \rightarrow 0$ when $p_H \rightarrow 1$, the factor $\frac{1-p_H}{p_H} \frac{1-F_L(\tilde{\theta})}{1-F_H(\tilde{\theta})} \rightarrow 0$. Now we show that $\frac{1-p_H}{p_H} \frac{F_L(\tilde{\theta})}{F_H(\tilde{\theta})}$ is bounded above. First note that when $p_H \rightarrow 1$, $\tilde{\theta}/\frac{1-p_H}{p_H} \rightarrow 1$. Using L'Hôpital's rule,

$$\begin{aligned} &\lim_{p_H \rightarrow 1} \frac{1-p_H}{p_H} \frac{\int_0^{\frac{1-p_H}{p_H}} f_L(s) ds}{\int_0^{\frac{1-p_H}{p_H}} f_H(s) ds} \\ &= \lim_{p_H \rightarrow 1} \frac{-\int_0^{\frac{1-p_H}{p_H}} f_L(s) ds + p_L f_L(\frac{1-p_H}{p_H}) \frac{d\frac{1-p_H}{p_H}}{dp_H}}{\int_0^{\frac{1-p_H}{p_H}} f_H(s) ds + p_H f_H(\frac{1-p_H}{p_H}) \frac{d\frac{1-p_H}{p_H}}{dp_H}} \\ &\leq \lim_{p_H \rightarrow 1} \frac{p_L f_L(\frac{1-p_H}{p_H}) \frac{d\frac{1-p_H}{p_H}}{dp_H}}{p_H f_H(\frac{1-p_H}{p_H}) \frac{d\frac{1-p_H}{p_H}}{dp_H}} \\ &= \lim_{p_H \rightarrow 1} \frac{p_L f_L(\frac{1-p_H}{p_H})}{p_H f_H(\frac{1-p_H}{p_H})} = 1. \end{aligned}$$

Thus, $\lim_{p_H \rightarrow 1} \frac{1-p_H}{p_H} \frac{1-F_L(\tilde{\theta})}{1-F_H(\tilde{\theta})} \times \frac{1-p_H}{p_H} \frac{F_L(\tilde{\theta})}{F_H(\tilde{\theta})} = 0 < 1$, i.e., Equation (3) holds when $p_H \rightarrow 1$. □

Online Appendix of Memory Overload and Asymmetric Underreaction

Lorenz Goette¹, Hua-Jing Han², Benson Tsz Kin Leung³

July 4, 2026

A.2 The Distribution in the Guessing Task

We present the rationale for the parameterization of the two distributions. In particular, the parameters are chosen to satisfy the following criteria:

1. The log-odds ratios are monotone and approximately linear, as shown in Figure A1. That is, higher numbers provide stronger evidence that the “high” distribution is true, and the differences in the strength of adjacent signals are approximately constant.
2. After the first sequence of numbers, a substantial fraction of subjects hold confident beliefs (i.e., assign at least 75% probability to the high or low state). Table A.2 shows that more than 40% of subjects are confident after observing five signals. This ensures that there exists a sufficient number of confident individuals for bias against belief-challenging signals to operate.
3. After the first sequence of numbers, few enough subjects hold extremely confident beliefs (e.g., assign at least 95% probability to the high or low state). Table A.3 shows that fewer than 2% of subjects are extremely confident after observing five signals. This ensures a sufficient number of observations in which a switch occurs, as shown in Table A.1.

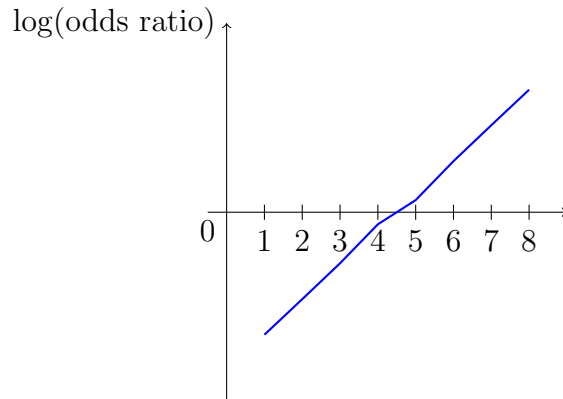


Figure A1: Log-odds ratios of the numbers 1-8.

¹National University of Singapore. Email: ecslfg@nus.edu.sg

²University of Bonn. Email: han@uni-bonn.de

³Hong Kong Baptist University. Email: btkleung@hkbu.edu.hk

min	mean	max
18.5%	24.2%	29.5%

Table A.1: Simulated proportion of observations where subjects should switch from believing “High” after Phase 1 to believing “Low” after Phase 2, or from believing “Low” to believing “High”, with 10000 simulations of 2000 observations.

		Belief				
		= 0.5	> 0.6 or < 0.4	> 0.65 or < 0.35	> 0.7 or < 0.3	> 0.75 or < 0.25
5 draws	Average proportion of subjects	0%	76.1%	64.7%	52.5%	41.7%
12 draws	Average proportion of subjects	0%	94.4%	88.8%	76.0%	68.9%

Table A.2: Distribution of Bayesian beliefs given 5 and 12 draws, with 10000 simulations of 2000 observations.

		Belief			
		> 0.8 or < 0.2	> 0.85 or < 0.15	> 0.9 or < 0.1	> 0.95 or < 0.05
5 draws	Average proportion of subjects	32.2%	19.2%	10.1%	1.65%
12 draws	Average proportion of subjects	60.9%	51.35%	40.05%	24.35%

Table A.3: Distribution of Bayesian beliefs given 5 and 12 draws, with 10000 simulations of 2000 observations.

A.2.1 Omitted Graphs and Analysis

Which Computer is more likely that has been selected?

☒ "high" ☐ "low"

Please specify your assessment:

Option 1			Option 2
Win 8€, if "high" is true	☒	☐	Win 8€ with probability 50 %
Win 8€, if "high" is true	☒	☐	Win 8€ with probability 55 %
Win 8€, if "high" is true	☒	☐	Win 8€ with probability 60 %
Win 8€, if "high" is true	☒	☐	Win 8€ with probability 65 %
Win 8€, if "high" is true	☐	☒	Win 8€ with probability 70 %
Win 8€, if "high" is true	☐	☒	Win 8€ with probability 75 %
Win 8€, if "high" is true	☐	☒	Win 8€ with probability 80 %
Win 8€, if "high" is true	☐	☒	Win 8€ with probability 85 %
Win 8€, if "high" is true	☐	☒	Win 8€ with probability 90 %
Win 8€, if "high" is true	☐	☒	Win 8€ with probability 95 %

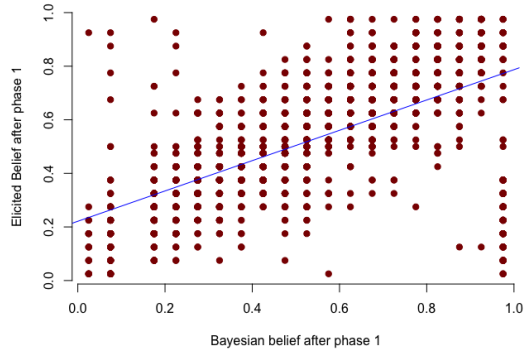
Next

Figure A2: The belief elicitation screen.

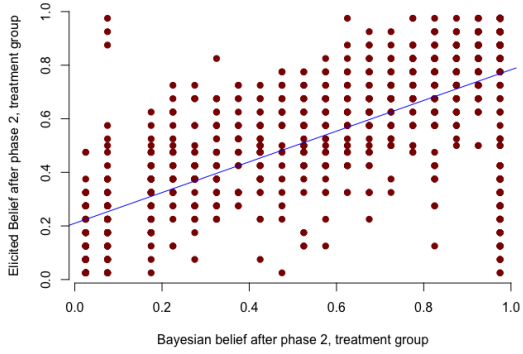
Figure A4a shows histograms of the absolute difference between elicited and Bayesian beliefs (updating errors) after Phase 1 in the treatment and control conditions, while Figures A4b and A4c show the corresponding histograms for beliefs after Phase 2 with belief-confirming and belief-challenging signals. The three graphs show that most errors (40-60%) are less than 10%, and the frequency of errors decreases with their magnitude. For example, for belief formation in Phase 1 in both treatment and control conditions, almost 60% of elicited beliefs are within 10% of the Bayesian beliefs, while only around 10% of the errors are as large as 20%. This again indicates that subjects understand the information structure and do not make large errors.

As expected, and as seen in Figure A4a, being in the treatment or control condition has no effect on errors in Phase 1, since the treatment and control settings are identical in Phase 1. In particular, a two-sided Kolmogorov-Smirnov test cannot reject equality of the error distributions between control and treatment ($p = 0.716$).

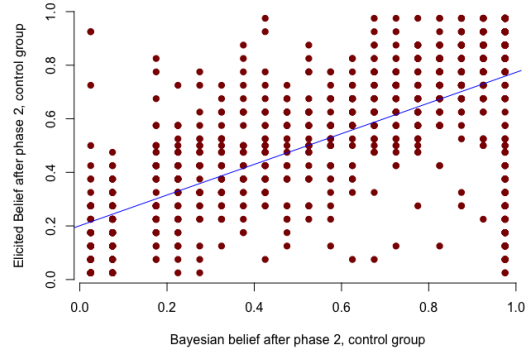
In contrast, we should expect subjects to make more errors when updating their beliefs in Phase 2 in the treatment than in the control condition. With belief-confirming information (Figure A4b), the Phase 2 updating-error distribution in the treatment condition is not significantly different from that in the control con-



(a) Beliefs after Phase 1.



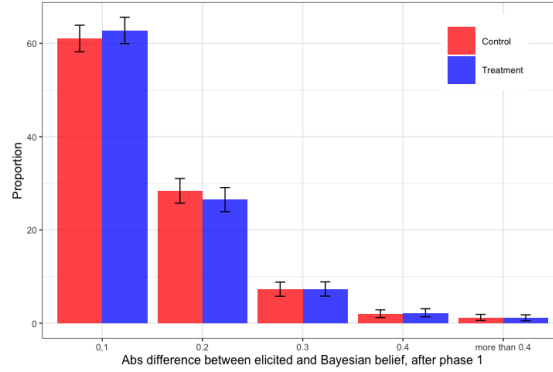
(b) Beliefs after Phase 2, Treatment.



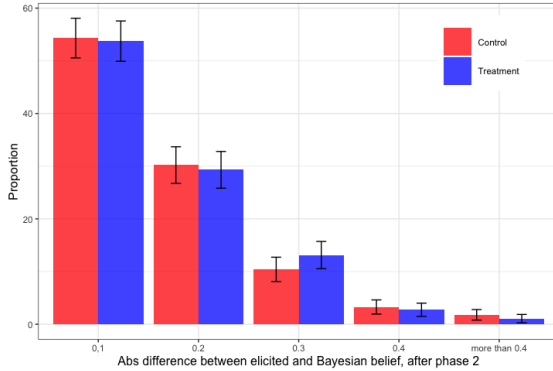
(c) Beliefs after Phase 2, Control.

Figure A3: Scatter plots and regression lines with Bayesian belief on the x-axis and elicited belief on the y-axis. Both Bayesian and elicited beliefs take values equal to the midpoint of their respective 5% bins.

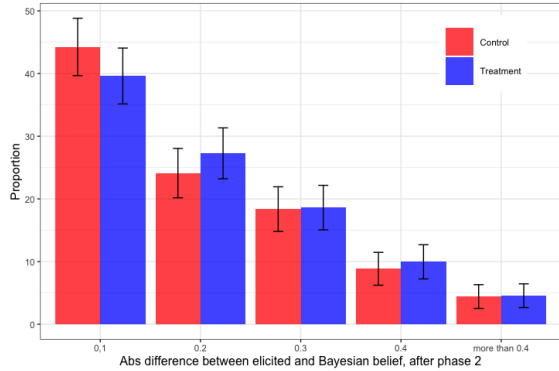
dition (two-sided Kolmogorov–Smirnov test $p = 0.361$). With belief-challenging information (Figure A4c), the treatment has a larger impact on the updating-error distribution—for example, the share of errors smaller than 10% is higher in the control condition (about 50% of observations) than in the treatment condition (about 40%), while the share of large errors is smaller in control. However, these differences are not strong enough to reject the null of equal distributions in a two-sided Kolmogorov–Smirnov test ($p = 0.189$), although this p-value is smaller than those for Phase 1 beliefs and Phase 2 beliefs with confirmatory signals.⁴



(a) Beliefs after Phase 1, treatment versus control.



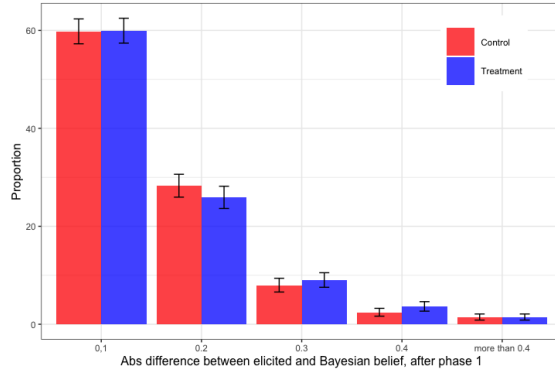
(b) Beliefs after Phase 2, treatment versus control with confirmatory signals.



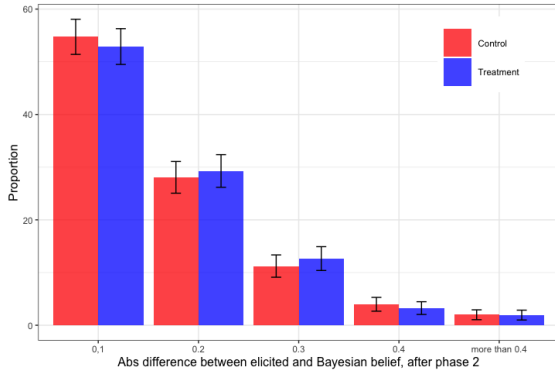
(c) Beliefs after Phase 2, treatment versus control with challenging signals.

Figure A4: Distributions of the absolute differences between elicited and Bayesian beliefs with 95% confidence intervals. Both Bayesian and elicited beliefs take values equal to the midpoint of their respective 5% bins and are normalized to lie in $[0, 1]$. Proportions are shown in percent.

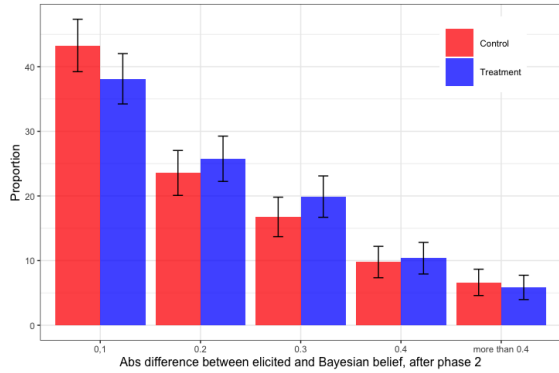
⁴When we include all observations, even when the first elicited belief does not equal 0.5, the two-sided Kolmogorov–Smirnov test indicates that subjects in the treatment condition make larger errors than those in the control condition with both belief-challenging and belief-confirming information ($p = 0.04767$ and $p = 0.06989$, respectively).



(a) Beliefs after Phase 1, treatment versus control.



(b) Beliefs after Phase 2, treatment versus control with confirmatory signals.



(c) Beliefs after Phase 2, treatment versus control with challenging signals.

Figure A5: Distributions of the absolute differences between elicited and Bayesian beliefs with 95% confidence intervals. Both Bayesian and elicited beliefs take values equal to the midpoint of their respective 5% bins and are normalized to lie in $[0, 1]$. Proportions are shown in percent. These correspond to Figures A4a, A4b and A4c.

A.3 Analysis with all data, including those with first elicited belief not equal to 0.5

This section presents the analysis with all data, including those with first elicited beliefs not equaling 0.5.

Figure A5a, A5b and A5c show the distributions of updating errors after Phase 1 and 2 in the control, and treatment conditions, with confirmatory and challenging signals. For updating errors after Phase 1, a two-sided Kolmogorov–Smirnov test shows that there is no significant treatment effect ($p = 0.8659$). For updating errors after Phase 2, a two-sided Kolmogorov–Smirnov test shows a treatment effect: subjects in the treatment condition make larger errors than those in the control condition with both belief-challenging and belief-confirming information ($p = 0.04767$ and $p = 0.06989$ respectively).

Table A.4 shows the regression analysis of how elicited beliefs correlate with Bayesian beliefs in Phase 1, and Table A.5 shows the corresponding analysis for

Phase 2. As in the main text, the first two columns of each table show that subjects' elicited beliefs after Phases 1 and 2 are positively and significantly correlated with Bayesian beliefs, although they underreact to signals. The last two columns of Table A.5 show that subjects underreact more to belief-challenging information than to belief-confirming information (i.e., $0.494 < 0.754$).

	(1) Elicited belief	(2) Elicited belief
Bayesian belief after Phase 1	0.774*** (0.016)	0.783*** (0.018)
Treatment		0.005 (0.012)
Treatment $\times$ Bayesian belief after Phase 1		-0.019 (0.020)
R^2	0.675	0.675
Observations	2856	2856
Subjects	260	260
Clustered standard errors at the subject level in parentheses		
[†] $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$		

Table A.4: Regression analysis of elicited beliefs after Phase 1. Both Bayesian and elicited beliefs take values equal to the midpoint of their respective 5% bins. Standard errors are clustered at the subject level, and we control for subjects' age, gender, and whether they have taken statistics/economics courses. Subjects generally underreact to information. Corresponds to Table 2.

More importantly, the treatment has no effect on how subjects update with confirmatory signals in Phase 2. On the other hand, the negative coefficient on the interaction between treatment and the Bayesian belief in Phase 2 suggests that treatment strengthens underreaction with challenging signals. The effect, however, is less significant than in the main text.

Table A.6 shows OLS analyses on the log-odds deviation. Columns 1–3 show that treatment induces underreaction to both confirmatory and challenging signals. However, underreaction to confirmatory signals is less statistically significant (column 1) and smaller in magnitude ($0.104 < 0.166$ in column 2 versus column 3). As in the main text, however, the difference is not statistically significant.

Next, Table A.7 tests whether attention spent on signals is substitutable. Similar to the main text, in Table A.7, the coefficient on the triple-interaction term between treatment, the challenging dummy for the 1st–6th signals, and the challenging dummy for the last signal is not significant ($p = 0.579, 0.910, 0.430$ in columns 1, 2, and 3, respectively). Thus, we do not find supportive evidence for the capacity mechanism. Lastly, in Table A.8, the coefficients on the interaction term between

	(1)	(2)	(3)	(4)
	Elicited belief	Elicited belief	Elicited belief (w confirmatory signals)	Elicited belief (w challenging signals)
Bayesian belief after Phase 2	0.686*** (0.013)	0.699*** (0.015)	0.754*** (0.014)	0.494*** (0.036)
Treatment		0.016 (0.010)	0.003 (0.010)	0.048 [†] (0.024)
Treatment $\times$ Bayesian belief after Phase 2		-0.025 (0.015)	-0.006 (0.013)	-0.082 [†] (0.044)
R^2	0.698	0.698	0.851	0.287
Observations	2856	2856	1685	1171
Subjects	260	260	260	259

Clustered standard errors at the subject level in parentheses

[†] $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A.5: Regression analysis of elicited beliefs after Phase 2. Both Bayesian and elicited beliefs take values equal to the midpoint of their respective 5% bins. Standard errors are clustered at the subject level, and we control for subjects' age, gender, and whether they have taken statistics/economics courses. As in Phase 1, subjects generally underreact to information. Moreover, they underreact more to challenging signals than to confirmatory signals. Moreover, treatment strengthens underreaction to challenging signals but not confirmatory signals. Corresponds to Table 3.

the treatment dummy and the number of Phase-2 belief-challenging signals are consistent with the per-signal cost mechanism, and are significant in column 1 where we pool the low-prior and high-prior data ($p = 0.005, 0.011, 0.164$ in columns 1, 2, and 3, respectively).

	(1)	(2)	(3)	(4)	(5)
	Pooled	Pooled (with confirmatory signals)	Pooled (with challenging signals)	High prior	Low prior
Treatment	-0.099 (0.068)	-0.104* (0.044)	0.166* (0.064)	0.106 (0.070)	-0.099* (0.045)
Challenging	1.028*** (0.102)			-1.031*** (0.102)	1.026*** (0.077)
Treatment $\times$ Challenging	0.270* (0.116)			-0.240 (0.127)	0.263** (0.082)
Pair Fixed Controls	No	Yes	Yes	No	No
R^2	0.224	0.713	0.720	0.226	0.230
Observations	2856	1685	1171	1308	2856
Subjects	260	260	259	259	258

Clustered standard errors at the subject level in parentheses

[†] $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A.6: Regression analysis of the log-odds deviation $\log \frac{p_2}{1-p_2} - \log \frac{p_2^B}{1-p_2^B}$. Standard errors are clustered at the subject level, and we control for subjects' age, gender, and whether they have taken statistics/economics courses. We control for pair fixed effects in columns 2 and 3. Columns 1–3 report pooled analyses (overall; with confirmatory signals; with challenging signals). Column 4 restricts to high priors ($p_1 > 0.5$); column 5 restricts to low priors ($p_1 < 0.5$). Corresponds to Table 4.

	(1) Pooled	(2) High prior	(3) Low prior
Treatment	-0.146* (0.064)	-0.182* (0.091)	0.112 (0.096)
Challenging 1 st -6 th	1.018*** (0.089)	1.126*** (0.121)	-0.865*** (0.115)
Challenging 7 th	0.268** (0.081)	0.272** (0.103)	-0.264* (0.118)
Challenging 1 st -6 th × Challenging 7 th	0.051 (0.138)	-0.162 (0.193)	-0.341 (0.187)
Treatment × Challenging 1 st -6 th	0.180 (0.112)	0.200 (0.153)	-0.165 (0.160)
Treatment × Challenging 7 th	0.230* (0.108)	0.268 (0.141)	-0.192 (0.151)
Treatment × Challenging 1 st -6 th × Challenging 7 th	-0.096 (0.172)	-0.026 (0.227)	0.198 (0.251)
R^2	0.238	0.245	0.248
Observations	2793	1512	1281
Subjects	260	259	258

Clustered standard errors at the subject level in parentheses

† $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A.7: Regression analysis of the log-odds deviation $\log \frac{p_2}{1-p_2} - \log \frac{p_2^B}{1-p_2^B}$ that analyzes whether attention is substitutable. Standard errors are clustered at the subject level, and we control for subjects' age, gender, and whether they have taken statistics/economics courses. Column 1 reports pooled analyses; column 2 restricts to high priors ($p_1 > 0.5$); column 3 restricts to low priors ($p_1 < 0.5$). Corresponds to Table 5.

	(1) Pooled	(2) High prior	(3) Low prior
Treatment	-0.265** (0.097)	-0.325* (0.136)	0.203 (0.149)
# Challenging	0.340*** (0.027)	0.352*** (0.034)	-0.323*** (0.037)
Treatment $\times$ # Challenging	0.086** (0.041)	0.104* (0.041)	-0.064 (0.046)
R^2	0.220	0.243	0.204
Observations	2793	1512	1281
Subjects	260	259	258
Clustered standard errors at the subject level in parentheses			
[†] $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$			

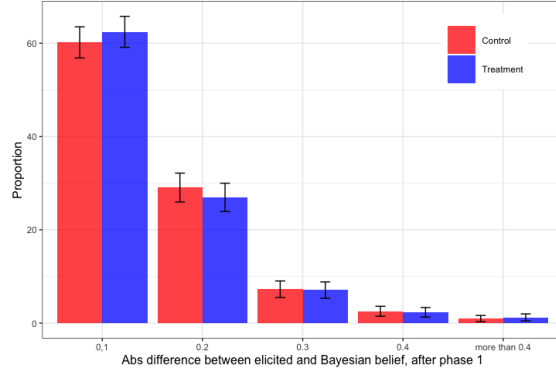
Table A.8: Regression analysis of the log-odds deviation $\log \frac{p_2}{1-p_2} - \log \frac{p_2^B}{1-p_2^B}$ that analyzes how the number of challenging signals affects the treatment effect. Standard errors are clustered at the subject level, and we control for subjects’ age, gender, and whether they have taken statistics/economics courses. Column 1 reports pooled analyses; column 2 restricts to high priors ($p_1 > 0.5$); column 3 restricts to low priors ($p_1 < 0.5$). Corresponds to Table 6.

A.4 Analysis with data excluding all observations in the first 4 sessions

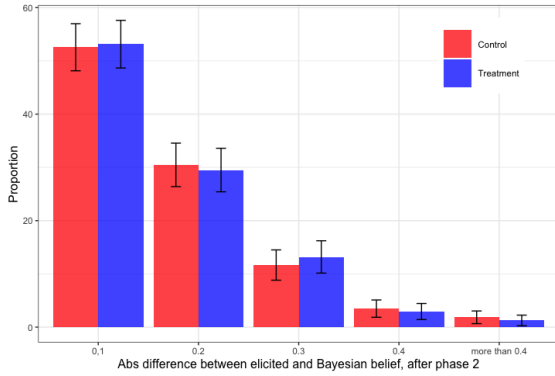
This section presents the analysis excluding all data from sessions 1 to 4, affected by the technical glitch.

Figure A6a, A6b and A6c show the distributions of updating errors after Phase 1 and 2 in the control, and treatment conditions, with confirmatory and challenging signals. For updating errors after Phase 1, a two-sided Kolmogorov–Smirnov test shows that there is no significant treatment effect ($p = 0.963$). For updating errors after Phase 2, a two-sided Kolmogorov–Smirnov test shows a marginal but insignificant treatment effect: subjects in the treatment condition make larger errors than those in the control condition with both belief-challenging and belief-confirming information ($p = 0.144, 0.295$ respectively).

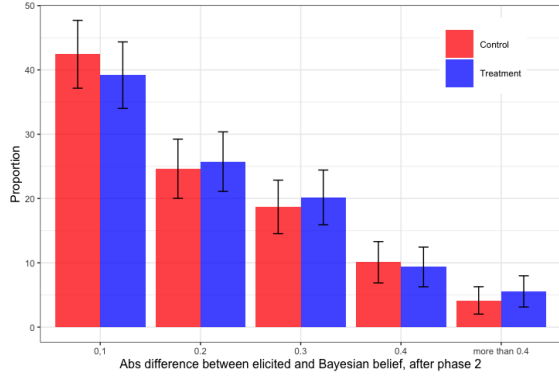
Table A.9 shows the regression analysis of how elicited beliefs correlate with Bayesian beliefs in Phase 1, and Table A.10 shows the corresponding analysis for Phase 2. As in the main text, the first two columns of each table show that subjects’ elicited beliefs after Phases 1 and 2 are positively and significantly correlated with Bayesian beliefs, although they underreact to signals. The last two columns of Table A.10 show that subjects underreact more to belief-challenging information than to belief-confirming information (i.e., $0.499 < 0.753$).



(a) Beliefs after Phase 1, treatment versus control.



(b) Beliefs after Phase 2, treatment versus control with confirmatory signals.



(c) Beliefs after Phase 2, treatment versus control with challenging signals.

Figure A6: Distributions of the absolute differences between elicited and Bayesian beliefs with 95% confidence intervals. Both Bayesian and elicited beliefs take values equal to the midpoint of their respective 5% bins and are normalized to lie in $[0, 1]$. Proportions are shown in percent. These correspond to Figures A4a, A4b and A4c.

	(1) Elicited belief	(2) Elicited belief
Bayesian belief after Phase 1	0.764*** (0.019)	0.779*** (0.022)
Treatment		0.016 (0.015)
Treatment $\times$ Bayesian belief after Phase 1		-0.031 (0.025)
R^2	0.699	0.700
Observations	1646	1646
Subjects	160	160

Clustered standard errors at the subject level in parentheses
[†] $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A.9: Regression analysis of elicited beliefs after Phase 1. Both Bayesian and elicited beliefs take values equal to the midpoint of their respective 5% bins. Standard errors are clustered at the subject level, and we control for subjects' age, gender, and whether they have taken statistics/economics courses. Subjects generally underreact to information. Corresponds to Table 2.

	(1) Elicited belief	(2) Elicited belief	(3) Elicited belief (with confirmatory signals)	(4) Elicited belief (with challenging signals)
Bayesian belief after Phase 2	0.682*** (0.016)	0.694*** (0.020)	0.753*** (0.018)	0.499*** (0.037)
Treatment		0.014 (0.013)	-0.001 (0.014)	0.059* (0.027)
Treatment $\times$ Bayesian belief after Phase 2		-0.025 (0.018)	-0.003 (0.017)	-0.111* (0.047)
R^2	0.716	0.716	0.856	0.336
Observations	1646	1646	967	679
Subjects	160	160	157	153

Clustered standard errors at the subject level in parentheses
[†] $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A.10: Regression analysis of elicited beliefs after Phase 2. Both Bayesian and elicited beliefs take values equal to the midpoint of their respective 5% bins. Standard errors are clustered at the subject level, and we control for subjects' age, gender, and whether they have taken statistics/economics courses. As in Phase 1, subjects generally underreact to information. Moreover, they underreact more to challenging signals than to confirmatory signals. Moreover, treatment strengthens underreaction to challenging signals but not confirmatory signals. Corresponds to Table 3.

More importantly, the treatment has no effect on how subjects update with confirmatory signals in Phase 2. On the other hand, the negative coefficient on the interaction between treatment and the Bayesian belief in Phase 2 suggests that treatment strengthens underreaction with challenging signals.

Table A.11 shows OLS analyses on the log-odds deviation. Columns 1–3 show that treatment induces underreaction to challenging signals but not to confirmatory signals. Moreover, underreaction to confirmatory signals is smaller in magnitude ($0.015 < 0.203$ in column 2 versus column 3; the p-value of a two-tailed t-test is 0.086).

	(1)	(2)	(3)	(4)	(5)
	Pooled	Pooled (with confirmatory signals)	Pooled (with challenging signals)	High prior	Low prior
Treatment	-0.066 (0.055)	-0.015 (0.068)	0.203* (0.086)	-0.108 (0.083)	0.050 (0.082)
Challenging	1.163*** (0.095)			1.161*** (0.127)	-1.161*** (0.102)
Treatment $\times$ Challenging	0.223* (0.102)			0.237 (0.147)	-0.214 (0.136)
Pair Controls	No	Yes	Yes	No	No
R^2	0.276	0.799	0.720	0.227	0.293
Observations	1646	967	679	909	737
Subjects	160	157	153	157	151

Clustered standard errors at the subject level in parentheses

$^\dagger p < 0.1$, $^* p < 0.05$, $^{**} p < 0.01$, $^{***} p < 0.001$

Table A.11: Regression analysis of the log-odds deviation $\log \frac{p_2}{1-p_2} - \log \frac{p_2^B}{1-p_2^B}$. Standard errors are clustered at the subject level, and we control for subjects' age, gender, and whether they have taken statistics/economics courses. We control for pair fixed effects in columns 2 and 3. Columns 1–3 report pooled analyses (overall; with confirmatory signals; with challenging signals). Column 4 restricts to high priors ($p_1 > 0.5$); column 5 restricts to low priors ($p_1 < 0.5$). Corresponds to Table 4.

Next, Table A.12 tests whether attention spent on signals is substitutable. Similar to the main text, in Table A.12, the coefficient on the triple-interaction term between treatment, the challenging dummy for the 1st–6th signals, and the challenging dummy for the last signal is not significant ($p = 0.596, 0.114, 0.480$ in columns 1, 2, and 3, respectively). Thus, we do not find supportive evidence for the capacity mechanism. Lastly, in Table A.13, the coefficients on the interaction term between the treatment dummy and the number of Phase-2 belief-challenging signals are consistent with the per-signal cost mechanism, although they are less signifi-

cant compared to the main text ($p = 0.051, 0.125, 0.146$ in columns 1, 2, and 3, respectively).

	(1) Pooled	(2) High prior	(3) Low prior
Treatment	-0.081 (0.083)	-0.102 (0.116)	0.065 (0.114)
Challenging 1 st -6 th	1.223*** (0.118)	1.421*** (0.153)	-0.954*** (0.144)
Challenging 7 th	0.333** (0.102)	0.451** (0.135)	-0.207 (0.149)
Challenging 1 st -6 th × Challenging 7 th	-0.131 (0.162)	-0.550* (0.238)	-0.384 (0.214)
Treatment × Challenging 1 st -6 th	0.063 (0.150)	0.015 (0.188)	-0.080 (0.215)
Treatment × Challenging 7 th	0.158 (0.136)	0.043 (0.187)	-0.291 (0.193)
Treatment × Challenging 1 st -6 th × Challenging 7 th	0.107 (0.201)	0.445 (0.280)	0.223 (0.314)
R^2	0.287	0.307	0.295
Observations	1601	883	718
Subjects	160	156	151

Clustered standard errors at the subject level in parentheses

† $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A.12: Regression analysis of the log-odds deviation $\log \frac{p_2}{1-p_2} - \log \frac{p_2^B}{1-p_2^B}$ that analyzes whether attention is substitutable. Standard errors are clustered at the subject level, and we control for subjects' age, gender, and whether they have taken statistics/economics courses. Column 1 reports pooled analyses; column 2 restricts to high priors ($p_1 > 0.5$); column 3 restricts to low priors ($p_1 < 0.5$). Corresponds to Table 5.

	(1) Pooled	(2) High prior	(3) Low prior
Treatment	-0.205 (0.121)	-0.269 (0.176)	0.172 (0.169)
# Challenging	0.384*** (0.035)	0.395*** (0.042)	-0.360*** (0.046)
Treatment $\times$ # Challenging	0.076 [†] (0.039)	0.082 (0.053)	-0.075 (0.051)
R^2	0.282	0.305	0.268
Observations	1601	883	718
Subjects	160	156	151
Clustered standard errors at the subject level in parentheses			
[†] $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$			

Table A.13: Regression analysis of the log-odds deviation $\log \frac{p_2}{1-p_2} - \log \frac{p_2^B}{1-p_2^B}$ that analyzes how the number of challenging signals affects the treatment effect. Standard errors are clustered at the subject level, and we control for subjects' age, gender, and whether they have taken statistics/economics courses. Column 1 reports pooled analyses; column 2 restricts to high priors ($p_1 > 0.5$); column 3 restricts to low priors ($p_1 < 0.5$). Corresponds to Table 6.