

WHAT YOU SEE IS ALL THERE IS^{*}

Benjamin Enke

May 30, 2019

Abstract

News reports and communication are inherently constrained by space, time, and attention. As a result, news sources often condition the decision of whether or not to share a piece of information on the similarity between the signal and the prior belief of the audience, so that the audience needs to draw inferences from a selected sample. 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 updates beliefs by exclusively taking into account information that is right in front of them, hence neglecting selection effects. A series of treatments aimed at identifying mechanisms document that neglect is largely driven by subjects maintaining a misspecified mental model of the known information environment, according to which selection does not even come to mind. This mental model of “what you see is all there is” is not constant but predictably responds to perhaps seemingly innocuous variations in the environment, including (i) the computational complexity of the decision problem, and (ii) whether selection problems arise due to a deliberate choice between information sources or as a mechanical by-product of another decision. These results point to the context-dependence of mental models and the resulting errors in belief updating.

JEL classification: D03; D80; D84.

Keywords: Bounded rationality; mental models; complexity.

^{*}I thank an editor and three very constructive referees. For helpful 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. In many contexts, these constraints induce the potential sender of a message to condition the decision of whether or not to report a piece of news on the *match between the signal realization 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) the “absence” of signals.

While recent theoretical literature has linked selection problems in belief updating to various economic applications,¹ empirical work on people’s reasoning in such contexts is scarce. Moreover, if people actually do neglect selection issues, 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. A candidate explanation is that neglect patterns reflect systematic computational errors. Another promising candidate mechanism – prominent in an active theory literature – is the idea that agents optimize but maintain a misspecified mental model of the environment.² Still, the extent to which misspecified mental models actually generate biased belief updating and whether they should be thought of as reasonably stable “neglect parameters” or are instead context-dependent, is open to discussion.

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

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

² See, e.g., Eyster and Rabin (2010), Jehiel (2005), Gennaioli and Shleifer (2010), Schwartzstein (2014), Gabaix (2014), Spiegel (2016), Bushong and Gagnon-Bartsch (2016), Bohren and Hauser (2017), Bordalo et al. (2017), Heidhues et al. (2017), Jehiel (2018), and Gagnon-Bartsch et al. (2018).

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 signals 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. Specifically, 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 infer some signals from the fact that they 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. This coarse information introduces a selection problem in the sense that only taking into account the signals that are "visible" introduces a biased sample.

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 are (colloquially) "visible." Thus, 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. Through a follow-up treatment, I verify that this pattern is robust against the short-run provision of feedback. Here, across 14 subsequent tasks, participants are provided with feedback about how their belief compares to the true state and resulting profits; yet, there is no indication that participants learn over this short period of time.

The average neglect patterns mask considerable heterogeneity. To characterize participants' belief formation rules, the analysis estimates an individual-level parameter that pins down updating rules in relation to Bayesian rationality. The distribution of updating types follows a bimodal structure, according to which 60% of participants

are either Bayesian or exhibit full neglect. In other words, many subjects state beliefs that correspond *exactly* to fully ignoring what they do not see, as if selection does not even come to mind. This bimodal pattern is reminiscent of the idea – prominent in much recent theoretical work – that the neglect types develop “optimal” beliefs but only conditional on maintaining a misspecified mental model of the data-generating process. According to this terminology, a participant’s mental model corresponds to the way they set up (rather than solve) the problem of computing posteriors from the information that is provided to them. Another perspective is that cognitive biases are generated through systematic computational errors, such as difficulties in computing conditional expectations. As argued by authors such as [Fudenberg \(2006\)](#), understanding the mechanisms behind biases is important both to develop appropriate debiasing strategies and to inform theoretical work.

Thus, in a next step, I explicitly study the roles of mental models, computations, and their interaction. The empirical analysis is based on a qualitative framework that assumes that people maintain an automatic mental default model of the environment of which selection is not a part. According to this mental model, “what you see is all there is,” and the invisible signals do not even come to mind. For example, this simplistic default model could be retrieved from memory as a normal version of the problem that people know how to solve ([Kahneman and Miller, 1986](#)). This mental default model can, in principle, be overturned, yet whether this happens depends on whether a participant’s focus is drawn towards the selection issue. The key hypothesis behind the analysis of mechanisms is that the decision maker’s focus and resulting mental model are endogenous to the structure of the decision problem, in particular its computational complexity. The idea is that high computational complexity may induce cognitive load and hence distract participants from the conceptual aspect of the problem. Put differently, when the updating problem is computationally complex, subjects may focus excessively on processing the numbers that are in front of them, as opposed to setting up a correct mental model in the first place.

To test this hypothesis, I devise three treatments that manipulate specific aspects of computational complexity. In these experiments, I purposefully construct the treatments such that they hold the difficulty of accounting for selection per se constant but vary how computationally cumbersome it is to compute beliefs. This approach has the attractive feature that it narrows down the pathways through which complexity can affect belief updating such that complexity can only matter to the extent that it distracts participants from developing a correct mental model.

The experiments operationalize complexity in two different ways: the complexity of the signal space and the number of signals. First, to vary the complexity of the

signal space, I implement two 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 invisible signal must be a 70. Thus, these treatments leave the difficulty of accounting for selection constant (if the first belief is above 100), but they manipulate the computational difficulty of computing a posterior belief. 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 a “missing” signal, complexity can only matter to the extent that it draws participants’ effort and attention away from the selection problem in the first place. Thus, if selection neglect could be adequately represented as a context-invariant neglect parameter, beliefs should be identical across conditions, or perhaps more noisy in the more complex treatments, but not systematically more biased. On the other hand, if the computational complexity of the environment induces cognitive load and hence distracts subjects from the selection problem, we would expect that the frequency of neglect varies across treatments.

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 expend more cognitive effort, as measured by response times. These results are arguably noteworthy because they provide the first empirical evidence that mental models and computations interact: mental models prescribe certain computations, but (anticipated) computational difficulty also appears to affect which mental model is formed in the first place.

The fact that variations in complexity matter for neglect even though the difficulty of accounting for selection is unchanged highlights the crucial role of (endogenous) misspecified mental models. To provide direct evidence for the role of misspecified models, I implement two treatments that either explicitly or implicitly draw subjects’ attention to the selection problem. First, I implement an experimental condition that includes a 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%.

Second, to shift participants’ attention to the selection problem in a more indirect way, I devise a treatment in which subjects are confronted with the same selection problem, yet they are explicitly asked to choose between an information source that supplies positively selected information and an information source that supplies negatively se-

lected information. Thus, in this treatment, the selection problem does not arise as a mechanical by-product of another decision by the subject but is deliberately induced by the subject in the quest to acquire further information. This arguably brings the selection problem exogenously top of mind. Again, the results show that neglect decreases by about 50% in this treatment. Thus, inducing subjects to develop a correct mental model has a significant effect on beliefs.

In summary, the key takeaways from the analysis of mechanisms are that (i) misspecified mental models play an important part in generating neglect and (ii) the computational complexity of the environment plays a role in bringing about these misspecified mental models in the first place. Thus, people’s tendency to neglect selection should be thought of as endogenous to the decision context. These insights are arguably relevant not only for modeling updating biases but potentially also for policy in terms of what will be an effective method to correct biased beliefs.

This paper ties into three strands of the literature. First, the results contribute to the recent experimental literature on boundedly rational reasoning ([Eyster et al., 2015](#); [Enke and Zimmermann, 2019](#); [Mormann and Frydman, 2016](#); [Bushong and Gagnon-Bartsch, 2016](#); [Abeler and Jäger, 2015](#); [Ngangoue and Weizsäcker, 2015](#); [Charness et al., 2018](#)), including contemporaneous work on endogenous sample selection problems ([Esponda and Vespa, 2018](#); [Jin et al., 2016](#); [Araujo et al., 2018](#)). 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 misspecified mental models for neglect; in particular (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 not only on which errors people make but also *why* ([Esponda and Vespa, 2016](#); [Martínez-Marquina et al., 2017](#)). I am not aware of prior work that has studied the effects of computational complexity on how people form mental models.

Second, the paper contributes to an active theory literature, discussed in greater detail in Section 6, that highlights the importance of misspecified mental models for belief formation. The results in this paper provide support for the idea that mental models play an important role for boundedly rational inference, and further suggest that these mental models could be modeled as dependent upon the complexity of the environment. The idea that cognition is context-dependent places this paper in the neighborhood of theories of context-dependence ([Bordalo et al., 2013, 2017](#); [Kőszegi and Szeidl, 2013](#); [Bushong et al., 2015](#)).

Finally, the paper relates to two distinct psychology literatures on inference from non-salient data, both of which are summarized in [McCaffery and Baron \(2006\)](#). First, work on the availability heuristic ([Tversky and Kahneman, 1973](#)) is related in its focus

on salient information. However, as pointed out by [McCaffery and Baron](#), 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 even though the entire data-generating process is transparent and known. Second, there is a more recent psychological literature on “isolation” or “focusing effects” that suggests that people mostly focus on information that is right before them, including work on selection problems ([Brenner et al., 1996](#); [Schkade et al., 2007](#); [Koechler and Mercer, 2009](#)). However, this work does not use the structured information-generating processes with objectively correct and financially incentivized answers that characterize modern experimental economics research. In addition, I am not aware of evidence in the psychology literature that studies how computational complexity affects the formation of misspecified mental models and neglect.

The paper proceeds as follows. Section 2 describes the experimental design, and Section 3 presents the results on selection neglect. The role of mental models and complexity are studied in Section 4, and a replication study is reported on in Section 5. Section 6 discusses the connection between selection neglect and related recent empirical findings, and Section 7 concludes.

2 Experimental Design

2.1 Setup

Studying belief formation in contexts of selection requires (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 present between-subjects design accommodates all these features.

The key idea of 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 $X = \{50, 70, 90, 110, 130, 150\}$. 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.

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

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, $b_1 \in \{low, high\}$.³

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. That is, in either case, the subjects were always shown 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 did not share with subjects as "missing" or as signals that subjects "do not see." This terminology is purely colloquial in nature. In an information-theoretic sense, these "missing" signals constitute coarse information. Still, I refer to these as signals that subjects "do not see" to make it salient that these signals do not appear on subjects' decision screens.

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 they were missing. This provides a crucial input into

³If the true state equalled 100, subjects received the full payment for either guess.

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 first guessed $b_1 = \text{high}$, a missing signal was 70, in expectation, while it was 130 when they first guessed $b_1 = \text{low}$.

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 a missing signal was in $\{50, 70, 90\}$, the information source communicated 70 to the subject, while if the missing signal was in $\{110, 130, 150\}$, the information source communicated 130. 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 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. Below, I discuss a treatment that verified 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

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 Section 2.3 for a derivation of the Bayesian and neglect benchmarks.

all control questions could they proceed to the experiment.⁴ Appendix G contains the experimental instructions and control questions.

2.2 Procedural Details

Apart from the two 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.⁵

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

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.

⁵The smaller sample size in *Control* was determined ex ante, which reflects the fact that it merely serves as a “straw man” with very little expected noise.

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>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
<i>Endogenous</i>	71	13.20

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.

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 in a given task j equalled $\pi^j = \max\{0; 18 - 0.2 \times (b^j - \theta^j)^2\}$, where b denotes the belief.

2.3 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 N denote the number of signals a subject actually sees on their computer screen. Given a set of signals, a Bayesian would compute the mean posterior belief b_B as

$$b_B = \frac{\sum_{v=1}^N s_v + \sum_{l=N+1}^6 \mathbb{E}[s_l \mid b_1]}{6}, \quad (2)$$

where s_v denotes a signal that appears on the decision screen, and s_l denotes an unobserved (coarse) signal.

Now imagine that people merely base their beliefs on “what they see” (i.e., the set of s_v). This introduces a selection problem in the sense that $E[s_i] = 100$, but $E[s_l|b_1] \neq 100$. That is, the set of signals that subjects do not observe is systematically different from the unconditional expectation. As an analog to the benchmark of rational inference above, I define a full neglect benchmark as:

$$b_N = \frac{\sum_{v=1}^N s_v}{4} \quad (3)$$

It is conceivable that individuals partially adjust for selection. Let $\chi \in [0, 1]$ parameterize the degree of neglect such that $\chi = 1$ implies full neglect. Then, people’s belief b can be expressed as a weighted average of b_B and b_N plus decision noise ϵ :

$$b = (1 - \chi)b_B + \chi b_N + \epsilon = b_B + \chi \frac{6 - N}{6} (\bar{s}_v - \bar{s}_l) + \epsilon. \quad (4)$$

$$\equiv b_B + \chi d + \epsilon, \quad (5)$$

where $\bar{s}_v \equiv 1/N \sum_{v=1}^N s_v$ is the average visible signal, $\bar{s}_l \equiv 1/(6 - N - 1) \sum_{i=N+1}^6 E(s_l|b_1)$ the average expected “non-visible” (coarse) signal, and ϵ is a mean zero random computational error. That is, 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 .

Throughout the paper, I will rely on this formal framework to compute estimates of neglect $\hat{\chi}$ and decision noise $|\hat{\epsilon}|$.

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

Table 4: Overview of beliefs across tasks

(1) True State	(2) First Signal	(3) Bayesian Belief	(4) Neglect Belief	(5) Median Belief <i>Control</i> Treatment	(6) Median Belief <i>Selected</i> Treatment	(7) p-value (Ranksum test)
96.67	High	103.33	122.00	103.00	110.00	0.0661
110.00	High	110.00	130.00	110.00	120.00	0.0001
93.33	Low	96.67	80.00	96.50	90.00	0.0130
90.00	High	90.00	100.00	90.00	90.00	0.0536
103.33	High	100.00	115.00	100.00	110.00	0.0001
116.67	Low	110.00	100.00	110.00	110.00	0.0635
116.67	High	120.00	130.00	120.00	123.00	0.0099
86.67	High	90.00	100.00	90.00	90.00	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 all tasks, average beliefs in *Selected* are biased away from the control condition benchmark in the direction of the full neglect benchmark, even if the medians are identical.

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 2.3, which has the additional advantage that going forward, all estimated quantities will have direct theoretical counterparts. Specifically, I use equation (5) to compute the estimated neglect implied in the belief of subject i in task j :

$$\hat{\chi}_i^j = E[\chi_i^j | b_i^j] = \frac{b_i^j - b_B^j}{d^j} = \frac{6(b_i^j - b_B^j)}{(6 - N^j)(\bar{s}_v^j - \bar{s}_l^j)}. \quad (6)$$

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 hence 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 ($N = 5,359$ 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 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 Accordingly, the standard errors are clustered at the subject level. All regressions include experimental session fixed effects, exploiting 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 control 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 (i.e., the average hypothetical profit from each belief) 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

3.2.1 Sequential Tasks

In the baseline design, subjects solved eight tasks, which were presented two at a time on one screen. This arguably mirrors the contexts that motivate the paper, such as when newspapers report on more than one topic on any given day. In *Control*, subjects also completed two tasks per screen, so that this feature cannot generate treatment differences. Still, 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. 75 subjects participated in this treatment.

Appendix D discusses the results from this treatment in detail, including a full econometric analysis akin to Table 5. Overall, the results are very similar to those in *Selected*. To illustrate, Figure 1 plots the median and mean of all subject-task-specific $\hat{\chi}_i^j$ 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* in terms of both median and average neglect; see Appendix D.

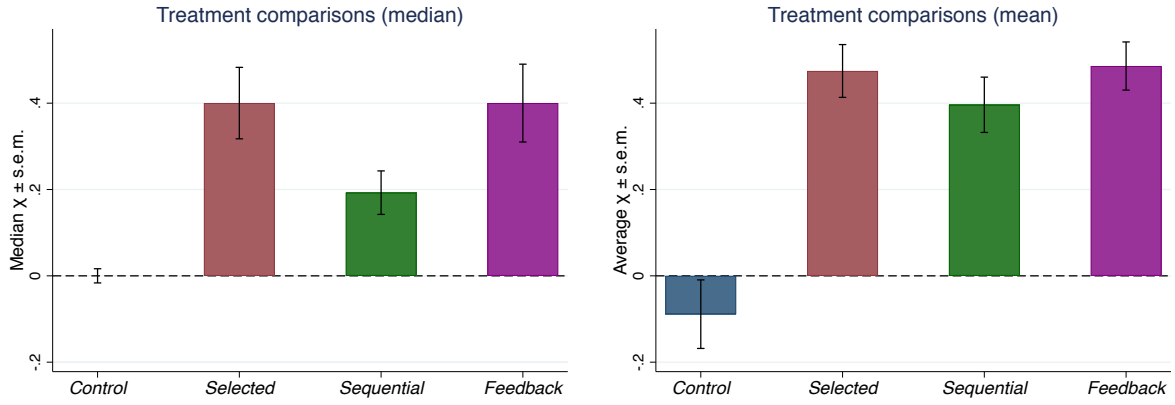


Figure 1: Overview of neglect $\hat{\chi}_i^j$ across treatments. The left panel shows the median $\hat{\chi}_i^j$ across all subject-task observations. The right panel shows the average $\hat{\chi}_i^j$ across all subject-task observations, where as in columns (3)–(4) of Table 5 the data are winsorized 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.

3.2.2 The Effect of Feedback

For applied contexts, a relevant question is whether people learn about their errors through experience and feedback. Naturally, studying the effects of feedback in laboratory experiments has limitations. On the one hand, laboratory experiments might understate the extent to which people learn from feedback because the number of repetitions is typically fairly small, and because increasing fatigue may stop subjects from learning after a while. On the other hand, experiments might overstate the extent to which people learn from feedback because feedback in the lab is usually much more “pure” than real-world feedback, where differences between beliefs and reality might be driven by a myriad of different errors.

Still, with these caveats in mind, the paper studies the effects of feedback as sensitivity check. In treatment *Feedback*, a new set of 75 subjects solved problems of the same type as those in *Selected*. This treatment was not randomized within the same sessions as *Selected* and *Control* but conducted in the same time period as the other treatments. In the experiment, 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 completed 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 randomly selected for payment.

Appendix E provides a detailed analysis of the data, including an econometric analysis. The key findings are twofold. First, comparing beliefs *across treatments* with those in *Selected* reveals that feedback does not lead to less neglect. The coefficient on the treatment dummy is statistically insignificant, quantitatively small, and even has the opposite sign as would be expected from the perspective of learning from feedback. Second, the analysis studies how subjects' beliefs in *Feedback* change *across periods* as they accumulate feedback, again relative to how beliefs evolve across periods in treatment *Selected*. Again, the data show no indication that feedback reduces the amount of neglect. Figure 1 illustrates these results.

3.3 Heterogeneity Analysis

Type distribution. To characterize subjects' belief patterns in more detail, I examine the subject-level distribution of neglect. For this purpose, 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. 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. (6)) satisfy $|t - \hat{\chi}_i^j| \leq 0.05$. That is, I count how often a subject's beliefs reflect an updating rule that is pinned down by candidate type $\hat{\chi}_i = t$. 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).⁷

The left panel of Figure 2 presents a histogram of these modal neglect types $\hat{\chi}_i$ in treatments *Selected*, *Sequential*, and *Feedback*.⁸ The data reveal a bimodal type distribution: 60% of all subjects are best characterized as Bayesian ($\hat{\chi}_i = 0$) or full neglect ($\hat{\chi}_i = 1$).⁹ For example, of those 150 subjects that are not approximately rational, one third (51) are classified as exactly or almost exactly full neglect types ($0.95 \leq \chi_i \leq 1.05$). In contrast, in treatment *Control*, 80% of all subjects are classified as exactly $\hat{\chi}_i = 0$; see Figure 11 in Appendix C.¹⁰

Because the left panel of Figure 2 shows modal types as computed across experimental tasks, the figure does not reveal whether the types $\hat{\chi}_i = 0$ or $\hat{\chi}_i = 1$ actually rationalize a larger number of underlying beliefs than the intermediate types. To ad-

⁷If more than one type rationalizes the maximal number of beliefs, I compute the average across t .

⁸The results are very similar within each of these treatments separately.

⁹Figure 8 in Appendix C provides a robustness check in which a subject's type is determined as the median χ_i^j across tasks. The distribution of these median neglect parameters is similarly bimodal with spikes at zero and one, respectively.

¹⁰Of course, in both treatments, some subjects may be thought of as noise types ([Fragiadakis et al., 2016](#)), which may generate some of the outliers in Figure 2. This would attenuate any treatment differences.

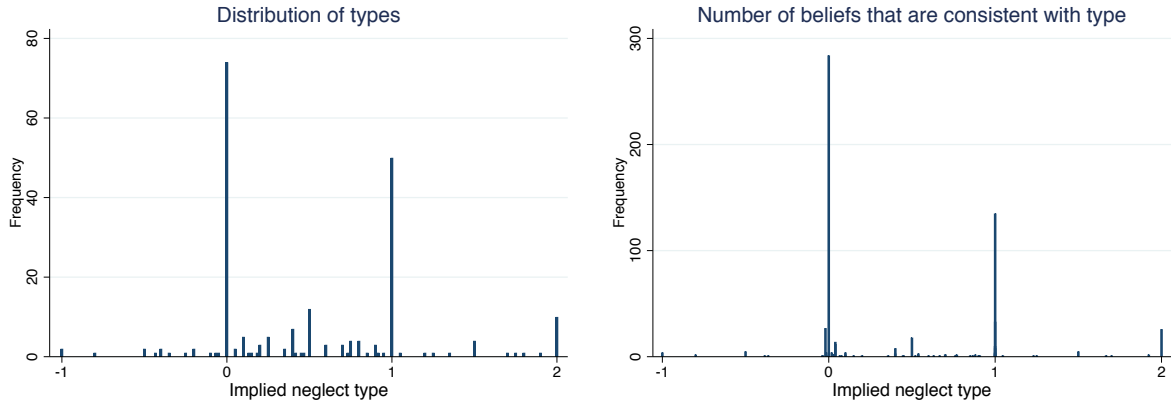


Figure 2: Distribution of modal neglect types $\hat{\chi}_i$ and number of classified beliefs in treatments *Selected*, *Sequential*, and *Feedback*. The left panel shows the distribution of estimated neglect types. A subject's type is determined based on the following procedure: for each subject i and 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. The right panel shows the total number of beliefs that are consistent with the estimated type of the respective subjects, separately for the implied neglect types. In other words, the number of type-consistent beliefs is computed by counting the number of beliefs that are consistent with a subject's type, across all subjects of a given type. Here, consistent with the estimated type means that $|\hat{\chi}_i - \hat{\chi}_i^j| \leq 0.05$.

dress this, the right panel shows the total number of beliefs that are consistent with a subject's overall estimated type. In other words, the number of type-consistent beliefs is computed by counting the number of beliefs that are consistent with a subject's type, across all subjects of a given type. Here, consistent with the estimated type means that the neglect parameter that is implied by a belief statement is close to the overall estimated type: $|\hat{\chi}_i - \hat{\chi}_i^j| \leq 0.05$.

The right panel reveals that the fully rational and full neglect types account for a substantially larger number of beliefs than the intermediate types. Across treatments, for the $\hat{\chi}_i = 0$ types, 4.5 beliefs are consistent with a subject's estimated type, 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. This suggests that the intermediate types partially adjust for selection but do so in a heuristic manner rather than in a quantitatively consistent fashion. Overall, these patterns suggest that the extreme types of $\chi = 0$ and $\chi = 1$ indeed organize a substantially larger part of the experimental data than the other estimated types.¹¹

As discussed in greater detail in Section 4, both cognitive scientists and economic theorists emphasize the importance of misspecified mental models for belief formation. Prima facie, the bimodal type distribution is consistent with such a view, as the two

¹¹An alternative way of showing that the bimodality of types is not an artifact of the aggregation procedure of the eight beliefs per subject into one type is to plot a histogram of the subject-task-specific $\hat{\chi}_i^j$, i.e., the underlying raw beliefs data. This is done in Figure 9 in Appendix C. Naturally, this distribution is noisier but also bimodal with spikes at zero and one.

Table 6: 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)	-0.072** (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. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

spikes in the data may be generated by the presence of two distinct types that differ in whether they entertain the correct model or a simplified subjective model that does not include selection at all.

Correlates of Neglect. Next, I investigate the correlates of neglect in treatments *Selected*, *Sequential*, and *Feedback*. In line with prior literature, the analysis focuses on cognitive ability and response times (Rubinstein, 2007, 2016). Cognitive ability is proxied by a participants’ high school grades. From participant’s self-reports, I construct a summary statistic as the average of the z-scores of a participant’s high school GPA and their high school math grade. Because subjects completed two tasks on the same decision screen, response times for a given task are taken to be one-half of observed response times.

Table 6 summarizes the results. To facilitate a quantitative interpretation of the regression coefficients, I standardize the high school grades variable into a z-score. All regressions include treatment and session fixed effects. Both higher (better) high school grades and longer response times are negatively correlated with neglect. At the same time, 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 strongly suggest that the type heterogeneity is not merely the product of the neglect types being lazier than the rational types.

4 Mental Models and Computational Complexity

4.1 Framework

Taking stock, we have seen that participants frequently state beliefs that reflect full neglect. This pattern is reminiscent of the idea that people entertain misspecified mental models of the environment, but optimize reasonably well conditional on these mental models. According to this terminology, a subject's mental model corresponds to the way they set up (rather than solve) the problem of computing posteriors from the information that is provided to them.

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 of mechanisms, I make use of a qualitative framework of participants' hypothesized reasoning process in the experiment. This framework distinguishes between mental models (and how they are formed) on the one hand and computational implementation on the other hand:

1. The participant develops a mental model (sets up the problem):
 - (a) The participant has an initial mental default model that neglects selection. According to this mental model, "what you see is all there is." This means that – as a default – subjects do not even pay attention to the signals that they (colloquially) "don't see." 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](#)).
 - (b) This initial mental model can be overturned if the presence of selection comes to mind. Whether or not the default model is overturned may be endogenous to the decision environment:
 - i. If the updating problem is computationally complex, then the participant may focus their mental energy on the computational aspect of the problem. This, in turn, may induce cognitive load that distracts the

participant from thinking through, or paying attention to, the selection problem.

- ii. Holding fixed the computational complexity of the updating problem, the default model may be more likely to get overturned if the participant’s focus is exogenously drawn to the selection problem.

2. The participant expends computational effort to solve the problem given the mental model developed before. For example, participants may correctly represent and set up the problem, but fail at mathematically backing out the signals that they “do not see” on their computer screens.

The distinction between mental models and computational implementation in this framework is in line with two distinct strands of the literature. First, cognitive scientists in the tradition of the computational theory of mind routinely partition thinking into mental models (often called mental representations) and computations on those models (Fodor, 1983; Thagard, 1996). Second, the importance of misspecified mental models is also highlighted by an active theoretical literature in economics (e.g., Eyster and Rabin, 2010; Jehiel, 2005; Gennaioli and Shleifer, 2010; Schwartzstein, 2014; Gabaix, 2014; Spiegel, 2016; Bushong and Gagnon-Bartsch, 2016; Bohren and Hauser, 2017; Bordalo et al., 2017; Heidhues et al., 2017; Gagnon-Bartsch et al., 2018). Yet while misspecified mental models have attracted substantial interest in the theory literature, corresponding empirical work is limited. For instance, work on the effects of computational complexity, and the pathways through which it operates is scarce.

Based on the framework above, the objective of the experiments to be reported below is to investigate (i) the importance of misspecified mental models for neglect and (ii) how computational complexity affects neglect through its impact on how mental models are formed. A key object of interest in this analysis will be the extent to which mental models and the resulting neglect patterns are context-dependent.

4.2 Computational Complexity

4.2.1 Experimental Design

The empirical analysis of mechanisms begins by studying how computational complexity affects selection neglect, in particular the ways in which it might induce cognitive load and hence distract participants from the selection problem. Thus, 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 problem (develop a mental model) in the first place. Thus, this paper takes a somewhat different approach to manipulating complexity than prior work by exogenously varying computational complexity in a way that is transparent about *how* and *why* complexity should matter.¹² 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

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

In *Simple*, it was

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

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

¹²For other work on complexity reductions see [Charness and Levin \(2009\)](#), [Enke and Zimmermann \(2019\)](#), and [Martínez-Marquina et al. \(2017\)](#).

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 “invisible” 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, the three treatments all hold the difficulty of accounting for selection constant but vary the computational burden of computing beliefs. An important 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 provide details on the signal realizations in these treatments.

4.2.2 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. These data arguably shed light on the complexity of the treatments in that higher computational complexity should translate into (i) longer response times and (ii) beliefs data that are noisier, or less consistent

Table 7: 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)
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)
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. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

across tasks.

To obtain an empirical estimate of decision noise that is invariant to the specific signal realizations in a given task j , I work with a subject-task-specific measure of decision noise $|\hat{\epsilon}_i^j|$ that follows the logic of the formal framework laid out in Section 2.3. Specifically, following equation (6), the estimate of decision noise is computed by comparing a subject’s belief in task j with the belief they “should have” stated given their estimated overall type $\hat{\chi}_i$:

$$|\hat{\epsilon}_i^j| = |\hat{\chi}_i^j - \hat{\chi}_i|, \quad (7)$$

where $\hat{\chi}_i$ is the overall estimate of i ’s type across tasks as derived in Section 3. Note that $|\hat{\epsilon}_i^j|$ is invariant to the overall level of neglect and only captures the consistency with which subjects state beliefs across tasks.

Table 7 shows that both response times and decision noise are indeed significantly lower in *Simple* and *Few*, as compared to *Complex*. In all regressions, the omitted base-line category is treatment *Complex*. The results provide reassuring evidence that the treatment variations actually induced meaningful variations in computational complexity as perceived by the experimental participants.¹³

¹³I have verified that quantitatively similar results hold when I implement median as opposed to OLS

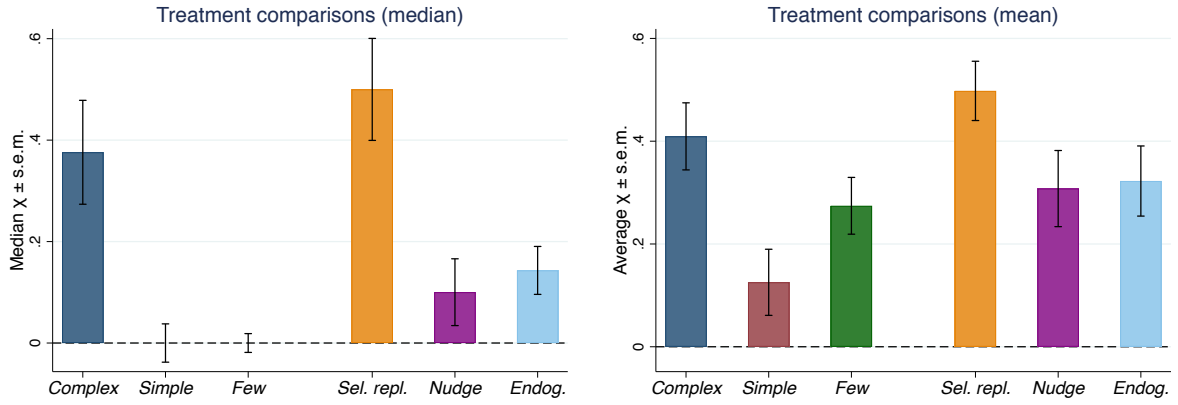


Figure 3: Overview of neglect $\hat{\chi}_i^j$ across treatments. The left panel shows the median $\hat{\chi}_i^j$ across all subject-task observations. The right panel shows the average $\hat{\chi}_i^j$ across all subject-task observations, where as in columns (3)–(4) of Table 5 the data are winsorized 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. Finally, also by design, in treatment *Endogenous* the analysis excludes those tasks in which a subject did not acquire information, so that no selection problem was present. As explained in the main text, these data exclusions follow mechanically from the construction of the different treatments.

4.2.3 Results

As explained above, by the design of the experiment, the analysis is restricted to those tasks in which subjects’ first signal was above 100 so that any “missing” signal had to be a 70 in all treatments.

Figure 3 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 8 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

regressions in the analysis of response times.

Table 8: Treatments *Complex*, *Simple*, and *Few*

	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. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

problem appears to have systematic effects on how participants approach the conceptual stage of forming a mental model to begin with.

4.3 Shifting People’s Mental Models

4.3.1 Variation I: Getting a Nudge

Experimental Design. If it is true that participants in *Selected* entertain a misspecified mental model, then nudging their attention towards (or reminding them of) the existence of the selection problem might attenuate the bias. I, hence, implement a treatment in which subjects are pointed to (or reminded of) the existence of information that is not visible on their decision screen. 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.

As detailed in Table 3 in Section 2, treatment *Nudge* was implemented along with a replication of treatment *Selected* to facilitate within-session randomization of subjects into treatments.¹⁴

Results. Figure 3 shows that treatment *Nudge* generates lower levels of neglect than *Selected replication*. Table 9 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 $\hat{\chi}_i^j = 0.30$.

4.3.2 Variation II: Endogenous Information Acquisition

Experimental Design. Treatment *Nudge* documents that *explicit* attentional nudges reduce neglect. A relevant question is whether the selection problem can also be brought to the top of mind through less stark and more indirect variations in the decision environment. In this respect, an arguably natural candidate is explicit endogenous information acquisition. Recall that in all treatments reported up to this point, the selection problem was induced *indirectly* through subjects stating a high or low first belief. Thus, participants were never required to *deliberately* choose between different information structures – rather, the prevailing information structure was a mechanical by-product of another financially incentivized choice. Arguably, the selection problem is likely to be more present in participants’ minds when they are explicitly asked to choose between receiving predominantly high or predominantly low signals.

To test this idea, treatment *Endogenous* implements a variant of treatment *Selected* in which subjects do not state a first belief that subsequently determines the behavior of the information source. Rather, after a subject observed their first signal, they were asked to choose between (i) acquiring information from information source I, (ii) acquiring information from information source II, or (iii) not acquiring any additional information. Acquiring either information source implies a cost of 0.50 Euros. Information source I, if chosen, delivered the same signals as those a subject saw in *Selected*, provided that they had stated a first belief higher than 100. That is, the experimental

¹⁴To additionally investigate whether subjects are, in principle, capable of computing the (admittedly relatively simple) 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. Figure 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.

instructions explained to subjects that if they were to choose information source I, they would get to observe all random draws above 100, but would observe at least three draws. Analogously, subjects knew that if they were to choose information source II, they would get to observe all random draws below 100, but would observe at least three draws. The option to not purchase any information source was provided to make the decision problem slightly more natural.¹⁵ Treatment *Endogenous* was implemented in separate experimental sessions with 71 subjects.

It is worth pointing out that the type of information acquisition problem that is studied here differs from other contexts that have attracted attention in the literature, in particular because there is no scope for motivated reasoning to protect cherished beliefs, and because there is no variation in the quality of the information sources.

Results. By design, the endogenous information choice induces a selection problem into the empirical analysis: it is now beyond the control of the experimenter whether subjects observe additional information signals or which ones. However, conditional on subjects choosing to purchase either information source, I can again compute an estimate of neglect $\hat{\chi}_i^j$ using the same techniques as before (following equation (6)).

In only 17% of all cases did subjects not acquire an information source. In these cases, by the construction of this treatment, I cannot compute a neglect parameter $\hat{\chi}_i^j$ because subjects do not face a selection problem when they state their second belief. The median subject acquires information in all eight tasks, and the average subject purchases information in 6.6 tasks. Among those cases where subjects do acquire an information source, 44% “follow their first signal” and choose the positively (negatively) selected information if they got a first signal above (below) 100.

Figure 3 shows that treatment *Endogenous* generates lower levels of neglect than *Selected replication*. The median level of neglect is $\hat{\chi}_i^j = 0.14$, while the mean is $\hat{\chi}_i^j = 0.34$. Table 9 provides a set of corresponding regressions. The comparison between the two treatments consistently goes in the same direction, but is sometimes only marginally statistically significant (the p-values in columns (9), (11), and (12) are 0.051, 0.053, and 0.083, respectively).

Figure 10 in Appendix C plots the distribution of $\hat{\chi}_i^j$, which has a clearly visible spike at exactly $\hat{\chi}_i^j = 0$. Moreover, the mass at $\hat{\chi}_i^j > 0$ and in particular at $\hat{\chi}_i^j = 1$ is reduced relative to *Selected replication*.¹⁶

¹⁵This setup is different from, but complements, a recent study on choices between biased information structures by [Charness et al. \(2018\)](#).

¹⁶As noted above, treatment *Endogenous* features a selection effect because assignment to an information structure (none, positively selected, negatively selected) is now endogenous and no longer mechanically determined by a subject’s first signal. Thus, as a robustness check, I leave the realm of analyses of the separate beliefs that subjects state in a given task, and again consider an estimate of their overall

Table 9: Treatments *Selected replication*, *Nudge*, and *Endogenous*

	<i>Dependent variable:</i> Neglect $\hat{\chi}_i^j$											
	<i>Selected replication and Nudge</i>						<i>Selected replication and Endogenous</i>					
	Median regression	OLS winsorized	OLS trimmed	Median regression	OLS trimmed	OLS winsorized	Median regression	OLS winsorized	OLS trimmed	Median regression	OLS winsorized	OLS trimmed
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
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)						
0 if <i>Selected repl.</i> , 1 if <i>Endogenous</i>							-0.36*** (0.10)	-0.34** (0.15)	-0.18* (0.09)	-0.23** (0.10)	-0.16* (0.08)	-0.18* (0.10)
Session FE	Yes	Yes	Yes	Yes	Yes	Yes	No	No	No	No	No	No
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	1174	1174	1174	1174	1154	1154	1068	1068	1068	1068	1051	1051
R^2	0.02	0.10	0.03	0.11	0.03	0.10	0.01	0.04	0.01	0.06	0.01	0.05

Notes. Regression estimates with robust standard errors (clustered at subject level) in parentheses. The sample in columns (1)–(4) includes treatments *Selected replication* and *Nudge*. In columns (5)–(8), the sample includes treatments *Selected replication* and *Endogenous*. In treatment *Endogenous*, the sample includes data from those subject-task combinations where an information source was purchased (since otherwise no selection problem exists). Columns (1)–(2) and (5)–(6) report median regressions, and all other columns OLS regressions. In columns (3)–(4) and (9)–(10), $|\hat{\chi}_i^j|$ is winsorized at 3. In columns (5)–(6) and (11)–(12), 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 Summary

Taken together, the evidence from the treatments aimed at shifting subjects' focus and manipulating computational complexity suggests that at least part of the reason why participants neglect selection in this context is that they do not focus on the selection problem in the first place. Instead, participants seem to be excessively focused on computationally implementing a solution strategy (which, however, rests on a wrong mental model). This being said, Figure 3 also shows that while selection neglect is sometimes substantially reduced through the various treatment variations, it rarely disappears, which provides further evidence for the robustness and strength of this error.

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 F 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.¹⁷

neglect type $\hat{\chi}_i$, as constructed in Section 3.3. This summary statistic is helpful because all but two subjects acquire an information in at least one task so that I can compute an estimate of neglect for almost all subjects. This allows me to gauge the relationship between neglect type and information acquisition to address the potential issue of sample selection bias in the analysis. Reassuringly, a subject's overall type $\hat{\chi}_i$ and the number of times they did purchase information are, in fact, weakly *positively* correlated ($\rho = 0.10, p = 0.42$). Thus, subjects that tend not to acquire information exhibit lower neglect than those that do acquire information, so that, if anything, the estimated treatment effect between *Selected* replication and *Endogenous* underestimates the true treatment effect.

¹⁷Apart 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

6 Placing the Results in the Context of the Literature

Recently, the topic of misspecified mental models has attracted considerable attention in the theory literature (e.g., [Eyster and Rabin, 2010](#); [Jehiel, 2005](#); [Gennaioli and Shleifer, 2010](#); [Schwartzstein, 2014](#); [Gabaix, 2014](#); [Spiegler, 2016](#); [Bushong and Gagnon-Bartsch, 2016](#); [Bohren and Hauser, 2017](#); [Bordalo et al., 2017](#); [Heidhues et al., 2017](#); [Gagnon-Bartsch et al., 2018](#)). In this vein, this paper has provided empirical evidence on the role of mental models in belief updating by examining the case of neglecting selection effects in information.

This section aims at providing a loose organizing framework to help place the results in this paper in the context of other recent empirical results on errors in belief updating. To do so, this section uses the language of directed acyclic graphs (DAGs, following [Spiegler, 2016](#)). Working with DAGs has the appealing feature that they allow for simple graphical representations of both objective data-generating processes and people’s mental models thereof. The objective is to informally suggest commonalities between different errors that are perhaps not obviously related from an *ex ante* perspective. Pointing out such commonalities may be helpful because if different errors are all generated by (loosely related) misspecified mental models, then they may be subject to similar comparative statics effects as those that are documented in this paper. In discussing these commonalities, the paper is loose and does not pretend to make an attempt at providing a formal model that unifies judgment errors.

The class of contexts that this section is concerned with includes belief-updating problems with two features: (i) they are backward-looking and (ii) they are characterized by data-generating processes that have “quirks.” First, “backward-looking” is intended to mean contexts in which economic agents draw inferences from signals that they have observed in the past, rather than inferences from future hypotheticals.¹⁸ Second, as discussed in greater detail below, by data-generating processes with “quirks” I mean situations in which people do not observe the underlying information signals *per se*, but rather messages that are functions of these signals. As will become clear below, the idea is that multiple errors in belief updating share the structure that the decision maker implicitly treats the messages he observes as if they *only* reflected the

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 F.6 for details.

¹⁸Other work seeks to understand interrelationships of a different class of errors that are related to conditioning on future hypotheticals. This includes cursedness in auctions and naïve (non-pivotal) voting, among others ([Esponda and Vespa, 2016](#)).

underlying signal of interest, rather than other factors as well. In the language of DAGs, this will amount to removing a single arrow from the DAG that represents the objective data-generating process.

Suppose a decision-maker (DM) is interested in estimating the realization of a random variable $\theta \sim f(\cdot)$. Nature draws θ (unobserved by the DM) and subsequently generates signals $s_i = \theta + \epsilon_i$, where the ϵ_i are all drawn independently from the same mean zero distribution function. The DM does not necessarily observe signal s_i but potentially message $m_i = g(s_i, x)$, where x and $g(\cdot)$ differ across the cognitive biases discussed below. The idea behind the next section is that, in processing m_i , people implicitly treat m_i as s_i and hence neglect the existence of x .¹⁹ This will amount to removing a particular causal arrow in a DAG.

Selection neglect. A stripped-down version of the setup in Section 2 can be represented by the true DAG on the left-hand side of Figure 4.

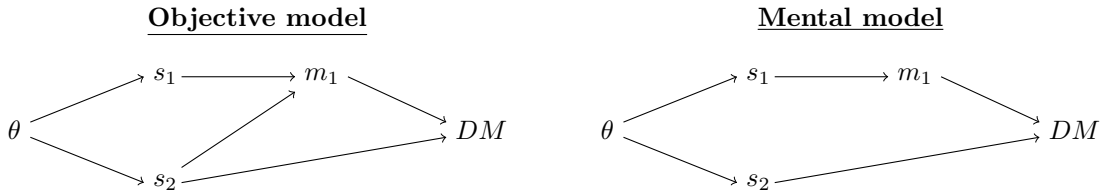


Figure 4: Selection neglect information structure represented as objective and subjective DAGs.

where $a \rightarrow b$ means that a (non-parametrically) causes b . Nature generates two signals, s_1 and s_2 , both of which are stochastically caused by the state θ . The DM observes s_2 and

$$m_1 = \begin{cases} s_1 & \text{if } (s_1 > c \wedge s_2 > c) \vee (s_1 \leq c \wedge s_2 \leq c) \\ \emptyset & \text{otherwise} \end{cases}$$

That is, the DM observes s_1 if it is similar to s_2 and “nothing” otherwise (again, \emptyset is purely colloquial here as the DM *does* observe a coarse version of the signal, albeit not on their screen). Thus, m_1 not only depends on s_1 but also on s_2 . However, we have seen above that, empirically, some people treat m_1 as if it reflected s_1 only. Neglecting the dependency of m_1 on s_2 amounts to removing an arrow so that a mental model can be represented by the subjective DAG on the right-hand side of Figure 4. This DAG corresponds to the belief patterns documented in Section 3.

¹⁹Mullainathan et al. (2008) call such inference “face value bias.”

Correlation neglect. Enke and Zimmermann (2019) present an experimental design to study correlation neglect, which refers to people’s propensity to double-count the informational content of signals that emerge through multiple messages. For example, people might neglect that the news articles of two different newspapers are both based on the same underlying report from a press agency and hence correlated conditional on the state.

A simplified version of Enke and Zimmermann’s setup can be represented by the true (objective) DAG on the left-hand side of Figure 5.²⁰ Nature again generates two signals. The DM observes $m_2 = s_2$ and $m_1 = \frac{s_1 + s_2}{2}$, two messages that are correlated conditional on θ because both contain s_2 . In the experiment, a large fraction of subjects neglects this correlation and updates beliefs as if m_1 only reflects s_1 . This can be represented by the subjective DAG on the right-hand side of Figure 5, i.e., by removing the causal arrow from s_2 to m_1 . Notice the similarity between the operation of removing the causal arrow $s_2 \rightarrow m_1$ and the case of selection neglect above: in both cases, exactly one arrow gets removed so that in the mental model, m_1 only depends on s_1 .

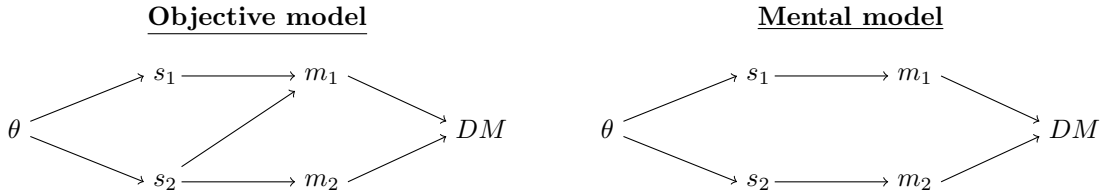


Figure 5: Correlation neglect information structure represented as objective and subjective DAGs.

Aggregation neglect. Experiments on learning in social networks routinely identify updating behavior that is best characterized by DeGroot’s rule, i.e., by averaging the beliefs of the neighbors (e.g., Grimm and Mengel, 2014; Chandrasekhar et al., 2015; Brandts et al., 2015). Among other aspects, DeGroot-type updating neglects the fact that the beliefs of some neighbors are more informative than the beliefs of others: because some agents have more neighbors than others, their beliefs reflect the average of a larger number of signals. For example, as illustrated by the DAG on the left-hand side of Figure 6, suppose Amy, Bob, and Charlie each receive a conditionally independent signal (nature generates s_1, s_2, s_3). Amy and Bob communicate their signals s_1 and s_2 , respectively, to Daniel, who averages these signals. Subsequently, both Daniel and Charlie communicate their opinions to the DM. Thus, the DM observes $m_2 = s_3$ and $m_1 = \frac{s_1 + s_2}{2}$.

²⁰See Eyster et al. (2015) for a related design.

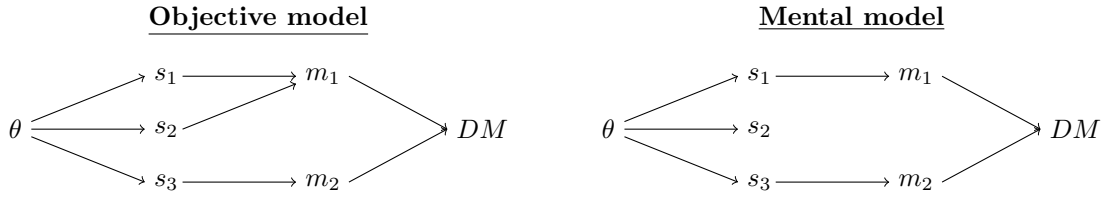


Figure 6: Aggregation neglect information structure represented as objective and subjective DAGs.

Here, the DM should place a higher weight on Daniel (m_1), but in network experiments, subjects typically average the messages they observe. This can be represented by the subjective DAG on the right-hand side of Figure 6, which again removes only one arrow from the objective DAG and hence eliminates the dependence of m_1 on s_2 .

Feature neglect, attribution bias, outcome bias, and failure to account for mean reversion. As highlighted by [Schwartzstein \(2014\)](#), a number of cognitive biases can be understood as variants of omitted variable bias. Again, suppose that a DM is interested in estimating θ and that they observe an outcome that is a function of both the state and other factors that are uncorrelated with the state: $m = s + x$, with $x \perp \theta$. Then, different types of x give rise to different biases that are discussed in the literature:

1. Attribution bias: The DM is interested in judging the quality θ of a good, but only observes the quality of the overall experience m , which also depends on external factors x . Attribution bias says that people overattribute the experience m to the state θ (e.g., people overattribute happiness in Disneyland to the park as opposed to the weather; see [Haggag et al., 2016](#)).
2. Outcome bias: The DM is interested in judging the type of another agent θ . The DM observes the outcome m of an agent's action, which not only depends on θ but also on luck x . Outcome bias says that people overattribute the outcome to the type of the other agent rather than luck ([Baron and Hershey, 1988](#)). A famous example of this is failure to account for mean reversion in the pilot story of [Kahneman \(2011\)](#): when asked to predict the future performance of a pilot based on a single extremely good (or bad) performance in the past, people overattribute the extreme performance to skill as opposed to luck and hence make excessively extreme predictions. See [Bertrand and Mullainathan \(2001\)](#) for related evidence in the context of CEO compensation.

Recently, [Graeber \(2018\)](#) presents a clean experimental design that shares the structure of these examples and identifies what is referred to as “feature neglect.” The DAG of the experiment is depicted on the left-hand side of Figure 7.

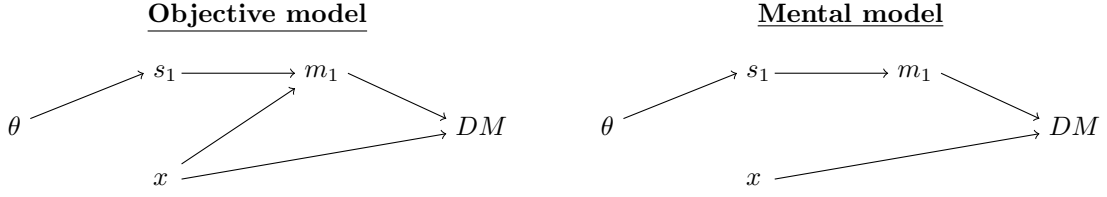


Figure 7: Feature neglect information structure represented as objective and subjective DAGs.

Nature generates one signal and one other random variable x that is uncorrelated with the state. The DM observes $m = \frac{s_1 + x}{2}$ as well as x and is asked to predict θ . In the experiment, subjects neglect the dependence of m_1 on x , i.e., when asked to predict θ they simply return m_1 even when it is not rational to do so. This corresponds to mentally removing the causal arrow from x to m_1 ; see the right-hand panel of Figure 7.

Summary. In line with the approach taken by an active recent theory literature, we have seen that various recent empirical results can potentially be understood by the idea that people can often entertain two potential models of the environment: the correct one and a simplified version in which the dependence of messages on factors other than the signal of interest is ignored. This review of the literature may be helpful if other errors in belief updating follow the same comparative statics patterns as the ones documented in Section 4. More theoretical and empirical work is needed to assess or formalize this possibility.

7 Conclusion

This paper has analyzed how people form beliefs when they need to learn from selected information samples. The results are twofold. First, the paper has provided a clean identification that people have a strong average propensity to neglect these selection problems when forming beliefs, even when the information-generating process is known and transparent.

Second, the paper has provided a detailed analysis of the mechanisms that give rise to biased belief updating. The key takeaway is the important role of misspecified mental models. As reflected by the type distribution of neglect, these mental models are binary in nature: subjects either employ a simplistic (and likely automatic) default model of the environment that ignores selection, 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 determined

by the structure of the environment, in particular the computational complexity of the decision problem, and the extent to which the decision maker's attention is drawn to the presence of selection. Working towards a goal of a hopefully unified theory of the occurrence of various judgment errors, more theoretical and empirical work is needed to understand the context-dependence of biases.

References

- Abeler, Johannes and Simon Jäger**, “Complex Tax Incentives,” *American Economic Journal: Economic Policy*, 2015, 7 (3), 1–28.
- 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.
- Baron, Jonathan and John C Hershey**, “Outcome Bias in Decision Evaluation,” *Journal of Personality and Social Psychology*, 1988, 54 (4), 569.
- 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.
- Bertrand, Marianne and Sendhil Mullainathan**, “Are CEOs rewarded for luck? The ones without principals are,” *The Quarterly Journal of Economics*, 2001, 116 (3), 901–932.
- 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.
- Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer**, “Salience and Consumer Choice,” *Journal of Political Economy*, 2013, 121, 803–843.
- , —, and —, “Memory, Attention, and Choice,” *Working Paper*, 2017.
- Brandts, Jordi, Ayça Ebru Giritligil, and Roberto A Weber**, “An experimental study of persuasion bias and social influence in networks,” *European Economic Review*, 2015, 80, 214–229.
- 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.

- Bushong, Benjamin and Tristan Gagnon-Bartsch**, “Learning with Misattribution of Reference Dependence,” 2016.
- , **Matthew Rabin**, and **Joshua Schwartzstein**, “A model of relative thinking,” *Unpublished manuscript, Harvard University, Cambridge, MA*, 2015.
- Chandrasekhar, Arun G., Horacio Larreguy, and Juan P. Xandri**, “Testing Models of Social Learning on Networks: Evidence From a Lab Experiment in the Field,” *Working Paper*, 2015.
- Charness, Gary and Dan Levin**, “The Origin of the Winner’s Curse: A Laboratory Study,” *American Economic Journal: Microeconomics*, 2009, pp. 207–236.
- , **Ryan Oprea**, and **Sevgi Yuksel**, “How Do People Choose Between Biased Information Sources? Evidence from a Laboratory Experiment,” *Working Paper*, 2018.
- Enke, Benjamin and Florian Zimmermann**, “Correlation Neglect in Belief Formation,” *Review of Economic Studies*, 2019, 86 (1), 313–332.
- Esponda, Ignacio 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.
- Fischbacher, Urs**, “z-Tree: Zurich Toolbox for Ready-Made Economic Experiments,” *Experimental Economics*, 2007, 10 (2), 171–178.
- Fodor, Jerry A.**, *The Modularity of Mind: An Essay on Faculty Psychology*, MIT press, 1983.
- 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.
- Haggag, Kareem, Devin Pope, Kinsey B Bryant-Lees, and Maarten W Bos**, “Attribution bias in consumer choice,” Technical Report, Working Paper 2016.
- 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*, 2016.
- 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.

- Kőszegi, Botond and Adam Szeidl**, “A model of focusing in economic choice,” *Quarterly Journal of Economics*, 2013, 128 (1), 53–104.
- Levy, Gilat and Ronny Razin**, “Segregation in Schools, the Echo Chamber Effect, and Labour Market Discrimination,” *Working Paper*, 2015.
- 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.
- McCaffery, Edward J. and Jonathan Baron**, “Isolation Effects and the Neglect of Indirect Effects of Fiscal Policies,” *Journal of Behavioral Decision Making*, 2006, 19 (4), 289–302.
- Mormann, Milica Milosavljevic and Cary Frydman**, “The Role of Salience and Attention in Choice Under Risk: An Experimental Investigation,” *Working Paper*, 2016.
- Mullainathan, Sendhil, Joshua Schwartzstein, and Andrei Shleifer**, “Coarse thinking and persuasion,” *The Quarterly journal of economics*, 2008, 123 (2), 577–619.
- 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.
- Schkade, David, Cass R. Sunstein, and Reid Hastie**, “What Happened on Deliberation Day?,” *California Law Review*, 2007, pp. 915–940.
- 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.
- Thagard, Paul**, *Mind: Introduction to Cognitive Science*, Vol. 4, MIT press Cambridge, MA, 1996.
- 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 10 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 11 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 10: Robustness: OLS regressions on full sample

	Dependent variable: Neglect $\hat{\chi}_i^j$							
	OLS regressions: Full sample							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
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)		
0 if <i>Selected repl.</i> , 1 if <i>Endogenous</i>							-0.17* (0.09)	-0.24** (0.10)
Session FE	Yes	No	Yes	Yes	Yes	Yes	No	No
Task FE \times prior	No	Yes	No	Yes	No	Yes	No	Yes
Controls	No	Yes	No	Yes	No	Yes	No	Yes
Observations	894	894	1177	1177	1174	1174	1068	1068
R^2	0.08	0.11	0.02	0.13	0.02	0.12	0.01	0.06

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

	Dependent variable: Neglect $\hat{\chi}_i^j$											
	Median regression		OLS winsorized		Median regression		OLS winsorized		Median regression		OLS winsorized	
	(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 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 2.3 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 2.3 for a derivation of the Bayesian and neglect benchmarks.

C Additional Figures

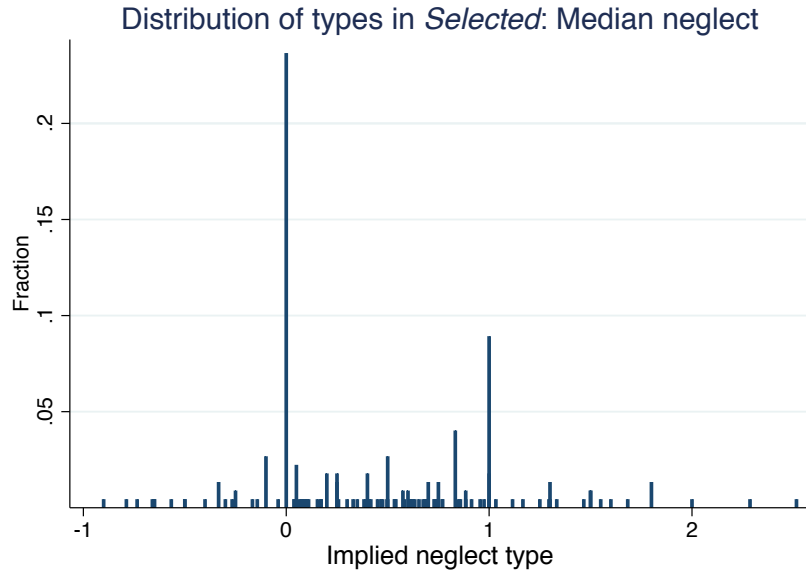


Figure 8: Distribution of neglect types $\hat{\chi}_i$ in treatment *Selected*. A subject's type is computed as median of $\hat{\chi}_i^j$ across tasks j .

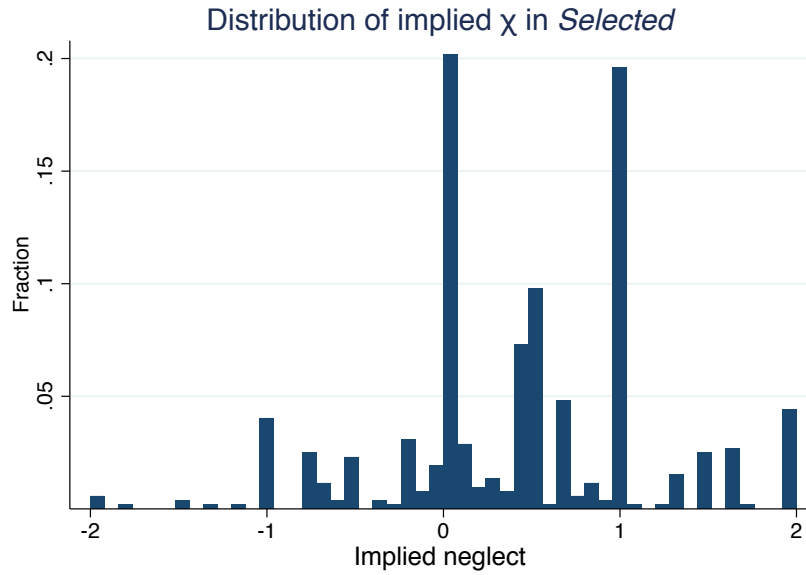


Figure 9: Distribution of implied $\hat{\chi}_i^j$ in treatment *Selected* in separate tasks. The figure plots the raw beliefs data across all subjects and tasks, normalized into units of χ according to eq. (6). That is, the data are not aggregated or rounded in any way. To ease readability, the plot excludes (i) beliefs from tasks in which a subject's first guess contradicted their private signal and (ii) beliefs with $|\hat{\chi}_i^j| > 2$.

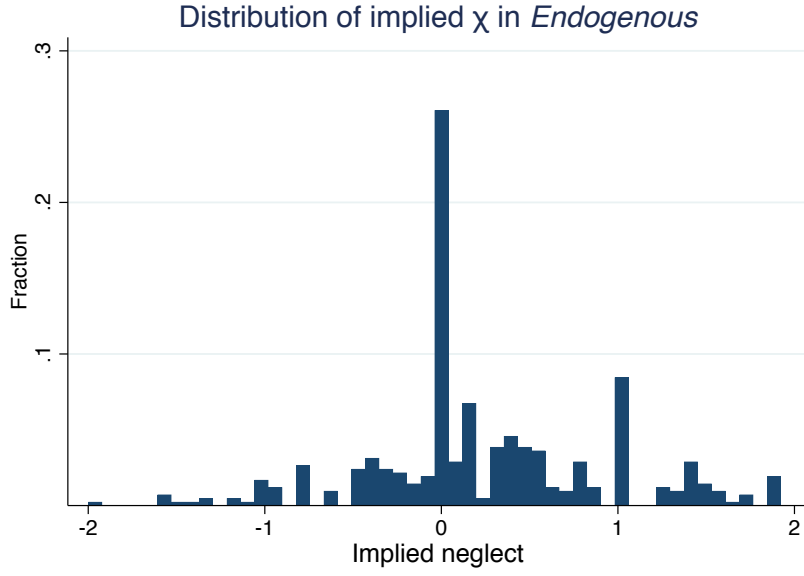


Figure 10: Distribution of implied $\hat{\chi}_i^j$ in treatment *Endogenous* in separate tasks. The figure plots the raw beliefs data across all subjects and tasks, normalized into units of χ according to eq. (6). That is, the data are not aggregated or rounded in any way. To ease readability, the plot excludes beliefs with $|\hat{\chi}_i^j| > 2$.

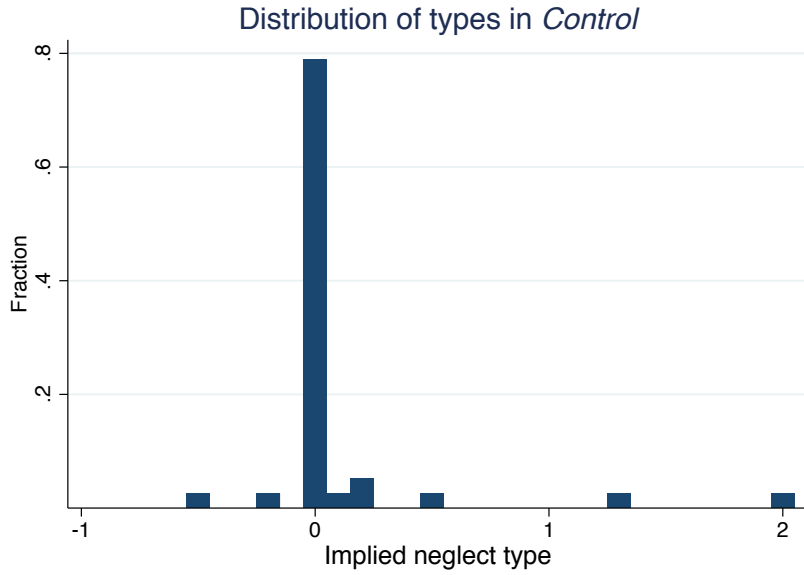


Figure 11: Distribution of neglect types $\hat{\chi}_i$ in treatment *Control*. A subject's type is determined based on the following procedure: for each subject i and candidate type $t \in \{-1, -0.9, \dots, 2\}$, I count how many of the implied $\hat{\chi}_i^j$ satisfy $|t - \hat{\chi}_i^j| < 1/20$.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

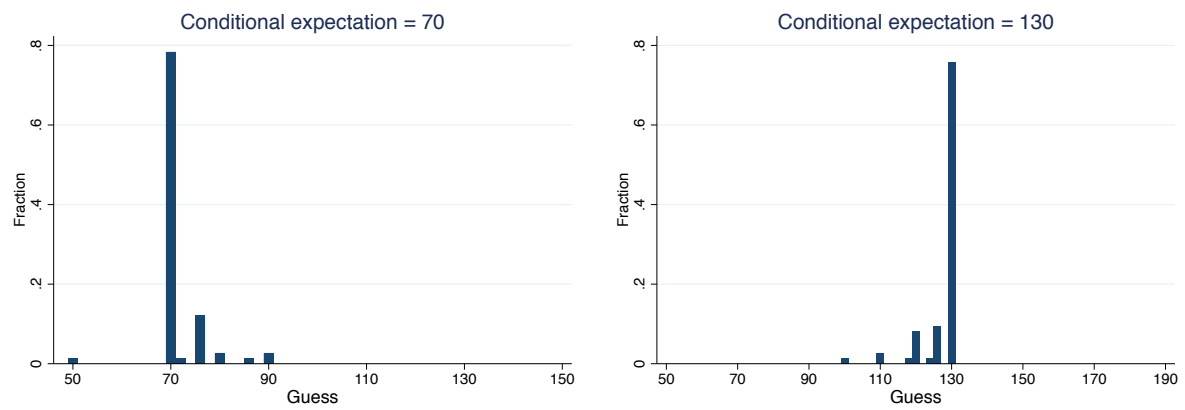


Figure 12: Guesses in conditional expectation tasks. The left panel presents the distribution of responses to a follow-up question in which subjects were asked to estimate the average of ten random draws, all of which are smaller than 100. The right panel follows an analogous logic, except that all ten 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 14 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 14: Treatment *Sequential*

	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>Sequential</i>	0.20*** (0.06)	0.25*** (0.09)	0.49*** (0.10)	0.53*** (0.11)	0.45*** (0.10)	0.45*** (0.10)
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	902	902	902	902	883	883
R^2	0.07	0.06	0.08	0.13	0.09	0.13

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 15 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 16 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 15: Comparing treatments *Feedback* and *Selected*

	Dependent variable: Neglect $\hat{\chi}_i^j$									
	Median regressions					OLS winsorized				
	(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)	-0.098 (0.10)	-0.012 (0.08)	0.026 (0.23)	-0.012 (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>			-0.0084 (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>				0.017 (0.10)			-0.0027 (0.09)			-0.039 (0.08)
Controls	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1190	1190	1190	1190	1190	1190	1190	1167	1167	1167
R^2	.	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 16: 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 Earlier Experiments: Replication

F.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 *Salience* 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>.

F.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.²¹ In the belief formation stage, all beliefs were restricted to be in $[0, 200]$ by the computer program.

²¹The 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.²² 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.

F.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

²²The 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

$$b_R = \frac{\sum_{v=1}^N s_v + \sum_{l=N+1}^6 E[s_l | g_l]}{6}. \quad (8)$$

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:

$$\begin{aligned} b_{SN} &= (1 - \chi)b_R + \chi b_N = (1 - \chi)b_R + \chi \frac{\sum_{i=1}^N s_v}{N} \\ &= b_R + \chi \frac{6 - N}{6} (\bar{s}_v - \bar{s}_l), \end{aligned} \quad (9)$$

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 .

F.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 9 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

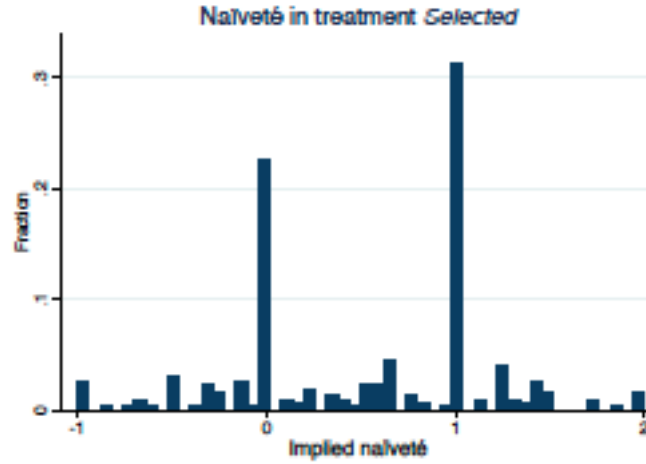


Figure 13: Distribution of naïveté in the *Selected* treatment. To ease readability, the figure excludes observations outside $\chi \in [-1, 2]$ (30 out of 336 obs.).

in columns (1) and (2) of Table 17 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).²³ 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 17 show that participants with better high school grades (a common proxy for cognitive ability) are significantly less likely to commit neglect ([Benjamin et](#)

²³Actual 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 17 investigate the relationship between subjects' naïveté χ (as implied in each belief, see eq. 9) 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).

F.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}.²⁴ 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-

²⁴To 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 17: 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 $|x_i^j| > 3$, 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*.²⁵ 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 18 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 18 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.²⁶

²⁵Note 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.

²⁶More 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 18: Representations, complexity and attention

	Dependent variable: Naïveté χ							
	<i>Intermediate and Simple</i>				<i>Saliency</i>			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
0 if <i>Simple</i> , 1 if <i>Intermediate</i>	0.31** (0.12)	0.32*** (0.12)						
0 if <i>Intermediate</i> , 1 if <i>Selected</i>			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 <i>Selected</i> , 1 if <i>Saliency</i>							-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 *Saliency* condition. All regressions exclude extreme outliers with $|\hat{\lambda}_i^j| > 3$, 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. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

F.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 19. 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 100.²⁷ Then, subjects stated a belief. Afterwards, they were shown the beliefs of two other randomly drawn subjects ("neighbors") from the same session.²⁸ 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.²⁹ 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

²⁷In 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.

²⁸This random matching was not constant across tasks.

²⁹The 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 19: 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

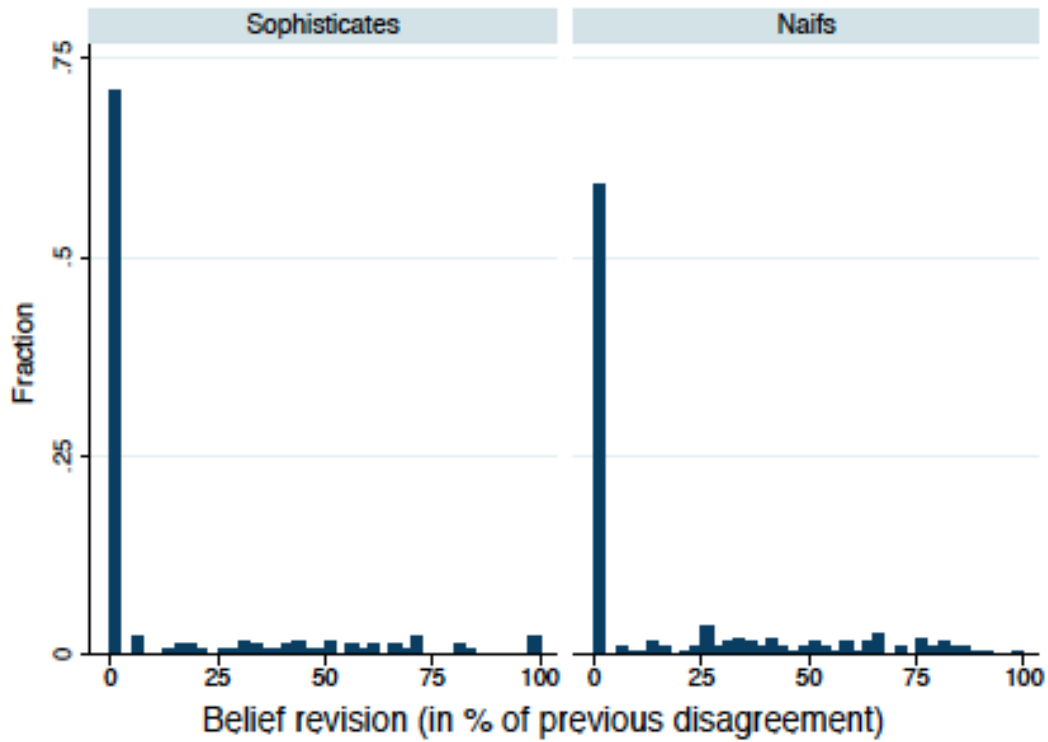


Figure 14: Magnitude of belief revisions. Each histogram depicts the belief revision between the first and second belief (expressed as percent of the difference between the first belief and the average belief of the two neighbors) conditional on the type of the subject (left / right panel). A subject 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.

G Experimental Instructions

G.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.

First Round

Information:
A-130
B-150

Please enter your first estimates.

Your estimate of variable A: ☐ Greater than 100 ☐ Less than 100

Your estimate of variable B: ☐ Greater than 100 ☐ Less 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.

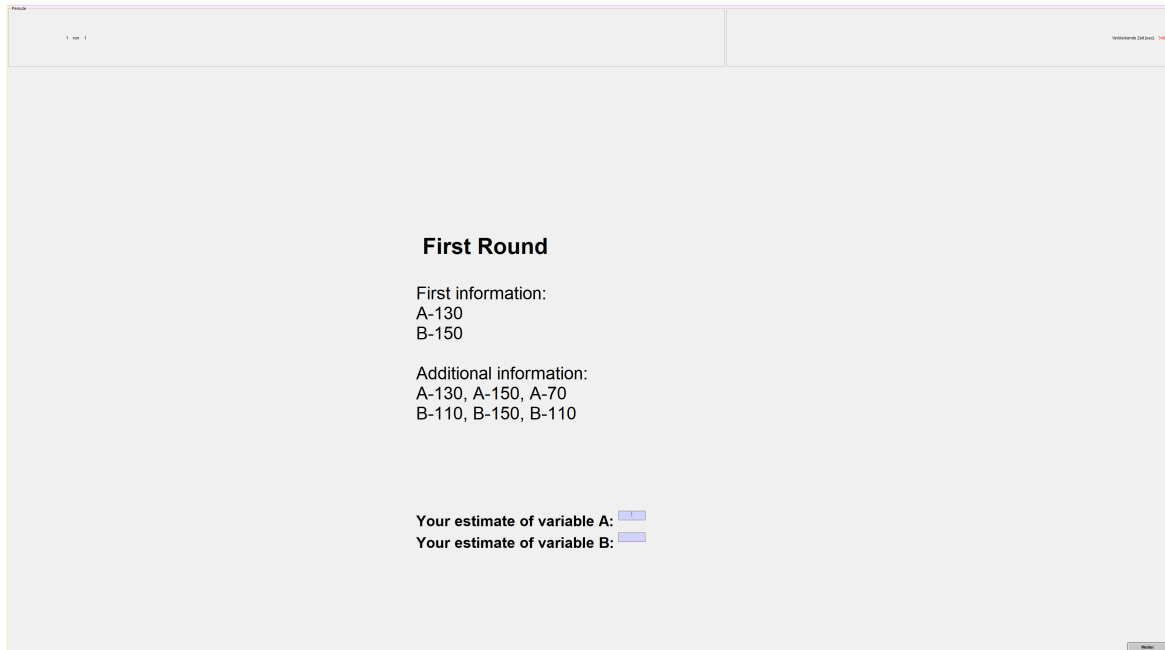


Figure 2: Example screenshot for the second estimates

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

G.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

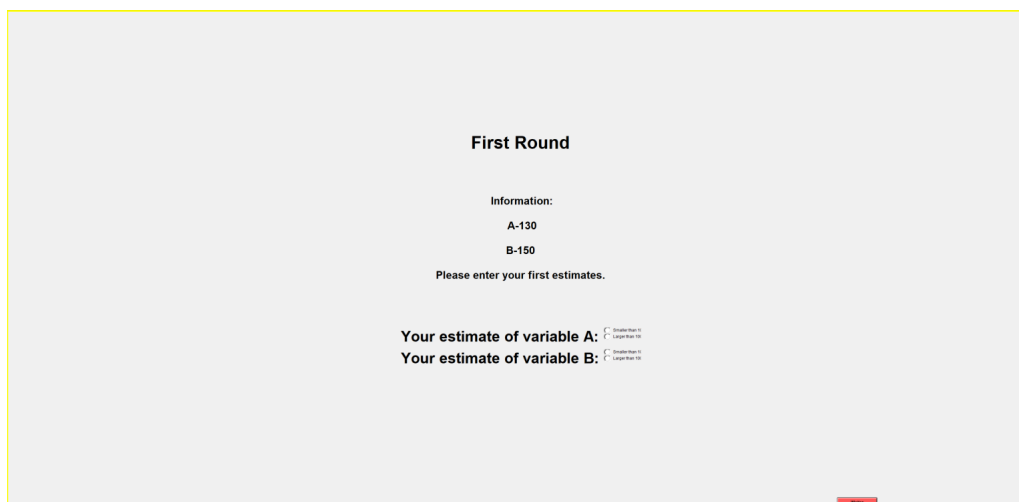
difference. For example, it is hence immaterial for your earnings whether you over- or underestimate the true value by 5.

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.



First Round

Information:
A-130
B-150

Please enter your first estimates.

Your estimate of variable A: ☐ Greater than 100 ☐ Smaller than 100
Your estimate of variable B: ☐ Greater than 100 ☐ Smaller than 100

Next

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

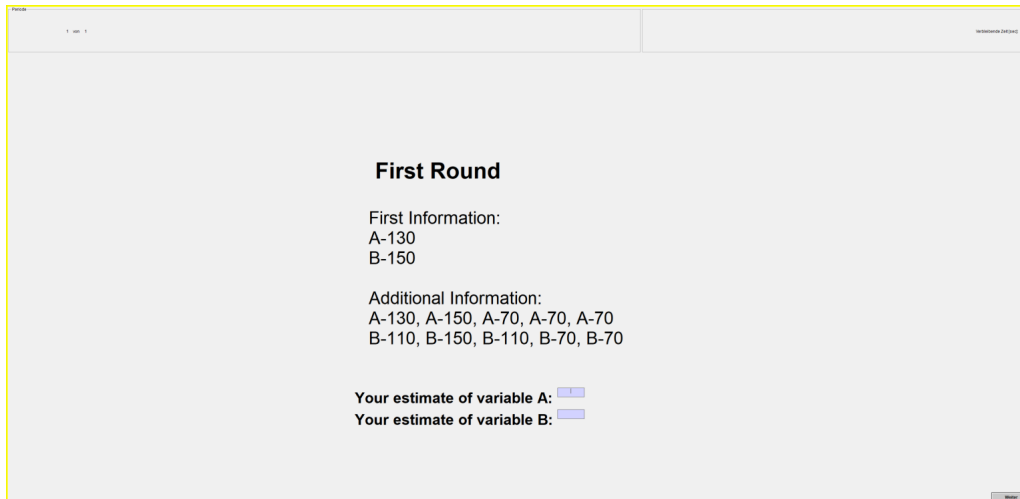


Figure 2: Example screenshot for the second estimates

instructions if you wish)

G.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.

G.4 Treatment *Feedback*

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

G.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.

G.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.

G.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.

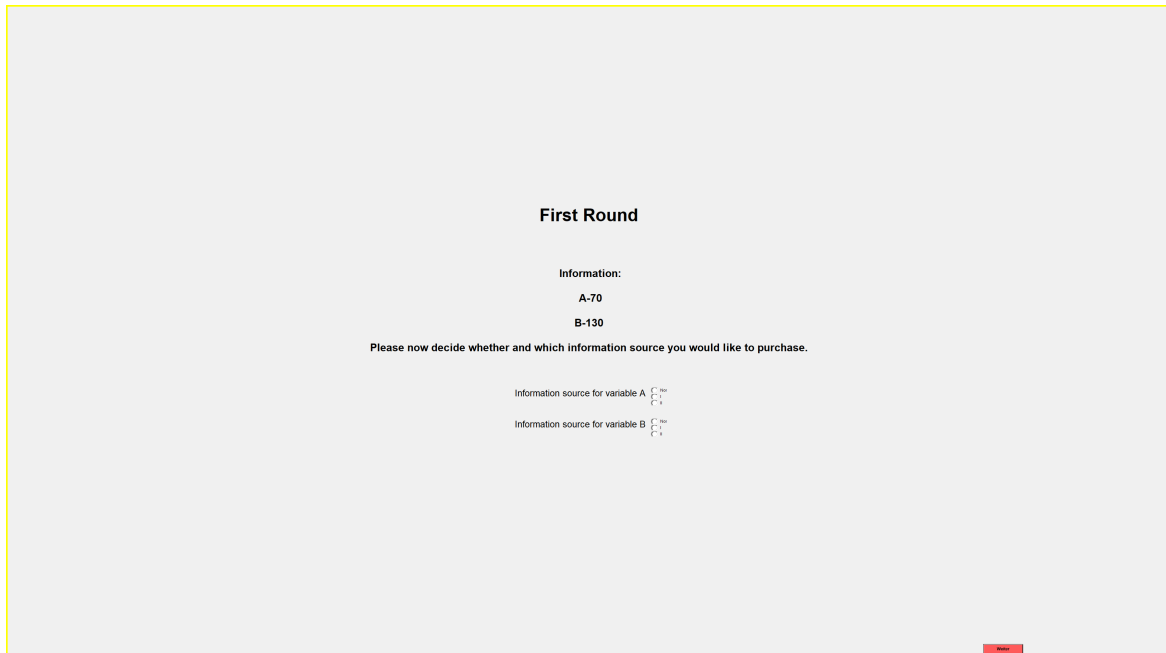


Figure 2: Example screenshot for the first estimates

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)

