Motivated Mislearning:
The Case of Correlation Neglect

Lukas Bolte[a] Tony Q. Fan[a,*]

[a]Department of Economics, Stanford University, 579 Jane Stanford Way, Stanford, CA 94305, USA

September 19, 2022

Abstract

We design an experiment to study the role of motivated reasoning in correlation neglect. Participants receive potentially redundant signals about an ego-relevant state—their IQ test performance. We elicit their belief that the signals came from the same source (and thus contain redundant information). Participants generally underappreciate the extent to which identical signals are more likely to come from the same source, but the bias is significantly stronger for good (ego-favorable) signals than for bad (ego-unfavorable) signals. This asymmetric effect disappears in a control treatment where the state is ego-irrelevant. These results suggest that individuals may neglect the correlation between desirable signals to sustain motivated beliefs. However, the estimated effect is not quantitatively large enough to generate significant asymmetric updating about own IQ test performance.

JEL CLASSIFICATION CODES: C91, D83, D91

KEYWORDS: Beliefs, cognitive errors, correlation neglect, motivated reasoning, belief-based utility

*Corresponding author at: Department of Economics, Stanford University, 579 Jane Stanford Way, Stanford, CA 94305, USA. *E-mail addresses*: lbolte@stanford.edu (L. Bolte), tonyqfan@stanford.edu (T.Q. Fan). The experiment in this paper was registered in the AEA RCT registry as trial AEARCTR0008836 and received IRB approval at Stanford University (Protocol 44866). We are grateful to Douglas Bernheim and Muriel Niederle for their guidance and support throughout this project, and to Gonzalo Arrieta, Matthew Gentzkow, Yucheng Liang, Kirby Nielsen, Collin Raymond, and seminar participants at Stanford University for helpful comments. This work has been supported (in part) by Grant # 2002-21719 from the Russell Sage Foundation. Any opinions expressed are those of the principal investigators alone and should not be construed as representing the opinions of the Foundation.

1 Introduction

Correlated or redundant information is prevalent in the modern information economy. For instance, marketers deploy similar advertising campaigns across multiple channels; different media outlets repeat the same news generated from the same underlying sources; and individuals often self-select into clusters of social networks with similar information and beliefs. Existing research suggests that individuals may neglect such correlation in information sources and mislearn from correlated information because of inherent cognitive limitations and the complexity of the learning problem (Fedyk and Hodson, 2019; Enke and Zimmermann, 2019; Hossain and Okui, 2020).

In this paper, we propose a different perspective and study whether *motivated reasoning* could affect the degree of correlation neglect. A large literature has documented that individuals could be motivated to hold certain beliefs across different domains, as such beliefs may provide various benefits (Bénabou, 2015; Zimmermann et al., 2019): For example, favorable beliefs about one's own ability may provide ego utility (Köszegi, 2006), incentivize individuals to overcome self-control issues and make more efforts (Bénabou and Tirole, 2002; Chen and Schildberg-Hörisch, 2019), or help them to better persuade others (Schwardmann and Van der Weele, 2019). Beliefs aligned with one's political ideology may help to convey one's political identity (Bénabou and Tirole, 2011; Kahan, 2015), reduce cognitive dissonance (Akerlof and Dickens, 1982), and justify certain political actions. Since correlated information potentially provides individuals with additional flexibility to reach motivated beliefs without conscious self-deception, we hypothesize that motivated reasoning may cause or exacerbate correlation neglect.

To test this hypothesis, we conduct a controlled online experiment where individuals receive potentially redundant signals about an unknown state and make inferences about both signal redundancy and the state. To isolate the role of motivated reasoning, we design a *Main* treatment where the state is ego-relevant and thus motivated reasoning could be evoked, and a *Control* treatment where the state is ego-irrelevant and thus motivated reasoning is plausibly absent.

In our *Main* treatment, participants first complete a short IQ test, which induces a binary ego-relevant state: whether they scored in the top half of all our test takers. We elicit participants' belief over the state after they take the test. To study learning from correlated information, participants then receive signals about the state as follows. For each participant, we first generate two independent binary signals about the state. The participant is then matched with two information sources, each showing her one signal; in particular, they show her either the same underlying signal or the two independent signals, each with 50% probability. Given the realization of the observed signal(s), we ask the participant to guess whether the two sources show her the same underlying signal and indicate how likely she thinks this is the case. Finally, we elicit her posterior belief about the binary state.

If the two sources show two signals of different values, then these are surely the two independent signals; thus, our main interest lies in how participants update about signal redundancy if the two observed signals are of identical values. Bayesian inference implies that observing two signals with the same value always increases the likelihood that they are the same underlying signal, and the extent of the increase depends on one's prior belief about the binary state. However, motivated reasoning may lead individuals to underestimate the likelihood that the signals are redundant when they are ego-favorable compared to when they are ego-unfavorable.

Indeed, we find that participants update their beliefs about signal redundancy *asymmetrically* (and insufficiently relative to the Bayesian benchmark): participants who see two identical ego-favorable signals believe that they are about 4 percentage points more likely to be two independent signals compared with those who see two identical ego-unfavorable signals, consistent with motivated reasoning driving participants' inferences about signal redundancy. The effect is moderate in magnitude, statistically significant, and cannot be accounted for by Bayesian updating. The estimated effect is not quantitatively large enough to generate significant asymmetric updating about own performance in the IQ test. Nonetheless, this may be a result of the remarkable simplicity of our information structure, which leads to power issues.

To further establish that our results are not driven by confounding factors, we design a *Control* treatment where the binary state is ego-irrelevant and present participants with signals generated from the same information structure as above. We observe no asymmetric updating about signal redundancy in the *Control* treatment, and we can reject the null hypothesis that there are equal degrees of asymmetric updating in the *Main* and *Control* treatments. This strengthens our interpretation of the effect in the *Main* treatment as driven by motivated reasoning.

This paper relates to a burgeoning literature in behavioral and experimental economics that has documented systematic patterns of mislearning from complex information structures, such as redundancy neglect in social learning settings (Kübler and Weizsäcker, 2004; Eyster et al., 2015), correlation neglect in private settings (Eyster and Weizsacker, 2016; Enke and Zimmermann, 2019; Hossain and Okui, 2020; Rees-Jones et al., 2020), selection neglect (Esponda and Vespa, 2018; Barron et al., 2019; Enke, 2020), and feature neglect (Graeber, 2020). While the prior literature has shown how cognitive limitations and complexity can lead to these "mistakes," we contribute by documenting motivated reasoning as an additional explanation for them. This (i) broadens the scope for when we expect them to happen, and (ii) adds nuance to their welfare interpretations given the potential role of beliefs as a source of utility (Caplin and Leahy, 2001; Brunnermeier and Parker, 2005). We thus join a nascent literature showing how "motivated" agents can make systematic decision errors of the kind typically attributed to cognitive limitations (Exley and Kessler, 2019; Wang et al., 2021).

Our paper also contributes to the literature on motivated reasoning and motivated beliefs in psychology and economics (Kunda, 1990; Bénabou and Tirole, 2016; Epley and Gilovich, 2016), which has found applications in overconfidence (Bénabou and Tirole, 2002), moral behavior (Babcock et al., 1995), and belief polarization (Kahan, 2012). On the "supply" side of motivated beliefs, there is now evidence for motivated information demand and avoidance (Oster et al., 2013; Ganguly and Tasoff, 2016; Golman et al., 2017; Castagnetti and Schmacker, 2022), asymmetric updating to noisy signals[1] (Eil and Rao, 2011; Ertac, 2011; Sharot et al., 2011; Drobner and Goerg, 2021; Drobner, 2022; Möbius et al., 2022), motivated memory management (Chew et al., 2018; Zimmermann et al., 2019), motivated recognition (Engelmann et al., 2019), motivated evaluation of information veracity (Taber and Lodge, 2006; Thaler, 2020, 2021b), and the social exchange of motivated beliefs (Oprea and Yuksel, 2022). We document a novel mechanism that could contribute to motivated beliefs and overconfidence: asymmetric inference about information redundancy.

In sum, the literature has provided abundant evidence for both mislearning from complex information structures (e.g., correlation neglect) and motivated reasoning. Our novel contribution is to provide a proof of concept for how motivated reasoning can generate or exacerbate correlation neglect and potentially other forms of mislearning.

The rest of the paper proceeds as follows. Section 2 outlines our experimental design. Section 3 presents results from the experiment, separately by the *Main* and *Control* treatments. Section 4 discusses the implications of our results and concludes.

[1]The evidence on asymmetric updating has been mixed in the motivated beliefs literature, with some papers finding significant asymmetric updating while others finding null effects; see Benjamin (2019) for a detailed discussion. Drobner (2022) shows that the anticipation of uncertainty resolution (or the lack thereof) is a key moderating factor for these mixed results: asymmetric updating is stronger when participants do not expect the resolution of uncertainty about the state at the end of the experiment. Since in our experiment such expectations are "ambiguous," one could argue that our results provide a lower bound for the role of motivated reasoning in interpreting correlated information and for the degree of asymmetric updating.

2 Experimental Design

2.1 Environment and Treatments

Our study of the role of motivated reasoning in correlation neglect requires an environment with the following features: (i) a treatment where motivated reasoning is prone to emerge, such as when people form ego-relevant beliefs, and a control treatment where motivated reasoning is absent; (ii) researchers' control over signals and their correlation, as well as participants' knowledge of the signal-generating process; and (iii) incentivized belief elicitation. A laboratory experiment allows us to create such an environment.

In our main treatment (henceforth referred to as *Main*), participants first take an IQ test which defines an ego-relevant state (*test stage*) and then receive possibly correlated (redundant) information about the state (*information stage*). In the *information stage*, we elicit participants' beliefs about the state both before and after they receive the information, as well as their beliefs about the redundancy of the information, with incentive-compatible mechanisms. At the end of the experiment, we randomly choose one of the two stages to determine a participant's bonus payment. Our main interest lies in how participants assess the redundancy of the information depending on whether it is ego-favorable or ego-unfavorable. In order to more convincingly attribute any effect we find to motivated reasoning, we also design a control treatment (referred to as *Control*) where the state is ego-irrelevant.

Figure 1 provides a diagrammatic illustration of the different stages of the experiment, which we now describe in greater detail.

2.1.1 Test stage

In both *Main* and *Control*, participants first complete an abridged IQ test. Participants are told about the relevance of the IQ test in measuring reasoning ability and fluid intelligence. The test consists of 20 Raven's Progressive Matrices problems that participants have 5 minutes to complete.[2] The final score of a participant is the number of correct answers minus the number of incorrect answers. If the *test stage* is chosen to determine a participant's bonus, she receives 10 cents for each point she scores, but she will not lose any money if her score is negative; this leads to a bonus between \$0 and \$2. After participants complete

[2]Each of the 20 Raven's Progressive Matrices consists of a visual pattern of eight symbols and a missing piece. Participants are given eight options to complete the pattern. An example problem can be found in the experimental instructions given in Section A.2 of the online appendix. We impose a time limit of 5 minutes both to reduce the duration of the experiment and to increase the difficulty of the test.

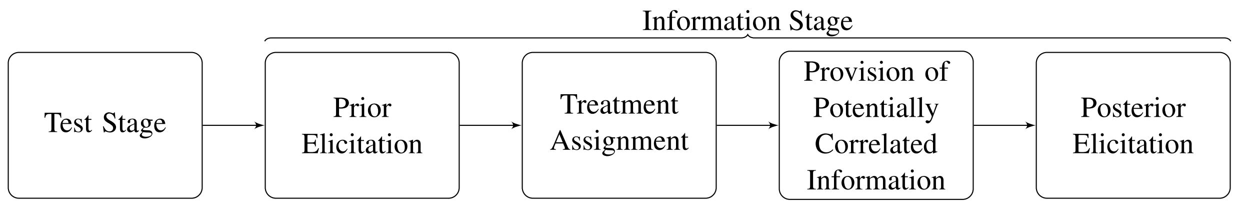


Figure 1: Experiment Stages

the IQ test, they answer the following question: "*How important is it to you to be able to perform well in the IQ test?*" on a scale from 0 to 10. This unincentivized measure is intended to be a proxy for the ego-relevance of the test for the participants.

2.1.2 Information stage

The *information stage* proceeds in several distinct steps, which we describe in order.

Prior elicitation. In both treatments, immediately after the *test stage*, we elicit each participant's beliefs about a binary ego-relevant state about her test performance: whether she scored in the top half of all participants (the *TOP* state) or the bottom half (the *BOTTOM* state).[3] In our experiment, we elicit all probabilistic beliefs over binary states using the crossover mechanism (Allen, 1987; Grether, 1992; Karni, 2009). For example, we ask each participant the probability of the *TOP* state according to her subjective beliefs, denoted by μ. We then draw a random number $\nu \sim U[0,1]$. If $\nu > \mu$, the participant receives a bonus of $2 with probability ν; if $\nu < \mu$, the participant receives a bonus of $2 if and only if she actually scored in the top half (i.e., if the *TOP* state is realized).[4] We explain to participants that they maximize their chances of getting a bonus by reporting their true beliefs, but only reveal details about the crossover mechanism if participants click on a button called "Payment Details".[5] We also inform participants truthfully that we will elicit beliefs several times but implement only one of them at random for payment if the *information stage* is chosen to determine the bonus payment.

Treatment assignment. After eliciting the priors, we assign each participant to either the *Main* treatment

[3] We use a binary state so that beliefs over the state can be summarized by a scalar probability, making them easier to elicit and analyze.

[4] This elicitation mechanism is truth-inducing under two mild assumptions: (i) Participants' preferences satisfy the monotonicity axiom in the sense that among lotteries that pay $2 with probability q and $0 with probability $(1-q)$, they strictly prefer those with higher q; (ii) Participants' marginal utility of earning $2 is independent of whether they scored in the top half (Möbius et al., 2022).

[5] Danz et al. (2022) shows that informing experimental participants of the incentive compatibility of the belief elicitation mechanism without providing precise quantitative information about the mechanism can lead to higher rates of truthful reporting. See Section A.2 of the online appendix for our complete experimental instructions.

or the *Control* treatment. Participants in *Main* will receive potentially correlated information about the ego-relevant state. For participants in *Control*, we induce a neutral ego-irrelevant state designed to be roughly comparable to the ego-relevant state (explained in detail below), and provide them with potentially correlated information about the *ego-irrelevant* state. Participants only learn about the ego-irrelevant state and the information structure after the prior elicitation, which minimizes the concern that they may distort their reports of their prior beliefs.

To make the priors approximately comparable across the two treatments, we define the priors over the ego-irrelevant state in *Control* using the following procedure: For a particular participant in *Control*, let μ denote her prior belief of having scored in the top half in the IQ test. For this participant, we define the ego-irrelevant state as a binary variable that takes the value of *TOP* with probability μ and *BOTTOM* with probability $(1-\mu)$, rounded to the nearest 5%. Specifically, we frame the ego-irrelevant state as a random draw from 20 balls consisting of $\left[\frac{\mu}{5}\right]$ TOP balls (balls with the word "TOP" written on them) and $(20-\left[\frac{\mu}{5}\right])$ BOTTOM balls (balls with the word "BOTTOM" written on them).[6] We then draw the actual realization of the ego-irrelevant state based on these probabilities. The participant knows the prior probabilities but not the realization of the state. One concern is that participants in *Control* may realize the relationship between the ego-irrelevant state and the ego-relevant state given the similar probabilities, which may lead to motivated reasoning even for the ego-irrelevant state and diminish the purpose of *Control*. However, we note that this should only reduce the differences in behavior between *Main* and *Control*, going against our hypothesis outlined in Section 2.2.

Provision of potentially correlated information. We next provide each participant with information about the state (ego-relevant in *Main* and ego-irrelevant in *Control*). To study correlation neglect, we design an information structure with potentially correlated information, which is used in both *Main* and *Control*. We then elicit both participants' beliefs about the redundancy of the information and their posterior beliefs about the state.[7]

Figure 2 illustrates the information structure we use, which is the simplest one that features correlation or redundancy. We first generate two independent "reports" (framed as "the blue report" and "the green

[6] See Section A.2.2 of the online appendix for detailed instructions. To avoid a degenerate distribution, we round very low priors to 5% instead of 0% and very high priors to 95% instead of 100%.

[7] We do not explicitly inform participants that the true state will *not* be revealed at the end of the experiment. Drobner (2022) argues that in this case, participants' expectations about the resolution of uncertainty are "ambiguous," which may attenuate the role of motivated reasoning in interpreting information. See Section 3.2 for more discussion.

report") which are binary noisy signals about the state, each matching the state with probability $\theta = \frac{2}{3}$. In other words, if the true state is *TOP*, then each report says *TOP* with probability $\frac{2}{3}$ and says *BOTTOM* with probability $\frac{1}{3}$, and vice versa. Thus, θ indicates the informativeness or diagnosticity of the reports. The participant does *not* directly observe the reports but receives two signals from computer players Ann and Bob. The participant knows that: (i) Ann's signal (henceforth the "first signal") is the blue report; (ii) Bob's signal (henceforth the "second signal") is either the blue report or the green report with equal probabilities, but Bob does not disclose which report it is. Thus, the second signal possibly contains redundant information. Participants receive extensive instructions and need to correctly answer a set of comprehension questions about the information structure before they can proceed.

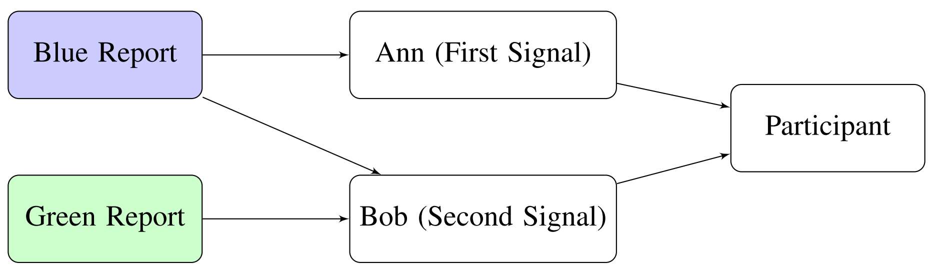


Figure 2: Information Structure with Correlation

We choose this simple information structure based on the following considerations: (i) By construction, correlation already adds to the complexity of the environment, so additional complexity risks confusing participants and adding noise to the data. (ii) A simple structure makes the environment easier to describe and analyze, so it serves as a natural starting point for our investigation.

Posterior elicitation. Finally, we ask participants to confirm the values of the two signals they observe and then elicit three beliefs from them, which are the main outcomes we analyze. (i) We first ask them to make a binary guess about whether the second signal is the blue report (and hence redundant) or the green report (and hence informative). They get a $2 bonus payment if their guess is correct and this question is randomly selected to determine their bonus. (ii) We further ask them to "indicate their confidence in their guess" by providing their subjective probability that the second signal is the report they guessed.[8] (iii) Finally, we elicit their posterior beliefs about the binary state, which is ego-relevant in *Main* and ego-irrelevant in *Control*. We again incentivize both probabilistic beliefs, (ii) and (iii), using the crossover mechanism described above.

[8]We divide the elicitation of beliefs about signal redundancy into two parts (the binary guess and the probabilistic beliefs) to help participants intuitively understand the object to be elicited.

Finally, the experiment ends with an unincentivized exit survey that asks about several demographic variables, which we control for in our analysis. At the end of the experiment, we randomly choose one of the two stages to determine a participant's bonus payment: if the *test stage* is chosen, the participant's performance in the IQ test determines her bonus payment; if the *information stage* is chosen, we randomly choose one of the incentivized belief elicitations to determine the participant's bonus payment.[9]

2.2 Theoretical Benchmarks and Hypotheses

We now derive the posteriors for a Bayesian agent in our experiment, which provide a set of benchmarks against which our participants' posteriors can be compared. We then discuss our main hypothesis in this paper — motivated reasoning may distort participants' inferences about information redundancy asymmetrically depending on whether the information is ego-favorable or ego-unfavorable.

A Bayesian agent who sees two different signals, i.e., one *TOP* signal and one *BOTTOM* signal, should clearly conclude that these are the two underlying reports. Figure A1 in the online appendix shows that most of our participants indeed do so. In the rest of the main text, we will focus on the case where participants see two identical signals: either two *TOP* signals or two *BOTTOM* signals. In such cases, the identity of the second signal cannot be inferred with certainty: Either it is a repetition of the first signal, or it is a different report which turns out to coincide with the first signal.

We first derive the Bayesian posteriors about the redundancy of the second signal. Recall that, unconditionally, the second signal is a repetition of the first signal with $\frac{1}{2}$ probability. When a participant sees two identical signals such as two *TOP* signals, the probability that the second signal is new (i.e., NOT redundant) decreases:

$$\begin{aligned}
&\Pr(\text{Second Signal New}|\text{Two } \textit{TOP} \text{ Signals})\\
=&\frac{\Pr(\text{Second Signal New, Two } \textit{TOP} \text{ Signals})}{\Pr(\text{Second Signal Redundant, Two } \textit{TOP} \text{ Signals}) + \Pr(\text{Second Signal New, Two } \textit{TOP} \text{ Signals})}\\
=&\frac{\frac{1}{2}[\mu\theta^2 + (1-\mu)(1-\theta)^2]}{\frac{1}{2}[\mu\theta + (1-\mu)(1-\theta)] + \frac{1}{2}[\mu\theta^2 + (1-\mu)(1-\theta)^2]} = \frac{3\mu+1}{6\mu+4} \in [25\%, 40\%]
\end{aligned}$$

where μ denotes the agent's prior belief that the state is *TOP* and $\theta = \frac{2}{3}$ indicates the diagnosticity of the

[9] Azrieli et al. (2018) discusses the rationales for paying for exactly one randomly selected task in an experiment with multiple tasks.

signals. Analogously,

$$\Pr(\text{Second Signal New}|\text{Two } \textit{BOTTOM} \text{ Signals}) = \frac{3(1-\mu)+1}{6(1-\mu)+4} \in [25\%, 40\%]$$

Thus, a Bayesian agent who sees two identical signals, regardless of their prior beliefs over the state, should infer that it is more likely that the second signal is redundant (and hence guess that the second signal is the blue report).

However, we hypothesize that ego-relevance may affect how individuals interpret possibly redundant information about the state in the following way: When the state is ego-relevant, as in *Main*, individuals may be motivated to underestimate redundancy when the information is "good," relative to the case in which the information is "bad," so as to update positively about the state. In other words, participants in *Main* who observe two ego-favorable signals may be motivated to assign a relatively high probability of the second signal being new, compared to participants who observe two ego-unfavorable signals. Since the motivated reasoning mechanism should be plausibly shut down in *Control*, we hypothesize that these effects diminish in that treatment. We refer to our hypothesis as the *Motivated Mislearning Hypothesis.*

Finally, we consider the Bayesian posteriors about the state after receiving the information. Let $\text{logit}(x) = \ln(\frac{x}{1-x})$ be the logit or log odds function. If a Bayesian agent observes two *TOP* signals, the logit of her posterior can be written as the sum of the logit of the prior and the log likelihood ratio of the signals:

$$\text{logit}(\eta) = \text{logit}(\mu) + \ln\left(\frac{\Pr(\textit{TOP} \text{ state, Two } \textit{TOP} \text{ Signals})}{\Pr(\textit{BOTTOM} \text{ state, Two } \textit{TOP} \text{ Signals})}\right)$$

where μ denotes the agent's prior belief that the state is *TOP* and η denotes the posterior belief. With a signal diagnosticity of $\theta = \frac{2}{3}$, the likelihood ratio of two *TOP* signals is 2.5. The symmetry of the signal structure implies that the likelihood ratio of two *BOTTOM* signals is given by $\frac{1}{2.5}$ or 0.4. Along the same lines, if someone receives one *TOP* signal and one *BOTTOM* signal, then they should exactly cancel each other out, and the Bayesian posterior exactly equals the prior. Thus, the general expression for the Bayesian posterior belief for the *TOP* state is:

$$\text{logit}(\eta) = \text{logit}(\mu) + \ln(2.5) \cdot \mathbb{1}_{\text{Two Good Signals}} - \ln(2.5) \cdot \mathbb{1}_{\text{Two Bad Signals}} + 0 \cdot \mathbb{1}_{\text{Mixed Signals}}$$

However, the *Motivated Mislearning Hypothesis* predicts that participants may react more to ego-favorable than to ego-unfavorable signals through the channel of motivated mislearning from correlated signals.

2.3 Procedures

We programmed our experiment using oTree (Chen et al., 2016). From March to May in 2021, we recruited our participants through Prolific, an online platform designed for social science research[10]. 601 participants participated in our *Main* treatment, among whom 444 got two identical signals. Another 601 participants participated in our *Control* treatment, among whom 454 got two identical signals. Our main sample in the rest of the paper consists of 898 participants, including the 444 participants who see two identical signals in *Main* and the 454 participants who see two identical signals in *Control*.

Participants spent on average about 20 minutes on the experiment and earned an average payment of $3.9, including a $3 base payment.

3 Results

This section presents the evidence for the *Motivated Mislearning Hypothesis* from our experiment. After confirming that our sample is balanced between *Main* and *Control*, we show that participants update asymmetrically about signal redundancy in *Main* but such asymmetric updating disappears in *Control*, supporting the *Motivated Mislearning Hypothesis*.

3.1 Descriptive Statistics and Treatment Balance

In Table 1, we provide some descriptive statistics of several pre-treatment variables (raw score in the IQ test, subjective importance of the IQ test, and prior belief of the *TOP* state) and demographic variables, separately for participants in *Main* and *Control*. These variables are evidently well balanced between *Main* and *Control*, as none of the p-values from unpaired t-tests of equal means is smaller than 0.10.

Table 1 also shows that there is a large variation among participants in all the pre-treatment IQ-test-related variables. The average participant attaches reasonable importance to being able to perform well in the IQ test (higher than 7 out of 10), which bolsters our confidence in inducing motivated reasoning in *Main*. Participants' average prior beliefs of the *TOP* state are only slightly higher than 50%, indicating

[10]See Palan and Schitter (2018) and Gupta et al. (2021) for using Prolific as a participant pool. We only recruited US participants who have completed more than 100 tasks on Prolific and have an approval rate of at least 95%.

that the difficulty of the test is well calibrated. For completeness, Table 1 also shows the statistics for the *posterior* belief of the *TOP* state: On average, the posteriors are close to the priors in both *Main* and *Control*, implying no significant asymmetric updating *on average* in either treatment. We expand on this point with more detailed results in Section 3.2.

3.2 Results from *Main*

We first use the data from our *Main* treatment to test our *Motivated Mislearning Hypothesis* — participants' inferences about signal redundancy may depend on whether the signals are ego-favorable or ego-unfavorable. We focus on the two outcomes that measure participants' beliefs about signal redundancy: the binary guess and the probabilistic belief about the identity of the second signal. We then look at how participants update their beliefs about their IQ test performance after receiving the signals.

As discussed above, if one's prior belief of scoring in the top half is μ, then the Bayesian inference about signal redundancy after seeing two identical signals should be

$$\Pr(\text{Second Signal is NOT Redundant}|\text{Both Signals are } \textit{TOP}) = \frac{3\mu+1}{6\mu+4} \in [25\%, 40\%]$$

and

$$\Pr(\text{Second Signal is NOT Redundant}|\text{Both Signals are } \textit{BOTTOM}) = \frac{4-3\mu}{10-6\mu} \in [25\%, 40\%]$$

Therefore, the correct binary guess after seeing two identical signals is always that the second signal is redundant.

Panel A of Figure 3 shows the percentages of participants who make the *wrong* guess (i.e., guess that the second signal is new or non-redundant) after seeing two identical signals. Although our main interest lies in comparing participants' reactions after seeing two *TOP* signals or two *BOTTOM* signals, we also break down the sample by participants' IQ test performance (top half vs. bottom half): Since the signals are informative about test performance, participants who see two *TOP* signals are more likely to have scored in the top half than those who see two *BOTTOM* signals, and thus may have higher cognitive ability or may be more attentive. Making the comparison *within* each performance group eliminates this confounding factor and exploits only the exogenous variation from signal randomness built into our design.

Indeed, we find that participants scoring in the top half are in general less likely to make the wrong guess than those in the bottom half (24% vs. 43%, $p < 0.01$), which necessitates our within-performance-group comparison. However, we fail to reject the null hypothesis that participants seeing two *TOP* signals are as likely to make the wrong guess as participants seeing two *BOTTOM* signals within each performance group. We thus find no evidence for our *Motivated Mislearning Hypothesis* in the binary guess outcome.

The binary guess is only a coarse proxy for the inference about signal redundancy, so we next look at participants' probabilistic belief that the second signal is new after seeing two identical signals. Note that given our design the Bayesian benchmark for this probability depends on the participant's prior belief about their test performance, and is generally different between people who see two *TOP* signals and people who see two *BOTTOM* signals. For example, if a participant's prior belief of scoring in the top half is higher than 50% (i.e., they are initially relatively confident about their performance), then they should think that the second signal is relatively more likely to be new after seeing two *TOP* signals; by contrast, if they see two *BOTTOM* signals, they should think that there is a high chance that the second signal is redundant. It is thus important that we control for the Bayesian benchmark when we compare the inferences about signal redundancy made by the "two *TOP*" group and the "two *BOTTOM*" group.

Panel B of Figure 3 shows both the average elicited posteriors and the average Bayesian posteriors that the second signal is new after seeing two identical signals, separately by IQ test performance group and signal valence. Focusing first on participants who scored in the top half, we notice that the average Bayesian posterior of the second signal being new is slightly ($\sim 3\%$) higher after seeing two *TOP* signals than after seeing two *BOTTOM* signals, consistent with this subgroup of participants being relatively confident about their performances initially. Their actual posteriors are much higher than the Bayesian posteriors on average, suggesting that they infer too little about signal redundancy from two identical signals. Importantly, the average elicited posterior about the probability that the second signal is new is around 5% higher for the two *TOP* group than the two *BOTTOM* group, a gap that is 2% wider than the corresponding gap for the Bayesian posteriors. This is directionally consistent with our *Motivated Mislearning Hypothesis*: Individuals who see repeated ego-favorable signals may be motivated to overestimate the chance that they contain independent information.

Analysis of the beliefs of participants who scored in the bottom half yields the same pattern. As these participants' prior beliefs about their performance are less confident, the average Bayesian posterior of the second signal being new is 1% lower with two *TOP* signals than with *BOTTOM* signals. However, the elicited posteriors from the participants show the opposite: Participants seeing two *TOP* signals on average decide that it is 49% likely that the second signal is new, which is 3% higher than participants seeing two *BOTTOM* signals. This provides even stronger evidence for the *Motivated Mislearning Hypothesis* that cannot be explained by Bayesian inference.

Appendix Figure A2 replicates the above results by showing the full distribution of elicited posterior beliefs (separately by IQ test performance and signal valence) and p-values from two-sided t-tests of the effect of signal valence: the effect is statistically significant ($p = 0.030$) for participants who scored in the top half but statistically insignificant ($p = 0.264$) for participants who scored in the bottom half, which is not unreasonable since the Bayesian prediction goes in the opposite direction for the latter group.

To increase power and formally control for the Bayesian posterior in the statistical test, we put these results into regression form by pooling participants who scored in the top and bottom halves and estimating the following specification:

$$\text{Posterior Belief} = \beta_1 \mathbb{1}_{\text{Two Good Signals}} + \beta_2 \mathbb{1}_{\text{Top Half in Test}} + \beta_3 \text{Bayesian Posterior} + \gamma X + \epsilon \qquad (1)$$

where the dependent variable is the posterior belief (in %) that the second signal is new, and X may be empty or include a set of controls. β_1 measures the degree of motivated bias in posterior beliefs, and the *Motivated Mislearning Hypothesis* predicts that $\beta_1 > 0$: a participant who sees two good signals is motivated to believe that the second signal is more likely to be new compared with someone who sees two bad signals and who has the same Bayesian posterior. It is also likely that $\beta_2 < 0$ as those who scored in the top half in the IQ test have on average higher cognitive ability and may thus be less susceptible to correlation neglect. [11] The test of the *Motivated Mislearning Hypothesis* here exploits the random variation in signals within the same IQ test performance group.

Columns (1) and (2) in Table 2 present the regression results using the full sample of 444 participants who see two identical signals. Controlling for test performance and the Bayesian posterior, we find that participants who see two good signals believe that the second signal is 4 percentage points more likely to be new compared with those who see two bad signals ($p = 0.036$), consistent with the *Motivated Mislearning Hypothesis*. This effect, while moderate in magnitude, is still notable as probabilistic beliefs elicited in experiments are often compressed towards 50% due to issues such as extreme-belief aversion (Benjamin,

[11] As has been pointed out before, since good signals are correlated with good IQ test performance, our test of the *Motivated Mislearning Hypothesis* will be invalid if we do not control for IQ test performance.

2019) and cognitive uncertainty (Enke and Graeber, 2019). The effect becomes slightly larger when we additionally control for personal characteristics including gender, race, age, education, and the raw score in the IQ test. Scoring in the top half in the IQ test is negatively and significantly associated with the posterior belief, but the coefficient switches signs and becomes insignificant once the raw score in the test is controlled for, since the two covariates are highly correlated. We also find that the coefficients before the Bayesian posterior are generally small and insignificant, which suggests that *across participants* their actual posteriors are only weakly correlated with the Bayesian posteriors.[12]

Table A1 in the online appendix explores potential gender differences in the motivated bias by reporting regression results separately for male and female participants. The motivated bias is statistically significant for female participants and is smaller and statistically insignificant for male participants; the magnitude and statistical significance of the gender difference are sensitive to the regression specification.[13] These results stand in contrast to certain findings in the literature that suggest males may be more susceptible to performance-related motivated reasoning than females (Thaler, 2021a).

We further examine the robustness of the results in two ways. First, in columns (3) and (4) of Table 2, we drop participants who report a higher than 50% posterior that the second signal is new after seeing two identical signals, which indicates that they update in the wrong direction about signal redundancy. The effect becomes even stronger in the restricted sample. Second, Table A2 in the online appendix presents an alternative regression specification using the *difference* between the actual posterior and the Bayesian posterior as the dependent variable, which generates similar results.

Why do we find an asymmetric effect in the probabilistic beliefs but no effect in the binary guess? In Appendix Table A3, we present a two-way tabulation of the probabilistic beliefs and the binary guess. The results suggest that receiving two good signals (instead of two bad signals) shifts around 10% of participants from providing a posterior belief of less than 50% to providing a posterior belief of exactly 50%. This is consistent with the *Motivated Mislearning Hypothesis*. However, since only around a fourth to a half of those reporting exactly 50% guess that the second signal is new in the binary guess, the effect size on the

[12]Note that the Bayesian posteriors only vary between 25% and 40%, and the variation is only driven by participants' prior beliefs in the binary state and the realizations of the signals (*TOP* or *BOTTOM*). More specifically, the Bayesian posteriors should be the same between someone with a prior belief of μ in the *TOP* state and sees two *TOP* signals, and someone with a prior belief of $(1-\mu)$ in the *TOP* state and sees two *BOTTOM* signals. What our regression essentially identifies is that the actual posterior of the first person is higher than that of the second person, controlling for other covariates.

[13]Note that the sample size of female participants is somewhat larger than that of male participants. Also, note that these results do *not* imply that females update more optimistically than males in general: in results unreported here, we in fact find that males *overall* update slightly more to good signals than to bad signals, and females *overall* update slightly more to bad signals than to good signals, although neither of these effects is statistically significant. Results are available upon request.

binary guess would be only a fourth to a half of 10%. There is also a small fraction of participants who indicate beliefs different from 50% but provide guesses inconsistent with their own beliefs, which further diminishes the effect on the binary guess. We conclude that the binary guess as a crude measure is not able to pick up the subtle effects on beliefs that we document.

Finally, we examine whether the asymmetric inference about signal redundancy that we document leads to significant asymmetric updating about participants' own IQ test performance. Recall that the Bayesian posterior about own test performance is as follows:

$$\mathrm{logit}(\eta) = \mathrm{logit}(\mu) + \ln(2.5) \cdot \mathbb{1}_{\text{Two Good Signals}} - \ln(2.5) \cdot \mathbb{1}_{\text{Two Bad Signals}} + 0 \cdot \mathbb{1}_{\text{Mixed Signals}} \tag{2}$$

where μ denotes the prior belief for the *TOP* state and η denotes the posterior belief. Equation (2) says that the log likelihood ratio of beliefs should go up by $\log(2.5)$ if one receives two (possibly redundant) good signals, should go down by $\log(2.5)$ if one receives two (possibly redundant) bad signals, and should not change if one receives one good signal and one bad signal.

In light of Equation (2), and focusing on the subsample of participants who receive two identical signals, we estimate the following specification (Grether, 1980):

$$\mathrm{logit}(\eta) = \delta\, \mathrm{logit}(\mu) + \alpha_G \ln(2.5) \cdot \mathbb{1}_{\text{Two Good Signals}} - \alpha_B \ln(2.5) \cdot \mathbb{1}_{\text{Two Bad Signals}} + \epsilon \tag{3}$$

We report the results in Table 3. We estimate $\alpha_G = 0.641$ and $\alpha_B = 0.621$, without a statistically significant difference between the two coefficients ($p = 0.802$), while both are significantly different from 1 ($p < 0.001$). In short, even in our setting of potentially redundant signals, the asymmetric inference about signal redundancy is not quantitatively large enough to generate significant asymmetric updating about participants' own IQ test performance. This stands in contrast to some existing evidence in the literature such as Möbius et al. (2022) which finds significant asymmetric updating with mutually independent signals.[14]

Why do we find an asymmetric updating effect on signal redundancy but in the end no significant asymmetric updating about test performance? Power issues arise because: (i) the effect on the inference about signal redundancy is moderate in magnitude to begin with; and (ii) all in all participants still underreact to the signals by a factor of 0.6 relative to the Bayesian benchmark, implying a large degree of conservatism,

[14] As mentioned in Section 1, there is mixed evidence on asymmetric updating in the literature. See Footnote 1 for a more detailed discussion.

even as they do not fully appreciate the potential redundancy of the second signal. For example, back-of-the-envelope calculations suggest that a 5% effect on the posterior probability that the second signal is redundant can generate no more than an effect of 0.03 on the α_G coefficient in (3). Our remarkably simple information structure, intended as the most transparent way to introduce correlation, may not be sufficient to ultimately generate a statistically detectable asymmetric effect on inference about the state. Nonetheless, the mechanism it illustrates may lead to larger effects in richer environments with a larger number of correlated signals, which we leave for future work.

3.3 Results from *Control*

Although evidence from the belief data in *Main* is consistent with the *Motivated Mislearning Hypothesis*, the need to control for the Bayesian posterior may lead some to worry that our results are sensitive to the chosen specification. To alleviate this concern, we conduct a "placebo" exercise by testing for a similar asymmetric effect in *Control* where participants receive potentially redundant signals about an *ego-irrelevant* state and the priors over the states are matched with participants' prior beliefs about their test performance.

Figure 4 exactly replicates Figure 3 using data from *Control*.[15] Specifically, Panel A shows the percentages of participants who make the *wrong* guess (i.e., guess that the second signal is new or non-redundant) after seeing two identical signals in *Control*. Here we find that participants seeing two *TOP* signals are actually *less* likely to guess that the second signal is new.[16] This implies that in *Main* the motivated bias may be overshadowed by an intrinsic cognitive bias working in the opposite direction, leading to a null result for the binary guess outcome. Panel B further shows the average elicited and Bayesian posteriors that the second signal is new after seeing two identical signals in *Control*, separately by IQ test performance group and signal valence. Again we find that participants seeing two *TOP* signals believe that it is slightly *less* likely that the second signal is new compared with participants seeing two *BOTTOM* signals, which goes in the opposite direction compared to *Main*.[17] Overall, the asymmetric updating effect on signal redundancy that we document in *Main* disappears or even reverses signs in *Control*. To ease the comparison between

[15]When interpreting Figure 4, note that in *Control* having scored in the top half in the IQ test is not equivalent to the ego-irrelevant state being *TOP*, because the ego-irrelevant state is randomly drawn based on prior probabilities matched to the participant's prior beliefs of having scored in the top half.

[16]We explore the source of this gap in Appendix Table A4, which shows a two-way tabulation of the probabilistic beliefs and the binary guess in *Control*. The gap seems to be partly driven by a higher propensity to provide a posterior belief of higher than 50% for those who see two *BOTTOM* signals.

[17]Figure A3 in the online appendix presents the full distribution of elicited posterior beliefs separately by IQ test performance and signal valence in *Control*.

Main and *Control* treatments, Figure A4 in the online appendix shows the above outcomes for *Main* and *Control* side by side. The evidence is consistent with motivated reasoning exacerbating correlation neglect in *Main* relative to *Control* when the two signals are *TOP*; in the case of two *BOTTOM* signals, the degree of correlation neglect is either similar or lower in *Main* relative to *Control*.[18] These patterns are broadly consistent with the *Motivated Mislearning Hypothesis*.

In Table 4, we present regression results pooling data from *Main* and *Control*, using the following specification:[19]

$$\begin{aligned}\text{Posterior Belief} \quad &= \beta_1 \mathbb{1}_{\textit{Main}\text{ Treatment}} \times \mathbb{1}_{\text{Two }\textit{TOP}\text{ Signals}} + \beta_2 \mathbb{1}_{\textit{Control}\text{ Treatment}} \times \mathbb{1}_{\text{Two }\textit{TOP}\text{ Signals}} \\ &+ \beta_3 \mathbb{1}_{\textit{Main}\text{ Treatment}} \times \mathbb{1}_{\text{Top Half in Test}} + \beta_4 \mathbb{1}_{\textit{Control}\text{ Treatment}} \times \mathbb{1}_{\text{Top Half in Test}} \\ &+ \beta_5 \mathbb{1}_{\textit{Main}\text{ Treatment}} + \beta_6 \text{Bayesian Belief} + \gamma X + \epsilon \end{aligned} \tag{4}$$

where the dependent variable is the posterior belief (in %) that the second signal is new, and X may be empty or include a set of controls.[20] Again the *Motivated Mislearning Hypothesis* predicts that $\beta_1 > 0$ and $\beta_2 \approx 0$ as the motivated bias in inference should diminish in *Control*.

Columns (1) and (2) in Table 4 present results using the full sample of 898 participants in the two treatments who see two identical signals. In column (1), the interaction term between the *Main* treatment and the "Two *TOP* Signals" dummy replicates the result that in the *Main* treatment participants' beliefs that the second signal is new is 4% higher if they see two *TOP* signals ($p = 0.030$). Importantly, the interaction term between the *Control* treatment and the "Two *TOP* Signals" dummy is estimated to be negative and insignificant, indicating that the asymmetric effect disappears in *Control*. We reject that the effect is the same in the two treatments at the 5% level ($p = 0.049$). In column (2), we additionally control for a battery of demographic and pre-treatment variables. In columns (3) and (4), we drop participants who report a higher than 50% posterior that the second signal is new. Finally, Table A7 in the online appendix presents an alternative regression specification using the *difference* between the actual posterior and the Bayesian

[18] Figure A5 in the online appendix pools the data from *Main* and *Control* treatments and presents the full distribution of elicited posterior beliefs separately by IQ test performance, signal valence, and treatment.

[19] Tables A5 and A6 in the online appendix present regression results with the *Control* treatment data only, using exactly the same specifications as those for the *Main* treatment in Section 3.2.

[20] In *Control*, participants who see two *TOP* signals have higher prior beliefs about their test performance and thus may have higher cognitive ability. To control for this, we add dummies for each possible level of the prior probability of the *TOP* state. To maintain symmetry in the specification, we define these dummies for participants in *Main* as well, rounding their prior beliefs about their own test performance to the nearest 5%, moving very low priors to 5% instead of 0% and very high priors to 95% instead of 100%. We then include all these dummies interacted with the treatment in the regressions.

posterior as the dependent variable. In all these alternative specifications, we obtain similar results.

In summary, the results from *Control* suggest that the asymmetric effect in *Main* is specific to the ego-relevant state of own test performance. Moreover, when the two signals are *TOP*, the degree of correlation neglect is consistently higher in *Main* compared with in *Control*. Overall, these results strengthen the empirical case for the *Motivated Mislearning Hypothesis*.

4 Conclusion

In this paper, we design an experiment to study whether motivated reasoning could affect individual learning from correlated information. We find that participants who receive identical ego-favorable signals indicate that the signals are 4 percentage points less likely to be redundant compared with participants who receive identical ego-unfavorable signals; this effect disappears or even reverses signs when the signals are framed as ego-irrelevant. This finding demonstrates the degree of flexibility in interpreting correlated information in settings prone to motivated beliefs. Although such asymmetric inference about signal redundancy does not lead to significant asymmetric updating about own performance in our experiment, we note that this may result from the remarkable simplicity of our design. In richer settings with a larger number of correlated signals, the asymmetric effect we document may well build up and ultimately generate more significant effects on the beliefs of interest.

More generally, our findings suggest that motivated reasoning can be an additional driver of learning "mistakes" that have typically been associated with cognitive limitations and complexity, which has a few further implications. First, we may expect these "mistakes" to emerge or intensify in settings where they help individuals reach certain desirable beliefs. For example, for managers and marketers considering persuading consumers or other stakeholders with repeated information, our findings suggest that this strategy may be more effective when such information caters to the information receivers' motivated beliefs (e.g., when it enhances their self-image). Second, the welfare interpretations of these "mistakes" become less clear-cut given the potential role of beliefs as a source of utility (Caplin and Leahy, 2001; Brunnermeier and Parker, 2005). Therefore, for policymakers interested in de-biasing consumers, we note that reducing these "motivated mistakes" may require different approaches compared to the case of purely cognitive mistakes (Exley and Kessler, 2019) and may not even be welfare-enhancing in the first place.

We highlight two avenues for future research. First, it is of interest to explore the effect of motivated reasoning on other types of learning biases. For example, motivated reasoning may lead individuals to learn narrowly from selected information without accounting for the information that is filtered out but can in principle be deduced from observed information (Enke, 2020). Second, our findings can be applied to other settings where motivated reasoning may play a role, such as political beliefs.

References

Akerlof, George A and William T Dickens, "The economic consequences of cognitive dissonance," *The American economic review*, 1982, *72* (3), 307–319.

Allen, Franklin, "Discovering personal probabilities when utility functions are unknown," *Management Science*, 1987, *33* (4), 542–544.

Azrieli, Yaron, Christopher P Chambers, and Paul J Healy, "Incentives in experiments: A theoretical analysis," *Journal of Political Economy*, 2018, *126* (4), 1472–1503.

Babcock, Linda, George Loewenstein, Samuel Issacharoff, and Colin Camerer, "Biased judgments of fairness in bargaining," *The American Economic Review*, 1995, *85* (5), 1337–1343.

Barron, Kai, Steffen Huck, and Philippe Jehiel, "Everyday econometricians: Selection neglect and overoptimism when learning from others," Technical Report, WZB Discussion Paper 2019.

Bénabou, Roland and Jean Tirole, "Self-confidence and personal motivation," *The Quarterly Journal of Economics*, 2002, *117* (3), 871–915.

_ and _, "Identity, morals, and taboos: Beliefs as assets," *The Quarterly Journal of Economics*, 2011, *126* (2), 805–855.

_ and _, "Mindful economics: The production, consumption, and value of beliefs," *Journal of Economic Perspectives*, 2016, *30* (3), 141–64.

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

Brunnermeier, Markus K. and Jonathan A. Parker, "Optimal Expectations," *American Economic Review*, September 2005, *95* (4), 1092–1118.

Bénabou, Roland, "The Economics of Motivated Beliefs," *Revue d'économie politique*, 2015, *125* (5), 665–685.

Caplin, Andrew and John Leahy, "Psychological expected utility theory and anticipatory feelings," *The Quarterly Journal of Economics*, 2001, *116* (1), 55–79.

Castagnetti, Alessandro and Renke Schmacker, "Protecting the ego: Motivated information selection and updating," *European Economic Review*, 2022, p. 104007.

Chen, Daniel L, Martin Schonger, and Chris Wickens, "oTree—An open-source platform for laboratory, online, and field experiments," *Journal of Behavioral and Experimental Finance*, 2016, *9*, 88–97.

Chen, Si and Hannah Schildberg-Hörisch, "Looking at the bright side: The motivational value of confidence," *European Economic Review*, 2019, *120*, 103302.

Chew, Soo Hong, Wei Huang, and Xiaojian Zhao, "Motivated false memory," *Available at SSRN 2127795*, 2018.

Danz, David, Lise Vesterlund, and Alistair J Wilson, "Belief elicitation and behavioral incentive compatibility," *American Economic Review*, 2022.

Drobner, Christoph, "Motivated beliefs and anticipation of uncertainty resolution," *American Economic Review: Insights*, 2022, *4* (1), 89–105.

_ **and Sebastian J Goerg**, "Motivated belief updating and rationalization of information," 2021.

Eil, David and Justin M Rao, "The good news-bad news effect: asymmetric processing of objective information about yourself," *American Economic Journal: Microeconomics*, 2011, *3* (2), 114–38.

Engelmann, Jan, Maël Lebreton, Peter Schwardmann, Joel J van der Weele, and Li-Ang Chang, "Anticipatory anxiety and wishful thinking," 2019.

Enke, Benjamin, "What you see is all there is," *The Quarterly Journal of Economics*, 2020, *135* (3), 1363–1398.

_ **and Florian Zimmermann**, "Correlation neglect in belief formation," *The Review of Economic Studies*, 2019, *86* (1), 313–332.

_ **and Thomas Graeber**, "Cognitive uncertainty," Technical Report, National Bureau of Economic Research 2019.

Epley, Nicholas and Thomas Gilovich, "The mechanics of motivated reasoning," *Journal of Economic Perspectives*, 2016, *30* (3), 133–40.

Ertac, Seda, "Does self-relevance affect information processing? Experimental evidence on the response to performance and non-performance feedback," *Journal of Economic Behavior & Organization*, 2011, *80* (3), 532–545.

Esponda, Ignacio and Emanuel Vespa, "Endogenous sample selection: A laboratory study," *Quantitative Economics*, 2018, *9* (1), 183–216.

Exley, Christine L and Judd B Kessler, "Motivated errors," Technical Report, National Bureau of Economic Research 2019.

Eyster, Erik and Georg Weizsacker, "Correlation neglect in portfolio choice: Lab evidence," *Available at SSRN 2914526*, 2016.

_ , Matthew Rabin, and Georg Weizsacker, "An experiment on social mislearning," 2015.

Fedyk, Anastassia and James Hodson, "When Can the Market Identify Old News?," *Available at SSRN 2433234*, 2019.

Ganguly, Ananda and Joshua Tasoff, "Fantasy and dread: The demand for information and the consumption utility of the future," *Management Science*, 2016, *63* (12), 4037–4060.

Golman, Russell, David Hagmann, and George Loewenstein, "Information avoidance," *Journal of Economic Literature*, 2017, *55* (1), 96–135.

Graeber, Thomas, "Inattentive inference," *Available at SSRN 3658112*, 2020.

Grether, David M, "Bayes rule as a descriptive model: The representativeness heuristic," *The Quarterly journal of economics*, 1980, *95* (3), 537–557.

_ , "Testing Bayes rule and the representativeness heuristic: Some experimental evidence," *Journal of Economic Behavior & Organization*, 1992, *17* (1), 31–57.

Gupta, Neeraja, Luca Rigotti, and Alistair Wilson, "The Experimenters' Dilemma: Inferential Preferences over Populations," *arXiv preprint arXiv:2107.05064*, 2021.

Hossain, Tanjim and Ryo Okui, "Belief formation under signal correlation," *Available at SSRN 3218152*, 2020.

Kahan, Dan M, "Ideology, motivated reasoning, and cognitive reflection: An experimental study," *Judgment and Decision making*, 2012, *8*, 407–24.

_ , "The Politically Motivated Reasoning Paradigm, Part 1: What Politically Motivated Reasoning Is and How to Measure It," *Emerging Trends in the Social and Behavioral Sciences: An Interdisciplinary, Searchable, and Linkable Resource*, 2015, pp. 1–16.

Karni, Edi, "A mechanism for eliciting probabilities," *Econometrica*, 2009, *77* (2), 603–606.

Köszegi, Botond, "Ego utility, overconfidence, and task choice," *Journal of the European Economic Association*, 2006, *4* (4), 673–707.

Kübler, Dorothea and Georg Weizsäcker, "Limited depth of reasoning and failure of cascade formation in the laboratory," *The Review of Economic Studies*, 2004, *71* (2), 425–441.

Kunda, Ziva, "The case for motivated reasoning.," *Psychological bulletin*, 1990, *108* (3), 480.

Möbius, Markus M, Muriel Niederle, Paul Niehaus, and Tanya S Rosenblat, "Managing self-confidence: Theory and experimental evidence," *Management Science*, 2022.

Oprea, Ryan and Sevgi Yuksel, "Social exchange of motivated beliefs," *Journal of the European Economic Association*, 2022, *20* (2), 667–699.

Oster, Emily, Ira Shoulson, and E Dorsey, "Optimal expectations and limited medical testing: evidence from Huntington disease," *American Economic Review*, 2013, *103* (2), 804–30.

Palan, Stefan and Christian Schitter, "Prolific.ac—A subject pool for online experiments," *Journal of Behavioral and Experimental Finance*, 2018, *17*, 22–27.

Rees-Jones, Alex, Ran Shorrer, and Chloe J Tergiman, "Correlation neglect in student-to-school matching," in "Proceedings of the 21st ACM Conference on Economics and Computation" 2020, pp. 467–468.

Schwardmann, Peter and Joel Van der Weele, "Deception and self-deception," *Nature human behaviour*, 2019, pp. 1–7.

Sharot, Tali, Christoph W Korn, and Raymond J Dolan, "How unrealistic optimism is maintained in the face of reality," *Nature neuroscience*, 2011, *14* (11), 1475.

Taber, Charles S and Milton Lodge, "Motivated skepticism in the evaluation of political beliefs," *American journal of political science*, 2006, *50* (3), 755–769.

Thaler, Michael, "The Fake News Effect: Experimentally Identifying Motivated Reasoning Using Trust in News," *arXiv preprint arXiv:2012.01663*, 2020.

_ , "Gender differences in motivated reasoning," *Journal of Economic Behavior & Organization*, 2021, *191*, 501–518.

_ , "The supply of motivated beliefs," *arXiv preprint arXiv:2111.06062*, 2021.

Wang, Ao, Shaoda Wang, and Xiaoyang Ye, "When Information Conflicts With Obligations: the Role of Motivated Cognition," 2021.

Zimmermann, Florian et al., "The dynamics of motivated beliefs," Technical Report, University of Bonn and University of Mannheim, Germany 2019.

Figures and Tables

Table 1: Descriptive statistics and tests of balance between *Main* and *Control*.

	Main Treatment	*Control* Treatment	*Main* = *Control* (p-value)
Raw Score in IQ Test	3.48	3.11	0.46
	(6.98)	(7.66)	
Importance of IQ Test (0 – 10)	7.17	7.10	0.67
	(2.36)	(2.38)	
Prior Belief of *TOP* State (%)	51.93	53.85	0.20
	(22.90)	(23.68)	
Posterior Belief of *TOP* State (%)	51.78	52.27	0.78
	(26.07)	(27.00)	
Male	0.44	0.40	0.25
	(0.50)	(0.49)	
White	0.68	0.72	0.22
	(0.47)	(0.45)	
Age ≤ 35	0.49	0.54	0.14
	(0.50)	(0.50)	
College Degree	0.51	0.54	0.29
	(0.50)	(0.50)	
Observations	444	454	898

Notes: This table provides the means and standard deviations of the demographic and individual-level experimental variables, for the 444 participants who see two identical signals in the *Main* treatment and the 454 participants who see two identical signals in the *Control* treatment. Entries in the last column are two-sided p-values from unpaired t-tests of equal means for the *Main* treatment and the *Control* treatment. *, **, and *** indicate statistical significance at the 0.10, 0.05, and 0.01 levels, respectively.

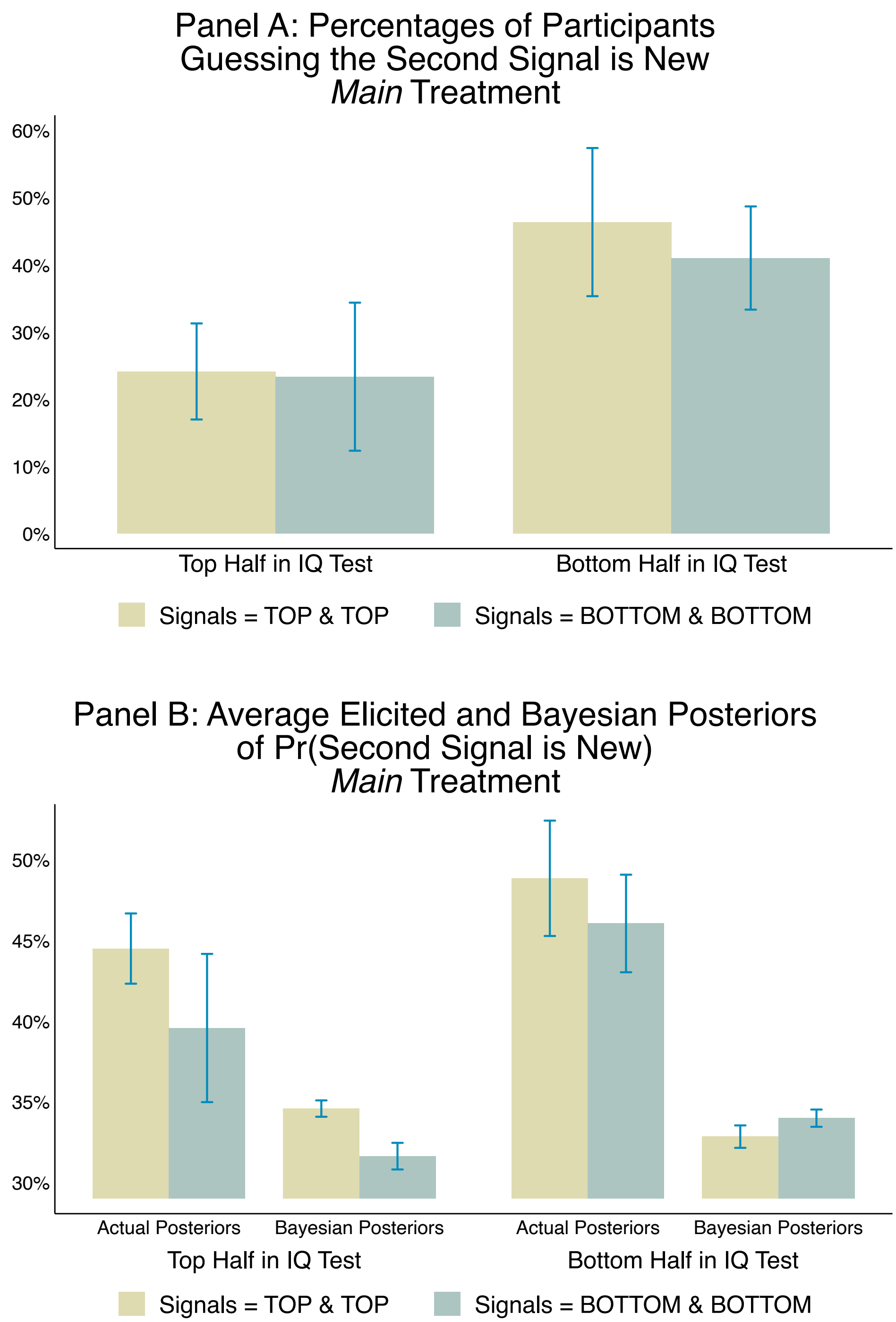


Figure 3: Beliefs about whether the second signal is new by IQ test performance group and signal valence in the *Main* Treatment. Panel A shows the percentages of participants who make the *wrong* guess (i.e., guess that the second signal is new) after seeing two identical signals, separately by IQ test performance group and signal valence. Panel B shows both the average elicited posteriors and the average Bayesian posteriors that the second signal is new after seeing two identical signals, separately by IQ test performance group and signal valence.

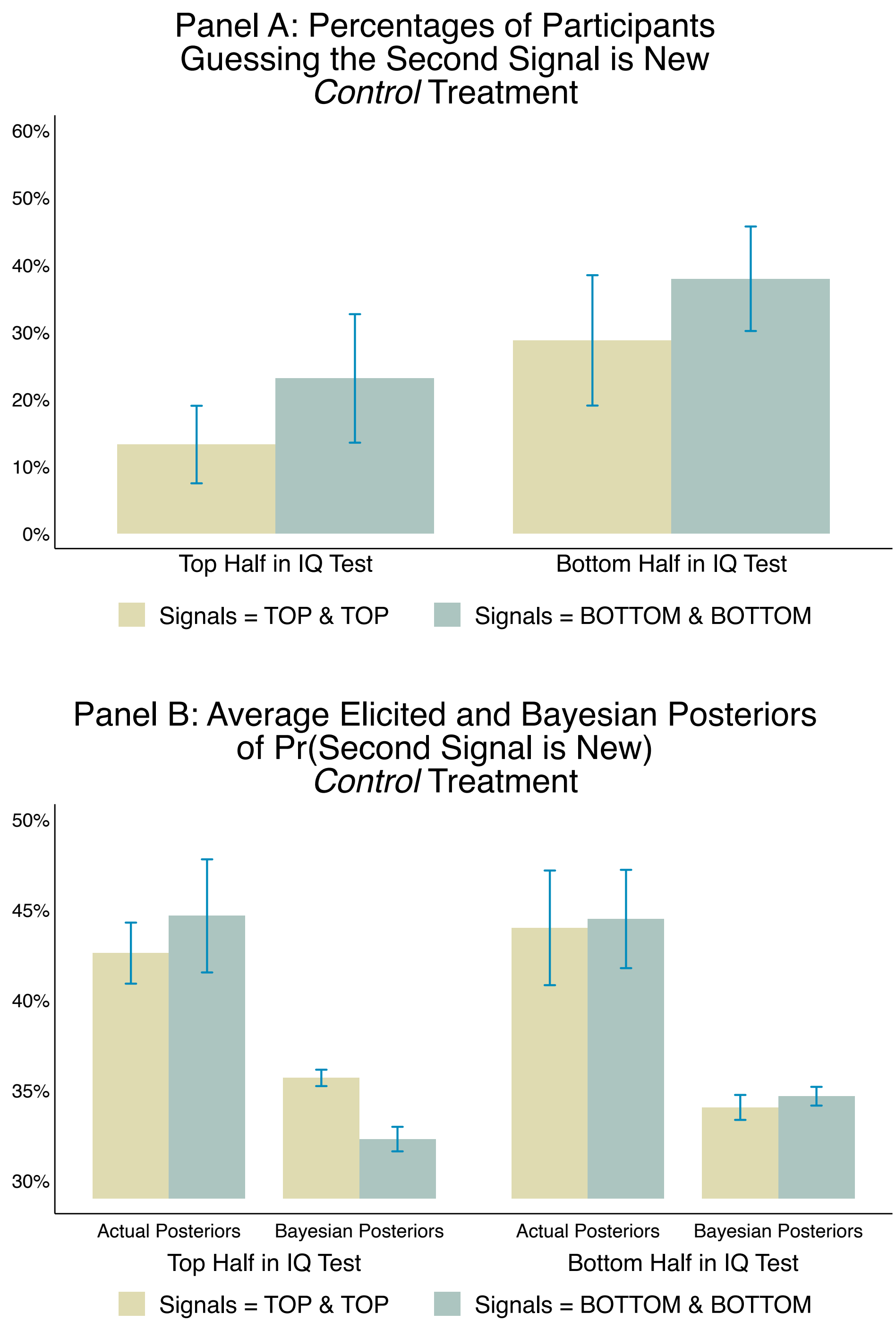


Figure 4: Beliefs about whether the second signal is new by IQ test performance group and signal valence in the *Control* Treatment. Panel A shows the percentages of participants who make the *wrong* guess (i.e., guess that the second signal is new) after seeing two identical signals, separately by IQ test performance group and signal valence. Panel B shows both the average elicited posteriors and the average Bayesian posteriors that the second signal is new after seeing two identical signals, separately by IQ test performance group and signal valence.

Table 2: Inference about signal redundancy after seeing two good signal or two bad signals, the *Main* treatment.

	Dep Var: Posterior of Pr(Second Signal is New) (%)			
	Full Sample		Drop Wrong Direction	
	(1)	(2)	(3)	(4)
Two Good Signals	3.625**	3.949**	4.637***	5.141***
	(1.722)	(1.643)	(1.414)	(1.356)
Top Half in Test	-5.303***	2.507	-2.055	-0.839
	(1.694)	(2.479)	(1.424)	(2.007)
Bayesian Posterior (in %)	0.157	0.033	-0.030	-0.063
	(0.284)	(0.281)	(0.223)	(0.211)
Controls	No	Yes	No	Yes
Observations	444	444	369	369
R^2	0.025	0.090	0.029	0.100

Notes: Robust standard errors are in parentheses. *, **, and *** indicate statistical significance at the 0.10, 0.05, and 0.01 levels, respectively. The sample in columns (1) and (2) includes the 444 participants who see two identical signals in the *Main* treatment. In columns (3) and (4), we drop participants who report a higher than 50% posterior that the second signal is new after seeing two identical signals, which indicates that they update in the wrong direction about signal redundancy. "Controls" include: dummy variables for Male, White, Age $\leq$ 35, and College Degree; the raw score in the IQ test; the subjective importance of the IQ test (0 – 10); and the prior belief of the *TOP* state.

Table 3: Inference about own IQ test performance after seeing two good signal or two bad signals, the *Main* treatment.

	Dep Var: Posterior logit(Pr(Top Half))
	(1)
δ	0.798^{***}
	(0.051)
α_G	0.641^{***}
	(0.052)
α_B	0.621^{***}
	(0.061)
Observations	420
R^2	0.696
p-value: $\alpha_G = \alpha_B$	0.802
p-value: $\alpha_G = 1$	<0.001
p-value: $\alpha_B = 1$	<0.001

Notes: Robust standard errors are in parentheses. We restrict the sample to the 420 participants in *Main* who receive two identical signals and whose prior and posterior beliefs are not 0% or 100%. The outcome in the regression is the log likelihood ratio of the posterior belief. δ is the coefficient on the log likelihood ratio of the prior belief; α_G and α_B are the estimated effects of the log likelihood ratio for two good signals and two bad signals, respectively. Bayesian updating corresponds to $\delta = \alpha_G = \alpha_B = 1$.

Table 4: Inference about signal redundancy, pooling data from the *Main* treatment and the *Control* treatment.

	Dep Var: Pr(Second Signal is New), Actual Posterior (in %)			
	Full Sample		Drop Wrong Direction	
	(1)	(2)	(3)	(4)
Main × Two Good Signals	3.749**	4.168**	4.680***	4.814***
	(1.727)	(1.660)	(1.428)	(1.391)
Control × Two Good Signals	-0.774	-0.003	1.387	1.250
	(1.519)	(1.601)	(1.210)	(1.286)
Main × Top Half in Test	-5.329***	-0.687	-2.063	-1.978
	(1.700)	(2.024)	(1.427)	(1.674)
Control × Top Half in Test	-1.582	1.802	0.096	-0.298
	(1.477)	(1.725)	(1.184)	(1.441)
Main Treatment	0.698	-1.944	-1.375	1.114
	(1.817)	(7.358)	(1.518)	(6.908)
Bayesian Posterior (in %)	-0.036	-0.094	-0.071	-0.002
	(0.190)	(0.206)	(0.151)	(0.160)
Controls	No	Yes	No	Yes
Observations	898	898	771	771
R^2	0.018	0.092	0.018	0.115
p-value (Equal Effects)	0.049	0.071	0.081	0.061

Notes: Robust standard errors are in parentheses. *, **, and *** indicate statistical significance at the 0.10, 0.05, and 0.01 levels, respectively. The sample in columns (1) and (2) includes the 444 participants who see two identical signals in the *Main* treatment and the 454 participants who see two identical signals in the *Control* treatment. In columns (3) and (4), we drop participants who report a higher than 50% posterior that the second signal is new after seeing two identical signals, which indicates that they update in the wrong direction about signal redundancy. "Controls" include: a dummy for each possible level of the prior probability of the *TOP* state, interacted with the treatment; dummies for Male, White, Age $\leq$ 35, and College Degree; the raw score in our IQ test; and the subjective importance of the IQ test (0 – 10).

A Online Appendix

A.1 Additional Figures and Tables

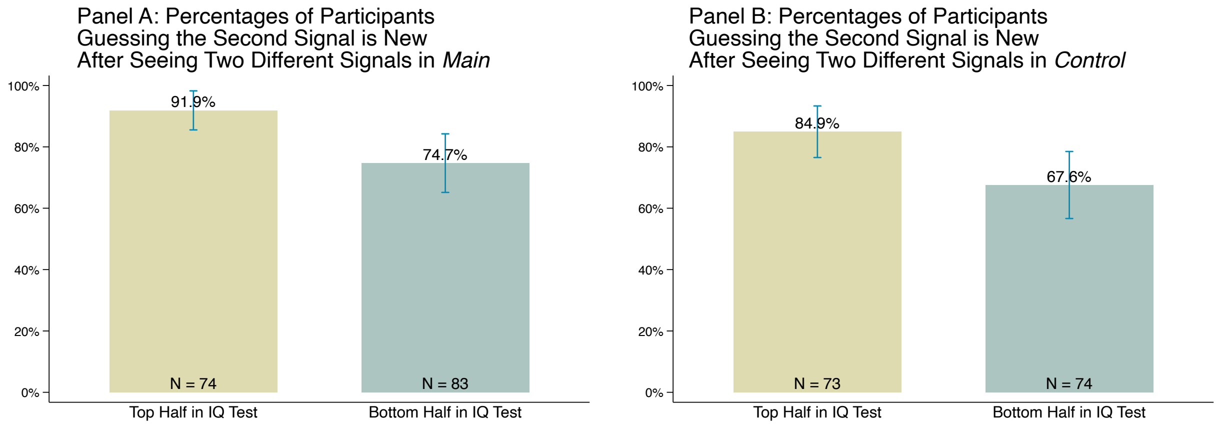
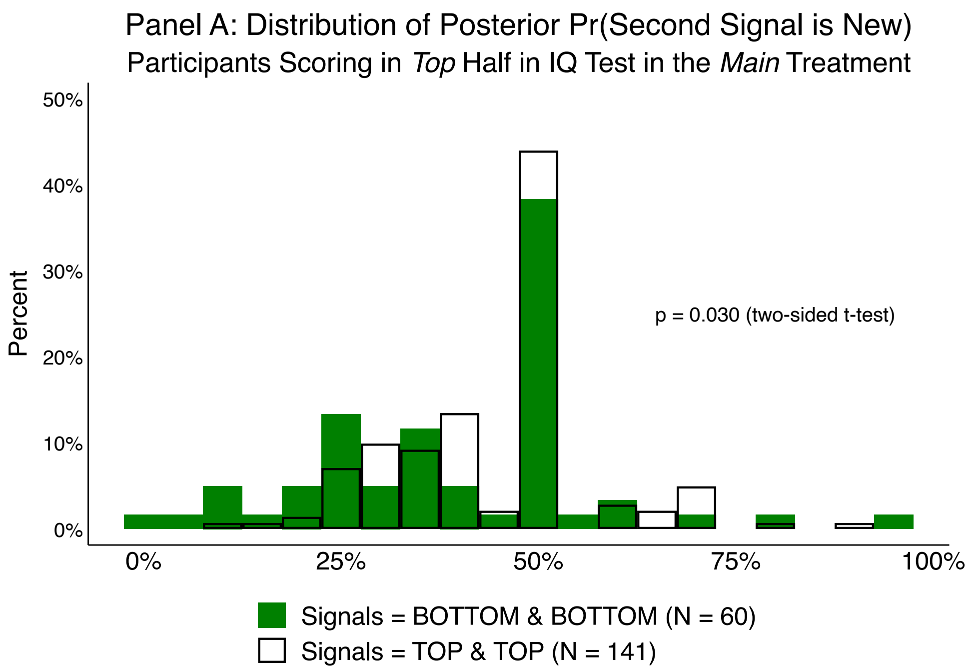


Figure A1: Percentages of participants guessing the second signal is new after seeing two different signals in *Main* and *Control*.



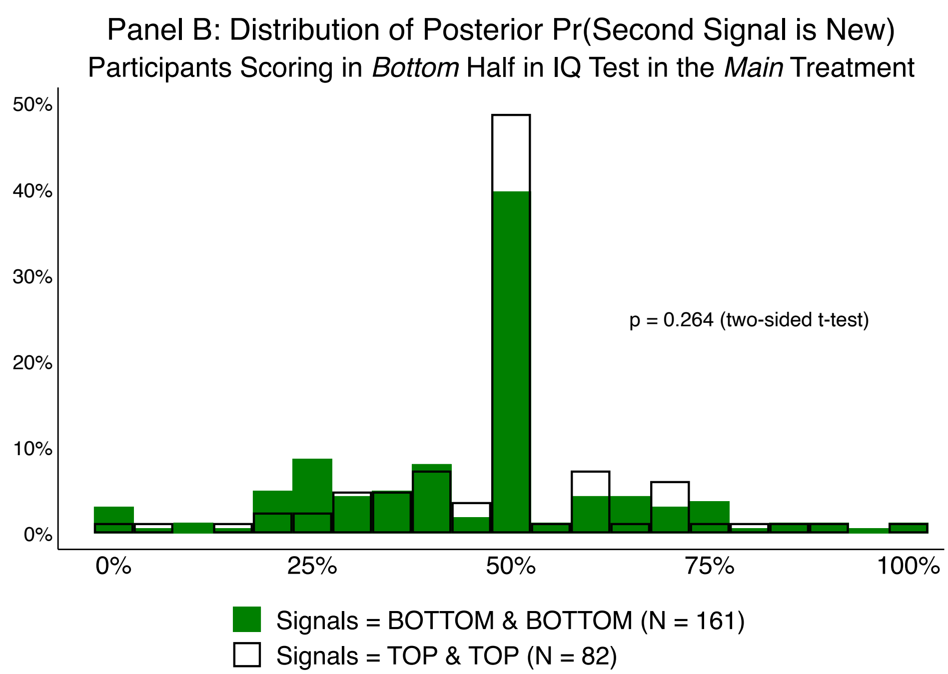


Figure A2: Full distribution of elicited posteriors that the second signal is new after seeing two identical signals in the *Main* Treatment, by IQ test performance group and signal valence. Panel A shows the distribution for participants scoring in the *top* half in the IQ test, separately for those seeing two *TOP* signals and those seeing two *BOTTOM* signals. Panel B shows the distribution for participants scoring in the *bottom* half in the IQ test, separately for those seeing two *TOP* signals and those seeing two *BOTTOM* signals. P-values from two-sided t-tests are also provided.

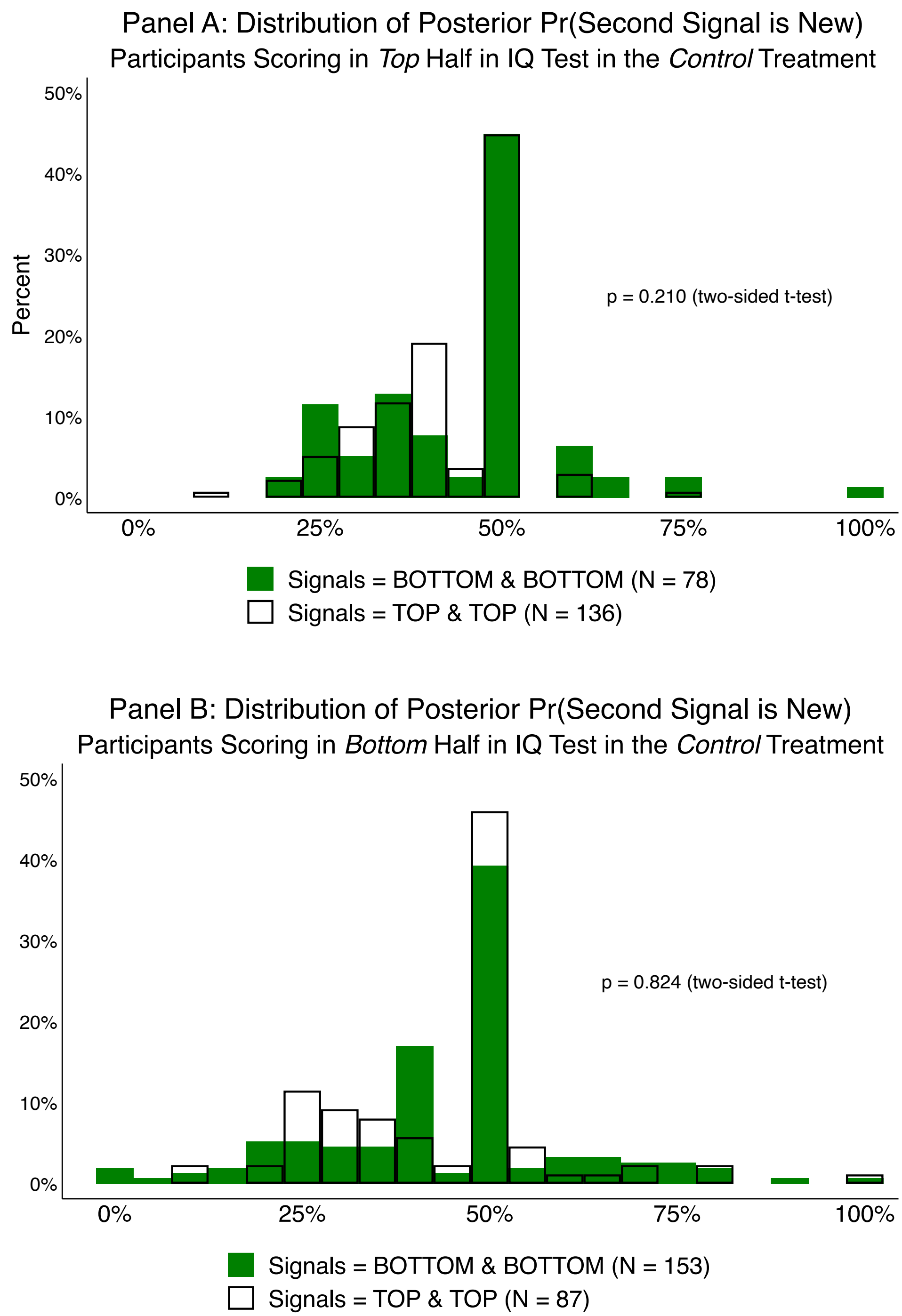


Figure A3: Full distribution of elicited posteriors that the second signal is new after seeing two identical signals in the *Control* Treatment, by IQ test performance group and signal valence. Panel A shows the distribution for participants scoring in the *top* half in the IQ test, separately for those seeing two *TOP* signals and those seeing two *BOTTOM* signals. Panel B shows the distribution for participants scoring in the *bottom* half in the IQ test, separately for those seeing two *TOP* signals and those seeing two *BOTTOM* signals. P-values from two-sided t-tests are also provided (note that the difference here goes in the opposite direction of the prediction of the *Motivated Mislearning Hypothesis*).

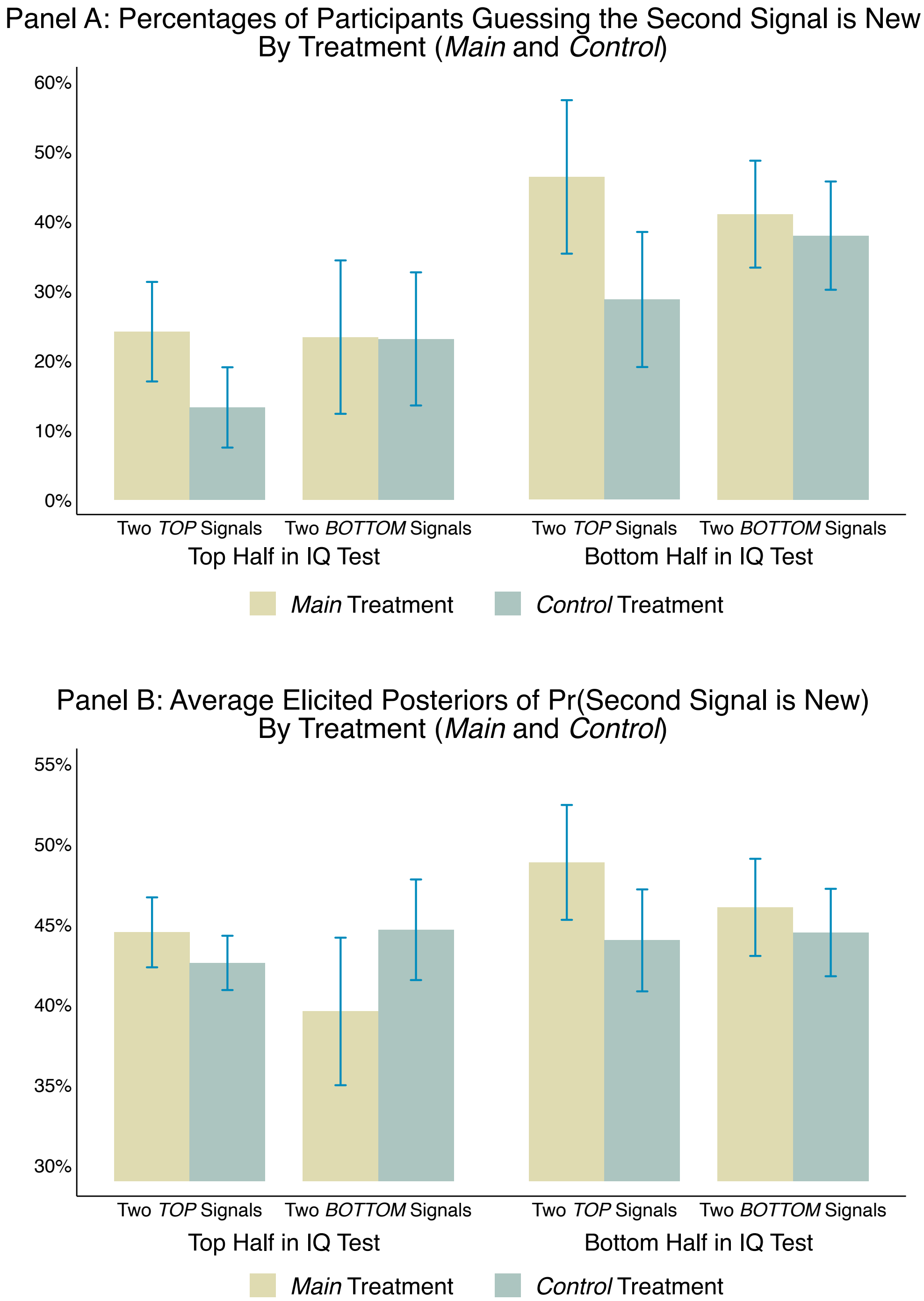


Figure A4: Beliefs about whether the second signal is new by IQ test performance group, signal valence, and treatment, pooling data from both the *Main* treatment and the *Control* treatment. Panel A shows the percentages of participants who make the *wrong* guess (i.e., guess that the second signal is new) after seeing two identical signals, separately by IQ test performance group, signal valence, and treatment. Panel B shows both the average elicited posteriors that the second signal is new after seeing two identical signals, separately by IQ test performance group, signal valence, and treatment.

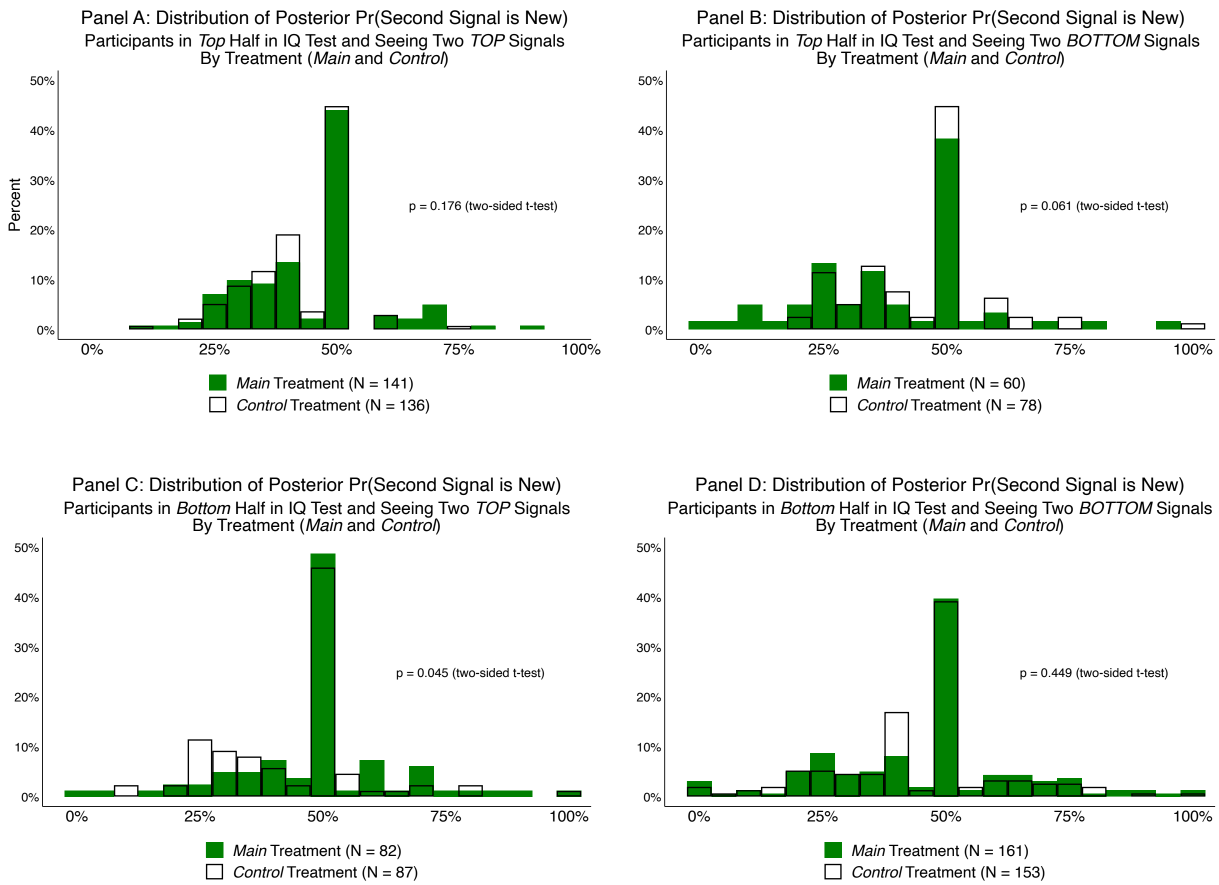


Figure A5: Full distribution of elicited posteriors that the second signal is new after seeing two identical signals by IQ test performance group, signal valence, and treatment, pooling data from both the *Main* treatment and the *Control* treatment. Panel A shows the distribution for participants scoring in the *top* half in the IQ test and seeing two *TOP* signals, separately for those in *Main* and those in *Control*. Panel B shows the distribution for participants scoring in the *top* half in the IQ test and seeing two *BOTTOM* signals, separately for those in *Main* and those in *Control*. Panel C shows the distribution for participants scoring in the *bottom* half in the IQ test and seeing two *TOP* signals, separately for those in *Main* and those in *Control*. Panel D shows the distribution for participants scoring in the *bottom* half in the IQ test and seeing two *BOTTOM* signals, separately for those in *Main* and those in *Control*.

Table A1: Inference about signal redundancy after seeing two good signals or two bad signals, separately by gender, the *Main* treatment.

	Dep Var: Posterior of Pr(Second Signal is New) (%)			
	Male		Female	
	(1)	(2)	(3)	(4)
Two Good Signals	1.923	3.047	5.267**	4.676**
	(3.003)	(3.003)	(2.063)	(1.988)
Top Half in Test	-5.115*	3.722	-5.067**	2.059
	(2.915)	(3.909)	(2.029)	(3.318)
Bayesian Posterior (in %)	-0.214	-0.357	0.513	0.343
	(0.499)	(0.494)	(0.318)	(0.303)
Controls	No	Yes	No	Yes
Observations	195	195	249	249
R^2	0.019	0.077	0.050	0.119

Notes: Robust standard errors are in parentheses. *, **, and *** indicate statistical significance at the 0.10, 0.05, and 0.01 levels, respectively. The sample in columns (1) and (2) includes the 195 male participants who see two identical signals in the *Main* treatment. The sample in columns (3) and (4) includes the 249 female participants who see two identical signals in the *Main* treatment. "Controls" include dummy variables for White, Age $\leq$ 35, College Degree, the raw score in the IQ test, the subjective importance of the IQ test (0 – 10), and the prior belief of the *TOP* state.

Table A2: Inference about signal redundancy in the *Main* treatment, alternative specification.

	Dep Var: Actual Posterior − Bayesian Posterior (in %)			
	Full Sample		Drop Wrong Direction	
	(1)	(2)	(3)	(4)
Two Good Signals	3.083*	3.250*	3.566**	3.974***
	(1.731)	(1.659)	(1.484)	(1.433)
Top Half in Test	-5.191***	1.813	-1.852	-1.640
	(1.700)	(2.541)	(1.472)	(2.098)
Controls	No	Yes	No	Yes
Observations	444	444	369	369
R^2	0.021	0.077	0.017	0.084

Notes: Robust standard errors are in parentheses. *, **, and *** indicate statistical significance at the 0.10, 0.05, and 0.01 levels, respectively. The sample in columns (1) and (2) includes the 444 participants who see two identical signals in the *Main* treatment. In columns (3) and (4), we drop participants who report a higher than 50% posterior that the second signal is new after seeing two identical signals, which indicates that they update in the wrong direction about signal redundancy. "Controls" include: dummy variables for Male, White, Age $\leq$ 35, and College Degree; the raw score in the IQ test; the subjective importance of the IQ test (0 – 10); and the prior belief of the *TOP* state.

Table A3: Detailed two-way tabulation of beliefs and guesses in the *Main* treatment.

Group	Posterior Belief Pr(Second Signal is New)	Number of Participants	Binary Guess Second = First	Binary Guess Second $\neq$ First
Top Half in IQ Test, Two Good Signals (N = 141)	Belief < 50%	63 (44.7%)	61 (43.3%)	2 (1.4%)
	Belief = 50%	62 (44.0%)	46 (32.6%)	16 (11.3%)
	Belief > 50%	16 (11.3%)	0 (0%)	16 (11.3%)
Top Half in IQ Test, Two Bad Signals (N = 60)	Belief < 50%	32 (53.3%)	30 (50%)	2 (3.3%)
	Belief = 50%	22 (36.7%)	16 (26.7%)	6 (10%)
	Belief > 50%	6 (10%)	0 (0%)	6 (10%)
Bottom Half in IQ Test, Two Good Signals (N = 82)	Belief < 50%	24 (29.3%)	20 (24.4%)	4 (4.9%)
	Belief = 50%	40 (48.8%)	18 (22.0%)	22 (26.8%)
	Belief > 50%	18 (22.0%)	6 (7.3%)	12 (14.6%)
Bottom Half in IQ Test, Two Bad Signals (N = 161)	Belief < 50%	63 (39.1%)	52 (32.3%)	11 (6.8%)
	Belief = 50%	63 (39.1%)	35 (21.7%)	28 (17.4%)
	Belief > 50%	35 (21.7%)	8 (5.0%)	27 (16.8%)

Notes: In this table, we first divide all participants in the *Main* treatment who see two identical signals into four groups depending on their performance in the IQ test (top half vs. bottom half) and whether their signals are good (ego-favorable). Within each group, we divide participants into subgroups according to whether their posterior belief that the second signal is new is lower than, equal to, or higher than 50%. In the third column, we provide the numbers and percentages of participants who fall into each subgroup. In the fourth and fifth columns, we provide the numbers of participants who provide each possible guess within the corresponding subgroup; the percentages are still calculated relative to the four broad groups.

Table A4: Detailed two-way tabulation of beliefs and guesses in the *Control* treatment.

Group	Posterior Belief Pr(Second Signal is New)	Number of Participants	Binary Guess Second = First	Binary Guess Second ≠ First
Top Half in IQ Test, Two *TOP* Signals (N = 136)	Belief < 50%	70 (51.4%)	70 (51.4%)	0 (0%)
	Belief = 50%	61 (44.9%)	47 (34.6%)	14 (10.3%)
	Belief > 50%	5 (3.7%)	1 (0.7%)	4 (2.9%)
Top Half in IQ Test, Two *BOTTOM* Signals (N = 78)	Belief < 50%	33 (42.3%)	31 (39.7%)	2 (2.6%)
	Belief = 50%	35 (44.9%)	27 (34.6%)	8 (10.3%)
	Belief > 50%	10 (12.8%)	2 (2.6%)	8 (10.3%)
Bottom Half in IQ Test, Two *TOP* Signals (N = 87)	Belief < 50%	36 (41.4%)	34 (39.1%)	2 (2.3%)
	Belief = 50%	40 (46.0%)	27 (31.0%)	13 (14.9%)
	Belief > 50%	11 (12.6%)	1 (1.1%)	10 (11.5%)
Bottom Half in IQ Test, Two *BOTTOM* Signals (N = 153)	Belief < 50%	68 (44.4%)	61 (39.9%)	7 (4.6%)
	Belief = 50%	59 (38.6%)	30 (19.6%)	29 (19.0%)
	Belief > 50%	26 (17.0%)	4 (2.6%)	22 (14.4%)

Notes: In this table, we first divide all participants in the *Control* treatment who see two identical signals into four groups depending on their performance in the IQ test (top half vs. bottom half) and whether their signals are *TOP* or *BOTTOM*. Within each group, we divide participants into subgroups according to whether their posterior belief that the second signal is new is lower than, equal to, or higher than 50%. In the third column, we provide the numbers and percentages of participants who fall into each subgroup. In the fourth and fifth columns, we provide the numbers of participants who provide each possible guess within the corresponding subgroup; the percentages are still calculated relative to the four broad groups.

Table A5: Inference about signal redundancy after seeing two good signal or two bad signals, the *Control* treatment.

	Dep Var: Posterior of Pr(Second Signal is New) (%)			
	Full Sample		Drop Wrong Direction	
	(1)	(2)	(3)	(4)
Two Good Signals	-0.485	0.436	1.458	1.836
	(1.549)	(1.577)	(1.228)	(1.263)
Top Half in Test	-1.692	0.414	0.072	-1.093
	(1.478)	(1.735)	(1.185)	(1.460)
Bayesian Posterior (in %)	-0.246	-0.267	-0.116	-0.124
	(0.248)	(0.238)	(0.201)	(0.196)
Controls	No	Yes	No	Yes
Observations	454	454	402	402
R^2	0.008	0.036	0.005	0.045

Notes: Robust standard errors are in parentheses. *, **, and *** indicate statistical significance at the 0.10, 0.05, and 0.01 levels, respectively. The sample in columns (1) and (2) includes the 454 participants who see two identical signals in the *Control* treatment. In columns (3) and (4), we drop participants who report a higher than 50% posterior that the second signal is new after seeing two identical signals, which indicates that they update in the wrong direction about signal redundancy. "Controls" include: dummy variables for Male, White, Age $\leq$ 35, and College Degree; the raw score in the IQ test; the subjective importance of the IQ test (0 – 10); and the prior belief of scoring in the top half in the IQ test.

Table A6: Inference about signal redundancy in the *Control* treatment, alternative specification.

	Dep Var: Actual Posterior − Bayesian Posterior (in %)			
	Full Sample		Drop Wrong Direction	
	(1)	(2)	(3)	(4)
Two Good Signals	-2.196	-1.487	-0.322	-0.123
	(1.541)	(1.612)	(1.239)	(1.279)
Top Half in Test	-1.039	0.592	0.664	-0.938
	(1.536)	(1.832)	(1.238)	(1.527)
Controls	No	Yes	No	Yes
Observations	454	454	402	402
R^2	0.009	0.034	0.001	0.038

Notes: Robust standard errors are in parentheses. *, **, and *** indicate statistical significance at the 0.10, 0.05, and 0.01 levels, respectively. The sample in columns (1) and (2) includes the 454 participants who see two identical signals in the *Control* treatment. In columns (3) and (4), we drop participants who report a higher than 50% posterior that the second signal is new after seeing two identical signals, which indicates that they update in the wrong direction about signal redundancy. "Controls" include: dummy variables for Male, White, Age $\leq$ 35, and College Degree; the raw score in the IQ test; the subjective importance of the IQ test (0 – 10); and the prior belief of scoring in the top half in the IQ test.

Table A7: Inference about signal redundancy in *Main* and *Control*, alternative specification.

	Dep Var: Actual Posterior − Bayesian Posterior (in %)			
	Full Sample		Drop Wrong Direction	
	(1)	(2)	(3)	(4)
Main × Two Good Signals	3.328*	3.372**	3.485**	3.723**
	(1.699)	(1.691)	(1.463)	(1.463)
Control × Two Good Signals	-1.523	-1.429	-0.104	-0.401
	(1.671)	(1.671)	(1.335)	(1.334)
Main × Top Half in Test	-3.361*	-0.850	0.349	-2.222
	(1.856)	(2.050)	(1.520)	(1.744)
Control × Top Half in Test	-0.052	2.026	1.359	-0.257
	(1.595)	(1.796)	(1.345)	(1.485)
Main Treatment	5.282	4.141	4.017	5.909
	(7.861)	(7.898)	(7.908)	(7.819)
Controls	Yes	Yes	Yes	Yes
Observations	898	898	771	771
R^2	0.080	0.093	0.098	0.121
P(Same Effect)	0.042	0.043	0.070	0.036

Notes: Robust standard errors are in parentheses. *, **, and *** indicate statistical significance at the 0.10, 0.05, and 0.01 levels, respectively. The sample in columns (1) and (2) includes the 444 participants who see two identical signals in the *Main* treatment and the 454 participants who see two identical signals in the *Control* treatment. In columns (3) and (4), we drop participants who report a higher than 50% posterior that the second signal is new after seeing two identical signals, which indicates that they update in the wrong direction about signal redundancy. "Controls" include: a dummy for each possible level of the prior probability of the *TOP* state, interacted with the treatment; dummies for Male, White, Age $\leq$ 35, and College Degree; the raw score in our IQ test; and the subjective importance of the IQ test (0 – 10).