

WHAT YOU SEE IS ALL THERE IS^{*}

Benjamin Enke

February 14, 2020

Abstract

News reports and communication are inherently constrained by space, time, and attention. As a result, news sources often condition the decision of whether to share a piece of information on the similarity between the signal and the prior belief of the audience, which generates a sample selection problem. This paper experimentally studies how people form beliefs in these contexts, in particular the mechanisms behind errors in statistical reasoning. I document that a substantial fraction of experimental participants follows a simple “what you see is all there is” heuristic, according to which they exclusively take into account information that is right in front of them, and directly use the sample mean to estimate the population mean. A series of treatments aimed at identifying mechanisms suggests that for many participants unobserved signals do not even come to mind. I provide causal evidence that the frequency of such incorrect mental models is a function of the computational complexity of the decision problem. These results point to the context-dependence of what comes to mind and the resulting errors in belief updating.

JEL classification: D03; D80; D84.

Keywords: Bounded rationality; mental models; complexity; beliefs.

^{*}I thank Andrei Shleifer and three extraordinarily constructive and helpful referees for comments that substantially improved the paper. For discussions and comments I am also grateful to Doug Bernheim, Thomas Dohmen, Armin Falk, Thomas Graeber, Shengwu Li, Josh Schwartzstein, Lukas Wenner, Florian Zimmermann, and seminar participants at Cornell, Harvard, Munich, Princeton, Wharton, BEAM 2017, SITE Experimental Economics 2017, and the 2016 ECBE conference. Financial support through the Bonn Graduate School of Economics and the Center for Economics and Neuroscience Bonn is gratefully acknowledged. Enke: Harvard University, Department of Economics, and NBER; enke@fas.harvard.edu.

1 Introduction

News reports and communication are both inherently constrained by space, time, and attention. As a result, news sources often condition the decision of whether to share a piece of information on the similarity between the signal and the prior belief of the audience. In some cases, news reports and communication disproportionately focus on events that are likely to *move* the audience’s priors, such as the occurrence of terror attacks, large movements in stock prices, or surprising research findings. While these types of events are routinely covered, the corresponding non-events are not: one rarely reads headlines such as “No terror attack in Afghanistan today.” In other cases, news providers supply news that *align* with people’s priors but omit those that do not. For example, social networks like Facebook exclude stories from newsfeeds that go against users’ previously articulated views. Regardless of the specific direction of the sample selection problem, all of these contexts share the feature that whether a specific signal gets transmitted depends on how this signal compares to the audience’s prior. In the presence of such selection problems, people need to draw inferences from (colloquially speaking) “unobserved” signals.

While an active theoretical literature has linked selection problems in belief updating to various economic applications,¹ empirical work on people’s reasoning in such contexts is more limited. Moreover, if people actually do fail to take into account unobserved information, a perhaps even more fundamental open question concerns the mechanisms behind such a bias. As reflected by a recent comprehensive survey paper on errors in statistical reasoning (Benjamin, 2018), researchers have accumulated a broad set of reduced-form judgmental biases. Yet despite early calls for empirical work on the micro-foundations of biases (Fudenberg, 2006), relatively little is known about the mechanisms that underlie judgment errors. In the present context, a promising candidate mechanism is the idea that agents maintain an incorrect mental model of the environment because selection does not even *come to mind* when a decision is taken: people may never even ask themselves what it is that is not directly in front of them.

This paper tackles these two sets of issues – how people process selected information and the role of mental models therein – by developing a tightly structured individual decision-making experiment that operationalizes the selection problems discussed in the opening paragraph. In the experiment, the entire information-generating process is computerized and known to participants. Subjects estimate an unknown state of the world and are paid for accuracy. The true state is generated as an average of six i.i.d. random draws from the simple discretized uniform distribution $\{50, 70, 90, 110, 130, 150\}$. I will refer to these random draws as signals. Participants observe one of these six sig-

¹See Levy and Razin (2015), Han and Hirshleifer (2015), Jehiel (2018), and Jackson (2016).

nals at random and subsequently indicate whether they believe the true state to be above or below 100. Thereafter, participants observe additional signals by interacting with a computerized information source. Just like in the motivating examples, this information source transparently conditions its behavior on the participant's first stated belief. On a participant's computer screen, the information source shares all signals that "align" with the participant's first stated belief (e.g., are smaller than 100 if the first belief is below 100) but not all signals that "contradict" the first belief (e.g., are larger than 100 if the first belief is below 100). Afterwards, participants guess the state.

Bayesian inference would require participants to draw an inference about signals that do not appear on their computer screens, just like readers should infer something from the fact that a news outlet does not report on the stock market on a given day. In what follows, I will colloquially say that participants "do not see" these latter signals, even though in an information-theoretic sense, this constitutes coarse information.

In a between-subjects design, I compare beliefs in this *Selected* treatment with those in a *Control* condition in which subjects receive the same objective information as those in *Selected* except that all signals physically appear on subjects' screens. Comparing beliefs across the two treatments allows us to causally identify participants' tendency to neglect selection problems in processing information, holding fixed the objective informational content of the signals.

The results document that beliefs exhibit large and statistically significant differences across the two treatments. Whenever participants' first signal is above 100, their final stated beliefs tend to be upward biased and conversely for initial signals below 100. I show that this pattern is robust against the provision of some feedback.

To disaggregate these cross-treatment differences, I analyze individual decision rules. Participants' responses are often heuristic in nature and reflect significant rounding to multiples of five or ten. Yet while individual decisions are noisy, these heuristics appear to have a systematic component. To identify this systematic part, the analysis estimates an individual-level parameter that pins down updating rules in relation to Bayesian rationality. Here, the distribution of updating types follows a bimodal structure: the modal responses of 60% of all participants are either Bayesian or reflect full neglect. In fact, even 87% of those participants that exhibit stable identifiable decision types can be characterized as exactly rational or exactly full neglect. Thus, a significant fraction of participants states beliefs whose stable component corresponds to fully ignoring what they do not see and averaging the visible data.

Economists are increasingly interested in the mechanisms behind reduced-form errors in statistical reasoning, probably due to the view that this may help develop appropriate debiasing strategies or inform theoretical work. In the present context, the

patterns are prima facie consistent with two alternative accounts of the data. A first is that – as posited in much recent theoretical work discussed below – neglect reflects an incorrect mental model of the data-generating process that arises because certain aspects of the problem do not even come to mind. Here, people may never even ask themselves which signals they do not see and why. Relatedly, a recent literature in cognitive psychology on the metaphor of the “naïve intuitive statistician” argues that people are reasonably skilled statisticians but often naïvely assume that their information samples are unbiased and that sample properties can be directly used to estimate population analogs (Fiedler and Juslin, 2006; Juslin et al., 2007). According to this “incorrect mental models” perspective, the probability that selection comes to mind may be a function of the computational complexity of the decision problem. Decision-makers need to allocate scarce cognitive resources between (i) setting up a mental model and (ii) computational implementation. Thus, people should be less likely to develop a correct mental model if they are cognitively busier with (or distracted by) computationally implementing a given solution strategy.

A plausible second view of the mechanisms behind neglect, however, is that people are aware of the unobserved signals but struggle with the conceptual or computational difficulty of correcting for selection. To investigate the relative importance of these two accounts, I implement three sets of follow-up treatments. Each of these treatment variations predicts a change in behavior under only one of these accounts.

First, I design a treatment in which the presence of a selection problem is eliminated, yet subjects still need to process unobserved signals. If neglect was largely about the conceptual or computational difficulty of correcting for selection, then neglect should disappear in this treatment. Operationally, subjects observe four *randomly selected* signals, while four additional signals are not directly communicated to them. As in the baseline condition, participants do have information about the unobserved signals, which in this case is their unconditional expectation. Nonetheless, a considerable fraction of subjects again follows a “what you see is all there is” heuristic of averaging the visible data. This shows that people struggle not (only) with conceptually thinking through a potential selection problem. Instead, they appear to have a more general tendency to estimate population means through sample means, where the “sample” is given by what is right in front of them and hence top of mind.

Second, I devise treatments that hold the conceptual difficulty of accounting for selection constant but vary the cognitive resources that participants have at their disposal to set up a correct mental model. To this effect, I vary the computational complexity of computing beliefs in such a way that it plausibly affects only the probability that the unobserved signals come to mind. The experiments operationalize complexity in two dif-

ferent ways: the complexity of the signal space and the number of signals. First, to vary the complexity of the signal space, I implement treatments *Complex* and *Simple*. In *Simple*, the signal space is given by {70, 70, 70, 70, 70, 70, 130, 130, 130, 130, 130, 130}. In *Complex*, it is {70, 70, 70, 70, 70, 70, 104, 114, 128, 136, 148, 150}. In both treatments, whenever a participant states a first belief above 100, the selection problem can be overcome by remembering that an unobserved signal must be a 70. Thus, these treatments leave the conceptual and computational difficulty of accounting for selection constant (if the first belief is above 100). At the same time, these treatments vary the computational difficulty of computing a posterior belief and problem-induced cognitive load. Second, to manipulate the number of signals, participants in condition *Few* were confronted with the same signal space as those in *Complex*, but the true state was generated as the average of only two, rather than six, random draws. Because all of these treatments fix the difficulty of backing out an unobserved signal, complexity can only matter to the extent that it induces cognitive load and reduces the probability that the unobserved signals come to mind.

The results show that increases in complexity (in terms of both the number of signals and the complexity of the signal space) lead to substantially more neglect than in the respective comparison treatments. This is even though participants in the more complex treatments work longer on the problems. The fact that variations in complexity matter for neglect even though the difficulty of accounting for selection is unchanged again highlights the role of (endogenous) incorrect mental models.

As a third test between the two alternative mechanisms explained above, I implement an experimental condition that includes a simple nudge on participants' decision screen to pay attention to, or remember, those signals that they "do not see." This intervention decreases neglect by about 50%, which again suggests that the unobserved signals did otherwise not come to subjects' minds in the first place.

In summary, the takeaways from the analysis of mechanisms are twofold. First, incorrect mental models play an important role in generating neglect. Unobserved signals do not seem to come to mind in the first place, which leads people to directly use the sample mean to estimate the population mean. Second, what comes to mind and the resulting mental models are not exogenously given "neglect parameters" – instead, they are context-dependent and endogenous to the computational complexity of the environment. These insights are potentially relevant not only for modeling updating errors but also for policy in terms of what will be an effective method to correct biased beliefs.

The paper proceeds as follows. Section 2 describes the experimental design. Sections 3–5 present the results and study mechanisms. Section 6 discusses related literature and offers concluding thoughts.

2 Experimental Design

2.1 Setup

The experiment was designed to achieve the following objectives: (i) full control over the data-generating process, (ii) exogenous manipulation of the degree of selection, (iii) a control condition that serves as a benchmark for updating without selected information, and (iv) incentive-compatible belief elicitation. Most importantly, a clean identification requires subjects' full knowledge of the data-generating process.

The main idea behind the design is to construct two sets of signals (two treatments) which result in the same Bayesian posterior, but only one information structure features a problem of selection. Subjects were asked to estimate an ex-ante unknown state of the world θ and were paid for accuracy. The computer generated θ by drawing six times, with replacement, from the set $X = \{50, 70, 90, 110, 130, 150\}$. Draws from X are uniform. The average of these six draws then constituted the true state θ , which in the experiment is referred to as the “variable” that subjects needed to estimate. Henceforth, I will refer to the random draws as signals.

In the course of the experiment, a subject interacted with a computerized information source that showed the subject (subsets of) the signals. An experimental task consisted of multiple stages, as summarized in Table 1. First, after the computer generated the true state, a subject observed one randomly selected signal. Second, based on this first signal, subjects provided an incentivized guess b_1 about whether they believed θ to be smaller or larger than 100, $b_1 \in \{low, high\}$.²

Third, the information source showed the subject additional signals. This is the only stage in which treatments *Selected* and *Control* differed, as detailed below. Finally, after subjects observed the messages of the information source, they stated an incentivized belief about the state $b_2 \in [50, 150]$, with at most two decimals.

In *Selected*, the information source faced a budget constraint and hence conditioned its decision of which out of the remaining five signals to show the subject on the subject's first guess. Specifically, if the subject's first guess was higher than 100, the information source showed the subject all signals above 100, but at least three signals. Likewise, if the subject's first guess was smaller than 100, the information source showed the subject all signals below 100, but at least three signals. For example, if a participant's first guess was above 100 and only two of the remaining five signals were above 100, the information source showed the subject these two signals and one randomly selected signal of those below 100. If four signals were above 100, the subject would be shown (only) these four. In what follows, I will refer to the signals that the information source

²If the true state equalled 100, subjects received the full payment for either guess.

Table 1: Overview of the experimental design

Stage 0	Stage 1	Stage 2	Stage 3	Stage 4
Computer determines state by drawing six signals	Subject receives one signal	First binary guess b_1 based on signal	Subject observes messages of information source	Continuous guess b_2

did not share with subjects as “unobserved” or as signals that subjects “do not see.” This terminology is purely colloquial in nature and meant to make it salient that these signals do not appear on subjects’ decision screens. In an information-theoretic sense, these “unobserved” signals constitute coarse information.

In summary, subjects in *Selected* faced a selection problem akin to the examples discussed in the Introduction in that the information source conditions its messages (whether or not to send a signal) on the subject’s prior. Given the simplified discretized uniform distribution over the signal space, it was rather straightforward for subjects to infer which types of signals were unobserved. Being sophisticated about selection requires subjects to understand that when they first guessed $b_1 = high$, an unobserved signal was 70, in expectation, while it was 130 when they first guessed $b_1 = low$.

Treatment *Control* was designed to deliver the same Bayesian posterior as *Selected* without the presence of a selection problem. In the *Control* condition, participants observed two types of signals on their decision screens. First, they observed those signals that subjects in the *Selected* treatment also observed. Second, they were also shown a coarse version of the signals that subjects in the *Selected* condition did not observe. Specifically, if an unobserved signal was in $\{50, 70, 90\}$, the information source communicated 70 to the subject, while if the unobserved signal was in $\{110, 130, 150\}$, the information source communicated 130.³ These coarse messages equal the expected signal conditional on a subject’s first guess in *Selected*. Thus, the informational content of the *Selected* and the *Control* treatments is identical.

Participants solved eight tasks with independent signal draws. To keep the experimental setup close to the motivating examples in which people need to process information about multiple variables of interest, the baseline experimental setup was such that subjects completed two tasks at the same time (on the same decision screen). In the instructions and in the computer program, this was referred to as estimating “variable A” and “variable B,” respectively. Accordingly, subjects observed a first signal for each variable, then provided a first guess for each variable, and were then shown the subsequent messages of the information source, again for both variables. To avoid confusion, both the experimental instructions and the computer program specified which variable

³On their computer screens, there was no way for subjects to tell apart a “realized” 70 and an “expected” 70. I made this design choice because telling them apart is redundant for rational inference.

Table 2: Overview of the experimental tasks

True State	First signal	Observed Signal A	Observed Signal B	Observed Signal C	Observed Signal D	Unobs. Signal E	Unobs. Signal F	Bayesian Belief	Neglect Belief
96.67	130	130	150	70	–	50	50	103.33	120.00
110.00	150	110	150	110	–	50	90	110.00	130.00
93.33	50	90	50	130	–	110	130	96.67	80.00
90.00	110	150	90	50	–	50	90	90.00	100.00
103.33	150	110	130	70	–	70	90	100.00	115.00
116.67	90	90	70	150	–	150	150	110.00	100.00
116.67	110	150	130	150	110	50	–	120.00	130.00
86.67	130	130	90	110	–	70	50	90.00	100.00

Notes. Overview of the belief formation tasks in order of appearance. The categorization into observed and unobserved signals applies to the case in which subjects follow their first signal, i.e., guess ≥ 100 if their signal was larger than 100, and $<$ otherwise. Subjects in the *Selected* treatment observed only their own signal and the “observed” signals. Subjects in the *Control* condition additionally had access to a coarse version of the “unobserved” signals, i.e., if the corresponding signal was less than 100, they saw 70, and if the signal was larger than 100, they saw 130. See equations (2) and (3) for the Bayesian and neglect benchmarks.

a signal belongs to by adding a capital letter. For example, subjects’ first signals in the first period (the first two tasks) would be given by $A-130$ and $B-150$. This procedure was the same in *Control* and *Selected*. In total, subjects completed four periods (eight tasks), summarized in Table 2. All subjects were exposed to the same sets of signal realizations. Below, I discuss a treatment that verifies that very similar results hold if subjects complete these eight tasks strictly sequentially.

The intrinsic interest of this study is in subjects’ second guesses; the first guess only serves the purpose of imposing a selection problem akin to the examples described in the Introduction. Thus, to reduce noise, the instructions mentioned that subjects’ earnings from the first guess would be maximized in expectation if they followed the first signal, i.e., stated a guess above (below) 100 if the signal was above (below) 100.

Control questions ensured that subjects understood the process generating their data. For example, subjects were asked, “Assume that you issued a first guess of larger than 100. Which draws will the information source show you no matter what? (a) None. (b) Those above 100. (c) Those below 100.” Only once subjects had correctly solved all control questions could they proceed to the experiment.⁴ Appendix H contains the experimental instructions and control questions.

⁴The control questions followed a multiple choice format with 3–4 questions per screen. Thus, trial-and-error was very cumbersome. Moreover, the BonnEconLab has a control room in which the experimenter can monitor the decision screens of all experimental subjects. Thus, whenever a subject appeared to have problems in answering the control questions, an experimenter approached the subject, clarified open questions (if any), and excluded the subject from the experiment if they did not appear to understand the instructions.

2.2 Theoretical Considerations

This subsection develops a simple, mechanical formal framework to fix ideas about the experimental design above. I will use this framework below for model-based empirical analyses. The true state of the world is given by $\theta = \sum_{k=1}^6 s_k/6$. Let $\mathbb{Z}(b_1)$ denote the set of signals a subject actually sees on their computer screen, which depends on b_1 . Denote $N = |\mathbb{Z}|$. Given a set of signals, a Bayesian would compute the mean posterior belief b_B as

$$b_B = \frac{\left[\sum_{k=1}^N s_{k \in \mathbb{Z}(b_1)} \right] + (6 - N) \cdot E[s_{k \notin \mathbb{Z}(b_1)} | b_1]}{6}, \quad (2)$$

where $s_k \in \mathbb{Z}(b_1)$ denotes a signal that appears on the decision screen. The second term in the numerator corresponds to the inference of a Bayesian of those signals that are not shown, which is the expectation conditional on the first belief.

I now introduce theoretical benchmarks for neglect. A first possibility is that the agent applies a heuristic of “what you see is all there is” and does not draw *any* inferences from unobserved signals but just averages the observed data:

$$b_{N,1} = \frac{\sum_{k=1}^N s_{k \in \mathbb{Z}(b_1)}}{N}. \quad (3)$$

Comparing this benchmark with equation (2), we see that averaging the visible data generates two potential sources of error. First, the sample may be biased: because only $s_k \in \mathbb{Z}(b_1)$ appear in the numerator, b_1 determines whether predominantly high or low signals are taken into account. This is the traditional sample selection problem.

A second source of error, however, arises because even if \mathbb{Z} did not depend on b_1 (if there were no systematic sample selection), equation (3) would still ignore the unobserved signals. This is important because even if \mathbb{Z} was determined at random, the decision maker has prior knowledge about the unobserved signals that he can make use of, which is that $E[s_k] = 100$.

A plausible alternative specification of a neglect benchmark eliminates the second type of error by positing that participants are aware of the signals they do not see but fail to understand the sample selection problem created in the process. Such a decision maker imputes the unconditional expectation of $E[s_k] = 100$ for any unobserved signal. The second neglect benchmark is given by

$$b_{N,2} = \frac{\left[\sum_{k=1}^6 s_{k \in \mathbb{Z}(b_1)} \right] + (6 - N)E[s_{k \notin \mathbb{Z}(b_1)}]}{6}. \quad (4)$$

It is perhaps helpful to provide an interpretation of the psychological difference between the two neglect benchmarks in equations (3) and (4). The agent in (4) only

struggles with understanding (or computing) conditional expectations. The agent in (3) ignores the unobserved signals altogether, plausibly because he never actively thinks about how many signals there are. Because the unobserved signals are not top of mind, he naïvely uses the (visible) sample mean to estimate the population mean. Indeed, a long literature in cognitive psychology on the metaphor of a “naïve intuitive statistician” posits that people have a tendency to directly use sample moments to estimate population analogs (Fiedler and Juslin, 2006; Juslin et al., 2007).

The main experiments were not designed to distinguish between these two neglect benchmarks. The correlation between $b_{N,1}$ and $b_{N,2}$ in my experimental tasks is $\rho = 0.99$, and they make quantitatively very similar predictions. However, in follow-up experiments to be discussed in Section 4, I use the distinction between the two benchmarks to tease out the mechanisms behind neglect. The results will show that a large majority of those subjects that are not Bayesian appear to follow the first neglect benchmark. I will hence use $b_{N,1}$ in what follows.⁵

Let $\chi \in [0, 1]$ parameterize the degree of neglect such that $\chi = 1$ implies full neglect. Then, a decision-maker’s belief b can be expressed as a weighted average of b_B and $b_{N,1}$ plus decision noise ϵ :

$$b = (1 - \chi)b_B + \chi b_{N,1} + \epsilon = b_B + \chi \underbrace{\frac{6 - N}{6} (\bar{s}_{k \in \mathbb{Z}(b_1)} - E[s_{k \notin \mathbb{Z}(b_1)} | b_1])}_{\equiv d} + \epsilon \quad (5)$$

$$= b_B + \chi d + \epsilon, \quad (6)$$

where $\bar{s}_{k \in \mathbb{Z}(b_1)}$ is the average visible signal and ϵ is a mean zero random computational error. The systematic component of a subject’s belief b can be expressed as Bayesian belief plus a distortion term d times the neglect parameter χ . I will use this formal framework to compute estimates of neglect $\hat{\chi}$ and decision noise $|\hat{\epsilon}|$.

2.3 Procedural Details

Apart from the treatments described above, I implemented eight additional treatments that will be discussed below. Table 3 provides an overview of all treatments; horizontal lines indicate which treatments were randomized within experimental sessions.

The experiments were conducted at the BonnEconLab of the University of Bonn and computerized using z-Tree (Fischbacher, 2007). Participants were recruited using hroot (Bock et al., 2014). After the written instructions were distributed, subjects had ten minutes to familiarize themselves with the task. Each period consisted of two computer

⁵Table 10 in Appendix B and Figure 10 in Appendix C reproduce the main results using the $b_{N,2}$ benchmark. The results are almost identical to those to be presented below.

Table 3: Treatment overview

Treatment	# of subjects	Ave. earnings (euros)
<i>Selected</i>	74	12.77
<i>Control</i>	40	17.83
<i>Sequential</i>	75	11.28
<i>Feedback</i>	75	15.08
<i>Random</i>	75	12.10
<i>Complex</i>	75	14.28
<i>Simple</i>	75	14.47
<i>Few</i>	75	17.43
<i>Nudge</i>	72	12.18
<i>Selected replication</i>	75	12.48

Notes. Horizontal lines indicate which treatments were randomized within the same experimental sessions. Payments included a show-up fee of € 10 in *Feedback* and of € 6 in all other treatments.

screens. On the first screen, subjects were informed of the first signal and issued a binary guess. On the second screen, participants received the messages from the information source and stated a point belief. Sessions lasted 50 minutes on average.

All decisions were financially incentivized, in expectation: in total, subjects took 16 decisions, one of which was randomly selected for payment. This constitutes an incentive-compatible mechanism in such a setup (Azrieli et al., 2018). The probability that a second (point) belief was randomly selected for payment was 90%, while one of the binary first guesses was chosen with probability 10%. The binary first guess was incentivized such that subjects received € 18 if the guess was correct and nothing otherwise. The continuous point beliefs were incentivized using a quadratic scoring rule with maximum variable earnings of € 18, i.e., variable earnings of subject i in task j equalled $\pi_i^j = \max\{0; 18 - 0.2 \times (b_i^j - \theta^j)^2\}$.

3 Results

3.1 Baseline Results

Preliminaries. The object of interest in the analysis is a potential treatment difference in the second beliefs that subjects state. For completeness, across the two treatments, 93% of all first binary guesses follow the first signal and enter a high (low) first guess

if the first signal is above (below) 100. Appendix A presents a set of robustness checks that restrict the analysis to observations that followed the first signal.

Beliefs across tasks. Table 4 presents an overview of the results in each of the eight tasks. For ease of comparison, I provide the benchmarks of full neglect and Bayesian beliefs, respectively. Reassuringly, beliefs in the *Control* condition follow the Bayesian prediction very closely, suggesting that the experimental setup was not systematically misconstrued by subjects: in the absence of selected information, people state rational beliefs. In the *Selected* treatment, however, beliefs are distorted away from the Bayesian benchmark towards the full neglect belief. In all eight tasks, beliefs significantly differ between treatments at least at the 10% level, and usually at the 1% level (Wilcoxon ranksum tests).

Econometric analysis. In the remainder of the paper, treatment comparisons will be conducted by pooling the data across tasks, both for brevity and to eliminate potential multiple testing concerns. Pooling the data requires transforming the beliefs data into a scale that has the same meaning across tasks. For this purpose, I make use of the simple belief formation rule introduced in Section 2.2, which has the additional advantage that going forward, all estimated quantities will have direct theoretical counterparts. I use equation (6) to estimate the neglect implied in the belief of subject i in task j :

$$\hat{\chi}_i^j = E[\chi_i^j | b_i^j] = \frac{b_i^j - b_B^j}{d^j} = \frac{6(b_i^j - b_B^j)}{(6 - N^j) \left(\bar{s}_{k \in \mathbb{Z}(b_1)}^j - E[s_{k \notin \mathbb{Z}(b_1)}^j | b_{i,1}^j] \right)}. \quad (7)$$

Note that this analytical tool corresponds to a simple linear transformation of the raw beliefs data (subtract the Bayesian belief and divide by the distortion term d , which is only a function of the signal realizations). This method only converts the data into a consistent interval, so that subjects' beliefs (i) are on the same scale across tasks and (ii) can be directly interpreted as reflecting Bayesian ($\hat{\chi} = 0$), full neglect ($\hat{\chi} = 1$), or intermediate levels.

While $\hat{\chi}_i^j$ should, in principle, be between zero and one, in the experimental data naturally not all observations lie within this interval, likely at least partly due to typing mistakes and random computational errors. This produces outliers that are partly severe. Across the treatments in Table 3 ($N = 5,416$ belief statements), the minimum implied $\hat{\chi}_i^j$ is -21 and the maximum 12.7 . To avoid arbitrary exclusion criteria while at the same time dealing with outliers, throughout the paper I present three different sets of regression specifications. First, I present an analysis with median regressions that includes the full sample of beliefs, including large outliers. Second, an OLS analysis

Table 4: Overview of beliefs across tasks

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
True State	First Signal	Bayesian Belief	Neglect Belief	Median Belief <i>Control</i>	Median Belief <i>Selected</i>	Mean Belief <i>Control</i>	Mean Belief <i>Selected</i>	p-value (Ranksum)
96.67	High	103.33	122.00	103.00	110.00	104.84	107.67	0.0661
110.00	High	110.00	130.00	110.00	120.00	109.79	119.36	0.0001
93.33	Low	96.67	80.00	96.50	90.00	96.58	90.88	0.0130
90.00	High	90.00	100.00	90.00	90.00	90.21	94.00	0.0536
103.33	High	100.00	115.00	100.00	110.00	98.79	107.78	0.0001
116.67	Low	110.00	100.00	110.00	110.00	110.29	108.08	0.0635
116.67	High	120.00	130.00	120.00	123.00	118.29	122.04	0.0099
86.67	High	90.00	100.00	90.00	90.00	89.16	95.89	0.0022

Notes. Overview of the estimation tasks in order of appearance. See Table 2 for details on the signals in each task as well as the computation of the Bayesian and full neglect benchmarks. High (low) private signals are defined as signals above (below) 100. The p -value refers to a Wilcoxon ranksum test between beliefs in *Selected* and *Control*.

in which I winsorize the data at $|\hat{\chi}_i^j| = 3$. That is, I replace each belief that is larger (smaller) than 3 (-3) by the corresponding value. This affects 3% of all observations. Third, I present an OLS analysis on a trimmed sample, where I drop all observations with $|\hat{\chi}_i^j| > 3$. For completeness, Appendix A presents an additional set of specifications in which I implement OLS regressions on the full sample, including all outliers. The results are similar to those reported in the main text.

Table 5 presents the results. In these analyses, the unit of observation is a subject-task, for a total of usually eight observations per subject.⁶ The standard errors are clustered at the subject level. All regressions include experimental session fixed effects, leveraging random assignment into treatments within sessions.

The results confirm a large and statistically significant aggregate treatment difference between *Control* and *Selected*. In column (1), the median regression only controls for session fixed effects. Column (2) adds a vector of controls: fixed effects for each experimental task interacted with the first guess (high / low) of the subject, as well as controls for individual characteristics. In columns (3)–(4), the dependent variable is winsorized at $|3|$, and I estimate OLS regressions. In columns (5)–(6), the sample excludes observations with $|\hat{\chi}_i^j| > 3$. Throughout, the coefficient is quantitatively large and suggests that – relative to the control treatment – subjects in *Selected* exhibit a neglect of 0.4 – 0.6 units of χ .

The bias implies lower earnings of subjects in the *Selected* condition. The expected profit from all eight belief formation tasks is € 6.33 in *Selected* and € 10.32 in *Control*. Actual profits, which include a show-up fee and depend on a random draw, are € 17.56 (\$20) in *Control* and € 12.73 (\$15) in *Selected*.

⁶In a few cases, subjects did not enter a belief on time, so these observations are missing.

Table 5: Baseline results: Treatments *Selected* and *Control*

	Dependent variable: Neglect $\hat{\chi}_i^j$					
	Median regression		OLS winsorized		OLS trimmed	
	(1)	(2)	(3)	(4)	(5)	(6)
0 if <i>Control</i> , 1 if <i>Selected</i>	0.40*** (0.08)	0.50*** (0.10)	0.54*** (0.09)	0.60*** (0.09)	0.51*** (0.09)	0.54*** (0.09)
Session FE	Yes	Yes	Yes	Yes	Yes	Yes
Task FE \times prior	No	Yes	No	Yes	No	Yes
Controls	No	Yes	No	Yes	No	Yes
Observations	894	894	894	894	874	874
R^2	0.07	0.10	0.09	0.11	0.10	0.11

Notes. Regression estimates, with robust standard errors (clustered at subject level) in parentheses. The dependent variable is the neglect $\hat{\chi}_i^j$ that is implied in a given belief. The sample includes each of subjects' eight beliefs in the *Selected* and *Control* conditions. Columns (1)–(2) report median regressions, and columns (3)–(6) are OLS regressions. In columns (3)–(4), the dependent variable is winsorized at $|\hat{\chi}_i^j| = 3$. In columns (5)–(6), the sample is trimmed at $|\hat{\chi}_i^j| = 3$. Controls include gender, high school grades, and log monthly disposable income. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

3.2 Robustness Treatments

Sequential Tasks. To assess the extent to which the simultaneous presentation of two variables induces neglect, I implemented treatment *Sequential*. This treatment was randomized along with *Control* and *Selected* within experimental sessions. *Sequential* is identical to *Selected*, except that all eight tasks were presented in eight, rather than four, consecutive rounds. Appendix D discusses the results from this treatment in detail. Overall, the results are very similar to those in *Selected*. To illustrate, Figure 1 plots the median and mean $\hat{\chi}_i^j$ across treatments, along with standard error bars. While the median neglect estimate is significantly lower in *Sequential* than in *Selected*, the averages are very similar ($\bar{\chi}_i^j = 0.49$ in *Selected* and $\bar{\chi}_i^j = 0.42$ in *Sequential*). Moreover, neglect in *Sequential* is significantly higher than in *Control*.⁷

Feedback. A relevant question is whether people learn about their errors through feedback. In treatment *Feedback*, subjects first solved six tasks (again two per period) that had the same structure as those in *Selected* but different signal realizations. Then, they completed the same eight tasks as subjects in *Selected*. Thus, I can compare beliefs across treatments for exactly the same tasks, yet subjects in *Feedback* have already com-

⁷As documented in Table 15 in Appendix D, median and average neglect are consistently lower in *Sequential* than in *Selected*. While these differences are usually not statistically significant, they provide some very tentative evidence that the simultaneous presentation of problems induces higher cognitive load, which in turn increases neglect. See Section 4 for a discussion along these lines.

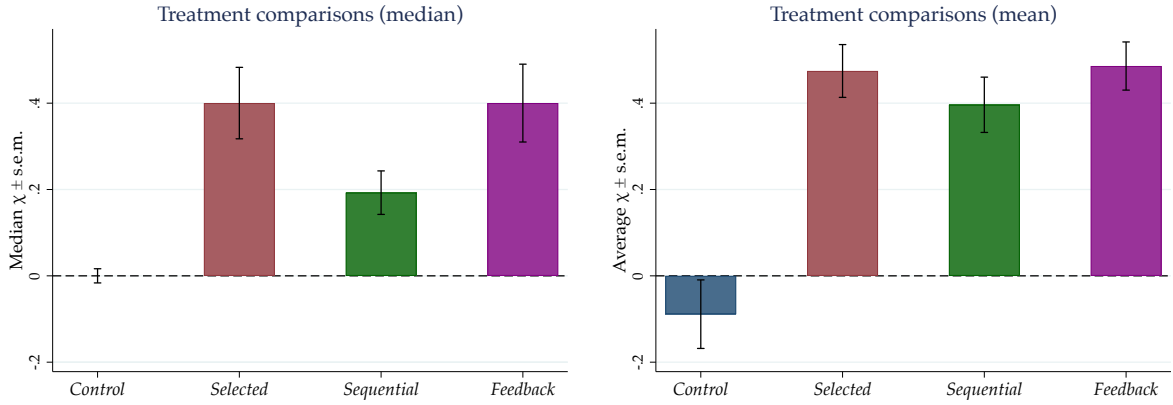


Figure 1: Overview of neglect $\hat{\chi}_i^j$ across treatments. The left panel shows the median $\hat{\chi}_i^j$ across all subject-task observations. The right panel shows the average $\hat{\chi}_i^j$ across all subject-task observations, where as in columns (3)–(4) of Table 5 the data are winsorized at $|\hat{\chi}_i^j| = 3$. For treatment *Feedback*, the sample median and average are computed for the last eight beliefs to keep the results comparable to the other treatments. Standard error bars are computed based on clustering at the subject level.

pleted six tasks and received feedback on each of them. After each period, subjects received feedback about their performance: (i) they were reminded of their continuous belief statement; (ii) they were informed of the corresponding true state; and (iii) they received information on the profits that would result from the respective task in case it would be selected for payment. Appendix E provides a detailed analysis of the data. The data show no indication that feedback reduces the amount of neglect. Figure 1 illustrates this result.

3.3 Decision Rules and Heterogeneity Analysis

Type distribution. To examine the subject-level distribution of neglect, I seek to identify a subject’s neglect type $\hat{\chi}_i$, i.e., an estimate of a subject’s solution strategy, net of computational errors and heuristic rounding. For this purpose, for each subject i and candidate type $t \in \{-1, -0.9, \dots, 2\}$, I count how many of the implied $\hat{\chi}_i^j$ (see eq. (7)) satisfy $|t - \hat{\chi}_i^j| \leq 0.05$. Then, I classify each subject as $\hat{\chi}_i = t_{max}$, where t_{max} is the candidate type that rationalizes the largest number of beliefs (see Fragiadakis et al., 2016, for a similar approach).⁸

The left panel of Figure 2 presents a histogram of these modal neglect types $\hat{\chi}_i$ in treatments *Selected*, *Sequential*, and *Feedback*. The data reveal a bimodal type distribution: 60% of all subjects are best characterized as Bayesian ($\hat{\chi}_i = 0$) or full neglect ($\hat{\chi}_i = 1$). For example, of those 150 subjects that are not approximately rational, one third (51) are classified as exactly or almost exactly full neglect types ($0.95 \leq \chi_i \leq$

⁸If more than one type rationalizes the maximal number of beliefs, I compute the average across t .

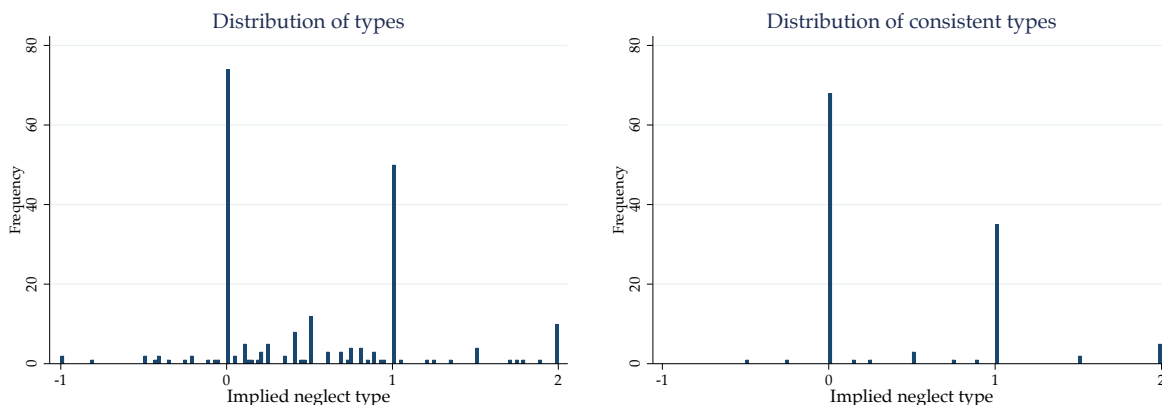


Figure 2: Distribution of modal neglect types $\hat{\chi}_i$ in treatments *Selected*, *Sequential*, and *Feedback*. The left panel shows the distribution of all estimated neglect types, and the right panel the distribution of neglect types for which at least three beliefs are type-consistent (53% of all subjects). For belief j to be consistent with the estimated type means that $|\hat{\chi}_i - \hat{\chi}_i^j| \leq 0.05$.

1.05).⁹ In contrast, in treatment *Control*, 80% of all subjects are classified as exactly $\hat{\chi}_i = 0$; see Figure 11 in Appendix C.

Across-task consistency and heuristic responses. Because the left panel of Figure 2 shows modal types, the figure does not take into account the within-subject-across-task consistency in stated beliefs. To address this, I look at the number of beliefs that are consistent with a subject’s modal type, where type-consistent means that the neglect parameter that is implied by a belief statement is close to the overall estimated type: $|\hat{\chi}_i - \hat{\chi}_i^j| \leq 0.05$. Figure 5 in Appendix C shows a histogram of the number of type-consistent beliefs. The average and median number of type-consistent beliefs are 3.2 and 3. The noisiness of the data in bounded rationality experiments – and the fact that a considerable fraction of subjects does not appear to behave according to a stable type – has recently been highlighted by Fragiadakis et al. (2016).¹⁰ They and Costa-Gomes and Crawford (2006) propose that a subject should be viewed as having a stable type if at least 40% of their experimental actions are type-consistent.

The right panel of Figure 2 shows the distribution of modal neglect types, restricting attention to those 53% of all subjects for which at least three beliefs ($\sim 40\%$) are type-consistent. We see that the two spikes at $\chi_i = 0$ and $\chi_i = 1$ largely remain, yet the

⁹Figure 9 in Appendix C plots a histogram of the subject-task-specific $\hat{\chi}_i^j$, i.e., the underlying raw beliefs data. Naturally, this distribution is noisier but also bimodal with spikes at zero and one.

¹⁰The noisiness of the beliefs data appears to be at least partly driven by heuristic rounding to the nearest multiple of five or ten, akin to the patterns documented in a large survey literature on subjective expectations about economic variables (Manski, 2004). In my data, 69% of reported beliefs are multiples of ten and 84% are multiples of five. These numbers are probably inflated because the Bayesian or full neglect benchmarks are themselves usually multiples of five or ten, compare Table 2. Yet when I exclude beliefs that correspond to the Bayesian or full neglect benchmarks, still 53% are multiples of ten and 75% multiples of five.

vast majority of all types $\hat{\chi}_i \neq 0, 1$ are relatively inconsistent across tasks. A perhaps remarkable 87% of those subjects that exhibit stable identifiable decision types can be characterized as exactly rational or exactly full neglect. Very few subjects exhibit a stable decision type of partial adjustment from neglect.

In the full sample of subjects, for the $\hat{\chi}_i = 0$ types, 4.5 beliefs are type-consistent, on average. For the $\hat{\chi}_i = 1$ types, 3.4 beliefs are type-consistent, on average. However, for all types $\hat{\chi}_i \neq 0, 1$, the average number of type-consistent beliefs is only 2.0. Overall, these patterns suggest that across-task consistency is relatively low, in particular for the $\hat{\chi}_i \neq 0, 1$ types.¹¹ Still, to the extent that there is within-subject consistency in my data, it points to the presence of two fundamentally different updating types.

As a final remark on within-subject consistency, it is worth pointing out that the relatively inconsistent subjects are *not* just random noise around the rational benchmark. The average and median task-level implied neglect parameters of relatively inconsistent subjects are $\chi_i^j = 0.35$ and $\chi_i^j = 0.40$. This shows that the inconsistent types *do* neglect selection – just in a quantitatively inconsistent fashion across tasks.

Correlates of Neglect. Table 11 in Appendix B investigates the correlates of neglect in treatments *Selected*, *Sequential*, and *Feedback*. I find that both better high school grades and longer response times are negatively correlated with neglect. The quantitative magnitude of the relationship between response times and neglect is small. Interpreted causally, the regression coefficients suggest that response times would have to increase by about four minutes per task to move a full neglect belief to a Bayesian belief. However, the average response time in the data in the three treatments that are considered here is only 48 seconds, and it is 52 seconds in treatment *Control*. These magnitudes suggest that the type heterogeneity is not merely the result of the neglect types being lazier than the rational types.

4 Mechanisms

4.1 Framework

Understanding the mechanisms behind errors in statistical reasoning is likely to be relevant not only for theorists who are interested in formalizing and endogenizing people’s errors, but also for policy in terms of what will be an effective method to correct biased beliefs. To structure the analysis, I pit two hypotheses against each other.

¹¹The intra-correlations between modal, median and average neglect types are all between 0.75 and 0.92. Figures 6–8 in Appendix C present histograms of (i) median subject-level neglect; (ii) average neglect parameters; and (iii) the subject-level standard deviation of implied neglect parameters.

Theory A: Incorrect mental model. Participants have an initial mental default model according to which the unobserved signals are not top of mind. This default model could result from intuitive system 1 reasoning (Kahneman, 2011), or it could be retrieved from memory as the “normal” version of a class of problems that people know how to solve (Kahneman and Miller, 1986). If the unobserved signals do not come to mind, participants directly use the (visible) sample mean to estimate the population mean, akin to the psychological metaphor of a “naïve intuitive statistician” who directly uses sample moments to estimate population analogs (Fiedler and Juslin, 2006; Juslin et al., 2007). This simple averaging process may be loosely summarized as “what you see is all there is.”

If selection does come to mind, the participant reasons about whether and how it needs to be corrected for. Whether this happens partly depends on how the decision-maker allocates cognitive resources between (i) setting up a mental model and (ii) computational implementation. In particular, people should be less likely to develop a correct mental model if they are cognitively busier with (or distracted by) computationally implementing a given solution strategy.

Linking this account to the literature, the importance of incorrect mental models is highlighted by an active theoretical literature (e.g., Esponda and Pouzo, 2016; Eyster and Rabin, 2010; Jehiel, 2005; Schwartzstein, 2014; Gabaix, 2014; Spiegler, 2016; Bohren and Hauser, 2017; Heidhues et al., 2017; Gagnon-Bartsch et al., 2018). For example, the model in Spiegler (2017) focuses on how an agent naïvely extrapolates from partial data, which is reminiscent of the sample selection problem in this paper. Indeed, incorrect mental models are often implicitly, and sometimes explicitly, motivated and modeled as resulting from attentional processes (Gennaioli and Shleifer, 2010).

Theory B: Conceptual or computational difficulty of accounting for selection. Participants are aware of the signals they do not see, but struggle with the conceptual or computational difficulty of correcting for selection.

It is worth highlighting that these two stories are not necessarily mutually exclusive but relate to two distinct steps of a sequential reasoning process. In the first step, it gets determined whether selection (the unobserved signals) are top of mind. In the second step, the decision maker reasons about how to correct for selection, if it comes to mind in the first place. It is in principle conceivable that selection does not come to mind, but even if it did come to mind, the participant wouldn't be able to account for it.

The experiments below test the relative importance of these two stories by exogenously manipulating parameters that should lead to changes in reported beliefs according to one theory but not the other. I conduct three such comparative statics exercises:

1. Holding fixed the presence of unobserved signals, I eliminate the presence of the selection problem. If neglect was largely driven by theory B, then it should disappear in this treatment. If neglect was largely driven by theory A, then it should remain roughly constant.
2. Holding fixed the conceptual and computational difficulty of accounting for selection, I increase the computational complexity of following a “what you see is all there is” averaging heuristic. If neglect was largely driven by theory B, then such complexity variations should have no effect. Under theory A, higher computational complexity should increase neglect because the decision maker is “distracted” by computational implementation and hence devotes less resources to thinking about what is not top of mind or visible.
3. Holding fixed the conceptual and computational difficulty of accounting for selection, I exogenously draw participants’ attention to the unobserved signals. If neglect was largely driven by theory A, it should substantially decrease. If neglect was largely driven by theory B, it should remain constant.

4.2 Eliminating the Selection Problem

4.2.1 Experimental Design

To test comparative statics prediction 1. above, I implemented treatment *Random*. *Random* closely follows treatment *Selected*. The true state to be estimated now consists of the average of eight random draws from the same simple discretized uniform distribution as before.¹² Deviating from the procedure in *Selected*, in Stage 3 of the experiment, a subject observed three signals that were selected *at random*, rather than based on a subject’s first guess. The timeline of this treatment was otherwise identical to that in treatment *Selected*. In this setup, the Bayesian belief is given by

$$b_B = \frac{[\sum_{k=1}^4 s_{k \in \mathbb{Z}(b_1)}] + 4 \cdot E[s_{k \notin \mathbb{Z}(b_1)}]}{8}, \quad (8)$$

while a “what you see is all there is” benchmark is given by the same equation as before:

$$b_{N,1} = \frac{\sum_{k=1}^4 s_k}{4}. \quad (9)$$

¹²In this treatment the true state was determined as average of eight rather than six random draws to allow for a larger number of invisible signals. With only two invisible signals, the Bayesian and full neglect benchmarks would have been too close to each other to allow for robust analyses that distinguish between these two updating types.

It is worth pointing out that this treatment also directly speaks to the two potential neglect benchmarks for treatment *Selected* that I discussed in Section 2.2: (i) a decision rule that assumes that subjects completely ignore information that is not visible on their computer screen and (ii) a decision rule that posits that participants are aware of the signals they do not see but wrongly assign them their unconditional rather than conditional expectation. If (ii) was the empirically correct benchmark, then subjects in *Random* should state Bayesian beliefs.

4.2.2 Results

The results are described in detail in Appendix F. To summarize, behavior in this treatment is very similar to behavior in treatment *Selected*. I again compute implied subject-level neglect parameters χ_i , where zero corresponds to the Bayesian and one to the full neglect benchmark noted above. As shown in Figure 3, the distribution of stated beliefs is again bimodal, with subjects either fully neglecting what they don't see or behaving rationally. Indeed, as shown in Appendix F, the distribution of neglect in this treatment is statistically indistinguishable from the one in treatment *Selected*.

I view the results of this treatment as suggesting two implications. First, a “what you see is all there is” heuristic describes behavior better than a theoretical benchmark in which subjects actively impute unconditional expectations for unobserved signals. This suggests that at least a majority, and probably a large majority, of those subjects that are classified as “neglect” types in treatment *Selected* do not at all take into account the unobserved signals. Second, the psychological mechanism behind neglect is likely not (just) a conceptual misunderstanding of selection problems but instead a general incorrect mental model according to which the unobserved signals do not even come to mind in the first place.

4.3 Computational Complexity as Distraction

4.3.1 Experimental Design

Next, I study how computational complexity affects selection neglect, in particular the ways in which it might induce cognitive load and hence distract participants from the unobserved signals. The experiments below exogenously manipulate the computational complexity of the updating problem but *hold fixed the difficulty of accounting for selection itself*. This thought experiment has the attractive feature that it narrows down the pathways through which complexity can affect belief updating: if the difficulty of correcting for selection remains unchanged, then differences in belief updating can plausibly be attributed to an effect of computational complexity on how participants approach the

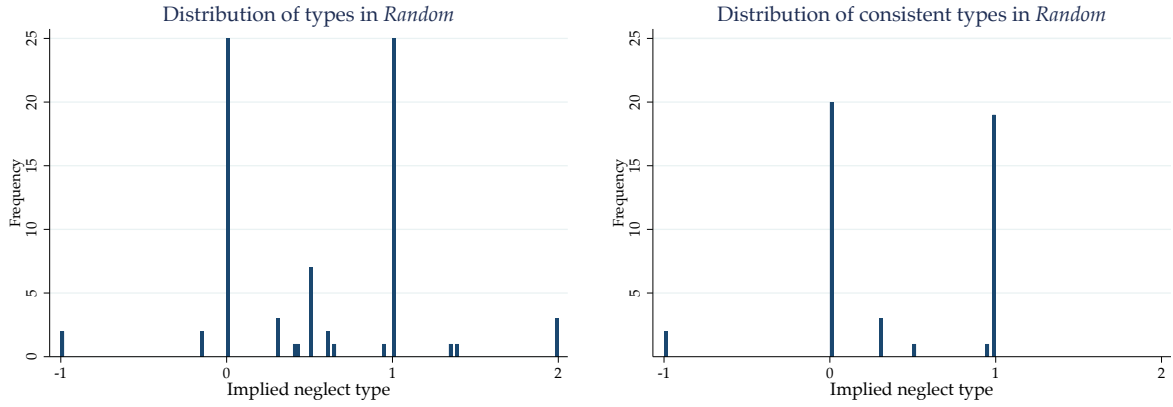


Figure 3: Distribution of modal neglect types $\hat{\chi}_i$ in treatment *Random*. See the main text for the specification of the Bayesian and full neglect benchmarks. The left panel shows the distribution of all estimated neglect types, and the right panel the distribution of neglect types for which at least three beliefs are type-consistent (61% of all subjects). For belief j to be consistent with the estimated type means that $|\hat{\chi}_i - \hat{\chi}_i^j| \leq 0.05$.

problem (develop a mental model) in the first place. Given the absence of a general theory of what is complex, the experiments operationalize computational complexity in two different and arguably intuitive ways: (i) the complexity of the signal space and (ii) the number of signals in a given updating problem.

Complexity I: The Complexity of the Signal Space. To exogenously vary the complexity of the signal space, I conducted two treatments, *Complex* and *Simple*. These two treatments were both identical to treatment *Selected* except that the set of numbers from which the true state was determined was varied. In *Complex*, the signal space was given by

$$\{70, 70, 70, 70, 70, 70, 70, 104, 114, 128, 136, 148, 150\}.$$

In *Simple*, it was

$$\{70, 70, 70, 70, 70, 70, 70, 130, 130, 130, 130, 130, 130\}.$$

These two treatments are identical in a number of ways: (i) the prior is 100; (ii) the conditional expectations of being above and below 100 are 130 and 70, respectively; (iii) most importantly, these two treatments leave the difficulty of accounting for selection constant if subjects state a first guess of above 100 (i.e., in practice, when they receive a first signal above 100). In such cases, accounting for selection only requires subjects to notice (remember) that they are missing a few 70's on their decision screens. Thus, in both treatments, people's potential problems in computing conditional expectations cannot drive any results. For example, in one task, subjects in *Complex*

observed 150, 104, 148, 114 on their decision screens, while those in *Simple* observed 130, 130, 130, 130.

Complexity II: The Number of Signals. Treatment *Few* was identical to *Complex* in almost all dimensions. The only difference is the number of random draws (signals) that determined the true state and were shown to subjects. In *Few*, the state was determined as the average of two, rather than six, random draws.

Subjects in *Few* also observed a first signal and then issued a first binary guess. Given that there are only two signals in total in this treatment, subjects then potentially observed one additional signal from the information source. Subjects only observed this second signal if it was above 100 and the subject's first guess was above 100, or if the second signal was below 100 and the subject's first guess below 100. Thus, in many tasks, subjects did not receive an additional (second) signal from the information source on the second decision screen. Notice that if subjects observe both signals, there is no selection problem, so that by design, the analysis of *Few* has to exclude the three experimental tasks for which this was the case.

Comparing treatments *Few* and *Complex* leaves the signal space and hence the difficulty of backing out unobserved signals unchanged. Still, the computational complexity of computing posteriors differs across treatments. For example, in one task, subjects in *Complex* observed 150, 104, 148, 114 on their decision screens, while those in *Few* observed 150.

In summary, all treatments hold the difficulty of accounting for selection constant but vary the computational burden of computing beliefs. A notable difference to earlier cognitive load experiments is that here cognitive load arises endogenously as feature of the decision problem, rather than being exogenously induced by the experimenter.

Finally, note that comparing treatments *Simple* and *Few* is not meaningful by design because these two treatments differ in two dimensions in ways that operate in opposite directions. Treatment *Simple* is simpler than *Few* in that it has a simpler signal space, but treatment *Few* is simpler in that it features a smaller number of signals. Thus, the analysis compares *Complex* to *Simple* and *Complex* to *Few*. Treatments *Complex*, *Simple*, and *Few* were all randomized within the same experimental sessions; compare Table 3. Tables 12 and 13 in Appendix B show the signal realizations in these treatments.

4.3.2 Manipulation Checks

Given that the treatment variations here are arguably relatively subtle and do not have immediate antecedents in the literature, it is worth performing a manipulation check to verify that the computational complexity is indeed meaningfully higher in *Complex*

than in *Simple* and *Few*. To provide such evidence, I consider data on (i) response times and (ii) the noisiness of responses across tasks. Higher computational complexity should translate into (i) longer response times and (ii) beliefs data that are noisier, or less consistent across tasks. Following equation (7), I estimate decision noise by comparing a subject’s belief in task j with the belief they “should have” stated given their estimated overall type $\hat{\chi}_i^j$: $|\hat{\epsilon}_i^j| = |\hat{\chi}_i^j - \hat{\chi}_i|$, where $\hat{\chi}_i$ is the overall estimate of i ’s type across tasks as derived in Section 3.

Table 14 in Appendix B shows that both response times and decision noise are indeed significantly lower in *Simple* and *Few*, as compared to *Complex*. This provides reassuring evidence that the treatment variations actually induced meaningful variations in computational complexity as perceived by the experimental participants.¹³

4.3.3 Results

By design of the experiment, the analysis is restricted to those tasks in which subjects’ first signal was above 100 so that any unobserved signal had to be a 70 in all treatments. Figure 4 plots median and average levels of $\hat{\chi}_i^j$ across treatments. Here, just like in the regression tables, $|\hat{\chi}_i^j|$ is winsorized at 3 when I compute treatment averages. As predicted, treatment *Complex* generates substantially higher levels of neglect than *Simple* and *Few*. The median implied neglect in *Simple* and *Few* is zero, though the averages are strictly positive ($\bar{\chi}_i^j = 0.13$ in *Simple* and $\bar{\chi}_i^j = 0.22$ in *Few*).

Table 6 provides a set of corresponding regression analyses. In all regressions, the omitted baseline category is treatment *Complex*. By including treatment dummies for *Simple* and *Few*, the regressions compare *Complex* with *Simple* and *Complex* with *Few*.

Both treatment dummies have negative coefficients that are statistically significant. These results hold both in the analysis with median regressions (columns (1)–(2)) and in robustness checks in which the dependent variable is winsorized or trimmed (columns (3)–(6)). In terms of quantitative magnitude, the coefficients suggest that both types of complexity reductions caused a reduction in neglect by about 0.2 – 0.3 units of $\hat{\chi}_i^j$. Thus, the increased cognitive load from the computational stage of the problem appears to have systematic effects on how participants approach the conceptual stage of forming a mental model to begin with. This provides further evidence that in this context selection neglect is not (just) driven by the conceptual or computational

¹³A potential issue with the interpretation that higher computational complexity increases decision noise is that it is impossible for me to formally disentangle the story that decision error is lower for less computationally complex tasks from a scenario where decision error conditional on type is independent of computational complexity, but the more complex treatment changes the type distribution and decision error are larger for neglect types. However, this alternative interpretation of the results is less plausible because the calculations that are required to be Bayesian are unambiguously more complicated than those required to follow the neglect benchmarks.

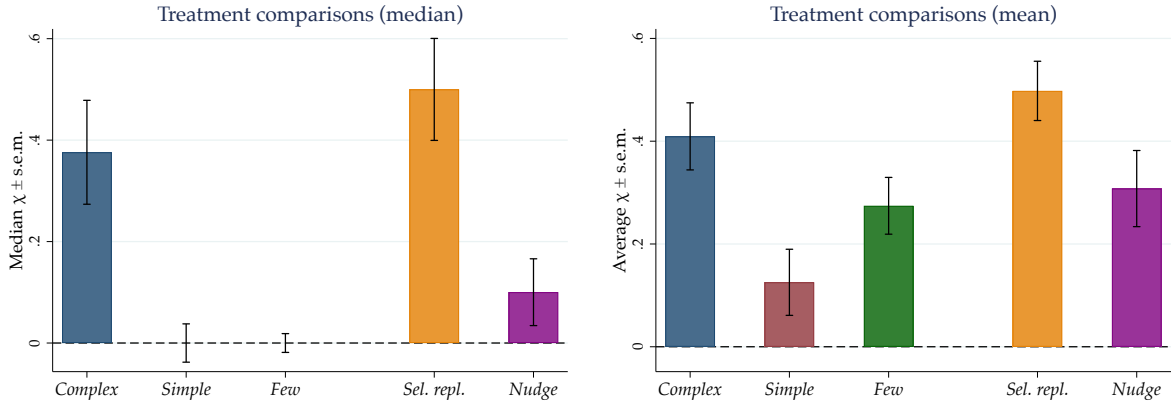


Figure 4: Overview of neglect $\hat{\chi}_i^j$ across treatments. The left panel shows the median $\hat{\chi}_i^j$ across all subject-task observations. The right panel shows the average $\hat{\chi}_i^j$ across all subject-task observations, where as in columns (3)–(4) of Table 5 the data are winsorized at $|\hat{\chi}_i^j| = 3$. Standard error bars are computed based on clustering at the subject level. As explained in the text, by design, the analysis of treatments *Complex*, *Simple*, and *Few* is restricted to those experimental tasks in which the first signal was above 100. Moreover, also by design, for treatment *Few* the analysis excludes those tasks in which subjects observed both (and hence all) signals, so no selection problem was present. As explained in the main text, these data exclusions follow mechanically from the construction of the different treatments.

difficulty of accounting for selection – as this was held constant across treatments – but by an incorrect mental model.

4.4 Nudge Evidence

4.4.1 Experimental Design

If it is true that participants in *Selected* entertain an incorrect mental model, then nudging their attention towards (or reminding them of) the existence of the selection problem might attenuate the bias. Specifically, treatment *Nudge* was identical to *Selected*, except that both the end of subjects’ written instructions and their decision screens contained the following hint:

HINT: Also pay attention to those randomly drawn balls that are not shown to you by the information source.

Treatment *Nudge* was implemented along with a replication of treatment *Selected* to facilitate within-session randomization of subjects into treatments.¹⁴

¹⁴To additionally investigate whether subjects are capable of computing the conditional expectations that are required in the present experiment, treatments *Selected* and *Sequential* contained two incentivized follow-up questions: “Suppose you knew that ten balls were randomly drawn and that all of these balls had numbers *GREATER* than 100. What would you estimate is the average of these ten numbers?” Subjects were asked the same question with *GREATER* replaced by *SMALLER*. For each question, subjects received €0.50 for a correct response and €0.20 if the response was within 5 of the correct response. Fig-

Table 6: Treatments *Complex*, *Simple*, and *Few*

Omitted category: <i>Complex</i>	Dependent variable: Neglect $\hat{\chi}_i^j$					
	Median regression		OLS winsorized		OLS trimmed	
	(1)	(2)	(3)	(4)	(5)	(6)
1 if <i>Simple</i>	-0.29** (0.12)	-0.26*** (0.10)	-0.28*** (0.09)	-0.27*** (0.09)	-0.25*** (0.08)	-0.25*** (0.09)
1 if <i>Few</i>	-0.29** (0.12)	-0.24** (0.10)	-0.17* (0.09)	-0.29*** (0.09)	-0.18** (0.08)	-0.22*** (0.08)
Session FE	Yes	Yes	Yes	Yes	Yes	Yes
Task FE \times prior	No	Yes	No	Yes	No	Yes
Controls	No	Yes	No	Yes	No	Yes
Observations	1177	1177	1177	1177	1138	1138
R^2	0.01	0.02	0.03	0.08	0.04	0.06

Notes. Regression estimates with robust standard errors (clustered at subject level) in parentheses. The sample includes treatments *Complex*, *Simple*, and *Few*. By the design of the experiment, the sample is restricted to those tasks in which following the first signal implies a first guess above 100. In treatment *Few*, experimental tasks in which subjects observe both signals are necessarily excluded because there is no scope for neglecting selection. Columns (1)–(2) report median regressions, and all other columns OLS regressions. In columns (3) and (4), $|\hat{\chi}_i^j|$ is winsorized at 3. In columns (5)–(6), the sample is trimmed at $|\hat{\chi}_i^j| = 3$. Controls include gender, high school grades, and log monthly disposable income. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

4.4.2 Results

Figure 4 shows that treatment *Nudge* generates lower levels of neglect than *Selected replication*. Table 7 provides a set of corresponding regression analyses. Treatment *Nudge* reduces neglect by about 0.2–0.4 units of $\hat{\chi}_i^j$, which corresponds to about half of the treatment difference between *Selected* and *Control*. In *Selected replication*, the median and average neglect are $\hat{\chi}_i^j = 0.50$ each, while in *Nudge* the median is $\hat{\chi}_i^j = 0.10$ and the average $\hat{\chi}_i^j = 0.30$.

4.5 Discussion

In summary, the evidence from the treatments aimed at identifying mechanisms suggests that at least a large part of the reason why participants neglect selection in my experiments is that the unobserved signals are not top of mind in the first place, so that participants operate with an incorrect mental model and directly use the sample mean

ure 12 in Appendix C presents histograms of subjects' responses to these two questions. A large majority (almost 80%) of subjects guess the correct conditional expectations.

Table 7: Treatments *Selected replication* and *Nudge*

	Dependent variable: Neglect $\hat{\chi}_i^j$					
	Median regression		OLS winsorized		OLS trimmed	
	(1)	(2)	(3)	(4)	(5)	(6)
0 if <i>Selected repl.</i> , 1 if <i>Nudge</i>	-0.40*** (0.11)	-0.20** (0.08)	-0.20** (0.09)	-0.21** (0.09)	-0.22*** (0.08)	-0.24*** (0.08)
Session FE	Yes	Yes	Yes	Yes	Yes	Yes
Task FE \times prior	No	Yes	No	Yes	No	Yes
Controls	No	Yes	No	Yes	No	Yes
Observations	1174	1174	1174	1174	1154	1154
R^2	0.02	0.10	0.03	0.11	0.03	0.10

Notes. Regression estimates with robust standard errors (clustered at subject level) in parentheses. The sample includes treatments *Selected replication* and *Nudge*. Columns (1)–(2) report median regressions, and all other columns OLS regressions. In columns (3)–(4), $|\hat{\chi}_i^j|$ is winsorized at 3. In columns (5)–(6), the sample is trimmed at $|\hat{\chi}_i^j| = 3$. Controls include gender, high school grades, and log monthly disposable income. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

to estimate the population mean.

At the same time, these results do not imply that the conceptual or computational difficulty of accounting for selection is unimportant. First, in treatment *Nudge*, neglect did not disappear despite the fairly strong hint, which suggests that some participants also struggle with the conceptual logic of selection. Second, this experiment was deliberately designed to make overcoming selection both conceptually and computationally reasonably simple, yet doing so is likely much more difficult in real-world applications.

5 Replication

The experiments reported above replace a set of similar experiments, on which an earlier working paper version of this paper was based. The earlier experiments followed a very similar logic to the ones described above. Subjects estimated an abstract true state and received computer-generated signals that induced a selection problem of the same kind as above. While there are a few differences between the earlier experiments and the ones discussed in the main text, the perhaps most important difference is that, in the earlier experiments, the true state was based on 15, rather than six, random draws. Thus, in the earlier experiments, subjects also needed to account for the base rate in processing selected signals. The new design eliminates this additional difficulty. Because the earlier experiments are very similar to the ones reported above, they can

be viewed as a replication or robustness exercise. In particular, the earlier experiments also contained versions of treatments *Selected*, *Control*, *Nudge*, *Complex*, and *Simple*. Appendix G summarizes these earlier experiments and the corresponding results. These experiments also show that (i) subjects neglect selection on average; (ii) the type distribution exhibits a bimodal structure; (iii) an experimental nudge to consider the off-screen signals has a significant effect on beliefs; and (iv) increasing the computational complexity of the decision problem – while holding the difficulty of accounting for selection constant – increases the frequency of neglect.¹⁵

6 Discussion and Related Literature

This paper has shown that people have a strong average propensity to neglect selection problems when forming beliefs, even when the information-generating process is known and transparent. A detailed analysis of the mechanisms that give rise to biased belief updating has highlighted the important role of what comes to mind and the resulting mental models. As reflected by the type distribution of neglect, these mental models appear to be binary in nature: subjects either employ a simplistic (and likely automatic) default model of the environment that ignores unobserved data, or they develop an objectively correct representation. An important result of the analysis is that this neglect should not be thought of as an exogenously given neglect parameter that is constant across individuals or even contexts. Rather, the extent to which subjects neglect selection is partly determined by the computational complexity of the decision problem, and the extent to which the decision maker’s attention is drawn to the presence of selection.

As discussed in the Introduction, the paper’s approach and results speak to the informal metaphor of a “naïve intuitive statistician” in cognitive psychology (see Fiedler and Juslin (2006); Juslin et al. (2007) for overviews and Brenner et al. (1996); Koehler and Mercer (2009) for applications to selection problems).¹⁶ This metaphor and a simple averaging heuristic also characterize much recent experimental economics work on

¹⁵Apart from providing a replication, the earlier experiments also allow for one extension: a study of the responsiveness of subjects’ wrong beliefs to observing others holding different beliefs, even though everybody received the same selected information. To investigate this, I implemented experiments that were similar to treatment *Selected*, except that after subjects had provided their continuous point belief about the true state, they were shown the beliefs of two randomly selected participants from the same experimental session who completed exactly the same task. Then, subjects were provided with an opportunity to revise their beliefs. However, in the data, subjects appear to be very confident in their own way of looking at the problem and largely abstain from revising their beliefs. See Appendix G.6 for details.

¹⁶Work on the availability heuristic (Tversky and Kahneman, 1973) is also related in its focus on salient information. However, experimental evidence for the availability heuristic usually involve showing that *irrelevant* information influences judgment such as in free-form cued recall problems, while in my experiments, *relevant* information is neglected.

information-processing (Eyster et al., 2015; Enke and Zimmermann, 2019; Grimm and Mengel, 2014; Graeber, 2018), including contemporaneous work on endogenous sample selection problems (Esponda and Vespa, 2018; Jin et al., 2018; Araujo et al., 2018; Charness et al., 2018). Indeed, a long line of work on network experiments has documented that a deGroot-style averaging heuristic often describes behavior in complex situations well. What sets this paper apart from these contributions is (i) the focus on selection problems under a known data-generating process; (ii) a detailed study of the role of incorrect mental models for neglect; including (iii) an exploration of the effect of computational complexity on how people form mental models. Thus, the paper is close to other work that focuses on *why* people make mistakes in contingent reasoning. Other such work has highlighted the importance of inferring from simultaneous vs. sequential data (Esponda and Vespa, 2016; Ngangoue and Weizsäcker, 2015) and of uncertainty (Martínez-Marquina et al., 2017).

The paper's results also contribute to an active theory literature that highlights the importance of incorrect mental models. Frequently, researchers motivate incorrect mental models by appealing to constraints on what is top of mind, and this paper has provided encouraging evidence in this regard. Going forward, a relevant issue for both the theory and the experimental literature will be to identify and describe (i) *which* incorrect mental models people form and (ii) how these depend on *contextual features* that are irrelevant under traditional theories, such as complexity, salience, and environmental cues that active different memory traces.

References

- Araujo, Felipe A, Stephanie W Wang, and Alistair J Wilson**, “The Times They are a-changing: Dynamic Adverse Selection in the Laboratory,” *Working Paper*, 2018.
- Azrieli, Yaron, Christopher P Chambers, and Paul J Healy**, “Incentives in Experiments: A Theoretical Analysis,” *Journal of Political Economy*, 2018, 126 (4), 1472–1503.
- Benjamin, Daniel J**, “Errors in Probabilistic Reasoning and Judgmental Biases,” in “Handbook of Behavioral Economics” 2018.
- Benjamin, Daniel J., Sebastian A. Brown, and Jesse M. Shapiro**, “Who is ‘Behavioral’? Cognitive Ability and Anomalous Preferences,” *Journal of the European Economic Association*, 2013, 11 (6), 1231–1255.
- Bock, Olaf, Ingmar Baetge, and Andreas Nicklisch**, “hroot: Hamburg Registration and Organization Online Tool,” *European Economic Review*, October 2014, 71, 117–120.
- Bohren, J Aislinn and Daniel Hauser**, “Bounded Rationality And Learning: A Framework and A Robustness Result,” *Working Paper*, 2017.
- Brenner, Lyle A., Derek J. Koehler, and Amos Tversky**, “On the Evaluation of One-Sided Evidence,” *Journal of Behavioral Decision Making*, 1996, 9 (1), 59–70.
- Charness, Gary, Ryan Oprea, and Sevgi Yuksel**, “How Do People Choose Between Biased Information Sources? Evidence from a Laboratory Experiment,” *Working Paper*, 2018.
- Costa-Gomes, Miguel A and Vincent P Crawford**, “Cognition and behavior in two-person guessing games: An experimental study,” *American Economic Review*, 2006, 96 (5), 1737–1768.
- Enke, Benjamin and Florian Zimmermann**, “Correlation Neglect in Belief Formation,” *Review of Economic Studies*, 2019, 86 (1), 313–332.
- Esponda, Ignacio and Demian Pouzo**, “Berk–Nash equilibrium: A framework for modeling agents with misspecified models,” *Econometrica*, 2016, 84 (3), 1093–1130.
- **and Emanuel Vespa**, “Hypothetical Thinking: Revisiting Classic Anomalies in the Laboratory,” *Working Paper*, 2016.

- **and** — , “Endogenous sample selection: A laboratory study,” *Quantitative Economics*, 2018, 9 (1), 183–216.
- Eyster, Erik and Matthew Rabin**, “Naïve Herding in Rich-Information Settings,” *American Economic Journal: Microeconomics*, 2010, 2 (4), 221–243.
- , — , **and Georg Weizsäcker**, “An Experiment on Social Mislearning,” *Working Paper*, 2015.
- Fiedler, Klaus and Peter Juslin**, *Information Sampling and Adaptive Cognition*, Cambridge University Press, 2006.
- Fischbacher, Urs**, “z-Tree: Zurich Toolbox for Ready-Made Economic Experiments,” *Experimental Economics*, 2007, 10 (2), 171–178.
- Fragiadakis, Daniel E, Daniel T Knoepfle, and Muriel Niederle**, “Who is Strategic?,” Technical Report, Working Paper. 1.1 2016.
- Fudenberg, Drew**, “Advancing Beyond “Advances in Behavioral Economics”,” *Journal of Economic Literature*, 2006, 44 (3), 694–711.
- Gabaix, Xavier**, “A Sparsity-Based Model of Bounded Rationality,” *Quarterly Journal of Economics*, 2014, 129 (4), 1661–1710.
- Gagnon-Bartsch, Tristan, Matthew Rabin, and Joshua Schwartzstein**, “Channeled Attention and Stable Errors,” *Working Paper*, 2018.
- Gennaioli, Nicola and Andrei Shleifer**, “What Comes to Mind,” *Quarterly Journal of Economics*, 2010, 125 (4), 1399–1433.
- Graeber, Thomas**, “Inattentive Inference,” *Working Paper*, 2018.
- Grether, David M.**, “Bayes Rule as a Descriptive Model: The Representativeness Heuristic,” *Quarterly Journal of Economics*, 1980, 95, 537–557.
- Grimm, Veronika and Friederike Mengel**, “An Experiment on Belief Formation in Networks,” *Working Paper*, 2014.
- Han, Bing and David Hirshleifer**, “Visibility Bias in the Transmission of Consumption Norms and Undersaving,” *Working paper*, 2015.
- Heidhues, Paul, Botond Koszegi, and Philipp Strack**, “Unrealistic expectations and misguided learning,” *Working Paper*, 2017.

- Jackson, Matthew O.**, “The Friendship Paradox and Systematic Biases in Perceptions and Social Norms,” *Working Paper*, 2016.
- Jehiel, Philippe**, “Analogy-based expectation equilibrium,” *Journal of Economic theory*, 2005, 123 (2), 81–104.
- , “Investment strategy and selection bias: An equilibrium perspective on overoptimism,” *American Economic Review*, 2018, 108 (6), 1582–97.
- Jin, Ginger, Mike Luca, and Daniel Martin**, “Is No News Perceived as Good News? An Experimental Investigation of Information Disclosure,” *Working Paper*, 2018.
- Juslin, Peter, Anders Winman, and Patrik Hansson**, “The naive intuitive statistician: a naive sampling model of intuitive confidence intervals.,” *Psychological review*, 2007, 114 (3), 678.
- Kahneman, Daniel**, *Thinking, Fast and Slow*, Macmillan, 2011.
- and **Dale T Miller**, “Norm theory: Comparing reality to its alternatives.,” *Psychological review*, 1986, 93 (2), 136.
- Koehler, Jonathan J. and Molly Mercer**, “Selection Neglect in Mutual Fund Advertisements,” *Management Science*, 2009, 55 (7), 1107–1121.
- Levy, Gilat and Ronny Razin**, “Segregation in Schools, the Echo Chamber Effect, and Labour Market Discrimination,” *Working Paper*, 2015.
- Manski, Charles F.**, “Measuring expectations,” *Econometrica*, 2004, 72 (5), 1329–1376.
- Martínez-Marquina, Alejandro, Muriel Niederle, and Emanuel Vespa**, “Probabilistic States versus Multiple Certainties: The Obstacle of Uncertainty in Contingent Reasoning,” Technical Report, National Bureau of Economic Research 2017.
- Ngangoue, Kathleen and Georg Weizsäcker**, “Learning from Unrealized Versus Realized Prices,” *Working Paper*, 2015.
- Rubinstein, Ariel**, “Instinctive and Cognitive Reasoning: A Study of Response Times,” *Economic Journal*, 2007, 117 (523), 1243–1259.
- , “A Typology of Players: Between Instinctive and Contemplative,” *Quarterly Journal of Economics*, 2016, 131 (2), 859–890.
- Schwartzstein, Joshua**, “Selective Attention and Learning,” *Journal of the European Economic Association*, 2014, 12 (6), 1423–1452.

Spiegler, Ran, “Bayesian Networks and Boundedly Rational Expectations,” *Quarterly Journal of Economics*, 2016.

—, ““Data Monkeys”: A Procedural Model of Extrapolation from Partial Statistics,” *The Review of Economic Studies*, 2017, 84 (4), 1818–1841.

Tversky, Amos and Daniel Kahneman, “Availability: A heuristic for judging frequency and probability,” *Cognitive psychology*, 1973, 5 (2), 207–232.

A Robustness Checks

This Appendix reports two further sets of robustness checks to show that the results are neither driven by outliers nor by sample exclusion criteria. Table 8 replicates all treatment comparisons reported in the main text, except that the regressions are estimated using OLS and the sample includes all observations, including extreme outliers.

Second, Table 9 provides an additional set of robustness checks. Here, in all specifications, the sample is restricted to beliefs in tasks where the respective subjects stated the “correct prior,” i.e., in which the subject’s first binary guess followed the private signal. Again, the results are very similar.

Table 8: Robustness: OLS regressions on full sample

	Dependent variable: Neglect $\hat{\chi}_i^j$					
	OLS regressions: Full sample					
	(1)	(2)	(3)	(4)	(5)	(6)
0 if <i>Control</i> , 1 if <i>Selected</i>	0.56*** (0.10)	0.66*** (0.11)				
0 if <i>Complex</i> or <i>Few</i> , 1 if <i>Simple</i>			-0.32*** (0.10)	-0.32*** (0.10)		
0 if <i>Complex</i> or <i>Simple</i> , 1 if <i>Few</i>			-0.28** (0.12)	-0.46*** (0.13)		
0 if <i>Selected repl.</i> , 1 if <i>Nudge</i>					-0.18* (0.10)	-0.19** (0.09)
Session FE	Yes	No	Yes	Yes	Yes	Yes
Task FE \times prior	No	Yes	No	Yes	No	Yes
Controls	No	Yes	No	Yes	No	Yes
Observations	894	894	1177	1177	1174	1174
R^2	0.08	0.11	0.02	0.13	0.02	0.12

Notes. OLS estimates, robust standard errors (clustered at subject level) in parentheses. The dependent variable is the neglect $\hat{\chi}_i^j$ that is implied in a given belief. In columns (1)–(2), the sample includes all observations from treatments *Control* and *Selected*. In columns (3)–(4), the sample includes all observations from treatments *Complex*, *Simple*, and *Few*. In columns (5)–(6), the sample includes all observations from treatments *Nudge* and *Selected replication*, and in columns (7)–(8) all observations from treatments *Selected replication* and *Endogenous*. Controls include gender, high school grades, and log monthly disposable income. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 9: Robustness: Restricting sample to beliefs with a “correct prior”

	Dependent variable: Neglect $\hat{\chi}_i^j$											
	Median regression (1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
0 if <i>Control</i> , 1 if <i>Selected</i>	0.40*** (0.09)	0.50*** (0.11)	0.52*** (0.09)	0.57*** (0.09)								
0 if <i>Complex</i> or <i>Few</i> , 1 if <i>Simple</i>					-0.32** (0.15)	-0.30*** (0.11)	-0.30*** (0.09)	-0.30*** (0.10)				
0 if <i>Complex</i> or <i>Simple</i> , 1 if <i>Few</i>					-0.32** (0.15)	-0.28** (0.12)	-0.26*** (0.08)	-0.32*** (0.09)				
0 if <i>Selected repl.</i> , 1 if <i>Nudge</i>									-0.40*** (0.12)	-0.20** (0.08)	-0.23*** (0.09)	-0.24*** (0.09)
Session FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Task FE \times prior	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Controls	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Observations	831	831	831	831	1062	1062	1062	1062	1116	1116	1116	1116
R ²	0.07	0.07	0.09	0.10	0.03	0.04	0.05	0.07	0.02	0.05	0.03	0.07

Notes. Regression estimates, robust standard errors (clustered at subject level) in parentheses. The dependent variable is the neglect $\hat{\chi}_i^j$ that is implied in a given belief. In all specifications, the sample is restricted to beliefs in tasks in which a subject stated a first guess that followed the first signal. In columns (1)–(4), the sample includes treatments *Control* and *Selected*. In columns (5)–(8), the sample includes treatments *Complex*, *Simple*, and *Few*. In columns (9)–(12), the sample includes from treatments *Nudge* and *Selected replication*. Controls include gender, high school grades, and log monthly disposable income. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

B Additional Tables

Table 10: Treatments *Selected* and *Control*: Using $b_{N,2}$ as neglect benchmark

	Dependent variable: Neglect $\hat{\chi}_i^j$					
	Median regression		OLS winsorized		OLS trimmed	
	(1)	(2)	(3)	(4)	(5)	(6)
0 if <i>Control</i> , 1 if <i>Selected</i>	0.67*** (0.11)	0.74*** (0.10)	0.57*** (0.10)	0.62*** (0.10)	0.67*** (0.11)	0.73*** (0.11)
Session FE	Yes	Yes	Yes	Yes	Yes	Yes
Task FE \times prior	No	Yes	No	Yes	No	Yes
Controls	No	Yes	No	Yes	No	Yes
Observations	894	894	894	894	845	845
R^2	0.08	0.11	0.09	0.11	0.10	0.14

Notes. Regression estimates, with robust standard errors (clustered at subject level) in parentheses. The dependent variable is the neglect $\hat{\chi}_i^j$ that is implied in a given belief, where this benchmark is computed using $b_{N,2}$ rather than $b_{N,1}$, see the discussion in Section 2.2. The sample includes each of subjects' eight beliefs in the *Selected* and *Control* conditions. Columns (1)–(2) report median regressions, and columns (3)–(6) are OLS regressions. In columns (3)–(4), the dependent variable is winsorized at $|\hat{\chi}_i^j| = 3$. In columns (5)–(6), the sample is trimmed at $|\hat{\chi}_i^j| = 3$. Controls include gender, high school grades, and log monthly disposable income. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 11: Correlates of neglect in treatments *Selected*, *Sequential*, and *Feedback*

	Dependent variable: Neglect $\hat{\chi}_i^j$					
	Median regression				OLS wins.	OLS trimmed
	(1)	(2)	(3)	(4)	(5)	(6)
High school grades [z-score]	-0.13*** (0.04)		-0.10*** (0.04)	-0.085** (0.03)	-0.071** (0.03)	-0.060** (0.03)
Response time [min.]		-0.25*** (0.04)	-0.22*** (0.04)	-0.24*** (0.06)	-0.22*** (0.04)	-0.18*** (0.04)
Treatment FE	Yes	Yes	Yes	Yes	Yes	Yes
Session FE	Yes	Yes	Yes	Yes	Yes	Yes
Task FE \times prior	No	No	No	Yes	Yes	Yes
Controls	No	No	No	Yes	Yes	Yes
Observations	2236	2230	2230	2230	2230	2148
R^2	0.02	0.03	0.03	0.16	0.12	0.07

Notes. Regression estimates with robust standard errors (clustered at subject level) in parentheses. The sample includes treatments *Selected*, *Sequential*, and *Feedback*. Columns (1)–(2) and (5)–(6) report median regressions, and all other columns OLS regressions. In column (5), $|\hat{\chi}_i^j|$ is winsorized at 3. In column (6), the sample is trimmed at $|\hat{\chi}_i^j| = 3$. Controls include gender and log monthly disposable income. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 12: Overview of the experimental tasks in treatments *Complex* and *Simple*

True State	First signal	Observed Signal A	Observed Signal B	Observed Signal C	Observed Signal D	Unobs. Signal E	Unobs. Signal F	Bayesian Belief	Neglect Belief
103.67	136	128	148	70	–	70	70	103.67	120.50
109.33	150	104	148	114	–	70	70	109.33	129.00
99.33	70	70	70	136	–	114	136	101.00	86.50
90.67	114	150	70	70	–	70	70	90.67	101.00
98.33	148	104	128	70	–	70	70	98.33	112.50
109.67	70	70	70	148	–	150	150	103.00	89.50
122.00	114	150	136	148	114	70	–	122.00	132.40
90.67	128	136	70	70	–	70	50	90.67	101.00

Notes. Overview of the belief formation tasks in treatment *Complex* in order of appearance. The signals in treatment *Simple* are obtained by replacing each signal that is larger than 100 by 130. The categorization into observed and unobserved signals applies to the case in which subjects follow their first signal, i.e., guess ≥ 100 if their signal was larger than 100, and ≤ 100 otherwise. Subjects in the *Complex* and *Simple* treatments observed only their own signal as well as the “observed” signals. See Section 2.2 for a derivation of the Bayesian and neglect benchmarks.

Table 13: Overview of the experimental tasks in treatment *Few*

True State	First signal	Observed Signal A	Unobs. Signal B	Bayesian Belief	Neglect Belief
132.00	136	128		132.00	132.00
110.00	150		70	110.00	150.00
100.00	70.00		136	100.00	70.00
92.00	114		70	92.00	114.00
109.00	148		70	109.00	148.00
70.00	70	70		70.00	70.00
132.00	114	150		132.00	132.00
99.00	128		70	99.00	128.00

Notes. Overview of the belief formation tasks in treatment *Few* in order of appearance. The categorization into observed and unobserved signals applies to the case in which subjects follow their first signal, i.e., guess ≥ 100 if their signal was larger than 100, and ≤ 100 otherwise. Subjects in the *Complex* and *Simple* treatments observed only their own signal as well as the “observed” signals. See Section 2.2 for a derivation of the Bayesian and neglect benchmarks.

Table 14: Manipulation checks for treatments *Complex*, *Simple*, and *Few*

	Response time [min.]		Dependent variable:			
			Decision noise $ \hat{\epsilon}_i^j $			
	OLS		Median regression	OLS wins.	OLS trimmed	
	(1)	(2)	(3)	(4)	(5)	(6)
0 if <i>Complex</i> or <i>Few</i> , 1 if <i>Simple</i>	-0.38*** (0.09)	-0.38*** (0.08)	-0.28*** (0.06)	-0.20*** (0.05)	-0.071 (0.07)	-0.077 (0.06)
0 if <i>Complex</i> or <i>Simple</i> , 1 if <i>Few</i>	-0.51*** (0.08)	-0.49*** (0.08)	-0.28*** (0.06)	-0.20*** (0.05)	-0.23*** (0.07)	-0.26*** (0.06)
Session FE	Yes	Yes	Yes	Yes	Yes	Yes
Task FE \times prior	No	Yes	No	Yes	Yes	Yes
Controls	No	Yes	No	Yes	Yes	Yes
Observations	1177	1177	1177	1177	1177	1138
R^2	0.15	0.21	0.00	0.15	0.19	0.09

Notes. Regression estimates with robust standard errors (clustered at subject level) in parentheses. The sample includes treatments *Complex*, *Simple*, and *Few*. By the design of the experiment, the sample is restricted to those tasks in which following the first signal implies a first guess above 100. In treatment *Few*, experimental tasks in which subjects observe both signals are necessarily excluded because there is no scope for neglecting selection. Columns (3)–(4) report median regressions, and all other columns OLS regressions. In column (5), $\hat{\epsilon}_i^j$ is computed after $|\hat{\chi}_i^j|$ is winsorized at 3. In column (6), the sample excludes observations with $|\hat{\chi}_i^j| > 3$. Controls include gender, high school grades, and log monthly disposable income. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

C Additional Figures

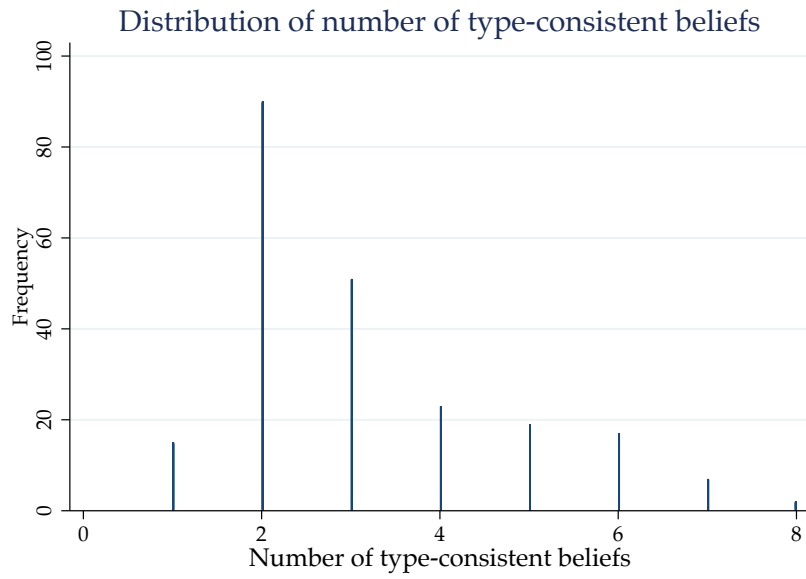


Figure 5: Distribution of number of type-consistent beliefs across subjects in treatments *Selected*, *Sequential* and *Feedback*. A belief j is type-consistent if $|\hat{\chi}_i - \hat{\chi}_i^j| \leq 0.05$.

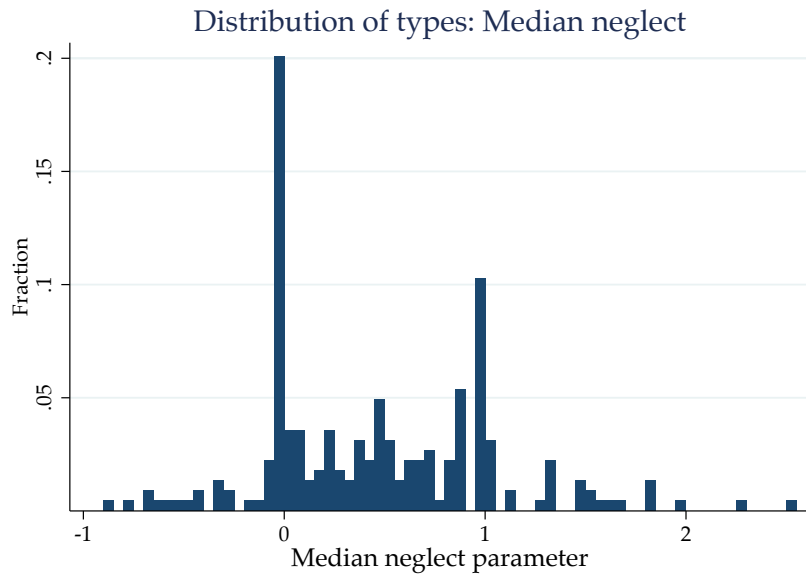


Figure 6: Distribution of neglect types $\hat{\chi}_i$ in treatments *Selected*, *Sequential* and *Feedback*. A subject's type is computed as median of $\hat{\chi}_i^j$ across tasks j .

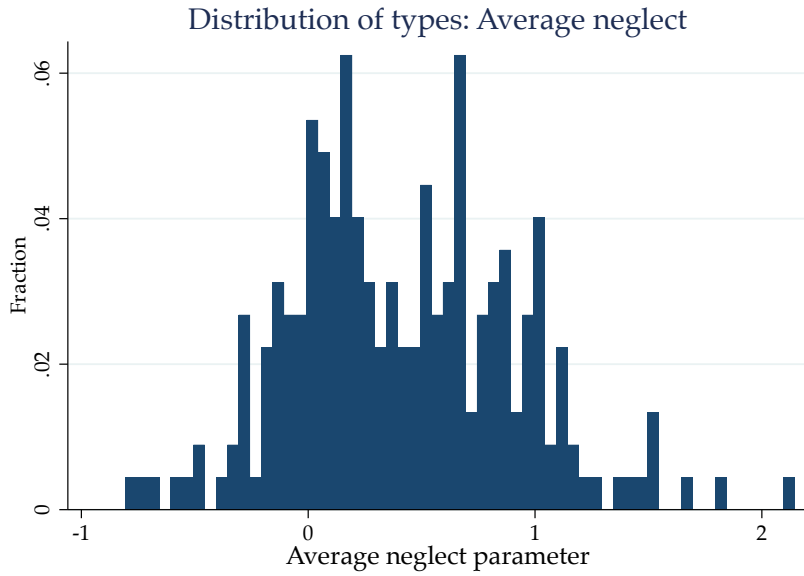


Figure 7: Distribution of neglect types $\hat{\chi}_i$ in treatments *Selected*, *Sequential* and *Feedback*. A subject's type is computed as mean of $\hat{\chi}_i^j$ across tasks j , where the distribution of $\hat{\chi}_i^j$ is trimmed at $|\hat{\chi}_i^j|=3$.

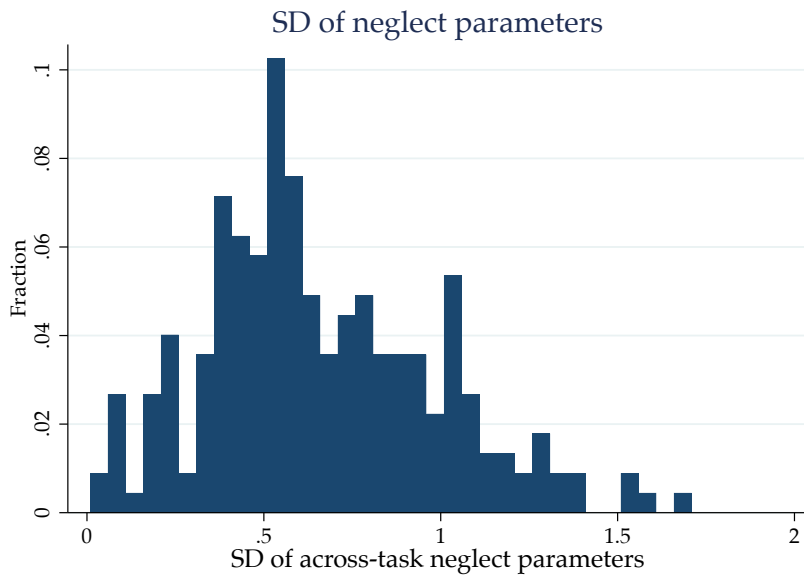


Figure 8: Subject-level standard deviation of implied neglect parameters $\hat{\chi}_i^j$ across tasks in treatments *Selected*, *Sequential* and *Feedback*. Before computing standard deviations, the distribution of $\hat{\chi}_i^j$ is trimmed at $|\hat{\chi}_i^j|=3$.

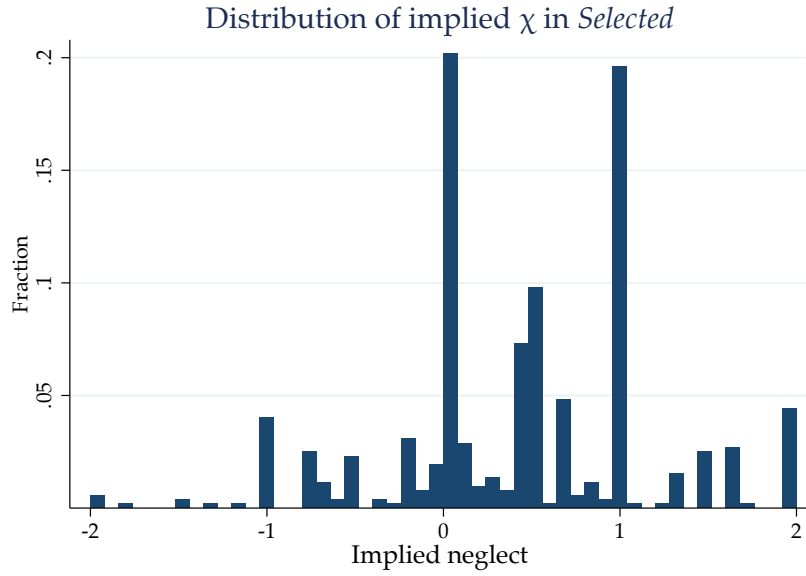


Figure 9: Distribution of implied $\hat{\chi}_i^j$ in treatment *Selected* in separate tasks. The figure plots the raw beliefs data across all subjects and tasks, normalized into units of χ according to eq. (7). That is, the data are not aggregated or rounded in any way. To ease readability, the plot excludes (i) beliefs from tasks in which a subject's first guess contradicted their private signal and (ii) beliefs with $|\hat{\chi}_i^j| > 2$.

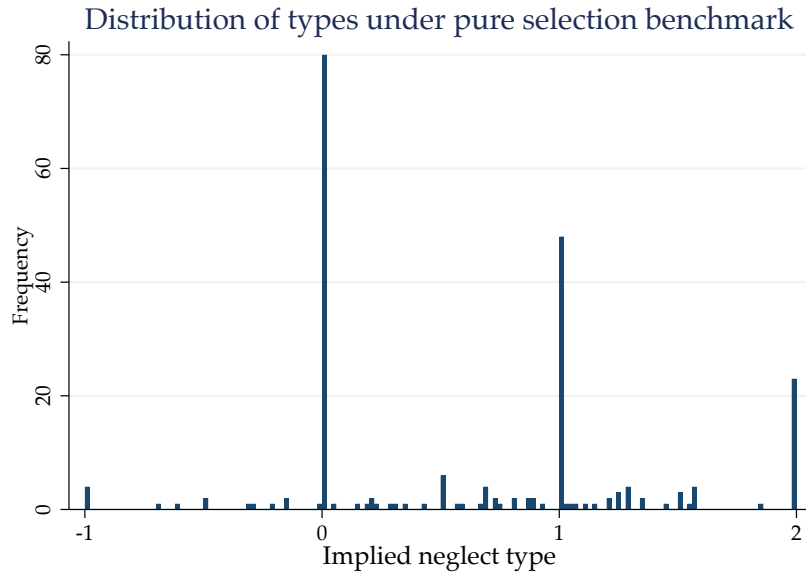


Figure 10: Distribution of neglect types $\hat{\chi}_i$ in treatments *Selected*, *Sequential*, and *Feedback*. A subject's type is determined based on the following procedure: for each subject i and candidate type $t \in \{-1, -0.9, \dots, 2\}$, I count how many $\hat{\chi}_i^j$ satisfy $|t - \hat{\chi}_i^j| \leq 0.05$. Then, I classify each subject as that candidate type that rationalizes the largest number of beliefs. Here, the neglect benchmark is given by $b_{N,2}$, see the discussion in Section 2.2.

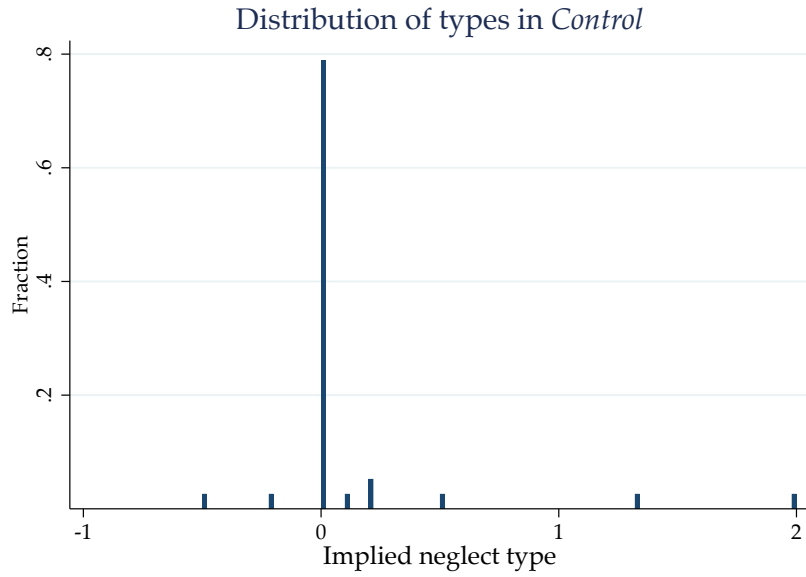


Figure 11: Distribution of neglect types $\hat{\chi}_i$ in treatment *Control*. A subject's type is determined based on the following procedure: for each subject i and candidate type $t \in \{-1, -0.9, \dots, 2\}$, I count how many of the implied $\hat{\chi}_i^j$ satisfy $|t - \hat{\chi}_i^j| < 1/20$. Then, I classify each subject as that candidate type that rationalizes the largest number of beliefs. Each $\hat{\chi}_i$ is computed using the same procedure as for treatment *Selected*, i.e., using equation (7).

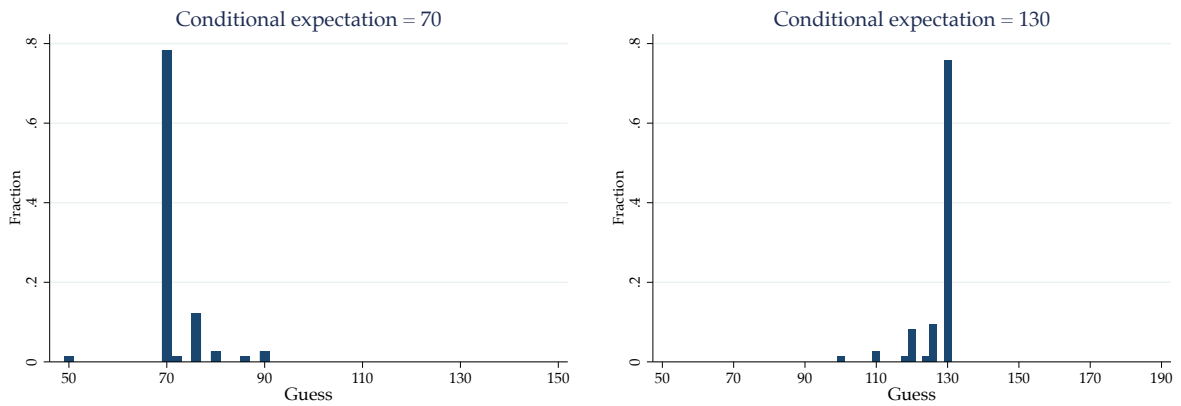


Figure 12: Guesses in conditional expectation tasks. The left panel presents the distribution of responses to a follow-up question in which subjects were asked to estimate the average of ten random draws, all of which are smaller than 100. The right panel follows an analogous logic, except that all ten random draws are above 100.

D Treatment *Sequential*

In treatment *Sequential*, 75 subjects went through the same procedures as those in *Selected*, except that each of the eight tasks was presented in a separate round / on a separate decision screen. This treatment was randomized along with treatments *Selected* and *Control* within the same experimental sessions.

Table 15 presents the results on the treatment comparison between *Sequential* and *Control*. Regardless of the regression specification, the treatment dummy is quantitatively large and statistically highly significant. In fact, the point estimates are similar to those in the treatment comparison between *Selected* and *Control*, compare Table 5 in the main text.

Table 15: Treatment Sequential

	Dependent variable: Neglect $\hat{\chi}_i^j$											
	Sequential vs. Control						Sequential vs. Selected					
	Median regression (1)	OLS winsorized (2)	OLS winsorized (3)	OLS winsorized (4)	OLS trimmed (5)	OLS trimmed (6)	Median regression (7)	Median regression (8)	OLS winsorized (9)	OLS winsorized (10)	OLS trimmed (11)	OLS trimmed (12)
0 if Control, 1 if Sequential	0.20*** (0.06)	0.25*** (0.09)	0.49*** (0.10)	0.53*** (0.11)	0.45*** (0.10)	0.45*** (0.10)						
0 if Selected, 1 if Sequential							-0.25* (0.13)	-0.15 (0.09)	-0.066 (0.09)	-0.084 (0.08)	-0.082 (0.09)	-0.099 (0.08)
Session FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Task FE \times prior	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Controls	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Observations	902	902	902	902	883	883	1188	1188	1188	1188	1167	1167
R^2	0.07	0.06	0.08	0.13	0.09	0.13	0.00	0.04	0.01	0.06	0.01	0.05

Notes. Regression estimates, robust standard errors (clustered at subject level) in parentheses. The dependent variable is the neglect $\hat{\chi}_i^j$ that is implied in a given belief. The sample includes each of subjects' eight beliefs in the *Sequential* and *Control* conditions, i.e., eight beliefs per subject. Columns (1)–(2) report median regressions, and columns (3)–(6) OLS regressions. In columns (3)–(4), the dependent variable is winsorized at |3|. In columns (5)–(6), the sample excludes $|\hat{\chi}_i^j| > 3$. Controls include gender, high school grades, and log monthly disposable income. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

E Treatment *Feedback*

In treatment *Feedback*, 75 new subjects went through a procedure that was very similar to that in *Selected*, except for two differences:

1. Before subjects completed the same eight tasks as those in *Selected*, they were asked to solve additional six tasks. Thus, in total, subjects worked on 14 tasks, spread over seven rounds. The six “new” tasks were of the same type as the other ones, they just had different signal realizations. These additional six tasks were meant to provide subjects with the possibility to receive feedback before they entered the tasks on which we compare behavior to treatment *Selected*.
2. After each round, subjects received feedback about their performance. This feedback included: (i) subjects were reminded of their continuous belief statement; (ii) they were informed of the corresponding true state; and (iii) they received information on the profits that would result from the respective task in case it would be randomly selected for payment.

Table 16 summarizes the results of this treatment by comparing belief patterns with those in treatment *Selected*. Throughout, to keep the comparison meaningful, the analysis focuses on those eight tasks that were identical across *Selected* and *Feedback*. Note that subjects in *Feedback* had already gathered feedback on six tasks before working on these eight tasks.

The table shows that the coefficient on the treatment dummy is statistically insignificant and very small in magnitude across all specifications. This provides evidence that, in terms of levels of beliefs across all tasks, there are no discernible differences between *Selected* and *Feedback*.

Second, the table also investigates changes in belief patterns over the course of the eight tasks (four rounds). In particular, it is conceivable that subjects in *Feedback* develop more rational beliefs over time, relative to those in *Selected*. To investigate this, the regressions in columns (3), (6), and (9) include an interaction term between the period (round) of the experimental task and the treatment condition. If feedback induced subjects to learn over the course of the experiment, then this coefficient should be negative. However, as the results show, the coefficient estimate is always very close to zero and statistically insignificant. Very similar results hold when I investigate learning over time by interacting the treatment dummy with a dummy for the last period, compare columns (4), (7), and (10). Thus, overall, there is no evidence that subjects learn over the course of the experiment.

Table 17 compares beliefs in *Feedback* with those in *Control*. Again, the analysis focuses on those tasks that both treatments share in common. Throughout, the treatment

Table 16: Comparing treatments *Feedback* and *Selected*

		Dependent variable: Neglect $\hat{\chi}_i^j$									
		Median regressions				OLS winsorized			OLS trimmed		
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
0 if <i>Selected</i> , 1 if <i>Feedback</i>		0 (0.10)	-0.12 (0.10)	-0.051 (0.23)	-0.097 (0.10)	-0.012 (0.08)	0.026 (0.23)	-0.011 (0.09)	-0.026 (0.08)	-0.13 (0.21)	-0.016 (0.08)
# of period				-0.017 (0.04)			0.038 (0.03)			0.010 (0.03)	
# of period \times 1 if <i>Feedback</i>				-0.0083 (0.05)			-0.0069 (0.04)			0.020 (0.04)	
1 if last period					0.12 (0.08)			-0.025 (0.07)			0.020 (0.06)
1 if last period in <i>Feedback</i>					0.016 (0.10)			-0.0027 (0.09)			-0.039 (0.08)
Controls	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1190	1190	1190	1190	1190	1190	1190	1190	1167	1167	1167
R^2	.	0.02	0.02	0.02	0.02	0.03	0.03	0.03	0.03	0.03	0.03

Notes. Regression estimates, robust standard errors (clustered at subject level) in parentheses. The dependent variable is the neglect $\hat{\chi}_i^j$ that is implied in a given belief. The sample includes each of subjects' eight beliefs in the *Sequential* and *Control* conditions, i.e., eight beliefs per subject. Columns (1)–(4) report median regressions, and columns (5)–(10) OLS regressions. In columns (5)–(7), the dependent variable is winsorized at $|\hat{\chi}_i^j| = 3$. In columns (8)–(10), the sample excludes $|\hat{\chi}_i^j| > 3$. Controls include gender, high school grades, and log monthly disposable income. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 17: Comparing treatments *Feedback* and *Control*

	Dependent variable: Neglect $\hat{\chi}_i^j$					
	Median regression		OLS winsorized		OLS trimmed	
	(1)	(2)	(3)	(4)	(5)	(6)
0 if <i>Control</i> , 1 if <i>Feedback</i>	0.40*** (0.09)	0.40*** (0.12)	0.57*** (0.10)	0.62*** (0.10)	0.53*** (0.09)	0.56*** (0.09)
Task FE \times prior	No	Yes	No	Yes	No	Yes
Controls	No	Yes	No	Yes	No	Yes
Observations	904	904	904	904	883	883
R^2	0.08	0.07	0.09	0.14	0.09	0.12

Notes. Regression estimates, robust standard errors (clustered at subject level) in parentheses. The dependent variable is the neglect $\hat{\chi}_i^j$ that is implied in a given belief. The sample includes each of subjects' eight beliefs in the *Sequential* and *Control* conditions, i.e., eight beliefs per subject. Columns (1)–(2) report median regressions, and columns (3)–(6) OLS regressions. In columns (3)–(4), the dependent variable is winsorized at $|\hat{\chi}_i^j| = 3$. In columns (5)–(6), the sample excludes $|\hat{\chi}_i^j| > 3$. Controls include gender, high school grades, and log monthly disposable income. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

dummy is statistically significant and comparable to the results in the baseline analysis in the main paper.

F Treatment *Random*

Table 16 summarizes the results of this treatment by comparing stated beliefs with a Bayesian and full neglect benchmark, which is computed as described in the main text. The table shows that median and average beliefs are always distorted away from the Bayesian benchmark in the direction of full neglect.

Treatments *Random* and *Selected* are not directly comparable because the true state was generated as average of six random draws in *Selected* but of eight draws in *Random*. Still, both of these treatments allow to compute the same task-level implied neglect parameters χ_i^j that reflect the location of a belief on the spectrum between full rationality and averaging the visible data. In a tentative analysis, Table 19 shows that implied neglect parameters in *Selected* and *Random* are statistically indistinguishable from each other.

In summary, all patterns reported here show that treatment *Random* produces neglect patterns that are comparable to those in *Selected*. This suggests that the reason for why subjects follow a “what you see is all there is” averaging heuristic is not the conceptual difficulty of selection but rather a more general tendency to estimate the population mean through the mean of the visible sample.

Table 18: Overview of the experimental tasks and results in treatment *Random*

First Signal	Observed A	Observed B	Observed C	Observed D	Unobs. E	Unobs. F	Unobs. G	Bayesian Belief	Neglect Belief	Median Belief	Average Belief
130	130	150	50	110	50	110	110	107.50	115.00	110.00	111.55
150	110	150	110	50	90	150	90	115.00	130.00	120.00	120.44
50	70	50	110	110	130	110	90	85.00	70.00	85.00	82.61
110	150	90	50	50	90	130	130	100.00	100.00	100.0	101.83
150	130	130	70	70	90	90	90	110.00	120.00	115.00	116.41
90	130	70	150	150	150	50	110	105.00	110.00	105.00	105.72
110	150	130	90	50	50	70	50	110.00	120.00	120.00	114.75
130	50	50	90	70	50	150	150	80.00	90.00	90.00	86.75

Notes. Overview of the belief formation tasks and stated beliefs in treatment *Random*. See Section 4.2 for a derivation of the Bayesian and neglect benchmarks.

Table 19: Treatments *Selected* and *Random*

	Dependent variable: Neglect $\hat{\chi}_i^j$					
	Median regression		OLS winsorized		OLS trimmed	
	(1)	(2)	(3)	(4)	(5)	(6)
0 if <i>Selected</i> , 1 if <i>Random</i>	-0.067 (0.18)	-0.069 (0.09)	-0.12 (0.09)	-0.13 (0.09)	-0.11 (0.08)	-0.12 (0.08)
Task FE \times prior	No	Yes	No	Yes	No	Yes
Controls	No	Yes	No	Yes	No	Yes
Observations	1115	1115	1115	1115	1083	1083
R^2	0.00	0.05	0.00	0.07	0.00	0.06

Notes. Regression estimates, with robust standard errors (clustered at subject level) in parentheses. The dependent variable is the neglect $\hat{\chi}_i^j$ that is implied in a given belief. The sample includes each of subjects' eight beliefs in the *Selected* and *Random* conditions. Columns (1)–(2) report median regressions, and columns (3)–(6) are OLS regressions. In columns (3)–(4), the dependent variable is winsorized at $|\hat{\chi}_i^j| = 3$. In columns (5)–(6), the sample is trimmed at $|\hat{\chi}_i^j| = 3$. Controls include gender, high school grades, and log monthly disposable income. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

G Earlier Experiments: Replication

G.1 Overview

All treatments reported in the main text were implemented in June – September 2018. These treatments replace a set of similar experiments that I ran in 2015 and 2016, on which an earlier working paper version of this paper was based. Because the earlier experiments are very similar to the ones reported in the main text, they can be viewed as replication and robustness exercise. In particular, the earlier experiments also contained versions of treatments *Selected*, *Control*, *Nudge* (called *Saliency* in the earlier experiments), *Complex* (called *Intermediate*), and *Simple*.

The earlier experiments followed a very similar logic to the ones described in the main text. Subjects estimated an abstract true state, and received computer-generated signals that induced a selection problem of the same kind as above. The most important difference between treatment *Selected* and these earlier experiments is that, in the earlier experiments, the true state was based on 15, rather than six, random draws, so that subjects also needed to account for the base rate in processing selected signals. The new design eliminated this additional difficulty.

This Appendix summarizes the design and results of the earlier experiments. First, I describe the basic experimental design and the results on selection neglect by comparing the previous versions of treatments *Selected* and *Control*. Then, I study the role of nudges and computational complexity. Finally, I describe the results of treatment *Disagreement* (briefly mentioned in the main text), in which subjects were given an opportunity to revise their beliefs after they had observed the beliefs of two peers who faced exactly the same decision problem and information signals. For completeness, the exposition of these earlier experiments largely follows the exposition in the previous version of the paper. The previous version of the paper, including all previous experimental instructions, can also be accessed at <https://sites.google.com/site/benjaminenke>.

G.2 Experimental Design

Subjects were asked to estimate an ex ante unknown state of the world μ and were paid for accuracy. First, the computer generated μ ; to this end, the computer drew 15 times with replacement from the set $X = \{50, 70, 90, 110, 130, 150\}$. The average of these 15 draws then constituted the true state μ . Second, the computer generated six signals about the state. Let Y denote the set of 15 numbers that determine the state. The computer generated six signals s_1, \dots, s_6 by randomly drawing from Y , without

replacement. Thus, ex ante, each signal is independently and uniformly distributed over the set X .

In the course of the experiment, a subject “interacted” with five computer players (called players I–V). The experimental task consisted of multiple stages. First, after the computer randomly generated the true state and the signals, a subject as well as each of the five computer players privately observed one of the six signals. In the second stage, the subject and the computer players each selected into a group based on the respective signal, which introduces an information-based selection problem. In the third stage, subjects observed the signals of some of the computer players, and finally stated a belief over μ in the fourth stage.

Specifically, in the first stage, subjects received a private signal. In the second stage, they had to decide upon their group membership (blue or red group) based on their signal. The payoff structure was such that subjects earned higher profits as member of the blue group if $\mu < 100$ and of the red group provided that $\mu > 100$, i.e., profits were €12 if the subject opted for the red (blue) group when $\mu > 100$ ($\mu < 100$), and €2 otherwise. Given this payoff structure, it was rather obvious for subjects which group to enter, and I show below that subjects indeed almost always entered the red group if their private signal was larger than 100 and the blue group otherwise. The five computer players similarly decided on their group membership using a decision rule that was *known* to subjects, i.e., these players opted for the blue (red) group if their private signal was smaller (higher) than 100. After this first stage, the two groups exhibit strong assortative matching on information, with all high signals being in the red group, and all low signals being in the blue group.

In the third stage, subjects observed the signals of some of the computer players to gather additional information about the state, i.e., subjects obtained the private signals of these computer players. The only difference between the *Selected* and the *Control* treatment consisted of the information subjects received from the computer players. In the *Selected* treatment, subjects talked to all computer players in their own group, but at least with three computers. Thus, for instance, if a subject’s group contained only one computer player, they obtained the signal of that player and of two randomly chosen players from the other group. If a subject’s group contained four players, a subject observed (only) these four. It was made clear to subjects that whenever they did not talk to a particular player, it would have to be that this player entered the opposite group. Thus, subjects could easily infer the number of players in each group. Note that given the simplified discretized uniform distribution over the signal space, it was rather straightforward for subjects to infer which types of signals they were missing. This provides a crucial input into the design, because it ensures that subjects can in principle

understand the statistical properties of the signals they do not see. In particular, being sophisticated about selection requires subjects to understand that when they are in the red group, a missing signal was 70, in expectation, while it was 130 when they were in the blue group. Finally, subjects stated a belief over μ .

In the *Control* condition, participants received the same signals as subjects in the *Selected* treatment, but additionally obtained a coarse version of the signals of the computer players that subjects in the *Selected* condition did not observe. Specifically, if the signal of these additional computer players was in $\{50, 70, 90\}$, the respective player communicated 70 to the subject, while if the signal was in $\{110, 130, 150\}$, the computer communicated 130. Given that these coarse messages equal the expected signal conditional on group membership, the informational content of the *Selected* and the *Control* treatments is identical.

Subjects completed seven independent tasks without receiving feedback in between. For instance, in the first task, subjects' private signal was 130, so that the optimal choice in the first decision was to opt for the red group. Here, subjects in the *Selected* condition would meet three computer players that obtained signals 110, 90, and 70, i.e., subjects observed the signal of one player from their own red group and two from the blue group. The remaining two computer players received private signals of 50 and 90, respectively. While subjects in the *Selected* condition did not observe the signals of these players, those in the *Control* condition observed coarse versions of these signals, i.e., 70 and 70.

A comprehensive set of control questions ensured that subjects understood the process generating their data. Most importantly, subjects were asked what they knew about a computer player's private signal if they were in the red group, but did not observe the signal of that computer player, i.e., that this computer player must have obtained a private signal of less than 100 and hence opted for the blue group. Only once subjects had correctly solved all questionnaire items could they proceed to the main tasks.¹⁷ In the belief formation stage, all beliefs were restricted to be in $[0, 200]$ by the computer program.

¹⁷The control questions followed a multiple choice format, with 3–4 questions per screen. Thus, trial-and-error was very cumbersome. Moreover, the BonnEconLab has a control room in which the experimenter can monitor the decision screens of all experimental subjects. Thus, whenever a subject appeared to have problems in answering the control questions, an experimenter approached the subject, clarified open questions (if any) and excluded the subject from the experiment if they did not appear to understand the instructions. Also notice that it turns out that one of the control questions was phrased suboptimally. This question asked subjects which signal a computer player must have gotten "on average" if that signal induced the computer player to enter the red group (i.e., 130). Here, roughly 25% of subjects indicated to the experimenter that they did not understand the concept of an "average signal" given that the question asked for the signal of one particular computer player; nevertheless, all of these subjects showed a clear understanding that the signal of that computer player must have been larger than 100. Given that an incentivized follow-up question explicitly investigated subjects' ability to compute conditional expectations, subjects were allowed to continue to the experiment after the experimenter privately explained how to interpret the phrase "average signal".

The experiments were conducted at the BonnEconLab of the University of Bonn and computerized using z-Tree (Fischbacher, 2007). Participants were recruited and invited using hroot (Bock et al., 2014). 78 student subjects participated in these two treatments (48 in *Selected* and 30 in *Control*) and earned an average of € 11.60 including a € 4 show-up fee.¹⁸ After the written instructions were distributed, subjects had 15 minutes to familiarize themselves with the task. Upon completion of the control questions, subjects entered the first task. Each task consisted of two computer screens. On the first screen, subjects were informed of their private signal and decided which group to enter. On the second screen, participants received the computer players' signals and stated a point belief. Both decisions were incentivized, in expectation: in total, subjects took 14 decisions (seven on which group to enter and seven belief statements), one of which was selected for payment, which constitutes an incentive-compatible mechanism in such a setup (Azrieli et al., 2018). The probability that a belief was randomly selected for payment was 80%, while a group membership was chosen with probability 20%. Beliefs were incentivized using a quadratic scoring rule with maximum variable earnings of € 18, i.e., variable earnings in a given task j equalled $\pi^j = \max\{0; 18 - 0.2 \times (b^j - t^j)^2\}$, where b denotes the belief and t the state. Across tasks, the average financial incentives to hold sophisticated (relative to fully naïve) beliefs were roughly € 12. Payments for the group entrance decision were € 12 if the subject opted for the red (blue) group when $\mu > 100$ ($\mu < 100$), and € 2 otherwise.

G.3 Baseline Hypothesis

Given true state $\mu = \sum_{k=1}^{15} m_k / 15$, for $m_k \in \{50, 70, 90, 110, 130, 150\}$ with probability $1/6$ each, the signals $s_i = m_k$ for some k and $i \in \{1, \dots, 6\}$ are unbiased. Let N denote the number of signals a subject actually sees, i.e., the number of “communication” partners. Denote by g_a the group membership of computer player a , i.e., $g_a \in \{\text{red}, \text{blue}\}$. In the present setup, $E(s_i | g_a = \text{red}) = 130$ and $E(s_i | g_a = \text{blue}) = 70$. Given some signals, a Bayesian would compute the mean posterior belief b_B as

$$b_B = E[\mu] = \frac{\sum_{v=1}^N s_v + \sum_{l=N+1}^6 E[s_l | g_l] + E[m] \times 9}{15}$$

where s_v denotes an observed signal and s_l an unobserved one. The second term in the numerator denotes the expectation of a signal conditional on the signal recipient entering a certain group. The third term in the numerator reflects the base rate $E[m] = 100$. However, starting with Grether (1980), a long stream of research has

¹⁸The unbalanced treatment allocation was determined ex ante, which reflects the fact that the *Control* condition merely serves as a “straw man” with very little expected noise.

shown that people tend to neglect the base rate. I thus define an alternative “sophisticated” benchmark (in the sense of absence of selection neglect) b_R as

$$b_R = \frac{\sum_{v=1}^N s_v + \sum_{l=N+1}^6 E[s_l | g_l]}{6}. \quad (10)$$

That is, the “sophisticated” benchmark ignores the base rate, but takes into account selection. This normalization only serves to illustrate the distribution of individual-level neglect: without assuming base rate neglect, any estimator for the naïveté parameter would be severely biased if people actually neglect the base rate. The assumption of full base rate neglect will be corroborated below using data from the *Control* treatment: here, people overwhelmingly state beliefs that reflect *full* base rate neglect, but are sophisticated otherwise. Still, the assumption of full base rate neglect is *only* used to identify naïveté parameters, while all treatment comparisons are conducted on the raw data. Below, I report upon a robustness treatment in which base rate neglect does not bias the estimates of selection neglect.

Now imagine that people neglect selection, so that they merely base their beliefs on “what they see”. Let $\chi \in [0, 1]$ parameterize the degree of naïveté such that $\chi = 1$ implies full neglect. Define a neglect posterior b_{SN} as a weighted average of b_R and a fully naïve belief b_N , which consists of averaging the visible signals:

$$\begin{aligned} b_{SN} &= (1 - \chi)b_R + \chi b_N = (1 - \chi)b_R + \chi \frac{\sum_{i=1}^N s_v}{N} \\ &= b_R + \chi \frac{6 - N}{6} (\bar{s}_v - \bar{s}_l), \end{aligned} \quad (11)$$

where $\bar{s}_v \equiv 1/N \sum_{v=1}^N s_v$ is the average visible signal and $\bar{s}_l \equiv 1/(6-N-1) \sum_{l=N+1}^6 E(s_l | g_l)$ the average expected “non-visible” signal. That is, the neglect belief b_{SN} consists of the sophisticated belief plus an intuitive distortion term that depends on χ .

G.4 Results on Selection Neglect

I will frequently work with a measure of subjects’ beliefs that is independent of the specific updating task. To this end, I use the analog of equation 11 to compute the naïveté implied in each belief of subject i in belief formation task j :

$$\hat{\chi}_i^j = \frac{6(b_i^j - b_R^j)}{(6 - N)(\bar{s}_v - \bar{s}_l)}. \quad (4)$$

Using this procedure, beliefs can be directly interpreted as reflecting sophisticated ($\chi = 0$), fully naïve ($\chi = 1$), or intermediate values. The OLS regressions reported

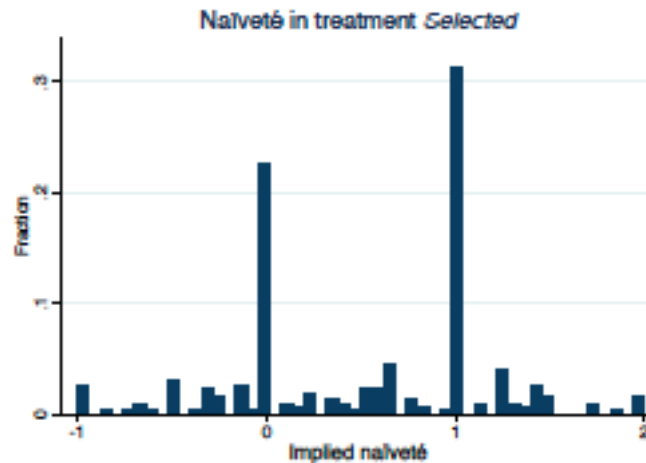


Figure 13: Distribution of naïveté in the *Selected* treatment. To ease readability, the figure excludes observations outside $\chi \in [-1, 2]$ (30 out of 336 obs.).

in columns (1) and (2) of Table 20 formally confirm that the full set of seven beliefs per subjects, expressed in units of χ , differs across treatments (the standard errors are clustered at the subject level). The large bias implies significantly lower earnings of subjects in the *Selected* condition. The expected profit from all seven belief formation tasks (i.e., the average hypothetical profit from each belief) is € 5.00 in *Selected* and € 10.50 in *Control* ($p < 0.0001$, Wilcoxon ranksum test).¹⁹ For comparison, the expected profit from being fully sophisticated in all tasks is € 12.70.

To develop a deeper understanding of subjects' precise belief patterns, I examine the distribution of estimated naïveté parameters χ . Figure 13 depicts the distribution of the implied naïveté in all separate beliefs, i.e., seven beliefs per subject. That is, the right panel plots the raw data, translated into units of naïveté, without any aggregation, rounding, or other reasons to expect beliefs to reflect one of the extreme predictions of $\chi = 0$ or $\chi = 1$. Nevertheless, the data exhibit two large spikes at *exactly* zero and one, i.e., the fully sophisticated and fully naïve benchmark. For example, more than 50% of the beliefs of all subjects with median $\chi_i > 0.5$ lie within a very small interval around the fully naïve belief, $0.95 \leq \chi_i^j \leq 1.05$. In addition, it is conceivable that this number would be even higher if we took into account that many of the beliefs close to one might reflect the same cognitive strategy plus decision noise.

Next, I examine basic correlates of biased updating within treatment *Selected*. Columns (3)–(4) of Table 20 show that participants with better high school grades (a common proxy for cognitive ability) are significantly less likely to commit neglect (Benjamin et

¹⁹Actual profits, which are partly based on group membership and include the show-up fee, are also significantly different from each other (€ 13.70 vs. € 10.10, $p = 0.0628$).

al., 2013). Columns (5) and (6) show that neglecting selection is significantly correlated with correlation neglect, measured as in Enke and Zimmermann (2019). When both high school grades and correlation neglect are inserted into the regression, the coefficient on subjects' high school grades drops in size and ceases to be significant.

Finally, I study the relationship between neglect and response times, which are often advocated for as proxy for cognitive effort in experiments (Rubinstein, 2007, 2016). In the data, the average response time across tasks and subjects in treatment *Selected* is 56 seconds. Columns (7)–(8) of Table 20 investigate the relationship between subjects' naïveté χ (as implied in each belief, see eq. 11) and the corresponding response time (in minutes). The results show that higher response times are significantly associated with less neglect. At the same time, the quantitative magnitude of this relationship is remarkably small: interpreted causally, the point estimate implies that response times would have to increase by four minutes per task to move a full neglect subject to fully sophisticated beliefs, which corresponds to roughly six standard deviations in the sample. Thus, it appears as if the relationship between response times and neglect is quantitatively much too small to be able to explain neglect purely as the result of low response times (cognitive effort).

G.5 Nudges and Computational Complexity

Design. I introduce treatments *Intermediate* and *Simple*. These experimental conditions follow the same procedures as those in *Selected*, except for one variation. Recall that in *Selected*, the true state (as well as the signals) were determined by random draws from the set $\{50, 70, 90, 110, 130, 150\}$. In *Intermediate*, this set is replaced by $\{70, 70, 70, 110, 130, 150\}$, and in *Simple* by $\{70, 70, 70, 130, 130, 130\}$.²⁰ Notice that whenever subjects' private signal is above 100, so that they enter the red group, the problem of backing out the missing observations from the blue group is both utterly simple and identical across the *Intermediate* and *Simple* treatments: subjects only need to remember that a computer player being in the blue group deterministically implies a signal of 70. That is, in both treatments, people's potential problems in computing conditional expectations cannot drive any results.

At the same time, treatment *Intermediate* is computationally more complex than *Simple* because the process of computing a (naïve) posterior from the visible signals involves averaging various different values, as opposed to mostly 130's. That is, just as required by the research hypothesis, these two treatments leave the difficulty of ac-

²⁰To implement these changes, the signal draws from *Selected* were simply replaced by the appropriate values, e.g., 50 became 70. Thus, subjects in *Intermediate* and *Simple* essentially solved the same tasks as those in *Selected*.

Table 20: Correlates of neglect

	Dependent variable: Naïveté χ								
	Selected vs. Control			Selected					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1 if Selected	0.55*** (0.09)	0.62*** (0.09)							
High school grades		-0.14*** (0.05)	-0.27*** (0.07)	-0.29*** (0.07)		-0.075 (0.09)			-0.28*** (0.07)
Correlation neglect parameter					0.29** (0.11)	0.26** (0.12)			
Response time (in minutes)							-0.25*** (0.09)	-0.24** (0.11)	-0.20* (0.10)
# of consistent beliefs									0.027 (0.05)
Constant	-0.019 (0.05)	-0.84* (0.49)	0.61*** (0.07)	0.025 (0.68)	0.34*** (0.11)	0.50 (1.11)	0.76*** (0.10)	0.054 (0.75)	0.17 (0.86)
Controls	No	Yes	No	Yes	No	Yes	No	Yes	Yes
Observations	526	526	323	323	215	215	323	323	323
R ²	0.11	0.17	0.10	0.14	0.06	0.11	0.04	0.08	0.16

Notes. OLS estimates, robust standard errors (clustered at subject level) in parentheses. In columns (1)–(2), the sample includes the naïveté implied in each of subjects' seven beliefs in the *Selected* and *Control* conditions, i.e., seven beliefs per subject. In columns (3)–(10), the sample includes subjects in treatment *Selected*. All regressions exclude extreme outliers with $|t_i^j| > 3$, but all results are robust to including these outliers. Controls include age, gender, log monthly income, and task fixed effects. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

counting for selection constant, but vary the extent to which the environment in general consumes mental resources, in particular *the extent to which people may be distracted by an aspect of the problem that is unrelated to accounting for selection*.²¹ In total, 89 subjects participated in *Intermediate* and *Simple*, which were randomized within session.

Results. In analyzing the data, I start by restricting attention to those experimental tasks in which subjects' private signal satisfies $s > 100$ so that the difficulty of backing out the missing signals is indeed identical across *Intermediate* and *Simple*. Columns (1) and (2) of Table 21 present the results of OLS estimations in which I regress the naïveté implied in subjects' beliefs (only in those tasks in which $s > 100$) on a treatment dummy, with the standard errors again clustered at the subject level. The coefficient on the dummy is large and statistically highly significant in both unconditional and conditional regressions.

Recall that the treatment comparison between *Intermediate* and *Simple* rests on the idea that the difficulty of backing out missing observations is identical as long as $s > 100$. A similar argument can be constructed for the case of $s < 100$. Here, subjects in both *Intermediate* and *Selected* had to back out missing signals from the set $\{110, 130, 150\}$, yet the difficulty of computing a fully naïve belief varies across these two conditions because subjects in *Intermediate* mostly had to process 70's as opposed to $\{50, 70, 90\}$. Accordingly, the research hypothesis would prescribe that subjects in *Selected* are more biased. Columns (3) and (4) of Table 21 report corresponding OLS regressions. As hypothesized, the point estimates are positive; at the same time, the coefficients are either only marginally significant or marginally not significant. A potential reason for the slight discrepancy between the results for the comparison *Intermediate–Selected* relative to *Intermediate–Simple* is that the mathematical steps of accounting for selection are harder in the first case, so that the data are potentially noisier.

In any case, columns (5) and (6) present a pooled analysis, in which I combine the observations from columns (1)–(4). Here, people exhibit significantly less neglect in the less complex tasks compared to the more complex ones, where again complexity is solely defined through the “distraction” of more cumbersome computations.²²

²¹Note that while the informational content of these two treatments is not identical, the differences are very small: a visible signal of 110 or 150 in *Intermediate* would turn into a 130 in *Simple*. In any case, backing out the absent observations is literally identical across conditions. Thus, by expressing all beliefs in terms of units of naïveté, we can evaluate the hypothesis that subjects in *Simple* will attend more to the absent observations and hence commit less neglect.

²²More precisely, in line with the specifications in columns (1)–(4), the complexity dummy assumes a value equal to zero if an observation is (i) from treatment *Simple* and $s > 100$, or (ii) from *Intermediate* and $s < 100$. It equals 1 if an observation is (i) from *Intermediate* and $s > 100$, or (ii) from *Selected* and $s < 100$.

Table 21: Representations, complexity and attention

	Dependent variable: Naïveté χ							
	Intermediate and Simple				Salience			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
0 if Simple, 1 if Intermediate	0.31** (0.12)	0.32*** (0.12)						
0 if Intermediate, 1 if Selected			0.17 (0.13)	0.21* (0.12)				
Pooled: 0 if low compl., 1 if high compl.					0.24** (0.10)	0.27*** (0.09)		
0 if Selected, 1 if Salience							-0.33*** (0.10)	-0.40*** (0.09)
Constant	0.29*** (0.09)	0.42 (0.50)	0.29*** (0.09)	0.73 (0.63)	0.30*** (0.08)	0.30 (0.39)	0.53*** (0.08)	0.41 (0.50)
Controls	No	Yes	No	Yes	No	Yes	No	Yes
Observations	341	341	272	272	604	604	654	654
R ²	0.04	0.13	0.01	0.06	0.02	0.11	0.04	0.11

Notes. OLS estimates, robust standard errors (clustered at subject level) in parentheses. In columns (1)–(2), the sample includes subjects in treatments *Intermediate* and *Simple*, but only beliefs in those experimental tasks in which subjects' private signal was above 100 (four beliefs per subject). In columns (3)–(4), the sample includes subjects in treatments *Intermediate* and *Selected*, but only beliefs in those experimental tasks in which subjects' private signal was below 100 (three beliefs per subject). In columns (5)–(6), the complexity dummy assumes a value equal to zero if an observation is (i) from treatment *Simple* and $s > 100$, or (ii) from *Intermediate* and $s < 100$. It equals 1 if an observation is (i) from *Intermediate* and $s > 100$, or (ii) from *Selected* and $s < 100$. In columns (7)–(8), the sample includes the naïveté implied in each of subjects' seven beliefs in the *Selected* and *Salience* condition. All regressions exclude extreme outliers with $|\hat{\chi}_i^j| > 3$, but all results are robust to including these outliers when employing median regressions. Controls include age, gender, high school grades, log monthly income, and task fixed effects. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

G.6 Disagreement

Hypothesis. While treatment *Salience* documented that shifting subjects' attention can have large effects on their beliefs, such *direct* attention manipulations are rare in practice. Instead, more natural contexts are likely to provide *indirect* hints that might induce people to reconsider their updating rule. A prime example is the presence of disagreement. After all, people are often exposed to the beliefs of others, and this may induce people to question their original strategy, and notice the selection problem.

Design. In treatment *Disagreement*, a new set of subjects solved the seven belief formation tasks from the *Selected* treatment reported above. The new treatment consisted of two parts, as illustrated by Table 22. In part one, subjects solved the first three belief formation tasks (without feedback). This allows me to compute an out-of-sample measure of subjects' type χ .

In part two, subjects solved the remaining four tasks. Here, similarly to treatment *Selected*, subjects received a private signal and were allocated to the red or blue group depending on whether their signal was above or below 100.²³ Then, subjects stated a belief. Afterwards, they were shown the beliefs of two other randomly drawn subjects ("neighbors") from the same session.²⁴ Importantly, all subjects not only solved the same tasks, they also *received the same private signal and observed the signals of the same computer players*. The written instructions placed heavy emphasis on the presence of identical information and a verbal summary was read out aloud to induce common knowledge. After subjects observed the beliefs of their neighbors, they were asked to state a second belief.²⁵ Subjects did not receive feedback between the different tasks, except for observing the beliefs of their neighbors. Subjects' decisions were financially incentivized such that either part one or part two of the experiment was drawn for payout with probability 50% each; conditional on either part being drawn, one of the respective decisions was implemented, just like in the baseline treatments.

Results. For the purposes of the empirical analysis, I again normalize the data across tasks by computing the naïveté χ that is implied by each belief and then pool the data across tasks and subjects. First note that the structure of the belief distribution in this

²³In these four tasks, subjects did not decide on their group membership. Rather, the computer allocated them into the red (blue) group when their private signal was higher (lower) than 100. This was done to ensure that subjects indeed had identical information.

²⁴This random matching was not constant across tasks.

²⁵The experimental procedures paid special attention to preserving anonymity between subjects to eliminate confounding effects of image concerns as arising from people feeling uncomfortable with stating and revising their beliefs in public.

Table 22: Basic timeline of treatment *Disagreement*

Part 1		Part 2		
Stage 0 – 4	Stage 0 – 3	Stage 4	Stage 5	Stage 6
As in <i>Selected</i> treatment	As in <i>Selected</i> , except that subjects do not choose their group membership, but rather get allocated depending on whether $s > 100$	Belief elicitation	Observe beliefs of two neighbors	Belief elicitation

Notes. Timeline of the treatments involving disagreement. In the first part, subjects completed three tasks from the *Selected* treatment. In the second part, they completed four additional tasks. Here, subjects again observed a private signal and were then allocated into the red and blue group according to their signal. Then, they observed the signals of a subset of the computer players as in *Selected*. After subjects stated a belief, they were shown the beliefs of two other subjects and then again stated a belief. Subjects did not receive any feedback between the different experimental tasks, except for observing the beliefs of their neighbors.

treatment is again bimodal with subjects being either fully naïve or sophisticated about the selection problem.

I investigate how subjects revised their beliefs as a function of their updating type. After all, sophisticated and neglect types may differ in how they respond to disagreement. To construct a measure of how much subjects revise their beliefs, I compute the difference between the beliefs subjects stated before and after observing the beliefs of their neighbors, expressed as percentage of the pre-communication disagreement (measured as simple difference between the subject’s pre-communication belief and the two neighbors’ average pre-communication belief):

$$\text{Belief revision of subject } i = \frac{\chi_i^2 - \chi_i^1}{\bar{\chi}_{-i}^1 - \chi_i^1} \times 100,$$

where $\bar{\chi}_{-i}^1$ denotes the average belief (naïveté) of i ’s two neighbors in their first belief statements. Thus, the belief revision measure quantifies by how much subjects altered their belief, relative to how much they could have changed their beliefs given the neighbors’ reports and their own first belief. Note that this belief revision measure takes into account that subjects might be confronted with zero, one, or two beliefs that substantially differ from their own assessment of the evidence.

Figure 14 presents histograms of subjects’ belief revisions as a consequence of the neighbors’ reports. To make matters interesting, I restrict attention to cases in which a subject’s first belief does not equal the average belief of the two neighbors. To visualize the results, I partition subjects into sophisticates and naïfs according to whether their out-of-sample median naïveté parameter from the first part of the experiment satisfies $\chi \leq 0.5$. The figure reveals that participants largely abstain from adjusting their beliefs

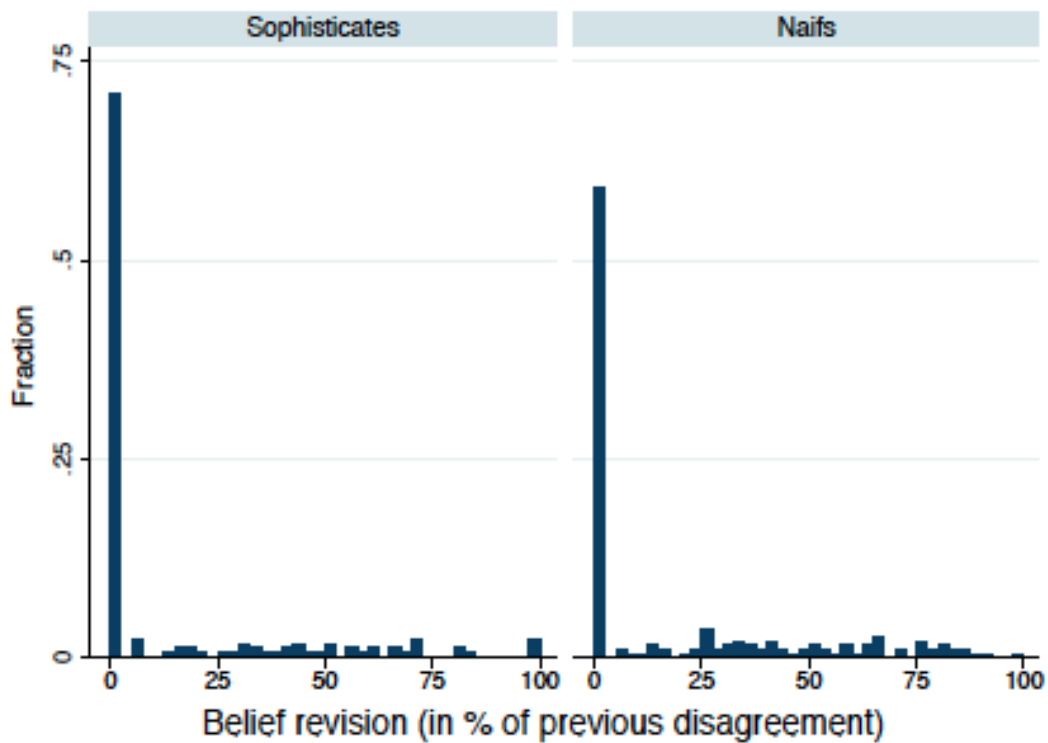


Figure 14: Magnitude of belief revisions. Each histogram depicts the belief revision between the first and second belief (expressed as percent of the difference between the first belief and the average belief of the two neighbors) conditional on the type of the subject (left / right panel). A subject is classified as sophisticated if the out-of-sample median naïveté parameter from the first part of the experiment satisfies $\chi \leq 0.5$ and conversely for naïfs. The figure includes all observations for which the first belief of a subject does not equal the average belief of the two neighbors. Adjustments $> 100\%$ and $< 0\%$ are excluded to ease readability (18 out of 374 obs.).

in response to the neighbors' assessments. While the patterns are slightly weaker for the neglect types, in both groups of subjects a large majority does not adjust their belief at all, i.e., subjects state exactly the same belief in the second question as in the first one. In addition, even those subjects that do adjust do so in a quantitatively small fashion.

H Experimental Instructions

H.1 Treatment *Selected*

Welcome. You will receive a fixed payment of 6 euros for participating in this experiment. This amount will be paid out to you in cash at the end of the experiment. How much money you earn on top of that depends on your decisions. In this experiment, you can earn points, where 100 points correspond to 10 euros. Your points will be converted into euros and paid out at the end of the experiment.

Your task

In this experiment, your task is to estimate two so-called “variables.” In what follows, we will refer to these variables as variable A and B. You will receive information about these variables from an information source. Based on this information, you will need to provide your estimates. In what follows, we will explain how these variables are generated and which type of information you will receive.

How variables A and B get determined

Every variable is determined through random draws from an urn. Figure 1 depicts these urns. Each urn contains exactly six balls with numbers 50, 70, 90, 110, 130, and 150. The letters are meant to help you in distinguishing between the urns.

The computer randomly determines variables A and B by drawing balls from the respective urn. From each urn, the computer draws six balls, i.e., six balls from urn A and six balls from urn B. Please note that, at each draw, each ball is equally likely to get drawn.

When a ball gets drawn, it gets replaced by another ball with the same number. That is, if the computer draws, say, a 130, then a new ball with number 130 is put into the urn before the next ball gets drawn. Thus, any given number can get drawn multiple times from the same urn.

Thus, the computer draws six balls from each urn. The average of the six balls then equals the respective variable:

- The average of the six balls from urn A equals variable A.
- The average of the six balls from urn B equals variable B.

In this experiment, you need to estimate the value of variables A and B. As you can see, these variables are fully independent from each other, so that you cannot learn anything from one variable about the other one. Thus, you should always distinguish

Urn A	Urn B
A-50	B-50
A-70	B-70
A-90	B-90
A-110	B-110
A-130	B-130
A-150	B-150

Figure 1: The urns, from which the computer draws six balls each. Please note that balls that get drawn get replaced by another ball with the same number, so that every number can get drawn multiple times. Only the numbers in this table can get drawn.

between these variables in the course of the experiment.

Your information

You receive your information from an information source. This information source does not draw balls from the urn itself. Rather, it observes all 12 balls that got drawn from urns A and B, i.e., all balls that determine the value of variables A and B. The experiment proceeds in multiple steps:

1. The information source observes the balls that got drawn from urns A and B.
2. For each variable, the information source shows you one of these randomly selected balls. Each ball is equally likely to be shown to you. You then have one piece of information about each variable, A and B.
3. Subsequently, you need to provide your first estimate about each variable. In this first step, you only need to estimate whether a variable is greater or smaller than 100. You can base this decision on the first ball that was shown to you by the information source. As will explained to you in greater detail below, you will earn the highest amount of money on average if:
 - You estimate that the variable is greater than 100 if the first ball had a number greater than 100.
 - You estimate that the variable is smaller than 100 if the first ball had a number smaller than 100.
4. Subsequently, you receive further information from the information source. It shows you some balls for variable A and some balls for variable B. In doing so, the information source depends on your first estimates:

- If you estimated that a variable is greater than 100, the information source definitely shows you all balls with numbers greater than 100. In case that there are less than three of such balls, the information source shows you additional randomly drawn balls with numbers smaller than 100, all of which previously got drawn from the urn. The information source continues with this process until you have seen three balls per variable.
 - If you estimated that a variable is smaller than 100, the information source definitely shows you all balls with numbers smaller than 100. In case that there are less than three of such balls, the information source shows you additional randomly drawn balls with numbers greater than 100, all of which previously got drawn from the urn. The information source continues with this process until you have seen three balls per variable.
5. This means that for each variable you will see at least three additional balls on your decision screen. In addition, as a reminder, you will also see the first ball that you had already seen in the first step.
 6. Then, you need to provide an estimate for each variable. These estimates can take on any value between 50 and 150. You have a total of up to six minutes to do so. As will be explained to you in greater detail below, you will maximize your earnings with your second estimate if your estimate is as close as possible to the value of the respective variable.

We will implement this entire procedure four times. These four “rounds” are entirely independent from each other: each time, the variables A and B get drawn anew and you receive new information about these variables. This means that variables A and B are determined in each round separately, so that you cannot learn anything from one round about another one.

Your payment

In addition to your fixed payment, you will be paid based on your estimates.

In each round, you can earn up to 180 points with your first estimate if you correctly estimate whether the respective variable is greater or smaller than 100. You receive 0 points if your estimate is not correct.

In each round, you can also earn up to 180 points with your second estimate. The further away your response from the truth, the less you earn. This is determined by the following equation (in points):

$$\text{Earnings} = 180 - 2 * (\text{Difference between estimate and truth})^2$$

This means that the difference between your estimate and the truth gets squared and multiplied by 2. This value then gets subtracted from the potential maximum earnings of 180. While this formula may look complicated, the underlying principle is very simple: **the smaller the difference between your estimate and the true value, the higher your earnings**. However, your earnings can never be less than zero, i.e., you cannot incur losses. You can also see that your earnings only depend on the absolute difference. For example, it is hence immaterial for your earnings whether you over- or underestimate the true value by 5.

In total, you will provide 16 estimates in the course of this experiment (two for each of the two variables, in each of four rounds). The computer will randomly determine one of these estimates, and your payment will then depend on this estimate. In every round, one of your first estimates gets randomly selected with probability 10% and one of your second estimates with probability 90%. Thus, you should work on each estimate as well as you can because each estimate may be relevant for your payment.

IMPORTANT: Please note that in this experiment you maximize your earnings on average if you always truthfully report your estimates! Because only one of your decisions gets selected for payment, there is no point for you in, say, strategizing by sometimes providing a high and sometimes a low estimate. You should simply try to make the best decision possible to maximize your earnings.

Example

Suppose that the computer has drawn six balls from each urn and has thereby determined the values of variables A and B. The information source now shows you a first randomly selected ball. As depicted in Figure 2, you then need to estimate, for each variable, whether it is greater or smaller than 100.

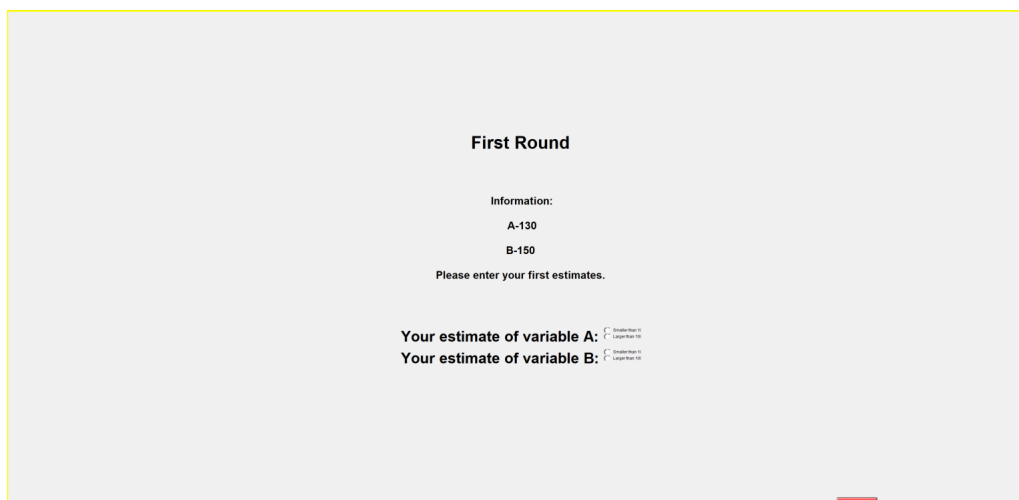


Figure 1: Example screenshot for the first estimates

Subsequently, the information source shows you additional balls. Figure 3 presents an example, As you can see, you also get reminded of the first ball that you have already seen on the previous screen.

Then, you need to provide an estimate about each variable.

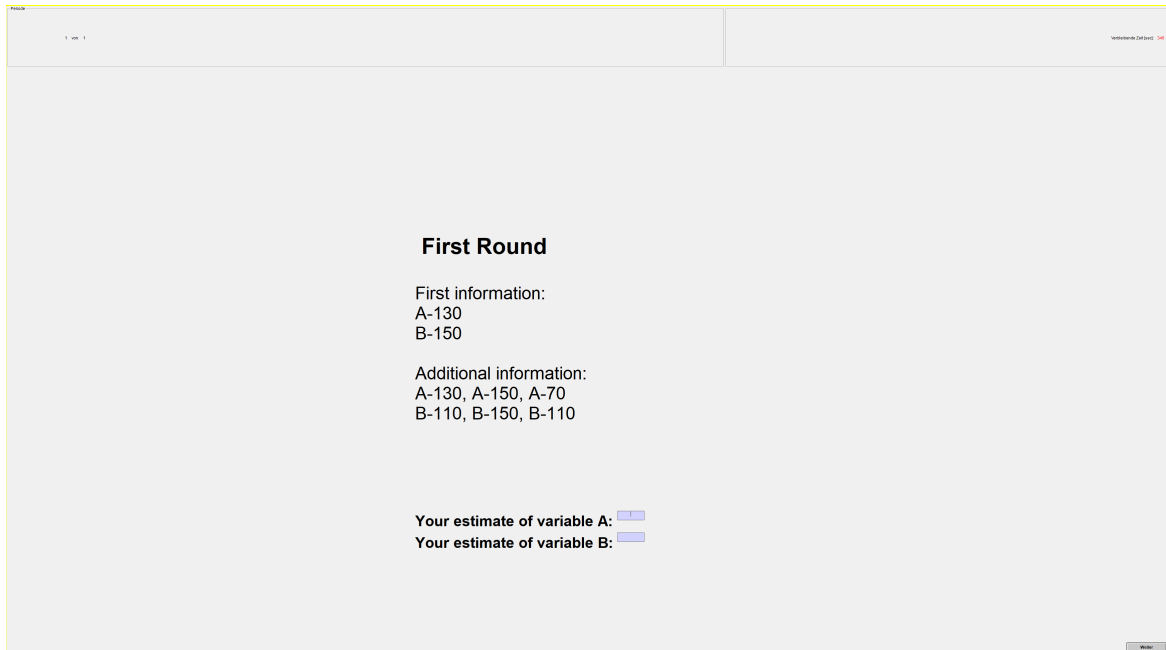


Figure 2: Example screenshot for the second estimates

Space for personal notes (you are welcome to write on or highlight on these instructions if you wish)

H.2 Treatment *Control*

Welcome. You will receive a fixed payment of 6 euros for participating in this experiment. This amount will be paid out to you in cash at the end of the experiment. How much money you earn on top of that depends on your decisions. In this experiment, you can earn points, where 100 points correspond to 10 euros. Your points will be converted into euros and paid out at the end of the experiment.

Your task

In this experiment, your task is to estimate two so-called “variables.” In what follows, we will refer to these variables as variable A and B. You will receive information about these variables from an information source. Based on this information, you will need to

provide your estimates. In what follows, we will explain how these variables are generated and which type of information you will receive.

How variables A and B get determined

Every variable is determined through random draws from an urn. Figure 1 depicts these urns. Each urn contains exactly six balls with numbers 50, 70, 90, 110, 130, and 150. The letters are meant to help you in distinguishing between the urns.

The computer randomly determines variables A and B by drawing balls from the respective urn. From each urn, the computer draws six balls, i.e., six balls from urn A and six balls from urn B. Please note that, at each draw, each ball is equally likely to get drawn.

When a ball gets drawn, it gets replaced by another ball with the same number. That is, if the computer draws, say, a 130, then a new ball with number 130 is put into the urn before the next ball gets drawn. Thus, any given number can get drawn multiple times from the same urn.

Urn A	Urn B
A-50	B-50
A-70	B-70
A-90	B-90
A-110	B-110
A-130	B-130
A-150	B-150

Figure 1: The urns, from which the computer draws six balls each. Please note that balls that get drawn get replaced by another ball with the same number, so that every number can get drawn multiple times. Only the numbers in this table can get drawn.

Thus, the computer draws six balls from each urn. The average of the six balls then equals the respective variable:

- The average of the six balls from urn A equals variable A.
- The average of the six balls from urn B equals variable B.

In this experiment, you need to estimate the value of variables A and B. As you can see, these variables are fully independent from each other, so that you cannot learn anything from one variable about the other one. Thus, you should always distinguish between these variables in the course of the experiment.

Your information

You receive your information from an information source. This information source does not draw balls from the urn itself. Rather, it observes all 12 balls that got drawn from urns A and B, i.e., all balls that determine the value of variables A and B. The experiment proceeds in multiple steps:

1. The information source observes the balls that got drawn from urns A and B.
2. For each variable, the information source shows you one of these randomly selected balls. Each ball is equally likely to be shown to you. You then have one piece of information about each variable, A and B.
3. Subsequently, you need to provide your first estimate about each variable. In this first step, you only need to estimate whether a variable is greater or smaller than 100. You can base this decision on the first ball that was shown to you by the information source. As will explained to you in greater detail below, you will earn the highest amount of money on average if:
 - You estimate that the variable is greater than 100 if the first ball had a number greater than 100.
 - You estimate that the variable is smaller than 100 if the first ball had a number smaller than 100.
4. Subsequently, you receive further information from the information source. More specifically, the information source in some way shows you all balls that determine variables A and B:
 - Case A: If your first estimated was greater than 100:
 - Then, the information source definitely shows you all balls with numbers greater than 100. In case that there are less than three of such balls, the information source shows you additional randomly drawn balls with numbers smaller than 100 until you have seen three balls per variable.
 - In addition, the information source shows you a “70” for all remaining balls with numbers smaller than 100, which corresponds exactly to the midpoint of this interval.
 - Case B: If your first estimated was smaller than 100:
 - Then, the information source definitely shows you all balls with numbers smaller than 100. In case that there are less than three of such balls, the information source shows you additional randomly drawn balls with numbers greater than 100 until you have seen three balls per variable.

- In addition, the information source shows you a “130” for all remaining balls with numbers greater than 100, which corresponds exactly to the midpoint of this interval.
 - Thus, in some way, you will see all variables that determine variables A and B. It is not important for you why you observe these balls in different ways.
5. This means that for each variable you will see five additional balls on your decision screen. In addition, as a reminder, you will also see the first ball that you had already seen in the first step.
 6. Then, you need to provide an estimate for each variable. These estimates can take on any value between 50 and 150. You have a total of up to six minutes to do so. As will be explained to you in greater detail below, you will maximize your earnings with your second estimate if your estimate is as close as possible to the value of the respective variable.

We will implement this entire procedure four times. These four “rounds” are entirely independent from each other: each time, the variables A and B get drawn anew and you receive new information about these variables. This means that variables A and B are determined in each round separately, so that you cannot learn anything from one round about another one.

Your payment

In addition to your fixed payment, you will be paid based on your estimates.

In each round, you can earn up to 180 points with your first estimate if you correctly estimate whether the respective variable is greater or smaller than 100. You receive 0 points if your estimate is not correct.

In each round, you can also earn up to 180 points with your second estimate. The further away your response from the truth, the less you earn. This is determined by the following equation (in points):

$$\text{Earnings} = 180 - 2 * (\text{Difference between estimate and truth})^2$$

This means that the difference between your estimate and the truth gets squared and multiplied by 2. This value then gets subtracted from the potential maximum earnings of 180. While this formula may look complicated, the underlying principle is very simple: **the smaller the difference between your estimate and the true value, the higher your earnings**. However, your earnings can never be less than zero, i.e., you cannot incur losses. You can also see that your earnings only depend on the absolute

difference. For example, it is hence immaterial for your earnings whether you over- or underestimate the true value by 5.

In total, you will provide 16 estimates in the course of this experiment (two for each of the two variables, in each of four rounds). The computer will randomly determine one of these estimates, and your payment will then depend on this estimate. In every round, one of your first estimates gets randomly selected with probability 10% and one of your second estimates with probability 90%. Thus, you should work on each estimate as well as you can because each estimate may be relevant for your payment.

IMPORTANT: Please note that in this experiment you maximize your earnings on average if you always truthfully report your estimates! Because only one of your decisions gets selected for payment, there is no point for you in, say, strategizing by sometimes providing a high and sometimes a low estimate. You should simply try to make the best decision possible to maximize your earnings.

Example

Suppose that the computer has drawn six balls from each urn and has thereby determined the values of variables A and B. The information source now shows you a first randomly selected ball. As depicted in Figure 2, you then need to estimate, for each variable, whether it is greater or smaller than 100.

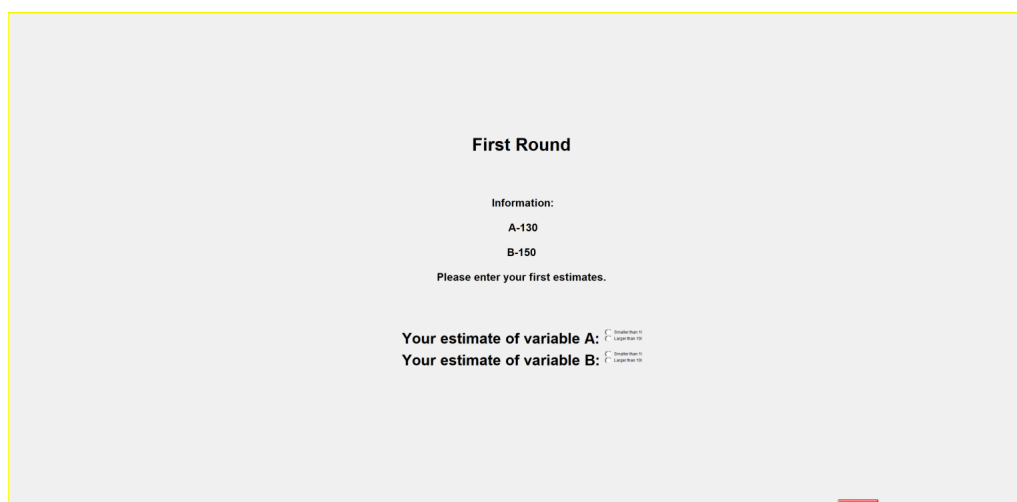


Figure 1: Example screenshot for the first estimates

Subsequently, the information source shows you additional balls. Figure 3 presents an example, As you can see, you also get reminded of the first ball that you have already seen on the previous screen.

Then, you need to provide an estimate about each variable.

Space for personal notes (you are welcome to write on or highlight on these

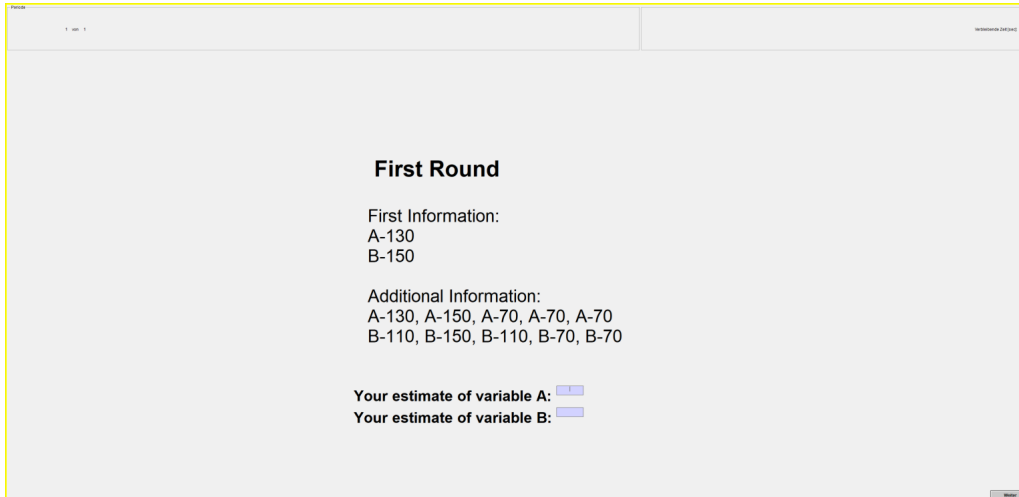


Figure 2: Example screenshot for the second estimates

instructions if you wish)

H.3 Treatment *Sequential*

The instructions were almost identical to those in *Selected*, except that the instructions mentioned only one variable to be estimated in a given round. Overall, subjects completed eight rounds.

H.4 Treatment *Feedback*

The instructions were identical to those in *Selected*, except that subjects completed 14 rather than 8 tasks.

H.5 Treatments *Complex, Simple, and Few*

The instructions in *Complex* and *Simple* were identical to those in *Selected*, except for the signal space.

The instructions in *Few* were identical to those in *Complex*, except that only two random draws determined the true state.

H.6 Treatment *Nudge*

The instructions were identical to those in *Selected*, except that there was a hint both at the end of the instructions and on subjects' decision screens:

HINT: Also pay attention to those randomly drawn balls that are not shown to you by the information source.

H.7 Treatment *Endogenous*

Welcome. You will receive a fixed payment of 6 euros for participating in this experiment. This amount will be paid out to you in cash at the end of the experiment. How much money you earn on top of that depends on your decisions. In this experiment, you can earn points, where 100 points correspond to 10 euros. Your points will be converted into euros and paid out at the end of the experiment.

Your task

In this experiment, your task is to estimate two so-called “variables.” In what follows, we will refer to these variables as variable A and B. You will receive information about these variables from an information source. Based on this information, you will need to provide your estimates. In what follows, we will explain how these variables are generated and which type of information you will receive.

How variables A and B get determined

Every variable is determined through random draws from an urn. Figure 1 depicts these urns. Each urn contains exactly six balls with numbers 50, 70, 90, 110, 130, and 150. The letters are meant to help you in distinguishing between the urns.

The computer randomly determines variables A and B by drawing balls from the respective urn. From each urn, the computer draws six balls, i.e., six balls from urn A and six balls from urn B. Please note that, at each draw, each ball is equally likely to get drawn.

When a ball gets drawn, it gets replaced by another ball with the same number. That is, if the computer draws, say, a 130, then a new ball with number 130 is put into the urn before the next ball gets drawn. Thus, any given number can get drawn multiple times from the same urn.

Thus, the computer draws six balls from each urn. The average of the six balls then equals the respective variable:

- The average of the six balls from urn A equals variable A.
- The average of the six balls from urn B equals variable B.

In this experiment, you need to estimate the value of variables A and B. As you can see, these variables are fully independent from each other, so that you cannot learn

Urn A	Urn B
A-50	B-50
A-70	B-70
A-90	B-90
A-110	B-110
A-130	B-130
A-150	B-150

Figure 1: The urns, from which the computer draws six balls each. Please note that balls that get drawn get replaced by another ball with the same number, so that every number can get drawn multiple times. Only the numbers in this table can get drawn.

anything from one variable about the other one. Thus, you should always distinguish between these variables in the course of the experiment.

Your information

You receive your information potentially from different information sources, information source I and information source II. These information sources do not draw balls from the urn themselves. Rather, they observe all 12 balls that got drawn from urns A and B, i.e., all balls that determine the value of variables A and B. The experiment proceeds in multiple steps:

1. The information sources observe the balls that got drawn from urns A and B.
2. For each variable, you will be shown one of these randomly selected balls. Each ball is equally likely to be shown to you. You then have one piece of information about each variable, A and B.
3. Subsequently, you need to decide whether you would like to purchase additional information about one or both of the variables. This information is helpful for making high-quality estimates. However, information is also costly.
 - For each variable, there are two potential information sources, information source I and information source II. These information sources send potentially different types of information, as explained in detail below. The information of an information source costs 5 points (0.50 euros) per variable. Thus, the price of information source I and II is identical.
 - Thus, you have three options for each variable: purchase no information, purchase information from information source I, or purchase information from information source II.

- You can also make different decisions for the two variables: for example, you could purchase the information from information source I for one variable, and for the other variable the information from information source II (or no information at all).
4. In case you decide to purchase information, the information sources will send you potentially different types of information:
 - Information source I definitely shows you all drawn balls with numbers greater than 100. In case that there are less than three of such balls, the information source shows you additional randomly drawn balls with numbers smaller than 100, all of which previously got drawn from the urn. The information source continues with this process until you have seen three balls per variable.
 - Information source I definitely shows you all drawn balls with numbers smaller than 100. In case that there are less than three of such balls, the information source shows you additional randomly drawn balls with numbers greater than 100, all of which previously got drawn from the urn. The information source continues with this process until you have seen three balls per variable.
 5. This means that in case you decide to purchase one of the information sources, you will see at least three additional balls on your decision screen for this variable. In addition, as a reminder, you will also see the first ball that you had already seen in the first step.
 6. This means that:
 - In case you do not purchase an information source, you ultimately only receive one piece of information (the one that you receive in the beginning).
 - In case you do purchase an information source, you ultimately receive at least for pieces of information for this variable.
 7. Then, you need to provide an estimate for each variable. These estimates can take on any value between 50 and 150. You have a total of up to six minutes to do so. As will be explained to you in greater detail below, you will maximize your earnings with your second estimate if your estimate is as close as possible to the value of the respective variable.

We will implement this entire procedure four times. These four “rounds” are entirely independent from each other: each time, the variables A and B get drawn anew and

you receive new information about these variables. This means that variables A and B are determined in each round separately, so that you cannot learn anything from one round about another one.

Your payment

In addition to your fixed payment, you will be paid based on your estimates.

In each round, you can earn up to 180 points with your first estimate if you correctly estimate whether the respective variable is greater or smaller than 100. You receive 0 points if your estimate is not correct.

In each round, you can also earn up to 180 points with your second estimate. The further away your response from the truth, the less you earn. This is determined by the following equation (in points):

$$\text{Earnings} = 180 - 2 * (\text{Difference between estimate and truth})^2 - \text{Cost for information}$$

This means that the difference between your estimate and the truth gets squared and multiplied by 2. This value then gets subtracted from the potential maximum earnings of 180. While this formula may look complicated, the underlying principle is very simple: **the smaller the difference between your estimate and the true value, the higher your earnings**. However, your earnings can never be less than zero, i.e., you cannot incur losses. You can also see that your earnings only depend on the absolute difference. For example, it is hence immaterial for your earnings whether you over- or underestimate the true value by 5.

In case you decide to purchase an information source, you only need to pay the price in case your corresponding estimate is selected for payment. This means that in this experiment you will have to pay the price for the information source at most once.

In total, you will provide 16 estimates in the course of this experiment (two for each of the two variables, in each of four rounds). The computer will randomly determine one of these estimates, and your payment will then depend on this estimate. In every round, one of your first estimates gets randomly selected with probability 10% and one of your second estimates with probability 90%. Thus, you should work on each estimate as well as you can because each estimate may be relevant for your payment.

IMPORTANT: Please note that in this experiment you maximize your earnings on average if you always truthfully report your estimates! Because only one of your decisions gets selected for payment, there is no point for you in, say, strategizing by sometimes providing a high and sometimes a low estimate. You should simply try to make the best decision possible to maximize your earnings.

Example

Suppose that the computer has drawn six balls from each urn and has thereby determined the values of variables A and B. The information source now shows you a first randomly selected ball. As depicted in Figure 2, you then need to decide whether and which information source you would like to purchase.

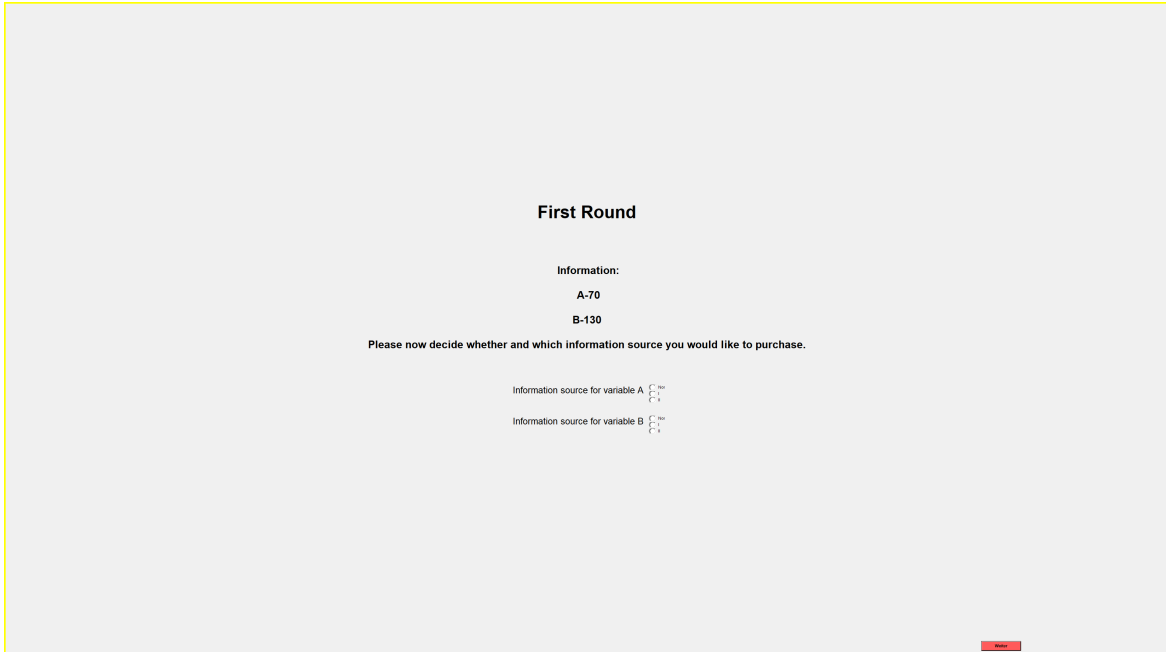


Figure 2: Example screenshot for the first estimates

Subsequently, the information source shows you additional balls. Figure 3 presents an example for a case in which you purchased an information source for both variables. As you can see, you also get reminded of the first ball that you have already seen on the previous screen.

Then, you need to provide an estimate about each variable.

Space for personal notes (you are welcome to write on or highlight on these instructions if you wish)

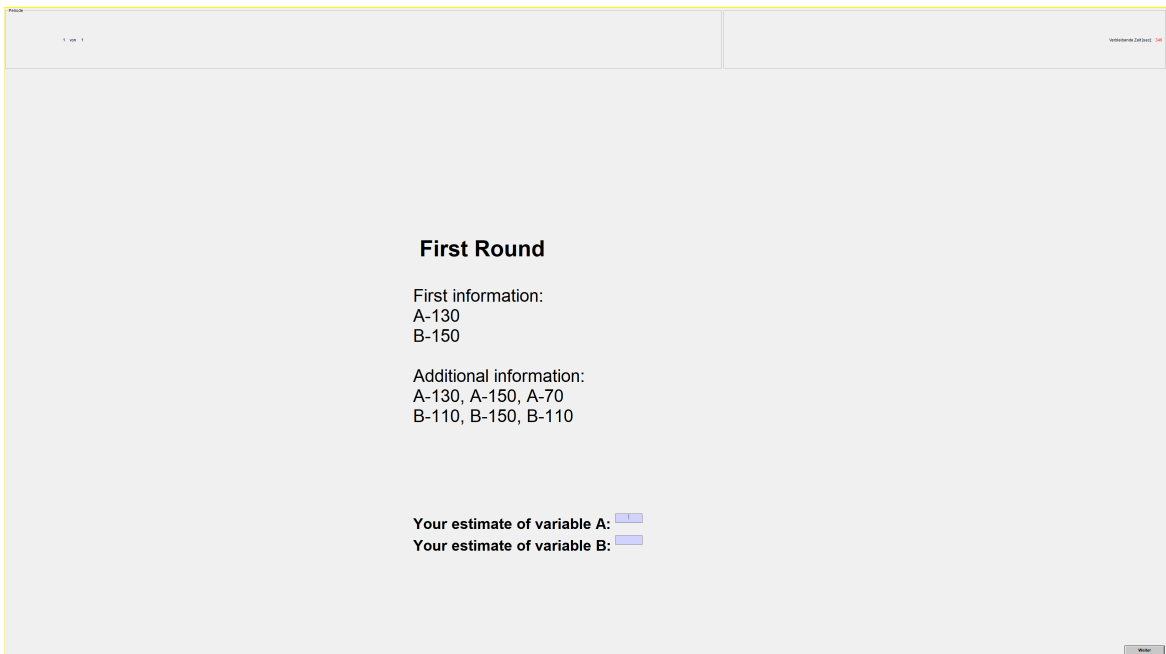


Figure 3: Example screenshot for the second estimates