What You See Is All There Is*

Benjamin Enke

September 23, 2017

Abstract

In many economic environments, people need to learn from systematic absences and non-occurrences. For example, news media differentially reports on events and corresponding non-events and social networks like Facebook selectively tailor their newsfeeds according to users' prior behavior. This paper studies people's belief formation rules in such selection contexts through a series of tightly structured experiments. Across various treatment variations, some subjects fully correct for selection, but many fail to take into account information they do not see and state beliefs that reflect *exactly* a full neglect benchmark. Follow-up treatments characterize the effects of environmental complexity on the presence of the updating bias and investigate the mechanism through which complexity operates. The results document that people's updating rule systematically depends on computational complexity and that complexity matters because it affects what people pay attention to: moderate complexity appears to induce cognitive busyness, which in turn distracts subjects from attending to the selection effect in the background of the process. Hence, most people can be debiased through a simple attentional nudge. However, exposing subjects to the conflicting beliefs of more rational participants is not sufficient to draw their attention to their mistaken reasoning.

JEL classification: D03; D80; D84.

Keywords: Bounded rationality; complexity; attention; beliefs.

*An earlier draft of this paper circulated under the title "Complexity, Mental Frames and Neglect". 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

This paper studies people's cognition in contexts that require *learning from something they do not see*. An important example of this phenomenon is selection problems. For instance, news media routinely cover certain events, but not the corresponding non-events: one rarely reads headlines such as "*No* terror attack in Afghanistan today." Related selection problems arise because people are often predominantly exposed to information that aligns with their priors. For example, people with similar beliefs mechanically tend to enter the same environments and are hence less likely to meet those with opposing views. Likewise, social networks like Facebook exclude stories from newsfeeds that go against users' previously articulated views, hence producing "echo chambers" (Bishop, 2009; Sunstein, 2009; Pariser, 2011; Mullainathan and Shleifer, 2005). Regardless of the specific mechanism, these contexts share the feature that *no news is news*, implying that people need to draw inferences from information they do not see. But even independent of selection effects, belief updating can require the need to comprehend the meaning of non-events. For example, when the stock market does not pick up after the announcement of a government stimulus, people again need to infer from "nothing" (Hearst, 1991).

Information structures with these characteristics pervade economic and social life and a nascent theoretical literature has indeed started linking the neglect of selection problems to issues such as belief polarization (Levy and Razin, 2015), savings behavior (Han and Hirshleifer, 2015), investment decisions (Jehiel, 2015), and network dynamics (Jackson, 2016). At the same time, little is known empirically about the cognitive strategies people employ to deal with information they do not see, whether and how belief updating depends upon the structure of the environment, and which psychological mechanisms may generate boundedly rational belief formation. After all, economists' interest arguably lies not only in documenting anomalies in statistical reasoning but also in the "first principles" underlying such errors (e.g., Fudenberg, 2006). This paper makes progress by (i) cleanly identifying people's updating rules in selection contexts and (ii) studying in detail the mechanisms underlying biased updating, including its dependence on environmental complexity and the reason why and how complexity matters.

Studying belief updating in the presence of absences and non-occurrences poses the challenge that people typically do not know the process that generates their information. This paper circumvents this problem by developing a tightly structured individual decision-making experiment that strips away all structural and strategic uncertainty. Subjects have to estimate an unknown state of the world and are paid for accuracy. In the beginning, one participant and five computer players each obtain a private signal over the state, drawn from the simple discretized uniform distribution $\{50, 70, 90, 110, 130, 150\}$. The true state that subjects need to estimate is given by the average of the random draws. Both the subject and the computer players then select into one out of two groups based on whether their respective signal is above or below 100. Thereafter, subjects are provided with the option to update their beliefs by observing the private signal of a subset of the computer players. This follows a simple and known rule: whenever subjects do not observe the signal of a given computer player, that player must have entered the opposite group. Thus, akin to many of the motivating examples, subjects predominantly observe the signals of computer players who received similar signals. Hence, subjects have to infer the expected signal of the other players from the fact that these players entered the other group and are not visible. Given the signal space, computing these conditional expectations is relatively straightforward. After subjects have observed the signals of the computer players, they guess the state. Crucially, the entire signal-generating process is computerized and known to subjects.

In a between-subjects design, I compare beliefs in this treatment with those in a control condition in which subjects receive the same objective signals as those in the main treatment condition, yet without a selection problem.

The results document that beliefs exhibit large and statistically significant differences across the two treatments, showing that people neglect information they do not see on average. Such neglect induces a specific form of path-dependence in beliefs: whenever a subject's private signal is relatively high (low), they select into the group in which they predominantly observe high (low) signals. Neglecting the resulting selection problem reinforces subjects' own signal and hence induces a belief pattern that is reminiscent of common notions of belief polarization across groups. These biased beliefs translate into significantly lower earnings in the treatment group.

These average patterns mask considerable heterogeneity. To characterize subjects' precise decision 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 pronounced bimodal structure: subjects either fully account for what they do not see or entirely neglect it. In particular, the vast majority of the neglect types compute beliefs that reflect *exactly* the "correct" solution, conditional on fully ignoring what they do not see. Thus, the neglect types appear to entertain a specific strategy and execute that strategy through effortful calculations. The bimodality of the type distribution is robust to a number of variations of the experimental design.

The pronounced bimodality in types and the observation that many subjects compute *exactly* the fully biased solution suggest that subjects form beliefs through effortful analytic calculations as opposed to quick, automatic, and intuitive responses. In line with this interpretation, the analysis documents that the relationship between neglect and response times – a commonly used proxy for cognitive effort in laboratory experiments (Rubinstein, 2007, 2016) – is quantitatively very small. In particular, the relationship is much too small to plausibly explain the observed difference in updating rules as a result of differences in effort.

The (reduced-form) identification of selection neglect constitutes only a first step in developing the empirical knowledge that is likely to be of use for economics. For example, Fudenberg (2006) and others have argued that the development of a set of empirical regularities about when and why a bias arises might support theoretical attempts to micro-found and unify updating biases. Hence, the paper next turns to studying the mechanisms behind the observed belief patterns. A plausible candidate explanation for the bias is that people "are bad at doing math" – e.g., that they cannot or do not want to compute the conditional expectations that are required in the present experiment. On the other hand, cognitive scientists routinely argue for the importance of mental representations for cognition. As discussed in greater detail in Section 4, researchers in the computational theory of mind partition thinking into (i) mental representations – the way in which people internally represent the external environment – and (ii) computations on those representations (e.g., Fodor, 1983; Thagard, 1996; Horst, 2011). Taking cognitive scientists' framework as point of departure, I study whether mental representations or computational skills generate biased updating in the present context and which factors make it more or less likely for people to develop a correct representation.

In the first step, I assess whether neglect is mostly driven by a lack of computational skills, i.e., problems in computing simple conditional expectations. For this purpose, I consider subjects' responses to an incentivized follow-up question that directly measures their ability to compute the same conditional expectations as in the baseline treatment. Here, a large majority of subjects provide the correct response and virtually all update in the right direction. Thus, subjects' problems in dealing with selection in this relatively simple setup are mostly not computational in nature. Rather – sticking with the terminology from the computational theory of mind – errors seem to arise due to a wrong mental representation.

How do such mental representations form and in which sense do they depend on the structure of the environment? To study this question, I investigate the dependence of the bias on the computational complexity of the environment and the mechanisms through which complexity shapes beliefs. After all, even abstracting from the context of selection, little is known about how and why complexity affects belief formation.

Following the spirit of work in cognitive psychology, the basic idea is that high computational complexity might induce “cognitive busyness” (Gilbert et al., 1988; Sweller, 1988) and hence “distract” people from attending to the selection problem that lurks in the background of the process. To test this idea, I develop two follow-up treatments. These treatments vary the overall computational complexity of the environment, while at the same time *fixing an extremely low difficulty of accounting for selection*. In particular, these treatments only differ in their signal space: $\{70, 70, 70, 110, 130, 150\}$ vs. $\{70, 70, 70, 130, 130, 130\}$. In both treatments, whenever a subject receives a signal above 100 and hence enters the “high signal” group, the selection problem can be overcome without quantitative reasoning by remembering that a missing signal deterministically implies a signal of 70. At the same time, the treatments differ in how computationally cumbersome it is to compute a full neglect belief. Because these treatments keep the difficulty of accounting for selection constant, complexity can only matter to the extent that it draws subjects' attention away from the selection problem in the first place. The results document that computational complexity indeed has a strong effect on stated beliefs. This suggests that complexity matters at least partly because it affects what people pay attention to and hence mentally represent decision problems.

If complexity indeed operated through attention allocation, it should be possible to debias subjects even in the more complex environment by directing their focus to the presence of the selection problem. To test this prediction, I conduct an additional treatment variation. Here, subjects are nudged towards the computer players whose signals they do not observe due to selection, while not providing any information about how to solve the selection problem. Even though the required computations are identical to those in the baseline treatment, this condition greatly reduces the share of neglect types and induces the majority of subjects to become fully sophisticated.

While this nudge evidence sheds light on the mechanisms underlying biased belief updating, such *direct* attention manipulations are rare in practice. Instead, in more natural contexts, people are likely to receive more *indirect* “hints” that might prod them to rethink their updating strategy. A prime example of such indirect nudges is the presence of disagreement: people are routinely exposed to the beliefs of others who might or might not share their own mental representation. To study whether observing disagreement induces subjects to reconsider their cognitive strategy, I ask a new set of subjects to solve the same belief formation task as those in the baseline treatment. Then, participants are provided with the beliefs of two other randomly drawn subjects who had access to the same information. All of this is publicly announced. Finally, beliefs are elicited again. The results show that both sophisticated and neglect types overwhelmingly abstain from revising their beliefs in response to their peers' assessments. Thus, in contrast to the more specific nudge treatment described above, disagreement as such does not shift subjects' attention to the missing pieces. Follow-up analyses suggest that this is because the neglect types are relatively confident in their own way of thinking about the problem.

In sum, (i) people neglect information they do not see on average; (ii) even a cognitively relatively homogeneous (student) subject pool exhibits systematic heterogeneity in the propensity to neglect; (iii) such neglect is driven by people's problem representations; (iv) environmental complexity affects how people represent and approach these types of problems, because it distracts them from the selection effect; so that (v) people are indeed capable of accounting for selection once their focus is exogenously steered towards this aspect. These results connect to the recent experimental literature on boundedly rational reasoning (Eyster et al., 2015; Bushong and Gagnon-Bartsch, 2016) that studies the roles of complexity (Abeler and Jäger, 2015; Esponda and Vespa, 2014, 2016a), attention (Enke and Zimmermann, 2015; Taubinsky and Rees-Jones, 2015; Mormann and Frydman, 2016), choice bracketing and framing (Rabin and Weizsäcker, 2009; Imas, 2016) and fast vs. slow thinking (Kessler et al., 2017).[1] What sets this paper apart from these contributions is the focus on selection problems under a known data-generating process as well as the detailed study of how complexity shapes beliefs through its effects on what people pay attention to and how they represent decision problems. While the presence of complexity is of course a key motivating observation for research on bounded rationality in general, very little, if any, work has identified a micro-foundation of why and how complexity affects reasoning.

The paper also connects to recent theoretical contributions in which mental representations induce biased beliefs (Gennaioli and Shleifer, 2010; Schwartzstein, 2014; Gabaix, 2014; Spiegler, 2016; Bohren and Hauser, 2017), or which emphasize the role of reminders and cues in decision-making (Bordalo et al., 2017). Even though such models cannot easily account for the evidence in this paper, the paper is also related to research on cursedness (Eyster and Rabin, 2005; Charness and Levin, 2009; Ivanov et al., 2010). Interestingly, recent work suggests that cursedness is pronounced in simultaneous-move environments but largely disappears in sequential contexts (Esponda and Vespa, 2014; Ngangoue and Weizsäcker, 2015). In contrast, this paper identifies boundedly rational updating in a purely sequential setup.

The paper proceeds as follows. Section 2 describes the experimental design. Section 3 presents the results and Section 4 studies mechanisms. Section 5 concludes.

[1] For evidence in selection setups in which people do not know the process that generates their data, see Esponda and Vespa (2016b) and Jin et al. (2016), or, in more qualitative non-incentivized setups, Brenner et al. (1996), Schkade et al. (2007), and Koehler and Mercer (2009).

2 Baseline Experiments

2.1 Experimental Design

Studying belief formation in contexts of selection requires (i) a task that allows fleshing out people's cognitive limitations and at the same time rules out affective reasons for holding certain beliefs, (ii) full control over the data-generating process, (iii) exogenous manipulation of the degree of selection, (iv) a control condition that serves as a benchmark for updating without selected information, and (v) incentive-compatible belief elicitation. Most importantly, a clean identification requires subjects' full knowledge of the data-generating process, i.e., a setup in which we know that subjects can in principle understand the statistical properties of those signals they do not see as a result of the selection mechanism. The present between-subjects design accommodates all these features.

The key idea of the design is to construct two sets of signals 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 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 $s_1, \ldots, s_6$ 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, as summarized in Table 1. 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

Table 1: Overview of the experimental design

Stage 0	Stage 1	Stage 2	Stage 3	Stage 4
Computer determines true state and generates six signals	Subject and five computer players each receive one private signal	Selection into blue or red group based on private signal	Subject learns signal of (subset of) computer players	Belief elicitation

Notes. Overview of the experimental design.

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.[2] 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

[2] I conducted a robustness check with a simplified selection rule. In treatment *Robustness*, subjects only observed the signals of all computer players in their own group. See Appendix C for details.

Table 2: Overview of the experimental tasks

True State	Private signal	Observed Signal A	Observed Signal B	Observed Signal C	Observed Signal D	Unobs. Signal E	Unobs. Signal F	Bayesian Belief	Sophisticated Belief	Naïve Belief
92.66	130	110	90	70	–	50	90	96.00	90.00	100.00
106.00	130	130	150	110	–	90	50	104.00	110.00	130.00
112.67	50	70	150	130	–	110	110	104.00	110.00	100.00
85.93	110	130	110	70	–	70	90	97.33	93.33	105.00
98.00	90	70	70	90	90	130	–	96.00	90.00	82.00
95.33	130	90	150	90	–	50	70	100.00	100.00	115.00
107.33	70	90	90	110	–	110	150	101.33	103.33	90.00

Notes. Overview of the belief formation tasks in order of appearance. The categorization into observed and unobserved messages applies to the case in which subjects follow their private signal, i.e., opt for the red group if their signal was larger than 100, and for the blue group otherwise. Subjects in the *Selected* treatment observed only their own signal as well as the "observed" messages. Subjects in the *Control* condition additionally had access to a coarse version of the "unobserved" messages, 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 2.2 for a derivation of the sophisticated and naïve belief benchmarks.

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. All subjects solved the same tasks, summarized in Table 2. 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.

Four features of this experimental environment are worth noting. First, the procedure induced a problem of absence and selection akin to the examples discussed in the introduction. This selection problem should intuitively be easy to understand, not just because subjects know the computer player's decision rules, but also because subjects actively select into a group themselves. Second, subjects' knowledge that they would talk to every computer player in their own group allowed participants to infer which types of observations they were missing. For example, if a subject was in the blue group and one computer did not talk to them, they knew that this computer had opted for the red group. Third, drawing signals from a simplified discretized uniform distribution ensures that computing the conditional expectation of the missing signals is rather straightforward and can be done, e.g., by choosing the middle option conditional on being above or below 100. Finally, the full data-generating process was exogenous and known, so that subjects knew how to interpret the computers' actions.

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.[3] In the belief formation stage, all beliefs were restricted to be in $[0, 200]$ by the computer program. Appendix F contains the experimental instructions and control questions.[4]

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.[5] 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., 2015). 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 $\pi^j = \max\{0; 18 - 0.2 \times (b^j - t^j)^2\}$,

[3]The 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 (see Section 4.2), subjects were allowed to continue to the experiment after the experimenter privately explained how to interpret the phrase "average signal".

[4]The instructions can also be accessed at `https://sites.google.com/site/benjaminenke/`.

[5]The 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.

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.

2.2 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 $i \in \{1, \ldots, 6\}$ 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, $E(s_i \mid g_a = \text{red}) = 130$ and $\mathrm{E}(s_i \mid g_a = \text{blue}) = 70$. Given some signals, a Bayesian would compute the mean posterior belief b_B as

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

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 $\mathrm{E}[m] = 100$. However, starting with Grether (1980), a long stream of research has 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

$$b_R = \frac{\sum_{v=1}^{N} s_v + \sum_{l=N+1}^{6} \mathrm{E}[s_l \mid g_l]}{6}. \tag{1}$$

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, see footnote 8 and Appendix B.1. 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:

$$\begin{aligned} b_{SN} &= (1-\chi)b_R + \chi b_N = (1-\chi)b_R + \chi \frac{\sum_{i=1}^{N} s_v}{N} \\ &= b_R + \chi \frac{6-N}{6} (\bar{s}_v - \bar{s}_l), \end{aligned} \tag{2}$$

where $\bar{s}_v \equiv 1/N \sum_{v=1}^{N} s_v$ is the average visible signal and $\bar{s}_l \equiv 1/(6–N–1) \sum_{i=N+1}^{6} E(s_l|g_l)$ 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.

Hypothesis. *Assuming that $\chi > 0$ (and $N < 6$), subjects' beliefs in the Selected condition are too high relative to the Control condition if the average of the visible signals is higher than the average expected non-visible signal, and vice versa.*

3 Results

3.1 Baseline Result

Result 1. *Beliefs significantly differ across treatments in the direction predicted by neglecting selection. Thus, beliefs in the Selected condition exhibit irrational path-dependence.*

Table 3 presents an overview of the results in each of the seven independent belief formation tasks. For ease of comparison, I provide the benchmarks of full neglect and sophisticated beliefs, respectively.[6] Reassuringly, beliefs in the *Control* condition follow the sophisticated prediction very closely, suggesting that the experimental setup was not systematically misconstrued by subjects: in the absence of selected information, people state sophisticated beliefs (abstracting from base rate neglect). In the *Selected* treatment, however, median beliefs are always distorted away from the sophisticated benchmark towards the full neglect belief. In all seven tasks do beliefs significantly differ between treatments at the 5% level (Wilcoxon ranksum test).[7]

To grasp the most basic implication of this updating bias, compare the second and seventh column of Table 3: whenever subjects' private signal is high ($s > 100$), the belief bias is positive. Conversely, when the private signal is low, the belief bias turns

[6]For completeness, note that, across tasks and treatments, virtually all subjects always enter the group that corresponds to their private signal realization. In total, in only 15 out of 546 group choice decisions did a subject enter the "wrong" group. In what follows, I exclude the beliefs from these particular subject-task combinations. All results are robust to including these observations or to excluding subjects that entered the wrong group at least once.

[7]Appendix B.3 visualizes the full distribution of beliefs in each task.

Table 3: Overview of beliefs across tasks

(1) True State	(2) Private Signal	(3) Sophisticated Belief	(4) Naïve Belief	(5) Median Belief *Control* Treatment	(6) Median Belief *Selected* Treatment	(7) Median belief bias	(8) p-value (Ranksum test)
92.66	High	90.00	100.00	90.00	100.00	10.00	0.0091
106.00	High	110.00	130.00	110.00	128.00	18.00	0.0001
112.67	Low	110.00	100.00	110.00	108.00	-2.00	0.0333
85.93	High	93.33	105.00	93.15	105.00	11.85	0.0001
98.00	Low	90.00	82.00	90.00	85.00	-5.00	0.0409
95.33	High	100.00	115.00	100.00	107.50	7.50	0.0001
107.33	Low	103.33	90.00	103.00	91.50	-11.50	0.0178

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 sophisticated and the naïve belief 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*.

out negative. Thus, in essence, neglecting information-based selection effects implies a form of irrational path-dependence: given a high prior belief (private signal), people select into an environment which on average reinforces their prior views, if selection is not taken into account. Thus, beliefs in the red and blue group end up being too extreme (on average), akin to common notions of belief polarization across groups.

In what follows, I will frequently work with a measure of subjects' beliefs that is independent of the specific updating task. To this end, I use equation 2 to compute the naïveté implied in each belief of subject i in belief formation task j:

$$\hat{\chi}_i^j = \frac{6(b_i^j - b_R^j)}{(6-N)(\bar{s}_v - \bar{s}_l)}. \tag{3}$$

Using this procedure, beliefs can be directly interpreted as reflecting sophisticated ($\chi = 0$), fully naïve ($\chi = 1$), or intermediate values.[8] Occasionally, I also work with a subject-level estimator of χ by computing the median naïveté across these seven values, i.e., $\hat{\chi}_i = med_j(\hat{\chi}_i^j)$. The OLS regressions reported in columns (1) and (2) of Table 4 then 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

[8] Recall that the precise identification of χ (though not the treatment comparison between *Selected* and *Control*) rests on the assumption of full base rate neglect. To justify this assumption, Appendix B.1 presents histograms of the naïveté χ implied in beliefs in treatment *Control*, separately computed with and without the assumption of base rate neglect. The figures show that – if full base rate neglect is assumed – the vast majority of beliefs reflect exactly $\chi = 0$. In contrast, when no base rate neglect is assumed, χ is extremely noisy and does not follow a clear pattern. In fact, without the assumption of full base rate neglect, *not a single* belief in *Control* reflects $\chi = 0$ when the rational belief does not equal the base rate. This set of results is to be understood as saying that *all* subjects in the *Control* treatment simply average the signals they observe, and do not take into account the base rate. Given this pattern, it seems very plausible to also assume full base rate neglect in the *Selected* condition.

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).[9] For comparison, the expected profit from being fully sophisticated in all tasks is €12.70.

3.2 Precise Belief Patterns and Correlates of Neglect

To develop a deeper understanding of subjects' precise belief patterns, I examine the distribution of estimated naïveté parameters χ. The left panel of Figure 1 plots the distribution of median naïveté parameters in *Selected*, and contrasts it with beliefs in the *Control* condition, where each subject is one observation. Visual inspection clearly shows that beliefs in the *Selected* condition exhibit a strongly bimodal distribution. While roughly 40% of participants are approximately sophisticated ($\chi = 0$), the majority fully neglects the selection problem. To show that the strong bimodality of types is not an artifact of the aggregation procedure of the seven beliefs per subject into one naïveté parameter, the right panel of Figure 1 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.[10]

These results are noteworthy because experimental beliefs data are typically quite noisy, due to, e.g., random computational errors or typing mistakes. But here, a large fraction of all beliefs are right at one of the extreme benchmark predictions, which is arguably inconsistent with quick, lazy, and intuitive thinking. This finding is also surprising in that it is sharply at odds with an ex ante plausible account in which people neglect selection problems on average, yet heuristically (partially) adjust from the fully

[9] Actual 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, $p = 0.0628$).

[10] The Appendix presents two additional – conceptually analogous – versions of the histogram in the right panel of Figure 1 by computing χ in two different ways. First, the construction underlying Figure 5 in Appendix B.1 does not assume base rate neglect. In such a framework, the implied χ are extremely noisy and very often outside the meaningful range, hence arguably providing support for the assumption of full base rate neglect. Second, Appendix B.2 describes an alternative updating rule in which people infer a signal of 100 from computer players whose signals they do not see. The evidence suggests that very few, if any, subjects followed such a rule.

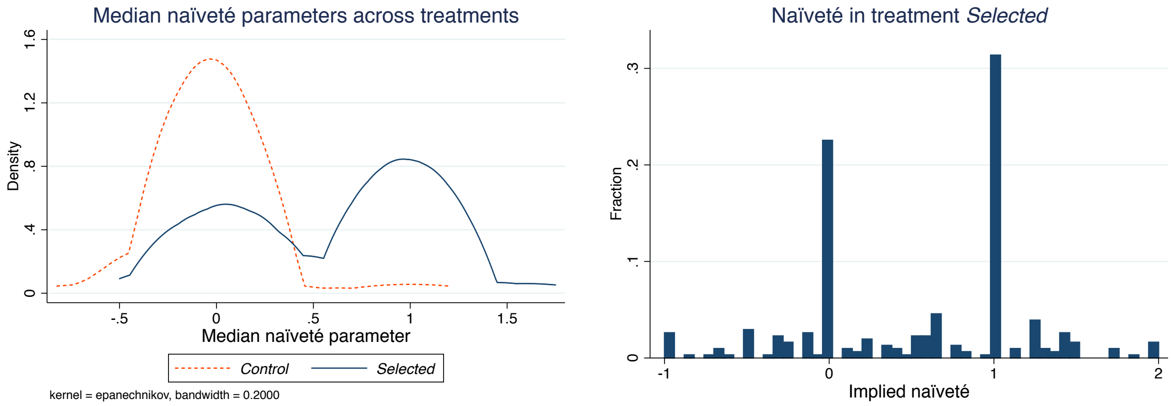

Figure 1: Distribution of naïveté in the *Selected* and the *Control* treatment. The left panel plots kernel density estimates of the median naïveté of each individual in both treatments, while the right panel illustrates the distribution of naïveté implied in all beliefs (7 per subject) in the *Selected* treatment. To ease readability, the right panel excludes observations outside $\chi \in [-1, 2]$ (30 out of 336 obs.).

naïve belief towards the sophisticated benchmark. Below, I will return to the observation of the pronounced bimodality in subjects' beliefs to build a case for the importance of (discrete) mental representations in dealing with selection problems.

Next, I examine basic correlates of biased updating within treatment *Selected*. Columns (3)–(4) of Table 4 show that participants with better high school grades (a common proxy for cognitive ability) are significantly less likely to commit neglect (Benjamin et al., 2013). Columns (5) and (6) show that neglecting selection is significantly correlated with correlation neglect, measured as in Enke and Zimmermann (2015).[11] 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. While these results need to be interpreted with care, they may suggest that correlation and selection neglect share common foundations that are more specific than low cognitive ability. In sum, these results suggest that people's tendency to neglect certain aspects of updating problems might be a somewhat stable trait that is linked to cognitive skills.

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). Indeed, a long literature in cognitive psychology has argued that updating biases are frequently the product of intuitive, quick, and mathematically effortless responses. Such thinking could either result from the intuitive reign of "system 1" in the sense of Stanovich and West (2000), Frederick (2005), and Kahneman (2003, 2011), or it could be driven by a conscious decision to economize on scarce cognitive resources, perhaps akin to costs of thinking models in economics (Gabaix, 2014; Caplin et al., 2006; Caplin

[11]32 out of 48 subjects in *Selected* agreed to take part in a follow-up study in which they solved five of the tasks used by Enke and Zimmermann (2015) to establish correlation neglect, see Appendix B.6.

Table 4: Correlates of neglect

	Dependent variable: Naiveté χ								
	Selected vs. *Control*		*Selected*						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1 if *Selected*	0.55***	0.62***							
	(0.09)	(0.09)							
High school grades		-0.14***	-0.27***	-0.29***		-0.075			-0.28***
		(0.05)	(0.07)	(0.07)		(0.09)			(0.07)
Correlation neglect parameter					0.29**	0.26**			
					(0.11)	(0.12)			
Response time (in minutes)							-0.25***	-0.24**	-0.20*
							(0.09)	(0.11)	(0.10)
# of consistent beliefs									0.027
									(0.05)
Constant	-0.019	-0.84*	0.61***	0.025	0.34***	0.50	0.76***	0.054	0.17
	(0.05)	(0.49)	(0.07)	(0.68)	(0.11)	(1.11)	(0.10)	(0.75)	(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*. See Appendix B.4 and footnote 13 for the derivation of the number of consistent beliefs per subject. All regressions exclude extreme outliers with $|\hat{\chi}_i^j| > 3$, but all results are robust to including these outliers. Controls include age, gender, log monthly income, and task fixed effects. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

and Dean, 2015) or as in the "cognitive miser" or "motivated tactician" metaphors of cognitive psychology (Fiske and Taylor, 2013).

In the data, the average response time across tasks and subjects in treatment *Selected* is 56 seconds. Columns (7)–(8) of Table 4 investigate the relationship between subjects' naïveté χ (as implied in each belief, see eq. 2) 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).[12] Instead, it seems possible that the difference in response times between sophisticated and neglect types reflects that these subjects have different solution strategies to begin with: sophisticates will need to work longer on the task because backing out the conditional expectation of the missing signals and computing the average of six numbers is more cumbersome than just computing the average of four numbers.[13]

3.3 Robustness Treatments

I conducted two robustness treatments to examine the extent to which the baseline result depends on particular features of the experimental design. Both treatments were simple variations of treatment *Selected*. First, in treatment *Robustness*, subjects went through the same procedures as in *Selected*, except that they *only* observed the signals of all computer players from their own group. Thus, in contrast to *Selected*, subjects never observed the signal of a computer player who did not enter their own group.

Second, in treatment *Base Rate*, the procedures were again identical to those in *Selected*, except that the true state was not determined by 15 random draws from the set X, but rather by only six draws. Given that a subjected interacted with five comput-

[12] Figure 14 in Appendix B.7 plots a histogram of the naïveté implied in subjects' beliefs, partitioned by whether the average response time is above or below 69 seconds, the median response time of sophisticated subjects ($\chi \leq 0.5$).

[13] If subjects' responses reflected careless answers, a perhaps natural conjecture is that such beliefs exhibit low within-subject consistency across tasks because they induce more noise (Choi et al., 2014). To evaluate this, I construct a measure of how often subjects state beliefs that are within a relatively small interval around some χ. Appendix B.4 discusses the construction of this measure in detail and shows that, overall, subjects exhibit an encouragingly high degree of consistency. As illustrated by the regression in columns (9) of Table 4, the index of consistency is virtually uncorrelated with subjects' updating type χ. This suggests that the neglect types do not state beliefs that wildly fluctuate across tasks or are otherwise more noisy than those of the sophisticates.

ers, this procedure implies that the number of signals equals the number of balls that determine the true state. Thus, in contrast to treatment *Selected*, there was no scope for selecting a base rate: for all balls that determined the true state, subjects either observed the ball or could in principle back out the conditional expectation from the group membership of the computer players. Thus, the "sophisticated" benchmark belief corresponds to Bayesianism in this treatment.

The results of both of these treatments are very similar to those in *Selected*, see Appendix C for details. In particular, the data again exhibit two pronounced spikes at $\chi = 0$ and $\chi = 1$, which shows that the bimodal type distribution hinges neither on assuming full base rate neglect, nor on the particular selection rule.

4 Mechanisms: Representation and Computation

4.1 Motivating Framework from Cognitive Science

Taking stock, we have seen that the neglect types consistently state beliefs that reflect exactly full neglect and spend almost as long on the experimental tasks as the sophisticated types. These patterns are suggestive that people *do* engage in specific and effortful mathematical calculations – just fundamentally wrong ones. The remainder of the paper is devoted to understanding why this is.

Casual accounts of cognitive biases frequently involve the notion that "people are notoriously bad at math". However, for the purpose of developing a set of empirical regularities that might provide inputs into theorists' attempts to micro-found and unify updating biases, the notion that "people are bad at math" is likely to be too vague. The paper thus proceeds by developing and experimentally testing a qualitative framework of cognition in selection contexts.

According to the dominant approach to conceptualizing cognition in cognitive science, in particular in the computational theory of mind, thinking can be partitioned into (i) mental representations and (ii) computations on those representations (Fodor, 1983; Thagard, 1996; Horst, 2011). While researchers in cognitive science do not offer a precise mathematical theory of these concepts, mental representations correspond to people's internal representations of the external environment, e.g., the way in which they subjectively perceive a data-generating processes, and what they pay attention to. Computations, on the other hand, refer to how the brain processes information within the aforementioned representational structures, i.e., computations are said to take representations as inputs or bases. In economics terminology, the distinction between representations and computations arguably comes close to the idea that people hold (potentially misspecified) subjective models of reality, and attempt to optimize conditional on these subjective models.

In the context of the present paper, people might fail at accounting for what they do not see either (i) because they mentally represent the problem in a wrong way and do not even pay attention to the missing pieces, or (ii) because they correctly represent the problem, but fail at optimizing appropriately, i.e., at mathematically backing out the missing signals. The pronounced bimodality in subjects' types strongly suggests that the processes of developing a correct representation and computationally backing out the missing signals can be thought of as binary.

4.2 Computational Skills

To investigate whether people possess the abstract skills that are necessary to compute rational beliefs *conditional* on being aware of the missing pieces, treatment *Selected* contained an incentivized follow-up question. This question allows to assess people's ability to compute simple conditional expectations:

> *In the course of this experiment, in total, you did not communicate with five computer players because you were part of the blue group, while these computer players opted for the red group. Based on this information, please estimate which signals these players in the red group have gotten, on average. You will receive an additional €2 if your guess is exactly right, 50 cents if your estimate is off by at most five, and nothing otherwise.*

46 out of 48 subjects provided a response above 100, which documents that virtually all subjects understood the setup and were capable of drawing qualitatively appropriate inferences from the behavior of the computer players. In addition, two thirds of all subjects provided *exactly* the correct conditional expectation of 130. But while participants do rather well in computing conditional expectations, their beliefs oftentimes exhibit full neglect in the actual experimental tasks, see Appendix B.5 for details. That is, even among those subjects that computed exactly the correct conditional expectations, many exhibit $\chi = 1$.

These patterns show that even the neglect types possess the computational skills to update correctly, at least directionally.[14] This result, in combination with the bimodality in subjects' types points to an important role of binary mental representations: after all, if subjects had the correct representation, why would they not at least partially adjust from the full neglect belief if they have the computational skills to do so?

[14]Of course, these findings should not be understood as suggesting that people *generally* do not fail at computing conditional expectations, in particular if the setup is more complex. However, the results do suggest that people already approach selection problems with a wrong representation of the problem in the first place, even if the underlying signal distribution is relatively simple.

4.3 Representations and Complexity

Hypothesis. If people's neglect of selection problems is generated through an incorrect mental representation, then that naturally raises the question of which environmental features make it more or less likely for people to develop the correct representation and hence state accurate beliefs. A prime candidate for a feature that might plausibly affect how people approach updating problems is environmental complexity. After all, different aspects of complexity have been shown to affect reasoning in a number of contexts (Charness and Levin, 2009; Enke and Zimmermann, 2015; Esponda and Vespa, 2016a, e.g.,), though it has not always been clear *why* complexity matters.

This paper takes a somewhat different approach than prior work by exogenously varying a particular type of complexity in a way that is clear about *how* and *why* complexity should matter. Specifically, consistent with psychological research, I investigate the hypothesis that high computational complexity might induce "cognitive busyness" (Gilbert et al., 1988; Sweller, 1988) and hence distract people from attending to the selection problem that lurks in the background of data-generating process. That is, I study whether increasing the computational complexity of the updating problem makes it less likely for people to attend to and correct for the selection problem, but *holding 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 is not changed, then differences in belief updating can plausibly be attributed to an effect of computational complexity on attention allocation.

Design. To test this mental comparative statics exercise, 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 $\{70, 70, 70, 130, 130, 130\}$.[15] 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.

[15]To 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*.

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 accounting 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*.[16] In total, 89 subjects participated in *Intermediate* and *Simple*, which were randomized within session.

To verify that computing a naïve belief is indeed perceived to be more complex in *Intermediate* than in *Simple*, I conducted a survey on Amazon Turk ($N = 209$, described in more detail in Appendix E. In this survey, I presented each participant with four pairs of math tasks that correspond to averaging the visible signals across the two treatments. I then asked participants to assess which task – if any – is harder. 6% of participants found the tasks in *Simple* harder, 73% those in *Intermediate*, and 21% found them equally difficult.

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*. Figure 2 depicts the results. The left panel visualizes the distribution of implied naïveté in subjects' beliefs in *Intermediate*, and the right panel shows the analogous patterns for *Simple*. The histograms suggest a stark difference: in *Simple*, subjects are still biased on average ($\bar{\chi} = 0.36$, $p < 0.01$), but the frequency of errors is much smaller compared to treatment *Intermediate*, even though the operation of accounting for selection is the same (and utterly simple, i.e., does not require any quantitative reasoning). To formally confirm this result, columns (1) and (2) of Table 5 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

[16]Note 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.

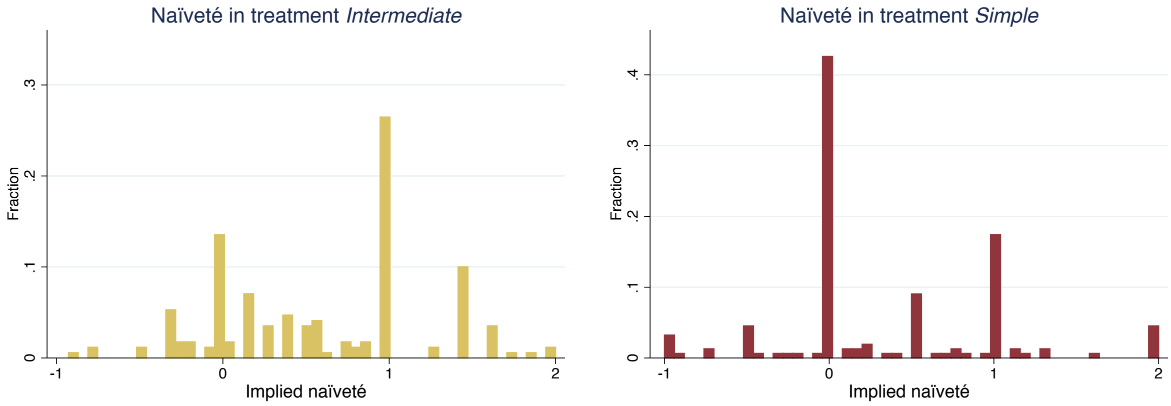

Figure 2: Distribution of naïveté in the *Intermediate* and the *Simple* treatments. The left panel plots the distribution of naïveté implied in beliefs in the *Intermediate* treatment, while the right panel illustrates the distribution of naïveté implied in beliefs in the *Simple* treatment. The sample is restricted to experimental tasks in which subjects' private signal satisfies $s > 100$. To ease readability, both panels exclude observations outside $\chi \in [-1, 2]$ (31 out of 356 obs.).

$\{110, 130, 150\}$, 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 5 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.[17]

Result 2. *Higher computational complexity leads to more neglect, holding fixed an extremely simple mental operation of accounting for selection.*

Of course, this result should not be interpreted as suggesting that the computational complexity of computing naïve beliefs is the *only* aspect of complexity that affects people's problem-solving approach. Rather, it should be viewed as a proof of concept: varying problem complexity affects what people pay attention to.

[17]More 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 5: Representations, complexity and attention

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

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

4.4 Attention: Nudge Evidence

Hypothesis. The aforementioned findings suggest that complexity affects reasoning in selection contexts through its effect on attention allocation (distraction). If this interpretation was correct, it should be possible to debias subjects even in the more complex environment by shifting their focus to the selection problem. This section seeks to provide such evidence.

Design. Treatment variation *Salience* is identical to treatment *Selected*, but additionally provided a hint both at the end of the instructions and on subjects' decision screens:[18]

> *HINT about the solution: Also think about the computer players whom you do not communicate with!*

This hint alerts subjects to reflect upon the missing computer players and the information they have gotten, but does not instruct them what to do about these missing signals. 48 subjects participated in this treatment and earned € 11.60 on average.

Results. The left panel of Figure 3 provides kernel density plots of subjects' median naïveté in this *Salience* treatment compared to the two baseline treatments, while the right panel plots the distribution of naïveté implied in all individual-level beliefs. As visual inspection suggests, this treatment had a large effect on subjects' beliefs relative to the *Selected* condition and reduced the fraction of neglect types by 60%. Notably, in this condition, most subjects develop beliefs that exactly reflect $\chi = 0$. To formally confirm the positive effect of this treatment variation relative to treatment *Selected*, columns (5) and (6) of Table 5 present the results of OLS regressions in which I regress the full set of seven beliefs per subject (expressed in units of naïveté χ) on a treatment dummy. The coefficient on the dummy suggests that increasing the salience of missing observations reduced the implied naïveté by about 0.3, relative to a level of about 0.5 in *Selected*.

Result 3. *Nudging subjects' attention to the missing pieces of information leads to a decrease in the fraction of neglect types by about two thirds.*

[18]The quote provided in the main text applies to subjects' decision screen. To prevent confusion, the hint at the end of the instructions reads: "HINT about the solution: When you estimate the number *X*, always also think about the computer players whom you do not communicate with!"

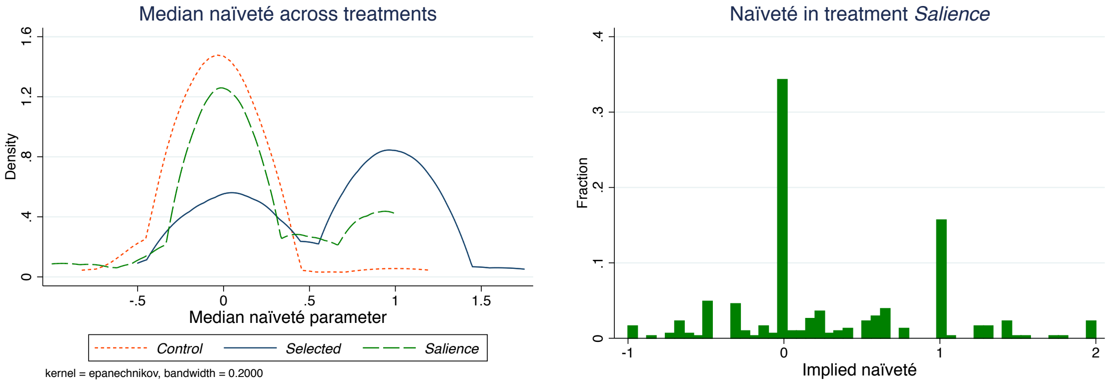

Figure 3: Distribution of naïveté in the *Selected*, *Control*, and *Salience* treatments. The left panel plots kernel density estimates of the median naïveté of each individual in all three treatments, while the right panel illustrates the distribution of naïveté implied in all individual-level beliefs in the *Salience* treatment. To ease readability, the right panel excludes observations outside $\chi \in [-1, 2]$ (30 out of 336 obs.).

4.5 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 6. 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.[19] Then, subjects stated a belief. Afterwards, they were shown the beliefs of two other randomly drawn subjects ("neighbors") from the same session.[20] 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

[19]In 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.

[20]This random matching was not constant across tasks.

Table 6: Basic timeline of treatment *Disagreement*

Part 1	Part 2			
Stage 0 – 4	Stage 0 – 3	Stage 4	Stage 5	Stage 6
As in *Selected* treatment	As in *Selected*, 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.

knowledge. After subjects observed the beliefs of their neighbors, they were asked to state a second belief.[21] 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 treatment is again bimodal with subjects being either fully naïve or sophisticated about the selection problem, see Appendix D.

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):

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

[21]The 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.

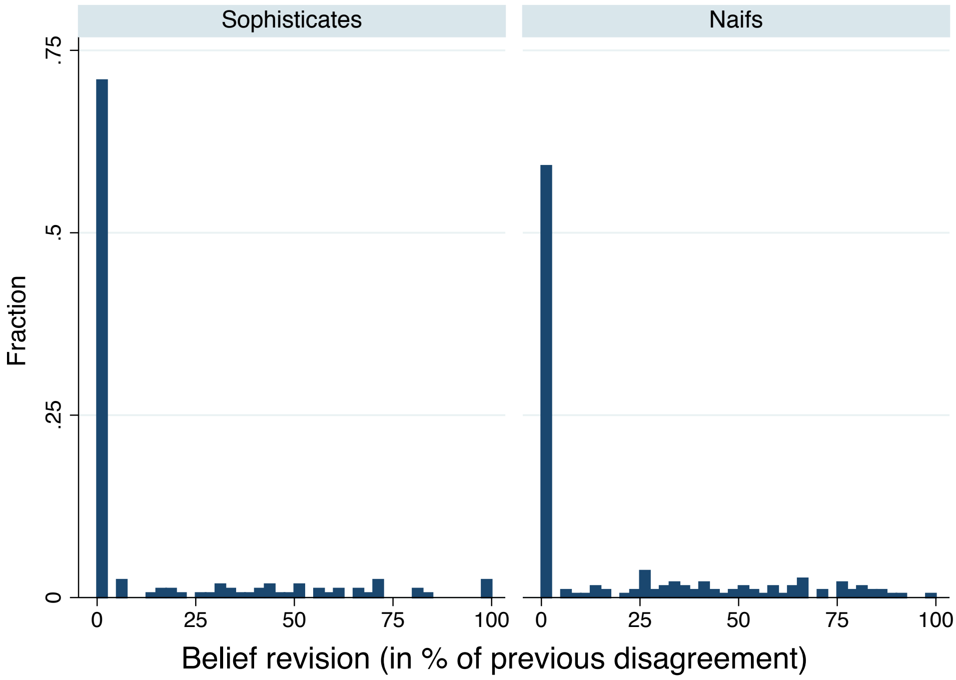

Figure 4: Magnitude of belief revisions. Each histogram depicts the belief revision between the first and second belief (expressed as percent of the difference between the first belief and the average belief of the two neighbors) conditional on the type of the subject (left / right panel). A subject 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. Appendix D.4 shows that very similar patterns hold if I classify subject-task observations based on both the out-of-sample naïveté parameter and the naïveté implied by the first belief in a given task. 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.).

where $\bar{\chi}^1_{-i}$ 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 4 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 $\chi \leq 0.5$.[22] The figure reveals that participants largely abstain from adjusting their

[22]The cutoff of $\chi = 0.5$ is arbitrary except that it denotes the midpoint of the naïveté interval. Appendix D.4 shows that very similar patterns hold if I classify subject-task observations based on both the out-of-sample naïveté parameter and the naïveté implied by the first belief in a given task.

beliefs 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.[23] Figure 20 in Appendix C shows that very similar patterns hold if I restrict attention to the subsample in which there is a large discrepancy between a subject's first belief and the average belief of the two neighbors. Appendix D.3 concludes this analysis by presenting an econometric analysis that facilitates an estimation of the weight that subjects implicitly assign to their own solution strategy relative to that of other subjects. The results show that the neglect types weight their own updating rule 6.7 times higher than that of another randomly drawn subject.

Result 4. *Observing disagreement by itself does not induce the neglect types to reconsider their problem representation.*

An interesting question is why the more direct nudge treatment succeeded in drawing subjects' attention to the selection issue, but disagreement as such did not. A plausible explanation (or interpretation) is that the neglect types appear to be relatively confident in their erroneous updating rule: when they observe others hold different beliefs, this does not immediately induce them to reconsider their solution strategy, which might in turn lead to noticing the selection problem. Appendix D.5 delves further into the question of how confident subjects are by discussing additional data on subjects' self-reported confidence levels. These data provide additional suggestive evidence that the neglect types are indeed almost as confident in their problem solving approach as the sophisticated types, hence bolstering the above explanation for why the more *indirect* "hint" of observing disagreement does not debias the neglect types.

5 Conclusion

This paper has provided an analysis of how people form beliefs when they need to learn from something they do not see. The results have shown that people have a strong average propensity to neglect the resulting selection problem, and that this average pattern masks strong heterogeneity: those types that neglect selection compute *exactly* the "correct" solution, conditional on fully ignoring what they do not see. This pattern suggests that people employ a specific strategy that they implement through effortful calculations. Folllow-up treatments provide evidence that this erroneous strategy does not reflect poor computational skills. Rather, people appear to develop a wrong mental

[23]Appendix D.6 investigates learning over time.

representation of the problem, but then optimize reasonably well within this representation. The paper has provided evidence that these representations are endogenous to the environment: in particular, the computational complexity of the environment affects how people mentally represent problems and what they pay attention to.

The findings in this paper may have an interesting subtle relationship to work on narrow bracketing (Gneezy and Potters, 1997; Rabin and Weizsäcker, 2009; Imas, 2016). In the present experiments, people appear to have a mental problem representation that is "too narrow" in the sense that it includes only those information pieces that are directly in front of them. While we do not have a formal theory of such behavior, it may be linked to behavior in contexts in which, say, prior decisions, are not taken into account. It would hence be interesting to explore whether the endogeneity of representations that is at the core of this paper extends to other choice contexts.

References

Abeler, Johannes and Simon Jäger, "Complex Tax Incentives," *American Economic Journal: Economic Policy*, 2015, *7* (3), 1–28.

Azrieli, Yaron, Christopher P. Chambers, and Paul J. Healy, "Incentives in Experiments: A Theoretical Analysis," *Working Paper*, 2015.

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.

Bishop, Bill, *The Big Sort: Why the Clustering of Like-Minded America is Tearing us Apart*, Houghton Mifflin Harcourt, 2009.

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, "Memory, Attention, and Choice," *Working Paper*, 2017.

Brenner, Lyle A., Derek J. Koehler, and Amos Tversky, "On the Evaluation of One-Sided Evidence," *Journal of Behavioral Decision Making*, 1996, *9* (1), 59–70.

Bushong, Benjamin and Tristan Gagnon-Bartsch, "Misattribution of Reference Dependence: Evidence from Real-Effort Experiments," 2016.

Caplin, Andrew and Mark Dean, "Revealed Preference, Rational Inattention, and Costly Information Acquisition," *American Economic Review*, 2015, *105* (7), 2183–2203.

_ , _ , and Daniel Martin, "Search and Satisficing," *American Economic Review*, 2006, *101* (7), 1043–1068.

Charness, Gary and Dan Levin, "The Origin of the Winner's Curse: A Laboratory Study," *American Economic Journal: Microeconomics*, 2009, pp. 207–236.

Choi, Syngjoo, Shachar Kariv, Wieland Müller, and Dan Silverman, "Who is (More) Rational?," *American Economic Review*, 2014, *104* (6), 1518–1550.

Enke, Benjamin and Florian Zimmermann, "Correlation Neglect in Belief Formation," *Working Paper*, 2015.

Esponda, Ignacio and Emanuel Vespa, "Hypothetical Thinking and Information Extraction in the Laboratory," *American Economic Journal: Microeconomics*, 2014, *6* (4), 180–202.

_ **and** _ , "Hypothetical Thinking: Revisiting Classic Anomalies in the Laboratory," *Working Paper*, 2016.

_ **and Emmanuel Vespa**, "Endogenous Sample Selection in Common Value Environments: A Laboratory Study," *Working Paper*, 2016.

Eyster, E. and M. Rabin, "Cursed equilibrium," *Econometrica*, 2005, *73* (5), 1623–1672.

Eyster, Erik, Matthew Rabin, 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.

Fiske, Susan T. and Shelley E. Taylor, *Social Cognition: From Brains to Culture*, Sage, 2013.

Fodor, Jerry A., *The Modularity of Mind: An Essay on Faculty Psychology*, MIT press, 1983.

Frederick, Shane, "Cognitive Reflection and Decision Making," *Journal of Economic Perspectives*, 2005, *19* (4), 25–42.

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.

Gennaioli, Nicola and Andrei Shleifer, "What Comes to Mind," *Quarterly Journal of Economics*, 2010, *125* (4), 1399–1433.

Gilbert, Daniel T., Brett W. Pelham, and Douglas S. Krull, "On Cognitive Busyness: When Person Perceivers Meet Persons Perceived," *Journal of personality and social psychology*, 1988, *54* (5), 733.

Gneezy, Uri and Jan Potters, "An Experiment on Risk Taking and Evaluation Periods," *Quarterly Journal of Economics*, 1997, *112* (2), 631–645.

Grether, David M., "Bayes Rule as a Descriptive Model: The Representativeness Heuristic," *Quarterly Journal of Economics*, 1980, *95*, 537–557.

Han, Bing and David Hirshleifer, "Visibility Bias in the Transmission of Consumption Norms and Undersaving," *Working paper*, 2015.

Hearst, Eliot, "Psychology and Nothing," *American Scientist*, 1991, *79* (5), 432–443.

Horst, Steven, "The Computational Theory of Mind," *Stanford Encyclopedia of Philosophy*, 2011.

Imas, Alex, "The Realization Effect: Risk-Taking After Realized Versus Paper Losses," *American Economic Review*, 2016, *106* (8), 2086–2109.

Ivanov, Asen, Dan Levin, and Muriel Niederle, "Can Relaxation of Beliefs Rationalize the Winner's Curse?: An Experimental Study," *Econometrica*, 2010, *78* (4), 1435–1452.

Jackson, Matthew O., "The Friendship Paradox and Systematic Biases in Perceptions and Social Norms," *Working Paper*, 2016.

Jehiel, Philippe, "Investment Strategy and Selection Bias: An Equilibrium Perspective on Overconfidence," *Working Paper*, 2015.

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, "Maps of Bounded Rationality: Psychology for Behavioral Economics," *American Economic Review*, 2003, *93* (5), 1449–1475.

_ , *Thinking, Fast and Slow*, Macmillan, 2011.

Kessler, Judd B., Hannu Kivimaki, and Muriel Niederle, "Thinking Fast and Slow: Generosity over Time," *Working Paper*, 2017.

Koehler, Jonathan J. and Molly Mercer, "Selection Neglect in Mutual Fund Advertisements," *Management Science*, 2009, *55* (7), 1107–1121.

Levy, Gilat and Ronny Razin, "Segregation in Schools, the Echo Chamber Effect, and Labour Market Discrimination," *Working Paper*, 2015.

Mormann, Milica Milosavljevic and Cary Frydman, "The Role of Salience and Attention in Choice Under Risk: An Experimental Investigation," *Working Paper*, 2016.

Mullainathan, Sendhil and Andrei Shleifer, "The Market for News," *American Economic Review*, 2005, *95* (4), 1031–1053.

Ngangoue, Kathleen and Georg Weizsäcker, "Learning from Unrealized Versus Realized Prices," *Working Paper*, 2015.

Pariser, Eli, *The Filter Bubble: What the Internet is Hiding From You*, Penguin UK, 2011.

Rabin, Matthew and Georg Weizsäcker, "Narrow bracketing and Dominated Choices," *American Economic Review*, 2009, *99* (4), 1508–1543.

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.

Stanovich, Keith E. and Richard F. West, "Individual Differences in Reasoning: Implications for the Rationality Debate?," *Behavioral and Brain Sciences*, 2000, *23* (05), 701–717.

Sunstein, Cass R., *Republic.com 2.0*, Princeton University Press, 2009.

Sweller, John, "Cognitive Load During Problem Solving: Effects on Learning," *Cognitive Science*, 1988, *12* (2), 257–285.

Taubinsky, Dmitry and Alexander Rees-Jones, "Attention Variation and Welfare: Theory and Evidence from a Tax Salience Experiment," *Working Paper*, 2015.

Thagard, Paul, *Mind: Introduction to Cognitive Science*, Vol. 4, MIT press Cambridge, MA, 1996.

A Treatment Overview

Table 7: Treatment overview

Treatment	# of subjects	Session length (mins)	Ave earnings (euros)
Selected	48	50	10.10
Control	30	50	13.70
Robustness	45	50	11.20
Base Rate	46	50	10.90
Salience	48	50	11.60
Intermediate	47	50	10.70
Simple	42	50	12.80
Disagreement	96	70	11.60

B Details and Robustness Checks for Baseline Treatments

B.1 The Assumption of Full Base Rate Neglect

The main text assumed full base rate neglect in order to be able to estimate χ as implied in each belief. This section justifies that assumption. To do so, I first derive the χ that is implied in subjects' stated beliefs without the assumption of base rate neglect. Following the notation introduced in Section 2.2, a belief can be expressed as:

$$b_{SN} = (1-\chi)\frac{9 \times 100 + \sum_{v=1}^{N} s_v + \sum_{l=N+1}^{6} \mathrm{E}[s_l \mid g_l]}{15} + \chi \frac{9 \times 100 + \sum_{v=1}^{N} s_v}{9+N}.$$

From this we can derive the implied χ of subject i in task j as

$$\hat{\chi}_i^j = \frac{15 \times (b_i^j - b_B^j)}{(900 + \sum_{i=1}^{N} s_v)/(9+N) - \sum_{l=N+1}^{6} \mathrm{E}[s_l \mid g_l]} \tag{4}$$

which may be contrasted with equation 3, the analogous equation when assuming base rate neglect. For both of these alternative ways of computing χ, I visualize the distribution in the *Control* treatment. The logic is that subjects' beliefs in *Control* serve as a natural benchmark for beliefs in *Selected* – if they exhibited full base rate neglect, then that would make the assumption of full base rate neglect in *Selected* arguably very plausible. The left and right panel of Figure 6 show that when one assumes base rate neglect, beliefs in the *Control* condition are tightly distributed around $\chi = 0$, with almost 60% of all beliefs at exactly that value. In contrast, without assuming base rate neglect (right panel), the implied χ are extremely noisy and span a wide range of meaningless values. In fact, the *only* few instances in which some subjects stated beliefs that reflect $\chi = 0$, the sophisticated benchmark with and the sophisticated benchmark without assuming base rate neglect are equal. These results show that subjects in the *Control* condition essentially only average the visible signals and fully neglect the base rate.

Figure 5 conducts an analogous exercise for the *Selected* treatment. Again, when one assumes full base rate neglect, as established in the main text, beliefs are tightly distributed around meaningful values and exhibit two large spikes at exactly $\chi = 0$ and $\chi = 1$. On the other hand, without assuming base rate neglect, the implied χ are again extremely noisy (note the support of the y-axis). Taken together, these results arguably strongly suggest that the assumption of full base rate neglect is confirmed by the data.

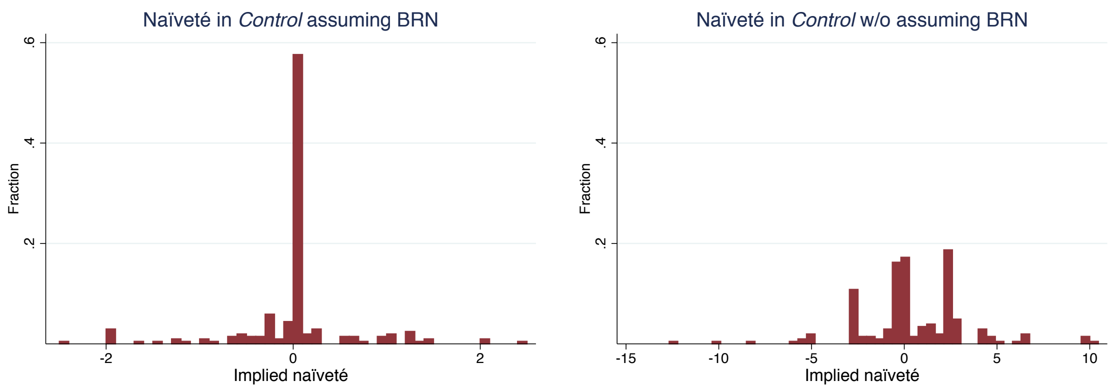

Figure 5: Distribution of naïveté in the *Control* treatment. The left panel depicts the distribution of the implied χ in equation 4 (i.e., when assuming base rate neglect), while the left panel depicts the analogous distribution without assuming base rate neglect, see equation 3.

B.2 Alternative Computation of χ

An alternative plausible updating rule in treatment *Selected* is to assume that subjects infer a signal of 100 from computer players whose signals they do not see. The logic behind 100 is that people might fail to take into account the informational content of the group entrance decisions of the computer players, but assign to them the base rate of 100, perhaps akin to cursed reasoning (Eyster and Rabin, 2005).

Figure 7 shows that the corresponding χ look overall reasonable, but the pattern are much weaker than when assuming that subjects infer nothing from the computer player's group entrance decisions. In particular, notice that sophisticated beliefs (i.e., fully taking into account the informational content of the computer player's actions)

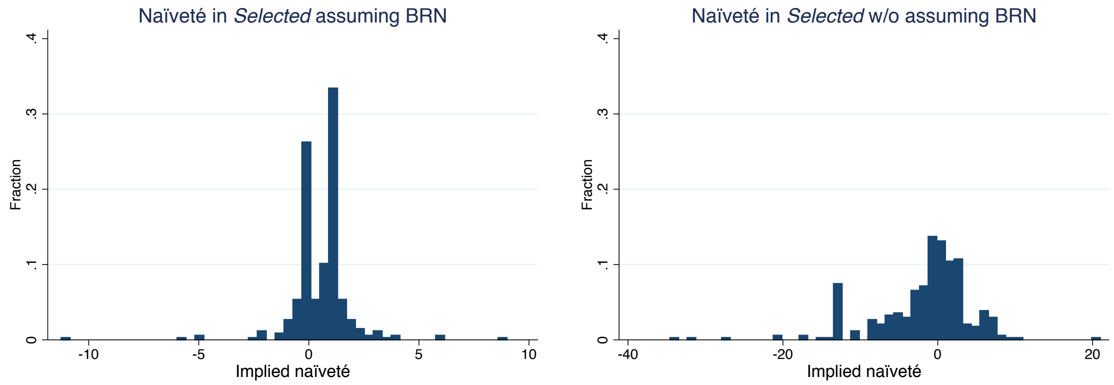

Figure 6: Distribution of naïveté in the *Selected* treatment. The left panel depicts the distribution of the implied χ in equation 4 (i.e., when assuming base rate neglect), while the left panel depicts the analogous distribution without assuming base rate neglect, see equation 3.

gives rise to $\chi = 0$ in both methods. Thus, the only way to separate the two updating rules is by considering beliefs that reflect naïve updating. Because fully naïve beliefs are often very close to each other in both methods of computing χ, it is conceivable that beliefs of $\chi = 1$ in Figure 7 actually reflect computations that were aimed at computing a fully naïve belief under the updating rule discussed in the main text.

The perhaps clearest way to see that (probably all) subjects employed the updating rule described in the main text as opposed to the one described here is by analyzing how many beliefs are in $|\chi - 1| < 0.05$, i.e., within a relatively narrow interval around full naïveté. It turns out that 96 beliefs satisfy this criterion when χ is computed as in the main text. Of these 96 beliefs, 69 are outside of the narrow interval when χ is computed using the method described in this Appendix. In contrast, only 29 beliefs are in $|\chi - 1| < 0.05$ when beliefs are computed using the alternative updating rule. However, of these 29 beliefs, 27 are also within the narrow interval when χ is computed as in the main text. The upshot of this discussion is that there is very little, if any, evidence for the alternative updating rule when the one described in the main text yields a different prediction.

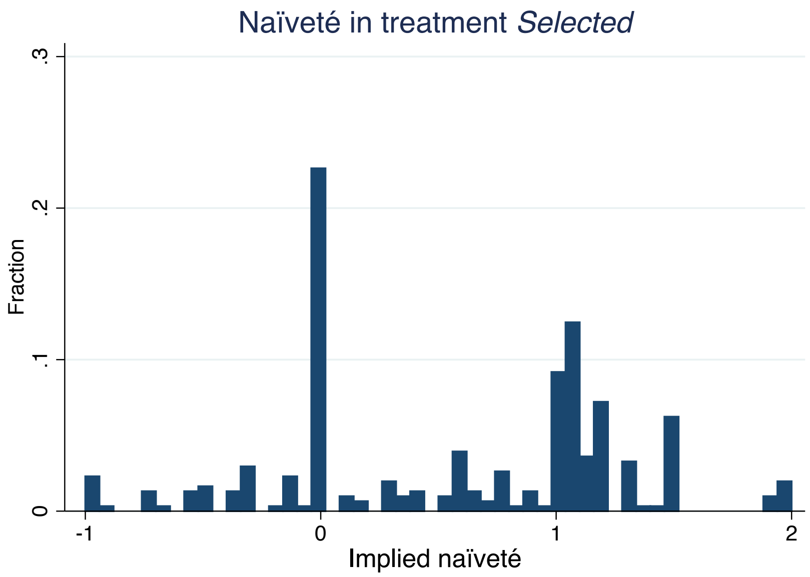

Figure 7: Distribution of naïveté in the *Selected* treatment. The distribution of naïveté is computed by assuming that subjects infer a signal of 100 from computer players whose signals they do not observe.

B.3 Kernel Density Estimates for each Task

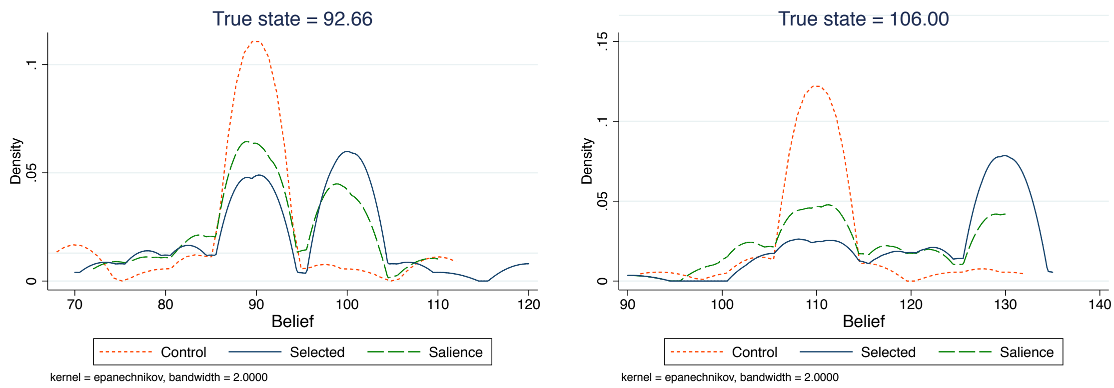

Figure 8: Distribution of beliefs by task (1/2). To ease readability, the plots exclude extreme outliers whose distance to both the fully naïve and sophisticated benchmarks is larger than 20.

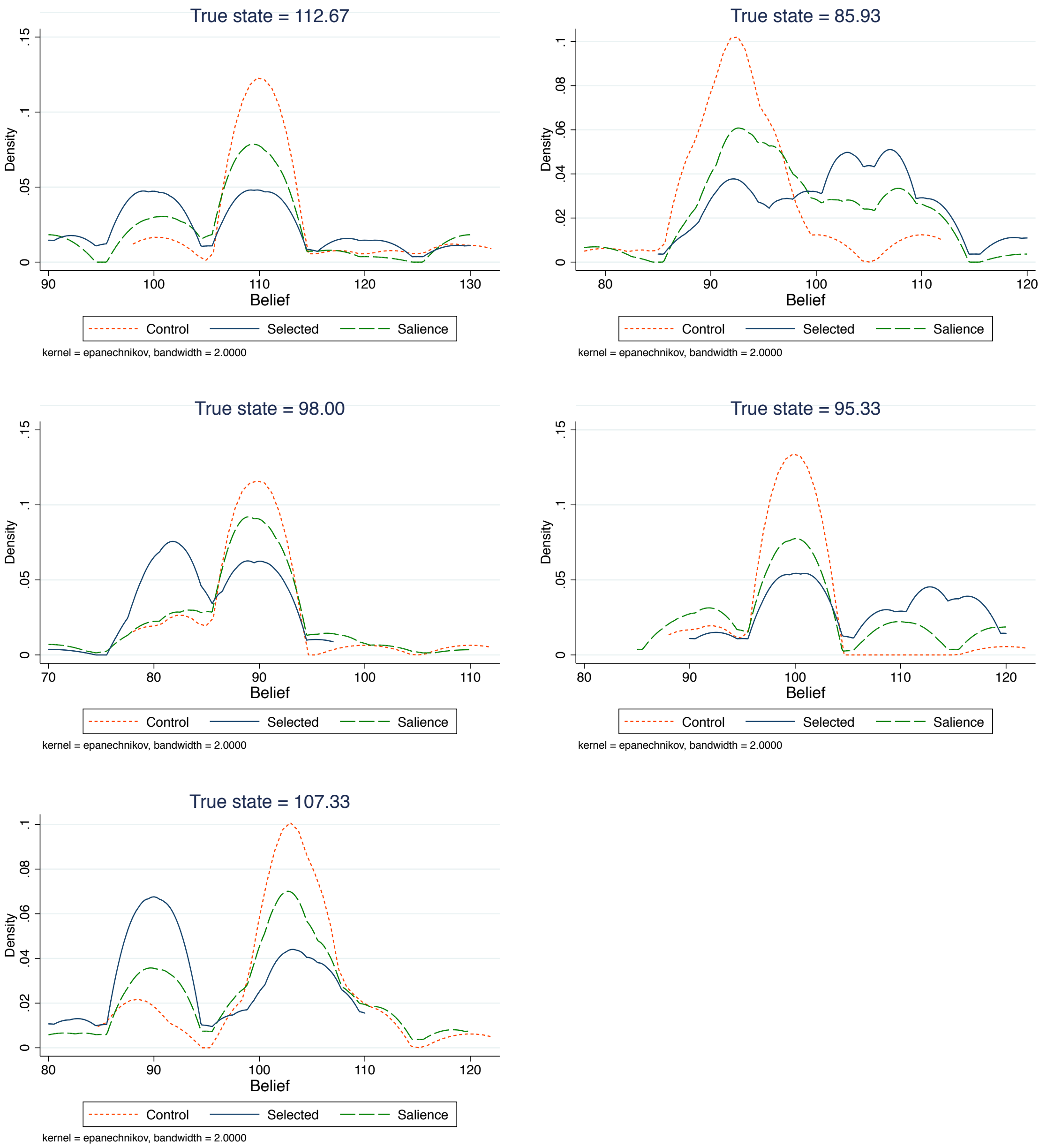

Figure 9: Distribution of beliefs by task (2/2). To ease readability, the plots exclude extreme outliers whose distance to both the fully naïve and sophisticated benchmarks is larger than 20.

B.4 Consistency of Beliefs Across Tasks

This section investigates the consistency with which subjects in *Selected* exhibit a certain degree of naïveté across tasks. To this end, I define a set of potential types $\chi = -2/3, -1/3, \ldots, 8/3$. Then, for each individual and each χ, I count the number of beliefs which reflect naïveté in $[\chi - 1/3, \chi + 1/3]$. Denote the number of beliefs in this interval as n_χ. Finally, I take the maximum over all n_χ, for each individual. This maximum represents the number of beliefs that exhibit a certain degree of consistency. Figure 10 presents a histogram of this measure, which reveals that the vast majority of subjects state at least three consistent beliefs, and 70% of all subjects state at least four consistent beliefs. Thus, overall, subjects' responses reflect a considerable degree of consistency.

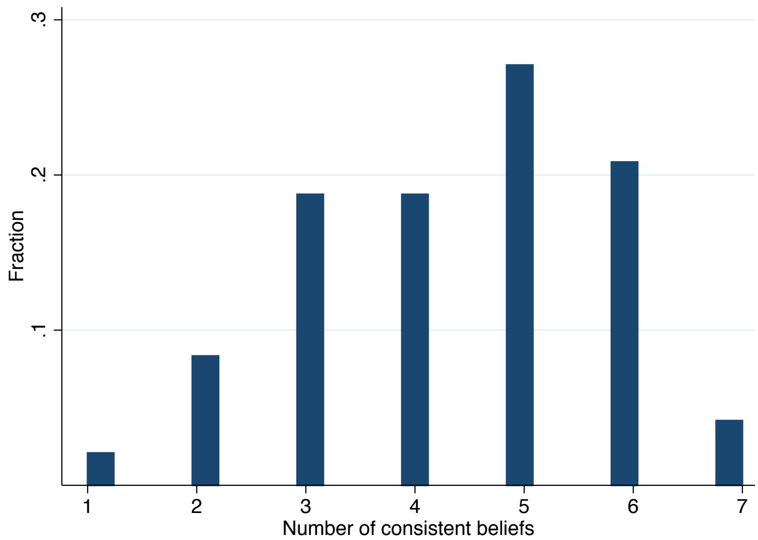

Figure 10: Number of consistent beliefs in treatment *Selected*.

B.5 Follow-Up Question

The left panel of Figure 11 plots the distribution of responses to the conditional expectation follow-up question. Almost all subjects provided a response above 100 (i.e., they understood at the direction in which they had to update), and roughly two thirds answered 130.

The left subpanel of the right panel of Figure 11 shows that the vast majority of subjects who provided a response larger than 100, but did not answer 130, exhibit full selection neglect ($\chi = 1$). Even more puzzling, those subjects that provided *exactly* the correct response of 130 (depicted in the right subpanel), also exhibit strong heterogeneity in their naïveté. While the fraction of sophisticated subjects is higher in this subgroup, many people still fully neglect the selection problem in the belief formation tasks.

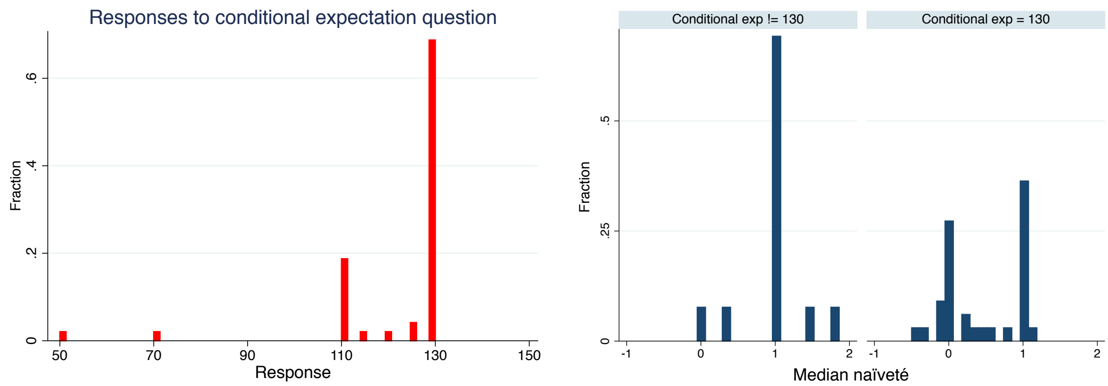

Figure 11: The left panel plots the responses to the follow-up question in the *Selected* treatment. The right panel illustrates the distribution of naïveté conditional on providing a response of larger than 100, but different from 130 (left subpanel), and conditional on answering exactly 130 (right subpanel).

B.6 Details for Correlation Neglect Follow-Up Study

B.6.1 Experimental Design

The design is taken from Enke and Zimmermann (2015). Subjects were asked to estimate a hypothetical true state μ, where I induced a prior belief by informing subjects that μ would be drawn from $\mathcal{N}(0; 250,000)$. Computers A-D generated four unbiased iid signals about μ by drawing from $s_h \sim \mathcal{N}(\mu; 250,000)$.

As illustrated by Figure 12, intermediary 1 observed the signal of Computer A and transmitted it to subjects. The intermediaries 2 to 4 observed both the signal of computer A and of computers B to D, respectively, and then reported the average of these two signals. Since subjects knew the signal of Computer A, they could extract the other

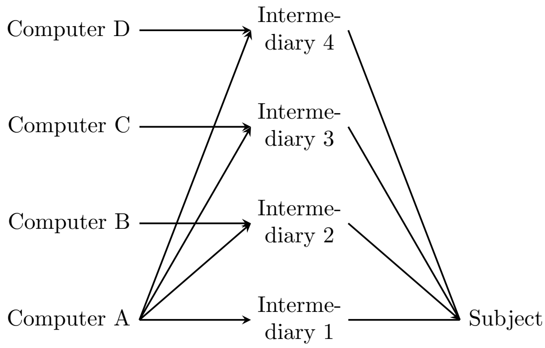

Figure 12: Correlation neglect information structure

independent signals from the intermediaries' reports.

As in the experiments designed to identify selection neglect, this treatment features an exogenous data-generating process wich is fully known to subjects. Control questions ensured that subjects understood the mechanics of this process. No feedback was provided between the five independent tasks. Earnings were computed through a quadratic scoring rule with maximum earnings of 12 euros: $\pi = max\{0; 12 - 0.01 \times (\text{Belief} - \text{True state})^2\}$. These experiments lasted 40 minutes on average, and subjects earned an average of 12.30 euros including a 7 euros show-up fee.

Table 8 presents details on the belief formation tasks as well as median beliefs in each task. As can be inferred from the rightmost column, median beliefs are always between the sophisticated and the full correlation neglect benchmark.

B.6.2 Computation and Distribution of Naïveté Parameters

Given the known data-generating process, one can again define and measure an individual-level naïveté parameter. As in the case of selection neglect, I assume full base rate neglect for this purpose, which is bolstered by the findings in Enke and Zimmermann (2015). The individual-level naïveté parameter is then computed as follows:

Subjects observed s_1 and $\tilde{s}_h = (s_1 + s_h)/2$ for $h \in \{2, 3, 4\}$. When prompted to es-

Table 8: Overview of correlation neglect tasks

True State	Computer A	Computer B	Computer C	Computer D	Sophisticated Belief	Correlation Neglect Belief	Median Belief
-241	249	-699	-139	70	-129.75	59.63	0.00
-563	-446	-1,374	-1,377	-1,475	-1,168	-807	-1,000
38	442	173	58	233	226.5	334.25	250.00
1,128	1,989	781	440	2,285	1,373.75	1,681.38	1373.75
-23	810	-822	-99	409	74.5	442.25	257

Notes. Overview of the correlation neglect estimation tasks in order of appearance. See Section B.6.2 for the derivation of the sophisticated and the full correlation neglect benchmarks.

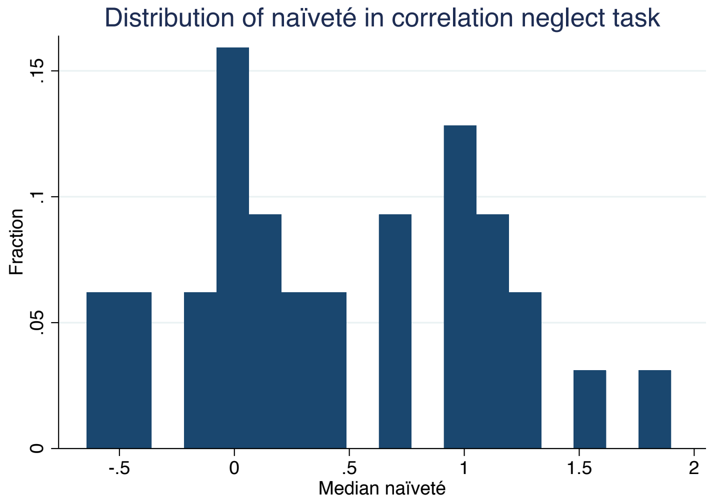

Figure 13: Distribution of median naïveté in correlation neglect task.

timate μ, a sophisticated decision maker would extract the underlying independent signals from the $\tilde{s}_h$ and compute the mean Bayesian posterior as $b_B = \sum_{h=1}^{4} s_h/4$. However, now suppose that the decision maker suffers from correlation neglect, i.e., he does not fully take into account the extent to which $\tilde{s}_h$ reflects s_1, but rather treats $\tilde{s}_h$ (to some extent) as independent. Call such a decision maker naïve and let his degree of naïveté be parameterized by $\chi \in [0,1]$ such that $\chi = 1$ implies full correlation neglect. A naïve agent extracts s_h from $\tilde{s}_h$ according to the rule

$$\hat{s}_h = \chi \tilde{s}_h + (1-\chi)s_h = s_h + \frac{1}{2}\chi(s_1 - s_h)$$

where $\hat{s}_h$ for $h \in \{2,3,4\}$ denotes the agent's (possibly biased) inference of s_h. He thus forms mean posterior beliefs according to

$$b_{CN} = \frac{s_1 + \sum_{h=1}^{3} \hat{s}_h}{4} = \bar{s} + \frac{3}{8}\chi(s_1 - \bar{s}_{-1})$$

where $\bar{s} = (\sum_{h=1}^{4} s_h)/4$ and $\bar{s}_{-1} = (\sum_{h=2}^{4} s_h)/3$.

Rearranging yields an individual- and task-specific naïveté parameter:

$$\chi = \frac{8 \times (b_{CN} - \bar{s})}{3 \times (s_1 - \bar{s}_{-1})}$$

For each individual, I then define their overall naïveté as the median χ across all tasks. Figure 13 plots the distribution of (median) naïveté in the follow-up study. As in Enke and Zimmermann (2015), this distribution exhibits a bimodal structure with some fraction of subjects fully accounting for the double-counting problem and others approximately fully ignoring the partial redundancy.

B.7 Response Times

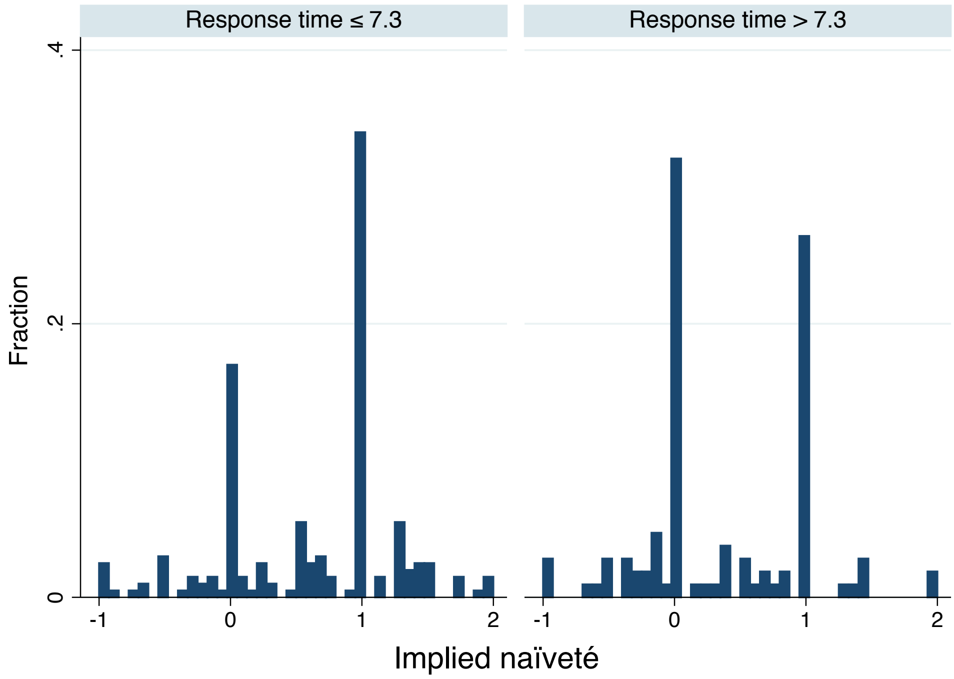

Figure 14: Distribution of naïveté in the *Selected* treatment, partitioned by whether the total response time is higher than 7.3 minutes (which is the median response time among subjects with a naïveté parameter of $\chi \leq 0.5$). To ease readability, both panels exclude observations outside $\chi \in [-1, 2]$ (30 out of 336 obs.).

C Robustness Treatments

To establish the phenomenon of selection neglect, the main text reported upon treatment *Selected*. This appendix reports on two robustness treatments.

In treatment *Robustness*, I replicated treatment *Selected*, except for one variation of the design: subjects only communicated with all computer players from their own group. Recall that in treatment *Selected*, subjects talked to all computer players in their own group, but at least with three. Thus, the *Robustness* treatment verifies that subjects also neglect systematic absences when their information sample is even more extremely skewed and contains either only relatively low or relatively high signals. 45 subjects took part in this treatment and earned an average of €11.20.

Treatment *Base Rate* also constitutes a simple variation of the baseline *Selected* treatment. In *Selected*, the true state was determined as average of 15 random draws from the set X. In *Base Rate*, the true state was determined as average of six random draws from X only. Given that a subject interacted with five computer players, this implies

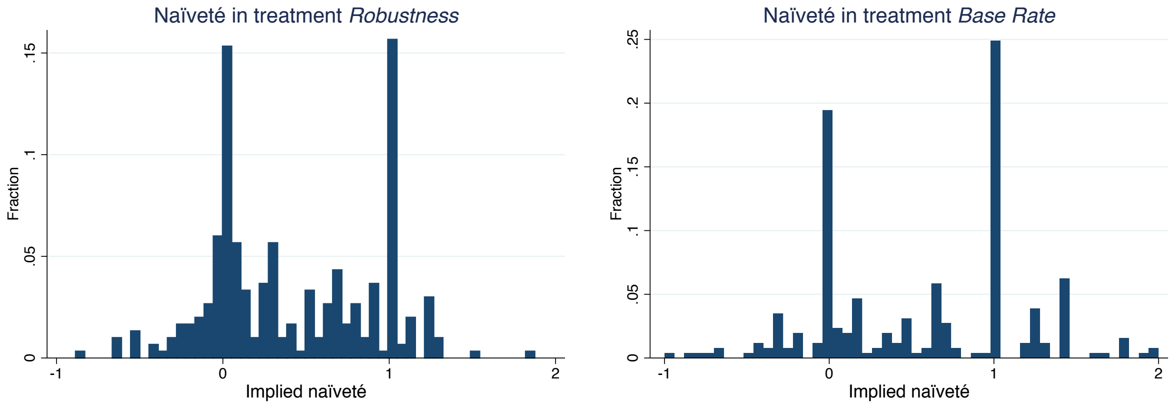

Figure 15: Distribution of naïveté in the *Robustness* and *Base Rate* treatments. The figure illustrates the distribution of naïveté implied in all beliefs (7 per subject) in the respective treatment. To ease readability, the figure excludes observations outside $\chi \in [-1, 2]$.

that all six balls that determine the true state are distributed among the subject and the computers. Thus, there was no scope for neglecting the base rate: subjects either saw a signal or had to infer it from the fact that the computer players entered the oposite group.

Figure 15 depicts the distribution of implied χ across the seven belief formation tasks in both treatments. In both treatments, the results are very similar to the baseline results. These findings highlight that the result established in the main text hinges neither on the particular selection mechanism nor on the possibility of neglecting the base rate.

D Treatment *Disagreement*

D.1 Distribution of Naïveté

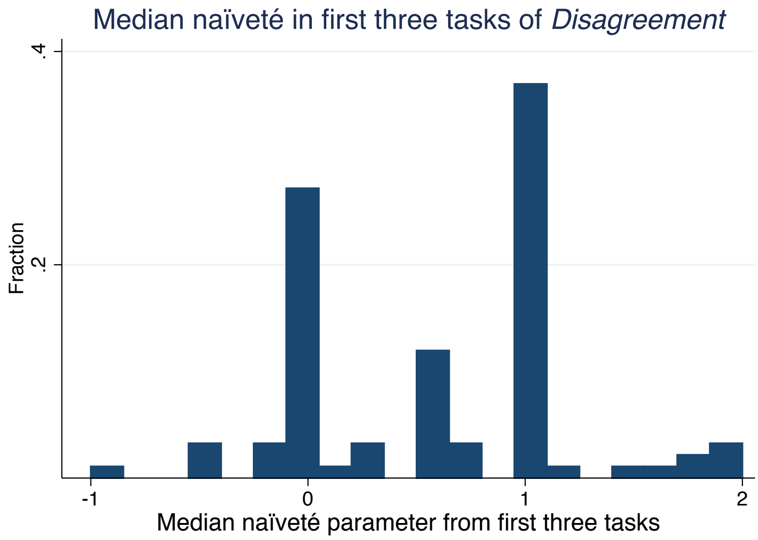

Figure 16: Distribution of median naïveté in the first three tasks (i.e., without seeing the beliefs of others). The density exclues observations outside [-1,2].

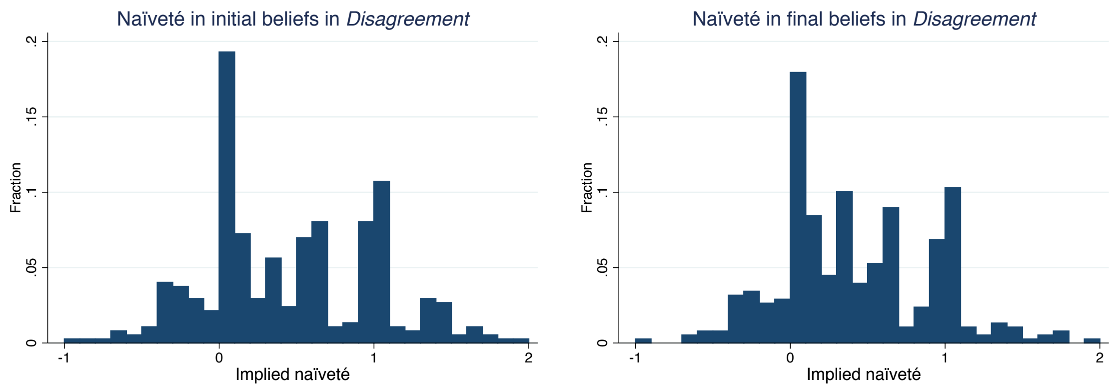

Figure 17: Distribution of decisions in the last four tasks (i.e., when seeing the beliefs of others). The left panel depicts the distribution of initial beliefs (before seeing the beliefs of the neighbors), and the right panel the distribution of post-communication beliefs. The histograms exclude observations outside [-1,2].

D.2 Raw Correlation Between Pre- and Post-Communication Beliefs

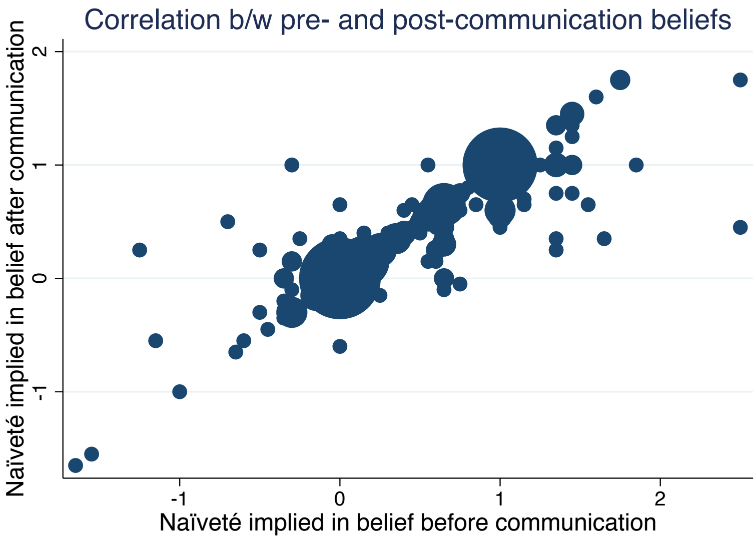

Figure 18: Raw correlation between the naïveté χ implied by first and second beliefs ($\rho = 0.86$). To construct this figure, subjects' pre- and post-communication naïveté is rounded to multiples of 0.05. The ball size then represents the number of observations in the respective bin. The scatter only includes observations for which there was at least partial disagreement (since otherwise people do not have a reason to revise their beliefs). I define disagreement as a binary variable which equals one iff the receiver's belief differs from the belief of at least one neighbor in the sense that the implied naïveté of the receiver is $\chi \leq 0.5$ and that of at least one neighbor $\chi > 0.5$, or vice versa. To ease readability, the scatter excludes observations for which the implied naïveté of at least one belief is outside [-1,2].

D.3 Estimation of Relative Weights

The results reported in column (1) confirm the visual impression developed above, i.e., subjects assign a much higher weight to their own belief than to that of their neighbors when they state their second beliefs. The last row of Table 9 reports the ratio of the weight that the average subject assigns to themselves relative to a random peer. To estimate these weights, one needs to keep in mind that the coefficient of the average naïveté of the neighbors picks up the beliefs of two participants. Thus, in column (1), the coefficient of each neighbor is 0.075, compared to a coefficient of 0.71 on subjects' own initial belief, which implies that the relative weight subjects assign to their own updating rule is 9.5. Column (2) shows that subjects assign a lower weight to their neighbors if their beliefs exhibit a larger degree of disagreement (defined as absolute difference between the naïveté implied in the neighbors' beliefs).

For this purpose, I return to using the full set of subjects × task observations. Table 9 presents a set of OLS regressions in which I regress subjects post-communication beliefs (expressed in units of naïveté) on (i) their own previous naïveté and (ii) the average naïveté of their neighbors. The resulting OLS coefficients can then be utilized to estimate the weight that subjects implicitly assign to their own belief formation rule relative to that of their peers.

The results reported in column (1) confirm the visual impression developed above, i.e., subjects assign a much higher weight to their own belief than to that of their neighbors when they state their second beliefs. The last row of Table 9 reports the ratio of the weight that the average subject assigns to themselves relative to a random peer. To estimate these weights, one needs to keep in mind that the coefficient of the average naïveté of the neighbors picks up the beliefs of two participants. Thus, in column (1), the coefficient of each neighbor is 0.075, compared to a coefficient of 0.71 on subjects' own initial belief, which implies that the relative weight subjects assign to their own updating rule is 9.5. Column (2) shows that subjects assign a lower weight to their neighbors if their beliefs exhibit a larger degree of disagreement (defined as absolute difference between the naïveté implied in the neighbors' beliefs).

Columns (3) through (6) break these patterns up into sophisticated and neglect types. As can be inferred from the last row, sophisticates assign a higher weight to themselves than neglect types (consistent with the visual patterns described above), but the neglect types still weight their own strategy 6.7 times higher than that of another randomly drawn subject. In addition, the results show that naïve subjects appear to react stronger to the consistency with which the neighbors state their beliefs.

D.4 Robustness Checks for Belief Revision

In the main text, sophisticated and naïve subjects were classified based on their out-of-sample median naïveté parameter from the first three tasks (the first part of the experiment). Figure 19 reports a robustness check for people's belief revisions in treatment *Disagreement* in which sophisticated and naïve subjects are classified based on both their out-of-sample median naïveté parameter from the first part of the experiment and the respective first belief in the fourth, fifth, sixth, or seventh task.

In the main text, the histogram of subjects' belief revisions included all observations for which there was at least minimal disagreement, i.e., in which a subject's first belief did not equal the average belief of the two neighbors. I now present a robustness check in which I show that similar results obtain when I restrict the sample to observations with large disagreement among subjects. To this end, I again normalize all beliefs into units of χ and then restrict attention to the subsample for which $|\chi_i - \bar{\chi}_{-i}| > 0.5$, i.e., for which the absolute difference between a subject's belief and the average belief of the two neighbors exceeds 0.5 units of χ. Figure 20 presents the results, which are

Table 9: Subjects' belief revision strategies

	Dependent variable: Naiveté implied in second belief					
	All subjects		Sophisticates		Naïfs	
	(1)	(2)	(3)	(4)	(5)	(6)
Naïveté implied in first belief	0.71***	0.72***	0.71***	0.72***	0.67***	0.68***
	(0.04)	(0.04)	(0.06)	(0.07)	(0.05)	(0.05)
Average naïveté of neighbors	0.15***	0.22***	0.099***	0.16***	0.20***	0.29***
	(0.03)	(0.04)	(0.03)	(0.05)	(0.05)	(0.06)
Disagreement among neighbors		0.061**		-0.017		0.15***
		(0.03)		(0.03)		(0.05)
Avg. naïveté × disagreement of neighbors		-0.067***		-0.034		-0.099***
		(0.02)		(0.02)		(0.03)
Constant	0.19	0.16	0.35	0.36*	0.050	0.0088
	(0.14)	(0.14)	(0.22)	(0.21)	(0.17)	(0.17)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	372	372	171	171	201	201
R^2	0.794	0.799	0.734	0.742	0.789	0.802
Relative weight on self	9.5		14.3		6.7	

Notes. OLS estimates, standard errors (clustered at subject level) in parentheses. The dependent variable is the second belief in a given task regressed on the respective first belief statement and the average belief of the two neighbors. All beliefs are expressed in units of naïveté. Controls include task fixed effects, age, gender, log monthly income, and high school grades. All regressions exclude observations with $|\hat{\chi}_i^j| > 3$; the results are robust to including these outliers. Disagreement among the neighbors is defined as the absolute difference between the naïveté implied by the neighbors' beliefs. The relative weight variable is computed as two times the ratio of the regression coefficients of a subject's own belief and that of their neighbors. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

slightly weaker, but overall very similar to those presented in the main text.

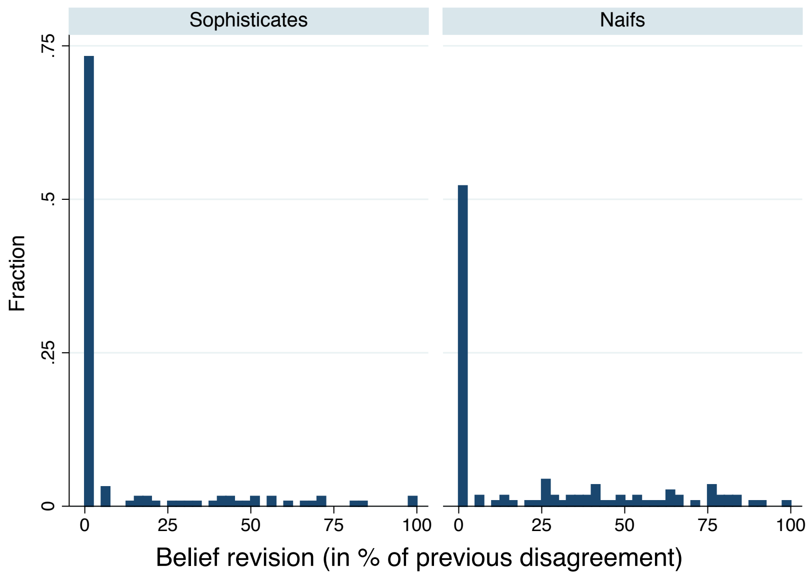

Figure 19: Magnitude of belief revisions. Each histogram depicts the belief revision between the first and second belief (expressed as percent of the difference between the first belief and the average belief of the two neighbors) conditional on the type of the subject (left / right panel). A subject is classified as sophisticated if both the out-of-sample median naïveté parameter from the first part of the experiment and the first belief in the respective task satisfy $\chi \leq 0.5$ 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.

D.5 Confidence

A different way to interpret the results from treatment *Disagreement* is that the neglect types are quite confident in their erroneous solution strategy. This section develops this argument a bit further, by considering an additional confidence proxy, i.e., a qualitative Likert scale question that I asked of subjects after they had completed the first part of the experiment: "On a scale from 1 (not certain at all) to 10 (very certain), how certain are you that your previous estimates (and the underlying strategy) were correct?". 96 subjects took part in this condition and earned €11.60 on average.

To formally analyze the relationship between subjects' updating type and their confidence, I compute the average belief revision measure by subject. Table 10 shows that the belief revision measure and the qualitative, non-incentivized confidence proxy are significantly correlated with each other as well as with being female, providing reassuring evidence for the meaningfulness of both of these constructs. However, neither of the two confidence proxies is significantly correlated with subject's median χ as estimated from the first part of the experiment. Thus, again, neglect types are almost as confident in their belief formation rule as the sophisticated types.

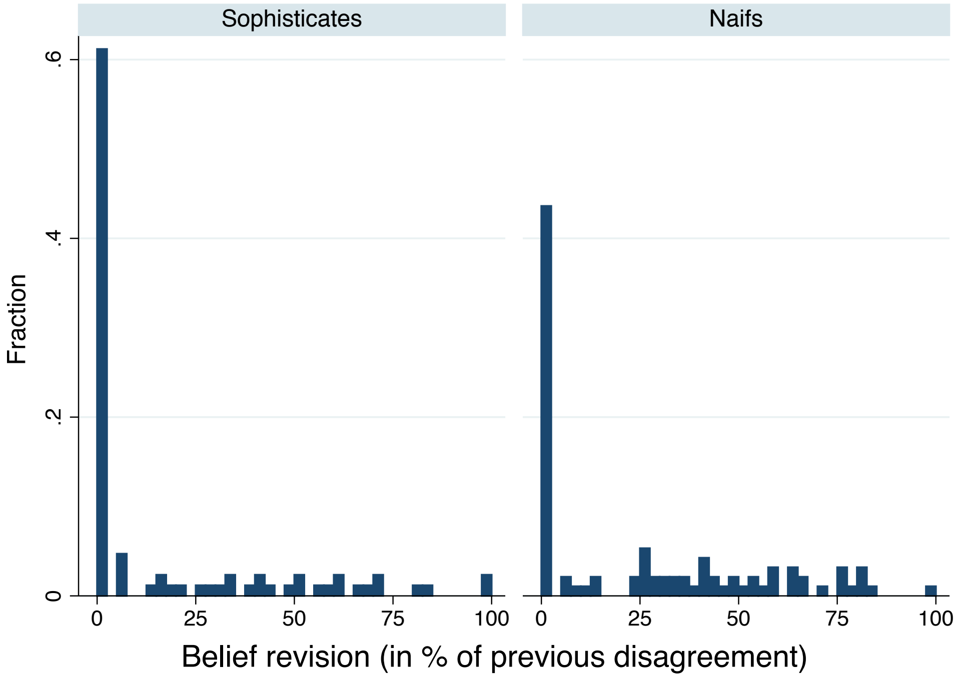

Figure 20: Magnitude of belief revisions. Each histogram depicts the belief revision between the first and second belief (expressed as percent of the difference between the first belief and the average belief of the two neighbors) conditional on the type of the subject (left / right panel). A subject 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 absolute difference between a subject's first belief and the average belief of the two neighbors (all expressed in units of χ) is larger than 0.5. Adjustments $> 100\%$ and $< 0\%$ are excluded to ease readability.

Table 10: Raw correlations between confidence, naïveté, and gender

	Belief revision	Confidence	Female	Median χ
Belief revision	1			
Confidence	-0.265***	1		
Female	0.250**	-0.240**	1	
Median χ	0.153	-0.131	-0.0113	1

Notes. Pearson correlations. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 11: Belief adjustment and learning

	Dependent variable: Naiveté implied in first belief	
	(1)	(2)
Naiveté in previous task	0.26^{***}	0.30^{***}
	(0.08)	(0.09)
Adjustment in previous task	0.094	0.058
	(0.09)	(0.10)
Age		-0.016
		(0.01)
1 if female		0.34^{***}
		(0.10)
Log [Monthly income]		-0.017
		(0.05)
Constant	0.41^{***}	0.83^{**}
	(0.07)	(0.37)
Task FE	No	Yes
Observations	156	156
R^2	0.133	0.428

OLS estimates, robust standard errors (clustered at subject level) in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

D.6 Learning?

It is conceivable that those naïve subjects who substantially revise their beliefs become less naïve in subsequent tasks. This could happen, for example, if subjects learn from the beliefs of more sophisticated subjects. Table 11 presents the results of OLS regressions of subjects' naïveté in a given task on the degree of adjustment towards the sophisticated belief in the previous task, conditional on the initial naïveté in the previous task. In these analyses, the sample is restricted to naïve subjects, i.e., to those participants whose out-of-sample median naïveté parameter from the first three tasks is larger than 0.5. Results show that those subjects who strongly revise their beliefs do not become more sophisticated over time. This suggests that some subjects may feel that their own problem-solving is incorrect, but have no superior way of solving the problem themselves.

E Amazon Turk Survey

In total, 209 participants completed the survey, which lasted less than five minutes. Each participant earned \$0.40 flat. Each participant was asked to assess the relative difficulty of four pairs of math problems. In particular, participants were asked:
"In your view, which of the following two mathematical tasks is more difficult?"

Task A: Compute the average of 130, 110, 70, and 70.
Task B: Compute the average of 130, 130, 70, and 70.

Task C: Compute the average of 130, 130, 130, and 130.
Task D: Compute the average of 130, 130, 150, and 110.

Task E: Compute the average of 70, 70, 130, and 130.
Task F: Compute the average of 70, 70, 150, and 130.

Task G: Compute the average of 110, 130, 110, and 70.
Task H: Compute the average of 130, 130, 130, and 70.

Respondents used a five-point scale to indicate which task, if any, they deemed harder. In analyzing the data, for each subject, I compute the average response across questions, i.e., the average extent to which subjects consider the tasks that relate to treatment *Intermediary* (A, D, F, G) more difficult. The main text reports the results.

F Experimental Instructions

F.1 Overview

F.2 Treatments *Selected* and *Control*

F.2.1 Instructions (Paper-Based)

Translated into English - whenever parts of the instructions differed between the selected and the control treatment, I always first provide the version from the selected treatment, followed by the respective part from the control treatment.

Welcome. You will now take part in an economics decision making experiment. You will receive a show-up fee of 4 euros, which will be paid out to you at the end of the experiment. How much money you earn on top of that depends on your decisions. In this experiment, we speak of points. 100 points equal 10 euros. At the end of the experiment, your points will be converted into euros and paid out to you.

Your task
In this experiment, there are two groups: group red and group blue. Being a member of a group is associated with a monetary reward for you. To be more precise, these two groups differ in their profitability. Sometimes one group if for profitable for you and sometimes the other. Whether one or the other group is more profitable depends on the so-called number X:

- If the number X is larger than 100, the red group is more profitable for you.
- If the number X is smaller than 100, the blue group is more profitable for you.

Your task will consist of (i) deciding which group you would like to enter and (ii) providing a precise estimate of the number X.

The number X is randomly determined by the computer. You will not know the number X. Rather, you will obtain information over X; then you need to take your decisions based on this information.

The number X is determined as follows. Imagine that in the room next door there is an urn which contains exactly 6 balls. These 6 balls have the following numbers:

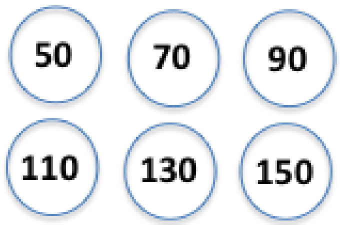

Figure 1: The set of balls from which the computer draws 15 times. Please note that balls that get drawn are replaced by another ball with the same number, so that each number can get drawn multiple times.

The computer randomly selects 15 balls from this urn, where drawn balls get replaced by a different ball with the same number before the next draw. I.e., if the computer draws a 130, a new ball with a 130 gets placed into the urn before the computer draws again. Thus, each number can get drawn multiple times! Also, this procedure implies that each time the computer draws, every ball in Figure 1 is equally likely to be drawn! The computer puts the randomly drawn 15 balls into a box. The average of the numbers on these 15 balls represents the number X.

As you can infer from the numbers on the balls, it is equally likely that the number X is larger or smaller than 100. The challenge for you is that you will neither see the number X nor the randomly 15 balls in the box. Thus, in this experiment, you will not get a chance to calculate the average of these 15 numbers. Rather, you receive hints over the number X on which you need to base your decisions.

The hints

In this experiment, you will interact with five other players which are all simulated by the computer. In what follows, we will call these five computer players I, II, III, IV, and V.

The computer generates six hints about the number X. The computer generates each hint by randomly drawing one of the 15 balls that determine the number X. Thus, again, each number in Figure 1 is equally likely to be drawn as a hint.

This implies that the hints are very helpful in guessing the number X! In particular, it holds true that these hints are on average correct, i.e., equal the number X. While ususally each hint deviates from the number X, these deviations are not systematic, so that these hints are on average correct (if we would draw a large number of times). You will have to base your decisions on these types of hints.

In total, the computer generates six hints and distributes these randomly among you and the five computer players such that everyone receives one separate hint. Each hint is of equal quality, but everybody sees a different (their own) hint.

In addition to your own hint, you will have access to the hints of some of the computer players. The next section explains how this works.

Course of events in this experiment

In total, we will implement 7 rounds. Please note that these 7 rounds are completely independent of each other, so that you cannot learn anything from the numbers in one round about another round. Each of these 7 rounds works as follows:

1. The computer determines the number X, whose realization you will not find out. The computer generates the six hints and distributes them among you and the five computer players I-V.

2. You take your first payoff-relevent decision: Based on your hint, you have to decide whether you would like to enter the blue or the red group. Here, you should note that you should enter the red group if you believe that the number X is larger than 100, and the blue group if you believe that the number X is smaller than 100. The computer players take the same decision: they enter the red group if their own hint is larger than 100 and the blue group if their own hint is less than 100.

3. (SELECTED TREATMENT): You "communicate" with some of the computer players, i.e., these players will tell you which hint they received in the beginning. More precisely, you will meet at least three computer players and sometimes more:

    - You will always obtain the hints of all computer players that opted for your own group, no matter what. This means that all players which opt for the same group as you tell you their own hint. Thus, it can never happen that a player is in your group and you don't talk to him.
    - If there are at least three computer players present in your group, you will only communicate with these players from your own group.

- If there are less than three computer players in your own group, you will additionally communicate with some players from the other group. To this end, in addition to the players from your own group, the computer randomly selects players from the other group until you have three communication partners in total.
- This procedure implies that you may not directly learn about all six hints because you may not communicate with all computer players from the other group.
- In communicating with you, the computer players never make mistakes and always truthfully tell you the hint they received.

(CONTROL TREATMENT): You "communicate" with the computer players, i.e., these players will tell you which hint they received in the beginning. Here, we distinguish between two different types of meeting a computer player. While these types differ, on average all communication pieces are equally helpful:

- You will always obtain the hints of all computer players that opted for your own group, no matter what. This means that all players which opt for the same group as you tell you their own hint. Thus, all players that decide for the same group as you will tell you their own hint in a precise way.
  - If there are at least three computer players present in your group, only these players will tell you their precise hints.
  - If there are less than three computer players in your own group, some players from the other group will also tell you their precise signal. To this end, in addition to the players from your own group, the computer randomly selects players from the other group until you have three communication partners that tell you their own hints in a precise way.
- In addition, you also communicate with the players that do not tell you their precise hint. These other players also provide you with useful information: if their own hint is 50, 70, or 90, they tell you "70", which equals the average of this set. If the hint of the respective computer player was 110, 130, or 150, he tells you "130", which also equals the average of the respective set.
- In communicating with you, the computer players never make mistakes and always truthfully tell you the hint they received. While it may be slightly confusing that you will communicate with the computer players in two different ways, the only important issue for you is to understand that all messages of the computer players are equally valuable and ON AVERAGE correct.

4. (BOTH TREATMENTS): You take your second payoff-relevant decision: You need to provide an estimate of the number X.

Your payment

In addition to your show-up fee, you will get paid according to your decisions. Since we will implement 7 rounds, you need to estimate the number X 7 times and you also need to decide 7 times which group to enter. The computer will randomly select ONE of these 14 decisions and you will then get paid according to that decision. The probability that an estimate will get selected is 80%, while the probability that your group membership will get paid out is 20%. This means that every single one of your decisions is potentially relevant for your earnings so that you should carefully think through all decisions.

In case your group membership gets paid, your earnings will get determined as follows:

- The number X is larger than 100:
  - You are in the red group: 120 points
  - You are in the blue group: 20 points
- The number X is smaller than 100:
  - You are in the red group: 20 points
  - You are in the blue group: 120 points
- The number X equals 100: You earn 120 points in both groups.

You should note that you should enter the red group if you believe that the number X is larger than 100 and the blue group if you believe that the number X is smaller than 100.

In case your estimate of the number X gets paid, you will receive more money the closer your estimate is to the number X. At most, you can earn 180 points with your estimate. The further away your estimate from the truth, the less you earn. This is determined according to the following formula (in points):

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

This means that the difference between your estimate and the true value will get squared and multiplied by 2. This value will then get subtracted from the maximum earnings of 180 points. While this formula may look complicated, the underlying principle is very simple: the smaller the deviation between your estimate and the true value, the higher your earnings. Note that your earnings in this task cannot be negative, i.e., you cannot make losses. You should also note that your earnings only depend on the absolute difference. Thus, it doesn't matter for your payment whether you overestimate or underestimate the true value by 5.

IMPORTANT: Please note that, in this experiment, on average you can earn the most money if you always truthfully enter your actual estimate. Since only one of your decisions will get paid, it doesn't make sense for you to "strategize" by, e.g., sometimes entering the blue and sometimes the red group, or by sometimes providing a high estimate and sometimes a low estimate. In order to earn as much money as possible, you should always try to take the best decision you currently have in mind.

Example

Suppose that the computer has determined the number X. Now the computer generates the six hints and distributes them among you and the five computer players. Based on your hint, you need to take a decision about your group membership, as illustrated in Figure 2. In this example, your own hint is 50. Suppose that you decide to opt for the blue group since the hint is smaller than 100.

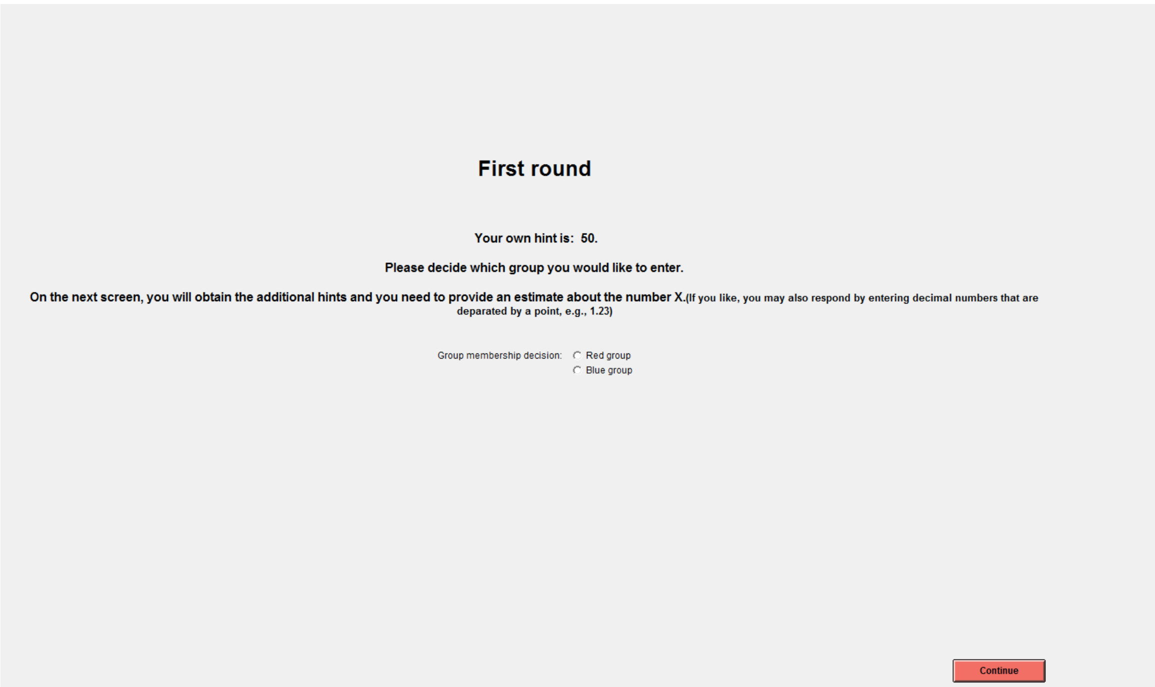

Figure 2: Exemplary screenshot for the first decision.

The computer players now take the same type of decision as you, i.e., they enter the red group if their hint is larger than 100 and the blue group if their hint is smaller than 100. Suppose that two computer players obtain a hint of less than 100 and hence enter the blue group, like you. The other three computer players see a hint of larger than 100 and hence enter the red group. However, we will not tell you all of this.

(SELECTED TREATMENT): Subsequently, you will communicate with the two computer players from your own group. These two obtained hints of 90 and 50, respectively. Since your group contains less than three computer players, the computer randomly selects one further communication partner from the other group, so that you have 3 communication partners in total. This computer player obtained a hint of 150. Figure 3 presents a screenshot of the second decision screen for this example. You then need to enter an estimate of the number X.

(CONTROL TREATMENT): Subsequently, you will obtain the precise hints of the two computer players from your own group (90 and 50) as well as of one other randomly selected player from the other group (150). In addition, you will obtain – in a somehwat more coarse way – the hints of the remaining two players in the other group (130 and 130). Figure 3 presents a screenshot of the second decision screen for this example. You then need to enter an estimate of the number X.

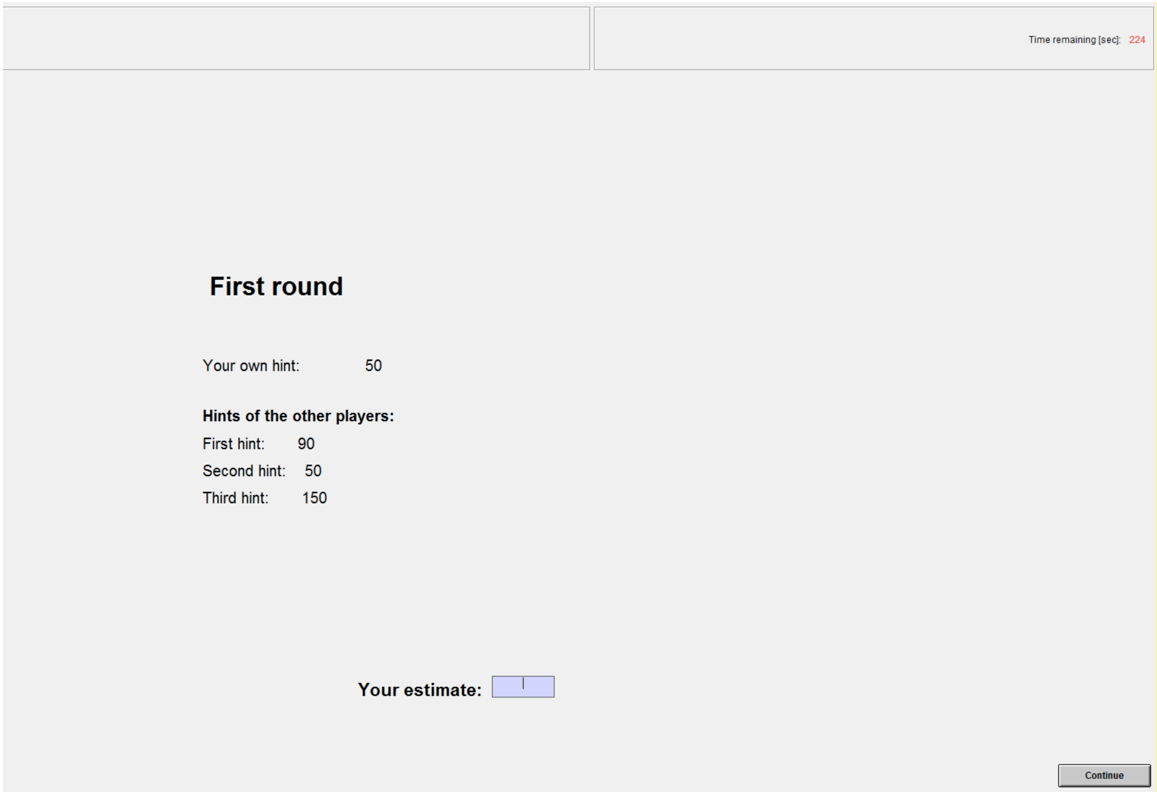

Figure 3: Exemplary screenshot for the second decision.

Space for personal notes (You may write on these instructions, if you like)

F.2.2 Control Questions (Computerized)

Note: Across all treatments, the control questions were presented on a computer screen such that a given decision screen contained usually five separate control questions. Subjects could only proceed to the next screen once they had correctly answered all questions. If at least one answer was incorrect, the subject was notified of this, but the program didn't tell subjects which question they got wrong. Also note that the BonnEconLab has a control room in which the decision screens of all 24 subjects can be monitored. Whenever a subject appeared to have problems in solving the control questions, one of the experimenters approached that subject, clarified open questions. Subjects which showed a clear lack of understanding of the experiment were excluded from the analysis, but were allowed to take part in the experiment so as to avoid noise due to subjects' leaving the room and getting paid while others were completing their tasks.

- What is your main task in this experiment?
    1. There are 20 numbered balls. I need to add these 20 numbers up.
    2. I need to estimate the number X.
- Please assess the following statement: "In total, 15 balls will be randomly selected and be put into a box. In each draw, each number is equally likely to be drawn."
    1. False. If, e.g., 110 gets drawn, then it is more likely that the next draw will not be a 110.
    2. Correct. If a number gets drawn, it will get replaced by the same number, so that all numbers are equally likely again.
- Your hint is larger than 100. Is this indicative that the red or the blue group is more profitable?
    1. I can't know.
    2. The blue one.
    3. The red one.
- A computer player receives a hint of 70. What does he do?
    1. He randomly enters a group.
    2. Because the hint indicates that X is smaller than 100, he enters the blue group.
    3. Because the hint indicates that X is smaller than 100, he enters the red group.

- In estimating the number X, how can you earn the maximum amount of money?
    1. By strategizing, i.e., sometimes providing and low and sometimes a high estimate.
    2. By always entering my estimate in the most precise way.
- Which of your decisions is payoff-relevant?
    1. Every decision gets paid.
    2. No decision.
    3. One randomly selected decisions gets paid.
- (SELECTED TREATMENT ONLY): What do you learn from the players that you communicate with?
    1. They tell me the number on their ball, but make mistakes in doing so.
    2. Nothing.
    3. Every player I communicate with truthfully tells me the number on his ball.
- (SELECTED TREATMENT ONLY): Which players tell you their hints?
    1. All.
    2. At least three, but all players from my group no matter what.
    3. At least four.
- (SELECTED TREATMENT ONLY): Suppose you're in the red group. What do you know about a computer player if he does not communicate with you?
    1. Nothing.
    2. He must be in the blue group and his hint hence be smaller than 100.
    3. He must be in the blue group and his hint hence be larger than 100.
- (SELECTED TREATMENT ONLY): Suppose a player is in the red group. Which hint must he have seen, ON AVERAGE?
    1. 50
    2. 70
    3. 90
    4. 110

5. 130
6. 150
7. I can't know.

NOTE Recall footnote 9 in the main text: This control question was phrased suboptimally. 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".

- (CONTROL TREATMENT ONLY): What do you learn from the computer players you communicate with?

  1. They tell me the number on their ball, but make mistakes in doing so.
  2. Nothing.
  3. Every player I communicate with truthfully tells me the number on his ball or the average of the respective set.

- (CONTROL TREATMENT ONLY): What do you learn from the players that do not tell you their precise hint?

  1. Nothing.
  2. They also provide me with useful hints, which are on average correct.
  3. They tell me useless things.

- (CONTROL TREATMENT ONLY): Which statement is correct?

  1. If I compute the average of the six hints, I'm correct on average.
  2. Only the precise hints I receive from some players are correct.
  3. I can't know which hints are on average correct.

F.3 Correlation Neglect Follow-Up Study

F.3.1 Instructions (Paper-Based)

You will now take part in an economic experiment. You will receive a show-up fee of 4 euros,[24] which will be paid out to you at the end of the experiment. You can earn additional money which will also be paid out at the end of the experiment. How much you earn depends on your decisions. In this experiment, we will talk about points. 100 points correspond to 10 euros. The points you earn during the course of the experiment will be exchanged into euros and paid out at the end of the experiment. During the experiment, communication with other participants is not allowed. The curtain of your cabin must be closed at all times. If you have questions, you can raise your arm out of your cabin and the experimenter will try to answer your questions.

Your task:

In this experiment, you will have to solve five estimation tasks. In these tasks, you will have to estimate an unknown number X. In each round, the computer randomly determines the number X, which will however be unknown to you. As will be explained in more detail below, you will receive some information about this number. Then you will be asked to provide an estimate about X. In total, there are 5 rounds; in each round, you will face a new estimation task, i.e., in each round the computer will determine a new number X and that number will be entirely independent from the numbers in the other rounds.

Your earnings will depend on how precisely you estimate, i.e., how close your estimate is to the actual number X. At the end of the experiment, one of the five tasks will be randomly selected and you will be paid according to the precision of your estimate in this task. This will be explained in more detail in the next section.

Your earnings:

In addition to your show-up fee you will be paid according to the precision of your estimates. You receive more money the closer your estimate is to the true number of items in the container. One of the five estimation tasks will be randomly selected for payment and you will be paid according to the precision of your estimate in that task. This means that every estimate is potentially relevant for your payment, so that you should think about each task carefully.

You earn more money the closer your estimate to the number X. You can earn at most 120 points. The further away your estimate lies from the true value, the lower will

[24] *Subjects received an additional 2 euros for filling out the sociodemographic questionnaire after the main part of the experiment.*

be your earnings. This will be determined according to the following formula:

$$\text{Payment } = 120 - 0.1001 \times (\text{Difference between estimate and truth})^2$$

This means that the difference between your estimate and the true value will be squared and multiplied by 0.001. This number will then be deducted from the maximum earnings of 120 points. 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 smaller then zero, i.e., you cannot make losses. You can also see that your earnings only depend on the absolute difference. For example, it does not matter whether you over- or underestimate the true value by 5.

IMPORTANT: Please note that, in this experiment, on average you can earn the most money if you always truthfully enter your actual estimate. Since only one of your decisions will get paid, it doesn't make sense for you to "strategize" by, e.g., sometimes providing a high estimate and sometimes a low estimate. In order to earn as much money as possible, you should always try to take the best decision you currently have in mind.

The estimation task:

In each round you will have to provide an estimate about an unknown number X. As already mentioned, for each round the computer will randomly determine a new number X. You will not know this number. The computer draws this number for each round from a probability distribution, that is displayed below.

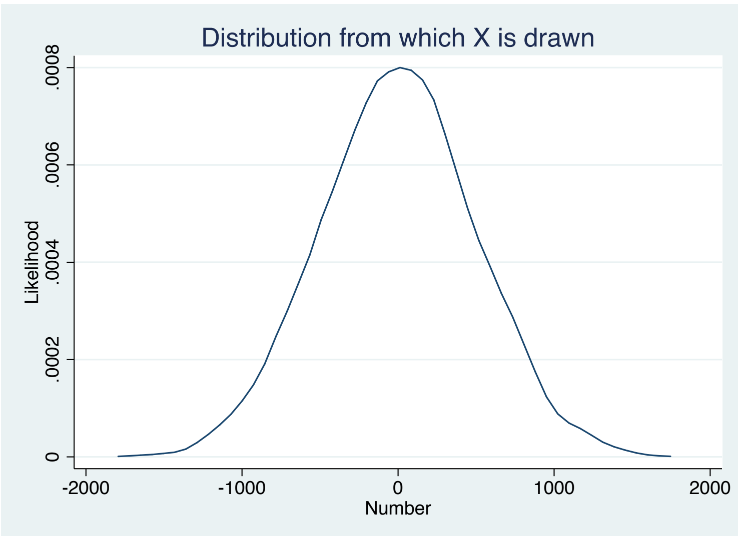

Figure 4: Distribution from which the computer draws X.

The distribution you see in Figure 4 is a so-called normal distribution. The distribution has a mean of 0 and a standard deviation of 500.[25] Although you will not know the number X, the graph tells you something about the range from which X is drawn.

After the computer has drawn X, you will need to provide an estimate about X. For that purpose, for every estimation task, you will receive different computer-generated pieces of information about the correct estimation result. For every task, you will see this information and then enter your own estimate. The information you receive will be explained in detail below.

Information regarding the estimation tasks:

Your task in this experiment is to provide an estimate about a randomly drawn number X (unknown to you), based on some information. For every estimation task, you will receive different computer-generated pieces of information about the number X. For every task, you will see this information and then enter your own estimate. The information you receive will be explained in detail in the following.

On computers, we simulate devices which solve exactly the same estimation tasks as you. There are two different types of devices. First, there are devices which themselves provide an estimate of X (these devices will be called estimation devices and are denoted by letters). Second, there are devices which observe the estimates of the estimation devices and compute their own estimate from these reports (these devices are referred to as communication devices and denoted by numbers).

The estimation devices provide an estimate about the number X, and the estimates of these devices are completely independent from each other. The estimation devices all have the same quality, i.e., they are equally good in determining estimates. Note that these estimation devices are good at solving these estimation tasks:
The estimation devices determine an estimate by randomly drawing a number from a normal distribution. Importantly, this distribution takes as mean the number X, and a standard deviation of 500. The figure below shows you an example of such a distribution. You can see that the highest point of the bell curve is at the number X, i.e., the correct value. The further you move away from X, the less likely it is that the corresponding numbers are drawn from the estimation devices.

This means that the estimation devices are good at solving the estimation task. If the estimation devices would provide a large number of estimates, then the average of these estimates would be correct. While almost every individual estimate will be incorrect, the average taken over many estimates will be very precise. In addition, many estimates

[25]The exact distribution of a normal distribution with mean 0 and a standard deviation of 500 is given by the following formula: $f(x) = \frac{1}{500\sqrt{2\pi}} exp(-\frac{x^2}{500000})$. Throughout the experiment, we round all drawn numbers to integers.

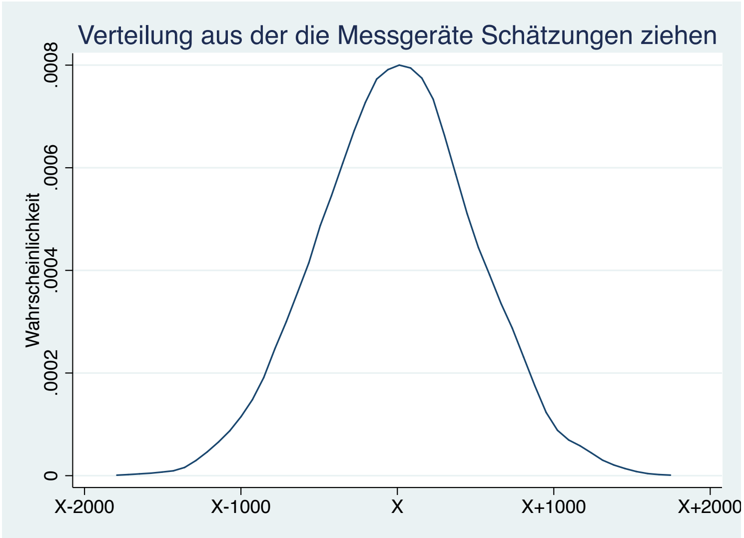

Figure 5: Distribution from which the estimation devices draw their estimates.

will be rather close to the correct value.

Summary: (1) The devices draw from a normal distribution with mean X. This means that the estimation devices are good at solving the estimation task. If the estimation devices would provide a large number of estimates, i.e., if they would draw many times from the normal distribution, then the average of these estimates would be correct (or very precise). (2) The devices make mistakes, but it is much more likely that the estimate is close to the true value, than that it is very far away.

For every estimation task, there are a total of four estimation devices (A, B, C, D). These four devices, which are completely independent from each other, each randomly draw an estimate from the normal distribution (with mean X and a standard deviation of 500).

Apart from the estimation devices, there are also four communication devices (1, 2, 3, 4). These communication devices do not determine an own estimate. Rather, they observe the estimation devices and compute an estimate from these observed estimates. Intermediary 1 only observes estimation device A, and simply transmits the estimate of estimation device A. Intermediaries 2, 3, 4 each observe the estimates from two of the estimation devices, and compute an estimate from these two estimates by computing the respective average.

You will receive the following information as described in the following Figure.

SUBJECTS SAW THE APPLICABLE FIGURE AS SHOWN IN THE MAIN TEXT.

This means that you will receive the following information: As is evident from the figure, communication device 1 receives the estimate from estimation device A and reports this estimate to you. The other communication devices all see the estimate of estimation device A and of one other estimation device. As you can see in the figure, communication device 2 receives the estimates of estimation devices A and B. Communication device 3 sees the estimates of estimation devices A and C. Communication device 4 sees the estimates of estimation devices A and D. The communication devices 2, 3, 4 take the average of these two estimates and report this average as their estimate.

Summary: You receive the estimates of the communication devices 1, 2, 3 and 4.

The following simple example illustrates this. We again assume that the correct number X is 150. Let's assume for this example that the estimates of the four estimation devices would be as follows:

Estimation device A: 81.0
Estimation device B: 127.0
Estimation device C: 209.0
Estimation device D: 176.0

Communication device 1 would then report the estimate of estimation device A. The communication devices 2, 3, 4 would take the average of the two estimates they see, as described above. The communication devices would thus report the following estimates:

Communication device 1: 81.0
Communication device 2: 104.0
Communication device 3: 145.0
Communication device 4: 128.5

Thus, for this estimation task, you would see the following information on your computer screen:

SUBJECTS SAW A SCREENSHOT ANALOGOUS TO THE ONE SHOWN IN THE BASELINE INSTRUCTIONS ABOVE.

Please read these instructions again carefully. Afterwards, you will answer a set of control questions at the computer in order to check your understanding of the instructions.

F.3.2 Control Questions (Computerized)

Questions 1 and 3 as in the individual baseline conditions

- In this experiment, you have to solve five estimation tasks. Which of these tasks will be relevant for your final profit?

    1. At the end of the experiment, one estimation task will be randomly selected. Profits will be paid out according to performance in this task.
    2. None of the estimation tasks will be paid out.
    3. At the end of the experiment, three estimation tasks will be randomly selected. Profits will be paid out according to performance in these tasks.
    4. All estimation tasks will be paid out.

- Your profit in this experiment will depend on the precision of your estimates. Suppose your estimate differs from the true value by 1000. How many points will you receive if this task is relevant for your profit?

    1. 100 points
    2. 50 points
    3. 0 points

- The estimation devices provide estimates for every estimation task. What can you say about the quality (regarding the probability of making errors) of the different estimation devices?

    1. The quality of the estimation devices is identical, i.e., the estimation devices do not differ in this respect.
    2. The quality of the estimation devices differs. Estimation device C is the best one.
    3. The quality of the estimation devices differs. Estimation device A is the best one.

- What is the mean of the distribution from which the number X is drawn?

    1. I cannot know this.
    2. 0.
    3. 100.

- What is the relation between the standard deviation of the distribution from which X is drawn and the standard deviation of the distribution from which the estimation devices draw their estimates?

    1. The standard deviation of both distributions is 500, i.e., identical.

2. The standard deviation of the distribution from which X is drawn is larger.
3. The standard deviation of the distribution from which X is drawn is smaller.

- Which of the following statements about your payment is correct?

  1. The closer my estimate is to the true value X, the smaller my earnings.
  2. The closer my estimate is to the true value X, the higher my earnings.

- Which of the following statements is correct?

  1. If estimation device B reports an estimate of 3160, then all other estimation devices will also report an estimate of 3160.
  2. The estimates of the estimation devices are independent of each other, so that they potentially report different estimates.

- Suppose estimation device A estimates 6. Estimation device B estimates 12 and estimation device C 16. Which estimate will communication device 2 report?

  1. 6
  2. 9
  3. 11
  4. 12
  5. 16

- Suppose estimation device A estimates 6. Estimation device B estimates 12 and estimation device C 16. Which estimate will the communication device 3 report?

  1. 6
  2. 9
  3. 11
  4. 12
  5. 16

- Which information will be provided to you for every estimation task?

  1. You will see the number X.
  2. You will see the estimates of all estimation devices.
  3. You will see the estimates of the communication devices 1, 2, 3 and 4.

F.4 Treatment *Robustness*

The paper-based instructions for this treatment were identical to those in *Selected*, except that the communication stage with the computer players was simplified. That is, the instructions are identical except for step 3. of the "course of events" section of the instructions for *Selected*. In this new treatments, step 3. reads as follows:

3. You "communicate" with some of the computer players, i.e., these players will tell you which hint they received in the beginning.

- You will obtain the hints of all computer players that opted for your own group, no matter what. This means that all players which opt for the same group as you tell you their own hint. Thus, it can never happen that a player is in your group and you don't talk to him.
- This procedure implies that you may not directly learn about all six hints because you will not communicate with the computer players from the other group.
- In communicating with you, the computer players never make mistakes and always truthfully tell you the hint they received.

F.5 Treatment *Base Rate*

The paper-based instructions for this treatment were identical to those in *Selected*, except that X was determined by six random draws, as opposed to 15 random draws.

F.6 Treatment *Disagreement*

F.6.1 Written Instructions

The paper-based instructions for the first part were essentially identical to those in *Selected*. The only difference is that subjects were told that the experiment consists of two parts and one of those parts will be randomly selected for payment. The first part then consisted of three rounds (the first three rounds from *Selected*). Subjects were then unexpectedly interrupted by the following computer screen:

Question: You just completed three rounds. Please answer the following question: "On a scale from 1 (not certain at all) to 10 (very certain), how certain are you that your previous estimates (and the underlying strategy) were correct?"

Afterwards, the experimenter distributed the written instructions for the second part:

Instructions – Second part

The second part of the experiment is very similar to the first part. You will now complete an additional four rounds, which work almost the same as previously: You will receive a hint, enter a group, see the hints of some of the computer players and then estimate the number X.

Nevertheless, there are some changes to the first part. More specifically, the next four rounds work as follows:

1. The computer determines the number X. You as well as the five computer players I-V receive a hint each, just as before.

2. After you have received your own hint, you will not decide whether to enter the blue or red group yourself. Rather, this decision will be made by the computer in a fashion that was also optimal for you in the first part of the experiment: if your own hint is higher than 100, you will enter the red group, and if your hint is below 100, you will enter the blue group. This is a small change, but the basic mechanism is still the same as before. The same holds true for the computer players, i.e., just as before, they enter the red group if their signal was above 100 and the blue group if their signal was below 100.

3. As before, you will learn the hints of some of the computer players. The procedure through which the computer players with whom you communicate get selected, is the same as before. That is, you will communicate with at least three computer players, but always with all players who are in your own group. Subsequently, you will have to estimate the number X, as before.

4. Here is the main change: After you have provided your estimate, we will randomly select two other participants from this session (where every participant who is present has the same probability of being selected).[26] Then, we will show you the estimates which these two players have provided. Please note that all participants in this room receive exactly the same information as you! That is, steps 1-3 from above are the same for all participants: everyone receives the same hint, enters the same group and obtains the hints of the same computer players. Put

[26]More precisely, this will be determined such that the estimate of every participant will be shown to two other participants. However, it can never happen that you will see the estimate of the same player twice in any given round.

differently: until you provide your own estimate, everyone sees the same information on their computer screen. Hence, please note again that the two participants whose estimates you will see, received exactly the same information as you!

5. Subsequently, you will provide a second estimate over X. This estimate can be the same as you provided before, but you can also change it, if you like to do so. The screenshot on the next page visualizes this situation.

We will implement this procedure in every one of the following four tasks. Please note that the computer randomly selects anew, whose estimates will be shown to you. Thus, it is not necessarily the case that you will always see the estimates of the same two participants!

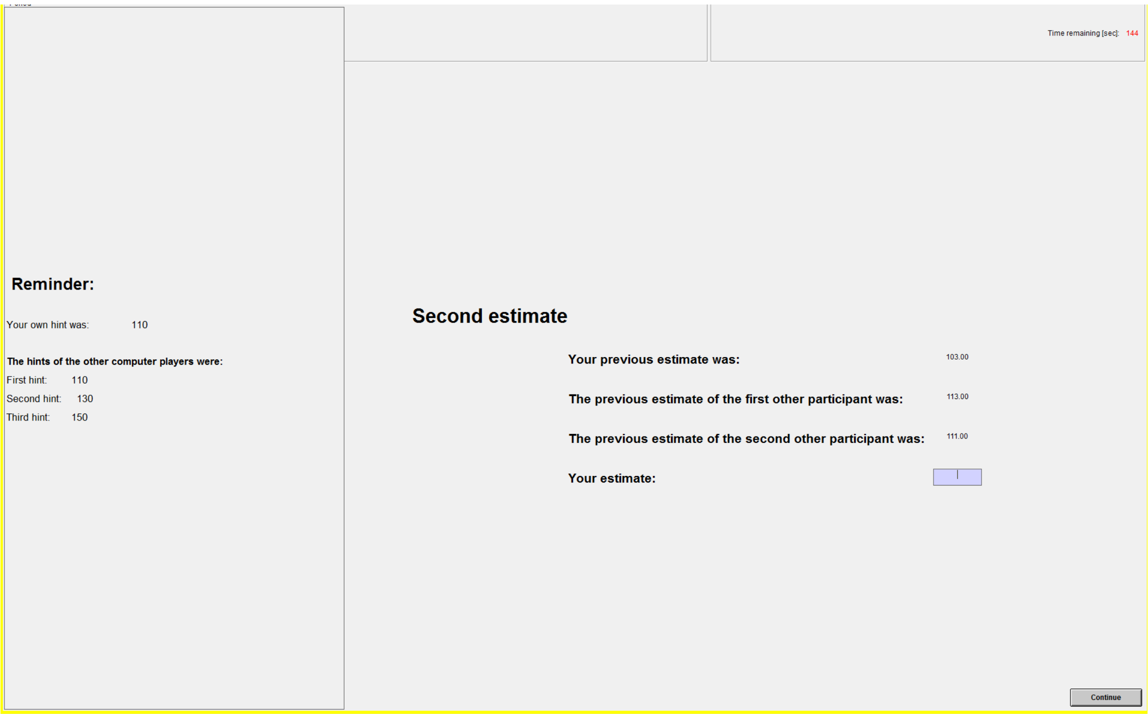

Figure 6: Screenshot for the second estimate, after both you and the other two participants have provided a first estimate

Your remuneration

In case this second part of the experiments gets selected for payment, you will receive the following remuneration, in addition to your show-up fee:

In the four rounds, you will provide $4 \times 2 = 8$ estimates. The computer randomly selects one of these estimates and you will be paid according to how precise this estimate was: The closer your estimate to the number X, the more money you receive. You can earn 180 at most, and your remuneration will be determined according to the formula in the first part of the instructions.

F.6.2 Verbal Summary (Read Out Aloud)

After subjects had read the second part of the instructions, the following summary was read out aloud:

Compared to the first part, there are two changes, one minor and one major. First, now you will not decide yourself which group to enter. Rather, the computer will make this decision for you. However, the mechanisms through which this happens is likely the same as your own decision rule in the first part: whenever your signal is above 100, you enter the red group and whenever your signal is below 100, you enter the blue group. Thus, this change is rather minor. More importantly, after you have provided your estimate, you will see the estimates of two other randomly selected participants who are present in this room at the moment. You then need to provide a new estimate, which can either be the same or a different one, as you wish. Just to be clear, every participant in this room receives exactly the same information, i.e., everyone gets the same hint, enters the same group, and talks to the same computer players. Thus, nobody in this room will see anything different on their computer screen than you.

F.7 Treatment *Salience*

This treatment was identical to the selected treatment, except for the hint provided in the main text.

F.8 Treatments *Simple* and *Intermediate*

The paper-based instructions for these treatments were identical to those in *Selected*, except for the set from which the true state and the signals were drawn.