Belief-Based Utility and Signal Interpretation*

Marta Kozakiewicz†

Job Market Paper

This version: October 17, 2020

Click here for the most recent version

Abstract

People tend to overestimate their abilities and chances of success, even though inaccurate beliefs lead to costly mistakes. How can these beliefs persist in the face of feedback? I propose a novel experiment to test whether people perceive favorable feedback as more informative. Using experimental data, I provide the first causal evidence that utility from beliefs affects perception of signal informativeness. To establish causality, I adopt a matching estimator approach and construct a counterfactual outcome of an agent who observed the same signal, but this signal didn't affect his belief-based utility. I find a strong and significant effect, with a positive asymmetry: subjects tend to interpret favorable signals as more informative. The results cast a new light on the origins of overconfidence and illuminate mechanisms that perpetuate it regardless of received feedback.

Keywords: overconfidence, belief formation, learning, experiment

JEL classification: C91, D83

*The author gratefully acknowledges funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) through CRC TR 224 (project A01).

†Bonn Graduate School of Economics; email: `martkozakiewicz@gmail.com`

1 Introduction

It is widely established that people tend to overestimate their abilities and chances of success, making costly mistakes as they hold on to their biased beliefs at the expense of accuracy. This tendency, commonly referred to as *overconfidence*, generates significant costs for both the individual and the society.[1] A long-standing question in behavioral economics is how it can persist in a world in which individual performance is constantly evaluated, and the results are reported back to the agent?

In this paper, I explore one possible explanation.[2] I consider an agent who does not know his ability and receives a signal that either reveals it or not. The agent is forming beliefs about both his ability *and* how informative the signal is. Importantly, he values his beliefs about his ability, so that any change in these beliefs directly affects his utility function (Brunnermeier and Parker, 2005; Caplin and Leahy, 2019; Kőszegi, 2006).

I attempt to answer the following questions: does the agent perceive a favorable signal to be more informative than an unfavorable one? Would he perceive the signal differently if the signal did not affect his utility function?

To this end, I designed a simple experiment in which subjects are learning about an ego-relevant parameter.[3] Participants observe a signal about their intelligence and report their beliefs about its informativeness. I compare these beliefs to a control condition, in which participants do not receive signals but make reports ex ante, for every possible signal realization. A signal received in the treatment condition affects the agent's beliefs about his ability and brings him additional utility (or, in case of unfavorable signals, disutility), while the one that is only being considered in the control does not. The drop in utility can be mitigated by distorting one's beliefs about the signal informativeness. I measure these distortions using an incentive-compatible elicitation method.

[1] Negative consequences of overconfidence include, among others, excessive selection into competitive environments (Camerer and Lovallo, 1999; Niederle and Vesterlund, 2007), excessive trading (Barber and Odean, 2001), suboptimal investment decisions (Malmendier and Tate, 2005, 2008), and political polarization (Ortoleva and Snowberg, 2015).

[2] Other explanations that are similar to my work, as they consider motivated reasoning rather than cognitive processes, can be divided into three broad categories: information avoidance (see Golman et al., 2017, for a comprehensive literature review), selective recall (Chew et al., 2019; Huffman et al., 2019; Zimmermann, 2020), and asymmetric updating (reviewed in Section 2).

[3] The experiment was pre-registered in the AEA RCT Registry (the registration number: AEARCTR-0006233). The details of the registration are provided in Appendix F.

The data support my main hypothesis that changes in belief-based utility affect subjects' perception of signal informativeness. I find a strong and significant effect, with a positive asymmetry: subjects tend to interpret favorable signals as more informative. In comparison to the control condition, participants' beliefs about a signal being entirely informative increase by 10.6% after a positive signal (a 27.9% increase in relative terms). Moreover, I show that the effect depends on the subject's expectations. In particular, it is entirely diminished if a subject assigned zero prior probability to the state of the world indicated by the signal.

I make the following contributions. My study provides the first clear evidence of a causal effect of belief-based utility on signal interpretation. While the research on updating about ego-relevant traits has a long tradition (I review the literature in Section 2), establishing causality has always been challenging. One common difficulty is a non-random assignment of signals to the agents. The problem arises because, with informative signals, high-ability or underconfident subjects are more likely to receive "good news". The resulting selection bias undermines the causal inference. I address this issue by proposing a matching estimator that enables me to compare participants in the treatment and in the control conditions that are of the same ability, reported similar beliefs, and considered the same signal.

Another challenge lies in introducing exogenous variation in the way signals affect belief-based utility. Previous work focused on comparisons between updating subjective beliefs about ego-relevant traits and updating objectively given probabilities of some unknown, ego-neutral parameter (Coutts, 2019; Eil and Rao, 2011; Ertac, 2011; Möbius et al., 2014).[4] I propose a control condition that is based on subjective beliefs over the same ego-relevant characteristic as in the treatment condition. It introduces exogenous variation in belief-based utility without changing other aspects of the treatment.

Moreover, my design introduces a richer state and signal space in a simple way. This allows me to relate signals to the subjects' expectations and capture the effect that was elusive for studies that were using different signal structures (Coutts, 2019; Eil and Rao, 2011; Grossman and Owens, 2012; Schwardmann and Van der Weele, 2019).

[4] One exception is a study of Buser et al. (2018) who compare updating about tasks that differ in how relevant they are to the agent's self-esteem. However, in their set-up, it is not possible to introduce exogenous variation in ego-relevance. Grossman and Owens (2012) propose a control condition in which participants are learning about the test result of another subject. In this case, subjects are updating their subjective beliefs about an unknown, ego-neutral variable.

The study was conducted in August 2020 in the BonnEconLab at the University of Bonn. In total, I collected data from 222 participants. The experiment consisted of several parts. Firstly, participants solved an IQ test and were incentivized to do their best. After the test, they were asked to report their subjective beliefs about their relative performance. I elicited, using an incentive-compatible mechanism, subjects' beliefs about their test score falling into the 1^{st}, 2^{nd}, ..., 10^{th} decile of the score distribution. I referred to the deciles as "ranks", with 1 denoting the highest and 10 denoting the lowest rank.

After the elicitation procedure, I explained the main task in the following way. There are two boxes. Box 1 contains 10 balls with numbers 1 to 10 written on them (each number occurs exactly once). Box 2 contains 10 balls with the same number written on them. That number is equal to your rank.

In the main task, one ball was randomly drawn from one of the boxes (either box could be selected with equal probability) and presented to the subject. After seeing the ball, the participant reported his beliefs about the event that the ball came from Box 2 (with his rank). The report was made by dividing 100 points between the two boxes. I incentivized truthful reporting with the Binarized Scoring Rule (Hossain and Okui, 2013). The method was explained to the participants and they were informed that their chances to win the highest reward are maximized when they divide their points in a way that corresponds to their true beliefs about the box. We explained in intuitive terms how one can arrive at a Bayesian update given one's prior beliefs about the rank.

An ideal counterfactual to the treatment would include an agent who has the same prior belief distribution (or the same set of prior distributions if the agent had multiple priors) and observes the same signal, but the signal has no effect on his belief-based utility function. To introduce exogenous variation in how signals affect subjects' belief-based utility, I designed the following control condition.

In the control condition, subjects do not see a ball being drawn but are asked to report their beliefs about signal informativeness ex ante, for every possible signal realization. The procedure, known as the Strategy Method, is commonly used in experiments investigating strategic interactions in games (Brandts and Charness, 2009). To alleviate concerns about the non-comparability of the two treatments, I adopted special procedures targeting the issues raised in the literature.[5] I argue that a participant in the control condition faces the same decision as a subject in the treatment condition but without the signal affecting his beliefs.

Note that, although the assignment of subjects to the treatment and control condition is random, the assignment of signals to subjects is not. Participants in the treatment condition make a report about one number that, with probability $\frac{1}{2}$, is their true rank. In the control condition, every participant reports his beliefs about all ten numbers. This leads to the covariance between the treatment status and numbers considered by subjects. To draw a causal inference, I follow Heckman et al. (1998) and construct a matching estimator.

For every participant in the treatment condition, I construct a counterfactual observation using decisions from the control condition regarding *the same number* as the one seen by the subject in the treatment condition. I weigh those decisions based on the similarity in subjects' ranks and reported prior beliefs. Those participants in the control condition whose true rank and prior belief distribution were closer to the rank and beliefs of the participant in the treatment condition receive a higher weight.[6] I estimate those weights using a kernel regression.[7]

The results strongly support my hypothesis. There is a significant difference in the reported probability of a signal being informative in the treatment condition in comparison to the control. The effect is driven by differential response to signals that

[5]One concern raised in the experimental game theory literature is that players may gain a better understanding of the game if they are induced to think about the best strategies from the perspective of other players. One can imagine that considering every possible signal the control condition may influence subjects' beliefs. I address this issue by presenting participants in the treatment condition with the screen-shots from the control condition and asking them to think about every possible draw before they proceed to the main task. While only participants in the control condition are allowed to enter their choices, both groups are required to consider every signal realization. Moreover, I hope to alleviate another concern, the problem of framing the answers in the strategy method with the order of options, by randomizing the order of the signals presented to the subjects in both conditions.

[6]I consider two alternative specifications: a matching based solely on the prior belief distribution and a matching based on one's rank and prior belief about the signal at hand. The results are very similar to those of the first specification.

[7]In addition to enabling causal inference, a non-parametric matching estimator is efficient in presence of nonlinearities. I chose this method to handle the nonlinear effects observed in the data.

are above and below one's median prior belief. Favorable signals are believed to be 10.6% more likely to be informative than unfavorable signals. The effect increases to 15.7% if we control for signals "too good to be true" – favorable signals that are outside of the subject's prior belief distribution.

I provide additional evidence, based on subjects' responses in questionnaires, to support my interpretation of the results as being driven by changes in belief-based utility. In the treatment condition, those participants who reported experiencing hopelessness (a negative anticipatory emotion) tend to deviate more from the Bayesian benchmark.

The effect is counteracted by the habitual use of emotion regulation strategies. Subjects who reported using more emotion regulation in their daily life tend to deviate less from Bayesian updating, even if they admitted to feeling more hopeless. While only suggestive, the evidence supports the view that the treatment effect is stemming from the visceral, emotion-based reaction to signals indicative of belief-based utility.

The paper is organized as follows. In the next section, I describe the relevant literature. In Section 3, I outline the experimental design. Section 4 presents the empirical results, and Section 5 describes the additional evidence. Section 6 concludes.

2 Literature Review

My work is based on the theoretical literature on overconfidence and belief formation. That literature postulates that people derive utility not only from physical outcomes but also from their beliefs about the current or future state (Brunnermeier and Parker, 2005; Caplin and Leahy, 2019; Kőszegi, 2006). The individual can choose his beliefs but faces a trade-off between their accuracy (necessary to take the optimal action) and their desirability (a consequence of the non-monetary value beliefs bring to the agent).[8] The tension is resolved by the agent manipulating his beliefs to the extent that he is not losing too much from action taken based on those beliefs.

[8]Some behavioral studies emphasize the consumption value of beliefs (due to pleasant or unpleasant emotional reactions they tend to induce), others stress the importance of non-classical instrumental value including motivational value, signaling value, or value from serving as a commitment device (see Bénabou and Tirole, 2016 for a comprehensive review of the literature).

The main difficulty in empirically investigating the belief-based component of the utility function is that not only we cannot observe *preferences* over different belief distributions but also, in opposition to the physical outcomes, we have limited information about the resulting *choices*, as we usually do not observe the choice set: all the distributions of beliefs the agent is choosing from.

Given these difficulties, it is unsurprising that we rarely model belief formation as a choice made, more or less consciously, by the agent. In fact, most studies conceptualize belief formation as beliefs updating assuming that beliefs, well-defined and probabilistically quantified, follow a pre-specified set of rules, Bayesian updating being the prime example. The Bayesian approach, bolstered by axiomatic derivation justifying its position of a rational benchmark, became the most prevalent model of beliefs updating (Gilboa and Marinacci, 2016).

Although it is a good approximation to reality in some contexts, the Bayesian model seems to be less adequate in others. One such an example is a situation in which the decision-maker has a clear preference over the states of the world, as the in case of learning about an ego-relevant trait. Several studies demonstrated that agents significantly deviate from Bayes' rule when forming beliefs about their own intelligence or beauty (Buser et al., 2018; Coutts, 2019; Eil and Rao, 2011; Ertac, 2011; Grossman and Owens, 2012; Möbius et al., 2014; Schwardmann and Van der Weele, 2019). The main conclusion emerging from this strand of literature is that belief formation over ego-relevant characteristics significantly differs from learning about ego-neutral variables. At the same time, the direction of the effect and its magnitude vary across studies.

The design presented in this paper differs from these experiments in important ways. First of all, subjects in my study observed only one signal. I aimed at disentangling the effect of attribution to noise from the way agents are aggregating information (which may also be affected by motivated reasoning, but is beyond the scope of the paper).[9]

[9]For that reason, my experiment is also related to the literature on self-serving attribution bias. It has been extensively studied by psychologists (see Mezulis et al., 2004, for a meta-analysis of the existing studies) and, more recently, by economists (Coutts et al., 2020; Hestermann and Yaouanq, 2020; Van den Steen, 2004). None of the studies, however, consider the counterfactual scenario discussed in my paper.

Secondly, in my set-up, the signal is either perfectly informative or entirely uninformative, with equal probability known to the subjects. This allows us to control for the extent to which subjects "compress" probabilities towards 50%, the effect observed in updating about ego-neutral variables (Ambuehl and Li, 2018; Enke and Graeber, 2019).

Moreover, I use a richer state and signal space compared to the above-mentioned studies. To understand why it is important, imagine a participant who believes that he is in the 80th percentile of the IQ test score distribution. Receiving a coarser signal, e.g. a signal indicating that his score was above the median, would not influence his beliefs as it merely confirms what he already knows. However, if the signal was more precise, e.g. it revealed that his score was only in the 60th percentile, it would affect his beliefs and, according to my hypothesis, induce a stronger reaction.

The idea presented in this paper is related to research on emotions and decision-making (Lerner et al., 2015). One conclusion from the psychological literature is that emotions may influence decisions via changes in the content of thought, and vice versa. A similar hypothesis has been tested in a recent study of Engelmann et al. (2019) who investigate the impact of anxiety on wishful thinking. Using data from a carefully designed experiment, they show a causal effect of anticipatory anxiety on belief formation. Although I cannot argue about the causal impact of anticipatory emotions in my experiment, the suggestive evidence is in line with their findings.

3 Experimental Design

The experiment consisted of two parts and is outlined in Figure 1. In the first part, subjects completed an IQ test intended to assess their cognitive ability. The second part included the elicitation of prior and posterior beliefs and a stage in which subjects received signals (or considered every possible signal realization in the control condition). I describe the procedures in detail in the following subsections.

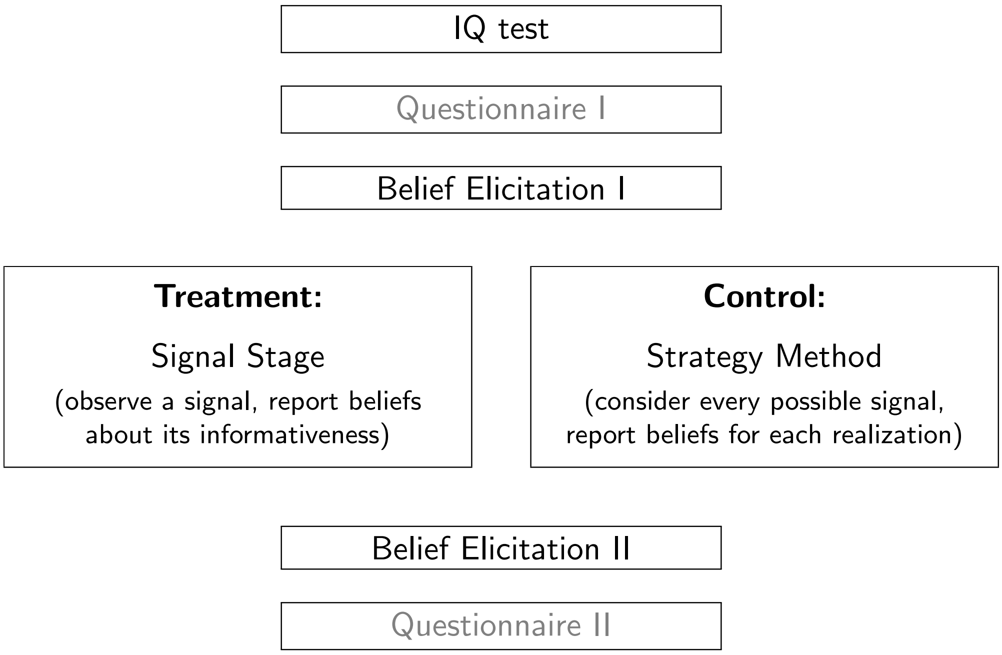


Figure 1: The outline of the experiment.

3.1 IQ Test

In the first part of the experiment, I evaluated the subjects' cognitive ability using an IQ test.[10] The test consisted of 29 standard logic questions and participants were asked to solve as many of them as possible in 10 minutes. Individual scores were calculated based on the number of correctly answered questions minus the number of incorrect answers, and subjects were paid 0.75 Euro for every point they obtained.

[10]I decided to use intelligence as a basis for the learning exercise for several reasons. Firstly, it is known that intelligence correlates strongly with educational achievement, success in the labor market, and income. Because of that, I expect people to care deeply about their cognitive ability. Therefore, IQ measure seems to be a good candidate for a genuine ego-relevant parameter. Secondly, the literature provides evidence that people have biased beliefs about their cognitive ability (with overconfidence prevailing among men), which suggests that learning about one's cognitive ability may be one of natural settings in which the mechanism is in play.

Participants were informed that their earnings from the IQ test will be added to their earnings from the remaining parts of the experiment and paid at the end of the session. They were also informed that, although they will receive the entire sum of money at the end of the study, they will not learn immediately the exact number of points they obtained in the IQ test, nor how much money they earned in each part. Participants were informed that their IQ test results and the details of their payoffs will be available to them in one week after the session. Each participant received a personal link to a website on which his individual information was posted one week later.

3.2 Belief Elicitation

At the beginning of the second part, participants were told that they have to complete 3 tasks, for which they can earn up to 12 Euro. They were informed that *one task* will be drawn at random at the end of the session, and they will be paid only for that task.

In the first task, I elicited subjects' beliefs about their test scores being in the $1^{st}, 2^{nd}, ..., 9^{th}$ and 10^{th} deciles of the distribution of the test scores of 300 participants who took the same test in the BonnEconLab in previous sessions. I introduced 10 "ranks", with Rank 1 denoting the highest rank (assigned to participants whose IQ test scores were higher than or equal to the test scores of $90 - 100\%$ of all participants), and Rank 10 denoting the lowest rank (defined analogously). The first task was to allocate 100 points among the ranks in a way that reflects one's beliefs about his relative performance in the IQ test.

The screen-shot of the computer interface used by subjects is presented in Figure 2. Participants were allocating points by dragging blue arrows to selected positions. They were informed that they can move the arrows back and forth to correct their choices. The text below the scales informed a participant how many points are being allocated to a given rank and the allocation was immediately appearing on the graph to the right. The number above the graph indicated how many points the participant still has to allocate before he can proceed to the next task.

Figure 2: The screen-shot of the interface used by subjects in belief elicitation.

To incentivize truthful reports, I used the Binarized Scoring Rule following Hossain and Okui (2013). The random variable X can take one of 10 values: (1,0,...,0,0), (0,1,...,0,0), ..., (0,0,...,1,0), (0,0,...,0,1); the position of 1 indicates in which decile subject's IQ test score fell. After receiving agent's report $x = (x_1, ..., x_{10})$, where x_i denotes the share of points allocated to decile $i \in \{1, ..., 10\}$, I observed his IQ test score in the k^{th} decile, and the agent won the prize if the QSR for multiple events,

$$s(x, k) = 2x_k - \sum_i x_i^2 + 1,$$

exceeded a uniformly drawn random variable with the support $[0, 2]$.

The formula was presented to the subjects in a simple way (avoiding mathematical notation). Importantly, I told participants the main implication of the method, that is, the probability of getting a large prize (12 Euro) is maximized when they allocate their points in a way that reflects their true beliefs about their rank.

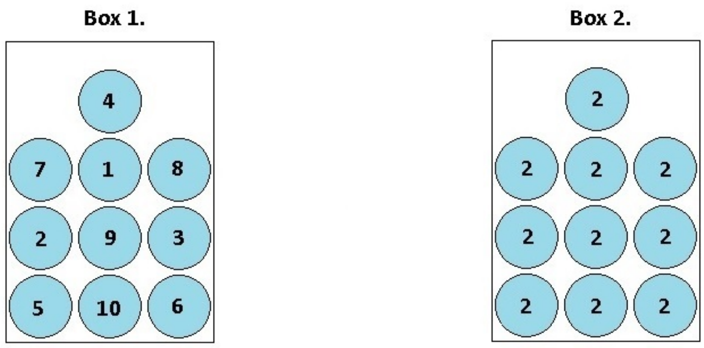


Figure 3: The composition of the boxes of a person whose rank was 2.

I followed the same procedure during the second belief elicitation, after the signal stage (after the strategy method in the control condition). However, during the first belief elicitation, subjects were not aware that they will be asked to state their beliefs one more time.

3.3 The Signal Stage

After eliciting the prior beliefs, Participants were given instructions for the second task. We explained the nature of the task in a simple language, using pictures and two illustrative examples. The task was framed in a neutral way and described as follows.

There are two boxes: Box 1 and Box 2. Each box contains 10 balls with numbers written on them. Box 1 contains balls with numbers from 1 to 10, and every number appears exactly once. The composition of the second box depends on the subject's rank in the IQ test. Box 2 contains 10 balls that all have one number written on them, and this number is equal to the individual rank. The composition of the boxes of a person assigned Rank 2 is presented in Figure 3.

For every participant, the computer program randomly selected one of the two boxes. Next, a ball was drawn from the selected box and displayed on the participant's screen. The participant did not know which box the ball was drawn from, but he knew that either box can be selected with equal probability. After seeing the ball, he had to state his beliefs about the box selected by the computer.

I used the same incentive-compatible elicitation method as for the prior and posterior belief elicitations. Participants had 100 points to allocate between Box 1 and Box 2 in proportions that reflect their beliefs about the source of the signal, and were rewarded for the truthful report with a higher probability of getting a large prize (12 Euro).

Importantly, subjects were instructed how to arrive at Bayesian posterior given one's prior belief distribution. I explained it on an example in two steps. Firstly, I demonstrated how a person should allocate her points after different signal realizations if she knew precisely her rank. Then, I showed how a person should allocate her points if she was not sure about her rank, but was assigning a certain probability to it.

Step 1: How should a person ranked 2 allocate her points if she knew for sure that her rank is 2, and saw a ball with a number "2" on it? There are 10-times as many balls with "2" in Box 2 as there are in Box 1, hence it is 10-times as likely that the ball came from the second box. Therefore, the person should allocate 9 points to Box 1, and 10-times as many, 90 points, to Box 2 (the remaining point should be allocated to the box with higher probability).

Step 2: What if a person did not know her true rank, but she believed that there is 30% chance that her rank is 2? The same logic applies to this case. One can visualize 30% chance as 3 out of 10 balls in Box 2 having a number "2" on them.[11] In this imaginary case, there are 3-times as many balls with the number "2" on them in Box 2 as in Box 1, implying an allocation of 25 points to Box 1 and 3-times as many (75 points) to Box 2.

The interface enabled subjects to split their points in desired proportions without calculating the respective ratios. The screen-shot of the interface used in the second task is presented in Figure 4. Crucially, the text below the scale informed subjects about their current allocation and the ratio between points allocated to the two boxes. By moving

[11] One reason why I decided to introduce 10 balls was the ease of exposition in a case when a person is uncertain about his rank.

the cursor participants could choose the number of points corresponding to allocating x-times as many points to one of the boxes (with $x \in \{1, 1.1, ..., 99\}$). The graph below was illustrating the current allocation.

Before proceeding to the signal stage, participants were required to answer a set of control questions, designed to check their understanding of the task (including the steps necessary for arriving at Bayesian posterior). The control questions also pointed out the aspects that participants may have missed at the first reading, but were necessary to fully comprehend the task.

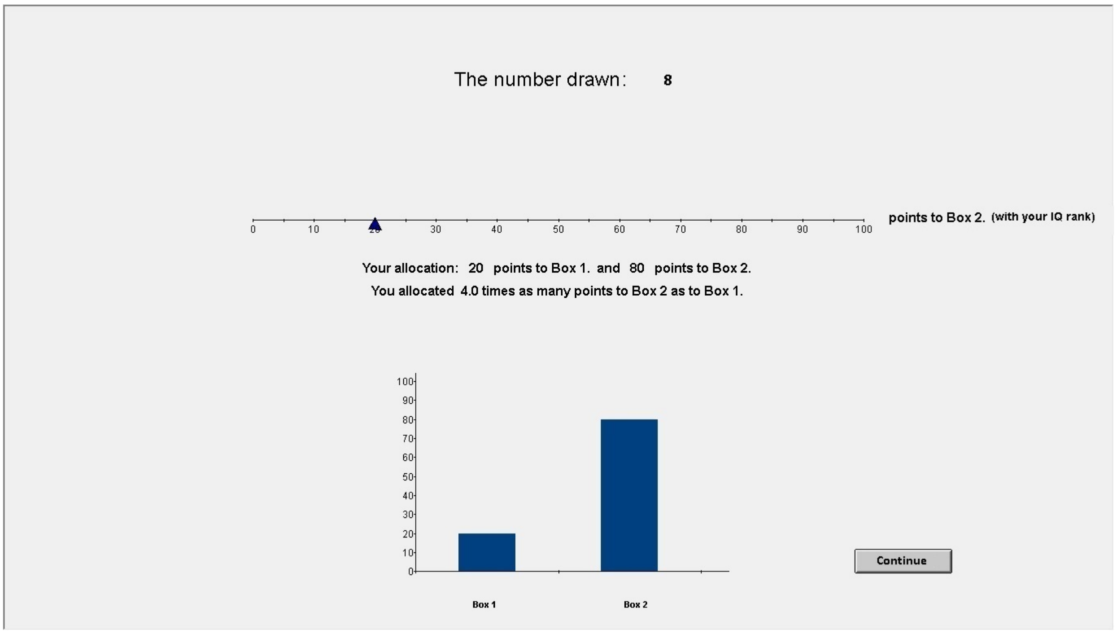


Figure 4: The screen-shot of the interface used in the second task (the signal stage).

3.4 Experimental Conditions

I introduced two experimental conditions: treatment and control. In the control condition, subjects did not see the number that was drawn but were asked to state their beliefs for every possible draw. The procedure, known as the Strategy Method, is commonly used in experiments investigating strategic interactions in games.

I informed participants in the control condition that the choices they are making are not entirely hypothetical. At the end of the session, one box was selected by the computer program and one ball was randomly drawn from the selected box. Subjects were paid as in the treatment condition, based on the decision that corresponded to the number drawn from the box. Note that the procedure is incentive-compatible as the probability of drawing any number is at least 5%.[12]

To alleviate concerns of the non-comparability of the two conditions, I adopted special procedures targeting the issues discussed in the literature. One concern raised in the experimental game theory literature is that players in the strategy method gain a better understanding of the game as a consequence of considering the problem from the point of view of different players. In my set-up, one can imagine that considering every possible signal realization may influence reported beliefs in the control condition.

For this reason, I asked the participants in the treatment condition to consider every possible signal realization *before* they saw the actual draw. Subjects were required to go through 10 slides, presented in random order, with the actual screen-shots of the interface displayed in the control condition. Participants were asked to contemplate a hypothetical decision in each slide before clicking on the button "Continue", which appeared on the screen only after 15 seconds. While only subjects in the control condition were allowed to enter their choices, both groups were required to go through the task.

Another problem that may arise in the Strategy Method is framing the answers with the order of options. I addressed the issue by randomizing the order of the numbers displayed to a subject in the control condition, and the order of slides presented to participants in the treatment.

[12]However, if subjects were weighting the cost of cognitive effort against the expected payoff, they may exert less effort in the control condition. In this case, one would expect subjects to behave *less* rationally: their decisions would be characterized by a higher variance and they would end up further away from Bayesian update. This is the opposite of what I found.

3.5 Questionnaires

After each part of the experiment, I asked participants to fill in a 3-page questionnaire. The first set of questions, displayed on individual computer screens after the IQ test, included a short version of the Big-5 personality test (Gerlitz and Schupp, 2005) and the state-trait anxiety inventory STAI (Spielberger, 1983).

The Big-5 personality test was designed to measure personality along five dimensions: extroversion, conscientiousness, openness to experience, neuroticism, and agreeableness. The STAI measures the current state of anxiety and anxiety level as a personal characteristic. The second set of questions, answered by the participants after the main task, comprised the Emotion Regulation Questionnaire (Gross and John, 2003) and a subset of questions from the Achievement Emotions Questionnaire (Pekrun et al., 2011).

The Emotion Regulation Questionnaire was designed to assess the habitual use of two strategies commonly used to alter emotions. To alleviate the emotional impact of a situation, one may try to reinterpret it in a different way. This emotion regulation strategy, broadly referred to as *reappraisal*, relies on "applying mental models to the often ambiguous and incomplete information" (Uusberg et al., 2019). The second emotion regulation strategy, *suppression*, involves "inhibiting ongoing emotion-expressive behavior" (Gross and John, 1998, cited in Uusberg et al., 2019).

People differ in their use of reappraisal and suppression, and these differences have implications for their experiences of emotions, behavior in response to those emotions, and general well-being (Gross and John, 2003). The habitual use of the two strategies is measured by the degree to which subjects agree with particular statements, e.g. "I keep my emotions to myself" or "When I want to feel less negative emotion, I change the way I'm thinking about the situation". I use the exact 10-item questionnaire developed by Gross and John (2003).

The Achievement Emotions Questionnaire was designed to measure *achievement emotions* (emotions that are directly linked to achievement activities or achievement outcomes) experienced by students in academic settings (Pekrun et al., 2011). I adopted part of the questionnaire to measure the following test-related emotions: enjoyment, hope, pride, relief, anger, anxiety, shame, and hopelessness.

Participants in both conditions were asked to report what they felt *after* learning the nature of the task, but *before* they saw the number(s). They had to indicate, using a 7-point Likert scale, how strongly they agree (or disagree) with various statements, e.g. "I was proud of how well the test went", or "I was angry about the task I had to do" (see Appendix G for the entire list of questions and the instructions).

4 Results

The experiment took place in August 2020 in the BonnEconLab at the University of Bonn.[13] I conducted 52 sessions, with 1 to 6 participants in each session. I collected data from 167 participants in the treatment condition and 55 participants in the control condition. The experiment lasted around 80 minutes and the participants earned 21.25 Euro on average.

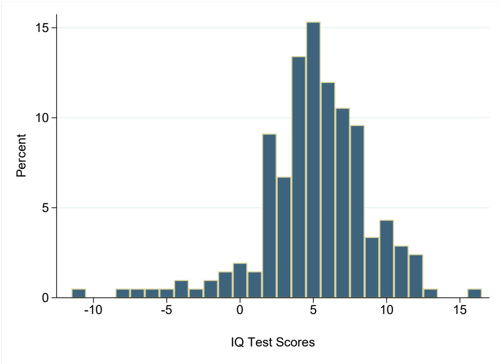
In the following section, I report the analysis based on the data from 209 participants who correctly answered at least half of the control questions (I excluded 13 participants, that is 5.8% of the sample).

4.1 IQ Test Results and Individual Ranks

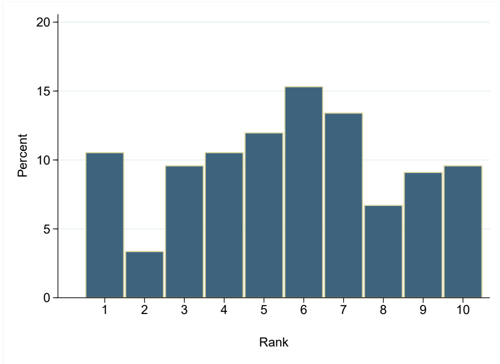
Figure 5 presents the distribution of the IQ test scores and ranks assigned to the participants based on the test results. The IQ test score distribution is fairly symmetrical (skewness −0.83), with a mean of 5.13 and a standard deviation of 3.73. The average rank is 5.65 with a standard deviation of 2.67.

[13]Due to the Covid-19 pandemic, I followed special procedures to ensure the safety of participants and others involved. The number of participants per session was restricted to 6 to ensure each participant a place in a separate room. Desks, chairs, and computer equipment were disinfected after every session and the rooms were aired before every session for at least half an hour. At the time of the experiment (August 2020), the Covid-19 pandemic was mostly under control in Germany; the lockdown restrictions were eased, allowing restaurants, schools, and public places to open with appropriate safety measures.

Figure 5: IQ Test Results and Individual Ranks.



(a) Distribution of the IQ test scores.



(b) Distribution of individual ranks.

Importantly, there is no significant difference between the mean IQ test result in the treatment and control conditions (p-value = 0.94), nor is there a difference between mean rank in the two conditions (p-value = 0.61).

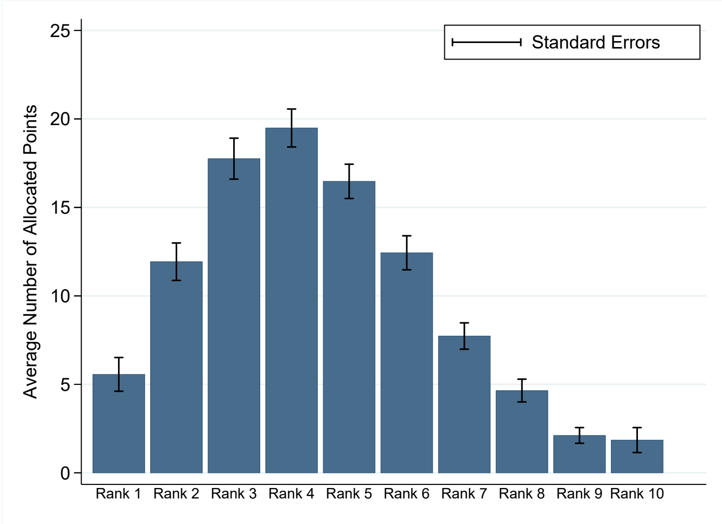
4.2 Prior Beliefs about Rank

Before the main task, I elicited from every participant his entire belief distribution. I analyze the data in two ways. Firstly, I look at the aggregate belief distribution. Then, I examine individual distributions and report the averages of individual measures (these include mean belief about rank, median and range).

To look at the aggregate of individual belief distributions, I treat separately every decision to allocate x points, $x \in \{0, ..., 100\}$, to rank k, $k \in \{1, ..., 10\}$. For each of 10 ranks, I calculate the average number of points allocated by the participants. The resulting aggregate distribution is presented on Panel a) in Figure 6 (each bar indicates the average $+/-$ standard errors). It is visibly skewed to the right, with the mean belief 4.47 and the median 4. On average, the subjects appear to be *overconfident* as they put a higher probability mass on lower (better) ranks.

In Table 1, I report the averages of individual measures of belief distribution. The averages, however, mask the fact that only 26 participants revealed symmetric belief

Figure 6: Average number of allocated points.



distribution. Almost half of all subjects (100 participants) revealed positively skewed belief distribution, and the remaining 83 participants revealed negatively skewed belief distribution (the average difference between mean and median in both groups was 0.21).

Importantly, there is no significant difference between the treatment and control conditions in the measures reported in Table 1, nor in the absolute differences between mean and median belief.

I define a person to be *overconfident* if his median belief is lower than his true rank. Similarly, I use a term *underconfident* to describe a person who assigns 50% or more probability mass to ranks higher than his true rank. A person is defined to be *unbiased* if his median belief matches his true rank. Note that, in common understanding, Rank 1 denotes "the highest" rank, while Rank 10 is "the lowest". To avoid confusion, I will not use the customary phrases, but the terms that match the values.

Table 1: Individual belief distributions.

	Mean Belief	**Q1**	**Median**	**Q3**	**Range**
Mean	4.47	3.71	4.45	5.16	4.89
(Std. Dev.)	(1.75)	(1.74)	(1.79)	(1.87)	(1.57)

Using this definition, there are 127 overconfident, 58 underconfident and 24 unbiased participants in my sample. In Appendix E, I report the average rank and beliefs separately for the three types and address the question of apparent overconfidence (Benoît and Dubra, 2011).

4.3 Beliefs about Signal Informativeness

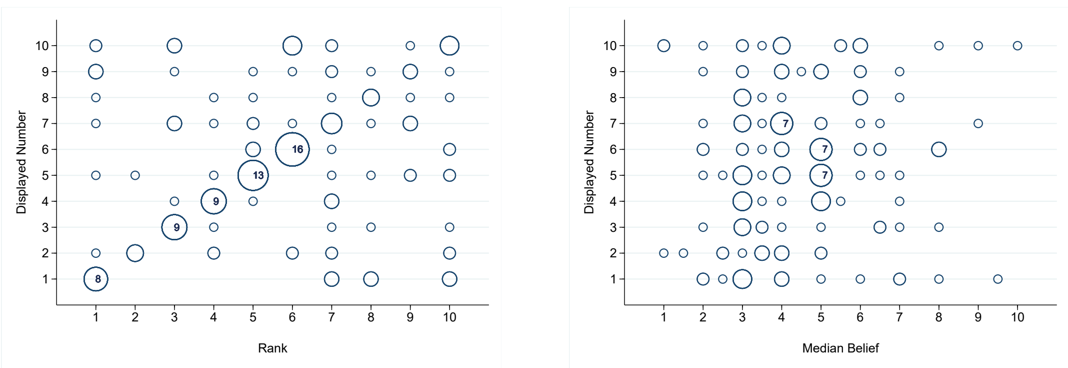
The main experimental task, neutrally framed in the instructions as “the second task”, differed depending on the condition. In the treatment condition, subjects observed one number drawn from the selected box and reported their beliefs about the box from which the number was drawn.

In the control condition, participants saw, in random order, numbers from 1 to 10, and stated a report for each one of them. In this subsection, I present the raw data on received signals and reported beliefs, as well as the results of the data analysis.

4.3.1 Received Signals

In Figure 7, I present the numbers displayed to participants in the treatment condition. The size of the hollow circles is proportional to frequencies. The numbers inside the circles denote the frequency. As expected, more than 50% of all participants saw a number equal to their actual rank.

Figure 7: Signals received in treatment (circle size reflects frequency).



(a) Depending on subject’s rank. (b) Depending on subject’s median belief.

Moreover, 69 participants (43% of all participants in the treatment condition) observed a number smaller or equal to their median belief. Those participants received a "good" signal, as a lower number denotes a better rank. 91 participants (57% of all) observed a "bad" signal – a number strictly higher than their median belief.

Of those who received a "good" signal in the treatment condition, 26 participants were classified as overconfident, 32 as underconfident, and 11 as unbiased. Among participants who obtained a "bad" signal, I classified 71 subjects as overconfident, 13 as underconfident, and 7 as unbiased.

4.3.2 Raw Data

In this section, I describe reports made by participants in the second task. Firstly, I look at raw data from the treatment and control conditions.

As presented in Table 2, there is a significant difference in the perceived signal informativeness in the treatment and the control conditions. On average, the participants in the treatment condition report a probability of 37.81% that the signal came from Box 2 (with numbers indicating their rank). Their report is 8.54% higher than the report of subjects in the control condition. The difference is significant at 0.01 level.

Table 2: Mean comparison of beliefs about the signal informativeness.

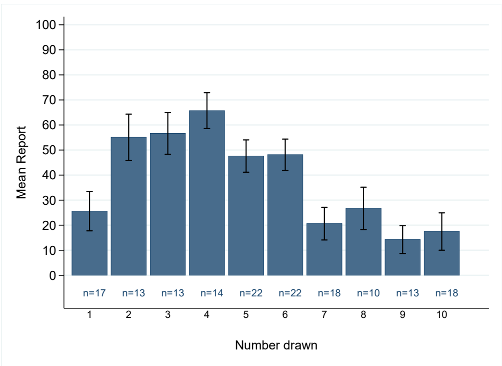
	Mean	**Std. Err.**	**N**
Control	29.27	1.42	490
Treatment	37.81	2.63	160
Difference	-8.54	2.91	650
	H_0: Diff < 0	H_0: Diff $\neq 0$	H_0: Diff > 0
p-value	0.0017	0.0034	0.9983

In the first two graphs in Figure 8, I present the average number of points allocated to Box 2 (the box with numbers equal to the subject's rank). Panel a) presents reports made by subjects in the treatment condition after receiving a signal (shown on the x-axis). Numbers above the x-axis indicate how many participants received a signal and, subsequently, stated a report. Panel b) shows reported by subjects in the control condition. The two graphs look differently, with subjects allocating more points to Box 2 in the treatment condition regardless of the signal realization.

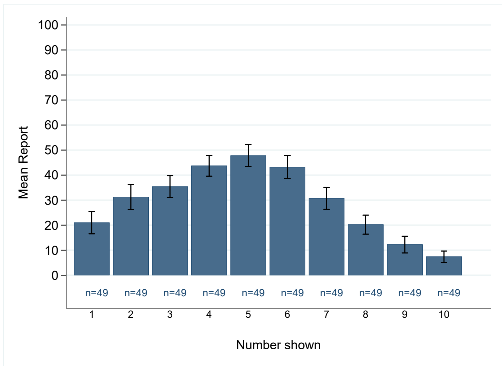
It is worthwhile to look at reports compared to the Bayesian benchmark. For every participant, I calculated a Baysian posterior about the box given the signal realization. I assume the Bayesian posterior to be zero if a participant put zero prior probability on the number displayed on his screen.

On Panel c) and d) in Figure 8, I present the average deviations from the Bayesian benchmark in the treatment and control conditions. The mean deviations are larger in the treatment for some signal realizations, although with large standard errors.

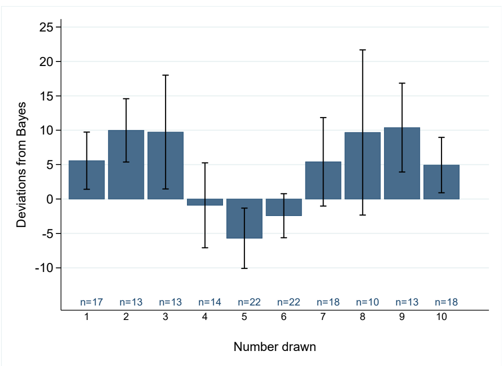
Figure 8: Beliefs about the signal informativeness in the two conditions.



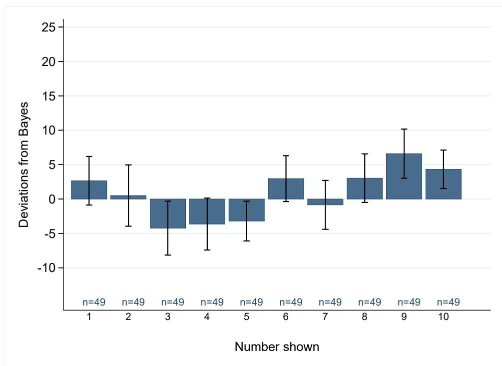
(a) Points allocated to Box 2 (treatment).



(b) Points allocated to Box 2 (control).



(c) Deviations from Bayes (treatment).



(d) Deviations from Bayes (control).

In Appendix H, I show similar graphs for participants who received a signal that was within their prior belief distribution (to which they attached non-zero prior probability), and those who received a signal that was outside of their priors. I also show graphs for participants who received "good" versus "bad" signals. The differences between the treatment and control conditions are visible in both cases, although a small sample and the resulting large standard errors make it hard to compare.

4.4 Data Analysis

In this section, I answer my research question: does perceived informativeness of a signal depend on the signal's valence? In other words, I aim at isolating the effect of receiving a good signal on the subject's belief about its informativeness.

4.4.1 Matching Estimator

Note that the research question does not refer to the treatment effect itself, but rather the heterogeneity in the treatment effects. Although the assignment into the treatment and control condition is random, the assignment of signals to agents is not. Participants in the treatment condition are presented one number which, with probability $\frac{1}{2}$, is their true rank. This is visible in Panel a) in Figure 7. The hollow circles are much larger on diagonal, meaning that participants are more likely to observe their true rank than any other number.

Imagine a subject who believes his rank is 1. In the control condition, he would consider all ten numbers, and 9 out of 10 decisions would pertain to an unfavorable signal. On the other hand, if he was in the treatment condition and his rank was indeed 1, he would receive a bad signal with much lower probability: $\frac{1}{2} \times \frac{9}{10}$. This leads to the covariance between the treatment status and signals considered by the participants. If there are reasons to believe that people with different beliefs or ranks respond differently to good signals, a simple comparison of means would not recover the treatment effect. I present this argument formally in Appendix B.1.

Moreover, the mapping from observables to outcomes is likely to be non-linear, and errors in the least squares regression are heteroskedastic (a graph of the residuals and formal tests are presented in Appendix B). As a consequence, the OLS estimates may not be efficient. Nevertheless, I conduct regression analysis and report the results in Appendix A.

For these reasons, I use a different approach to analyze the data. I follow Heckman et al. (1998) and construct a matching estimator:

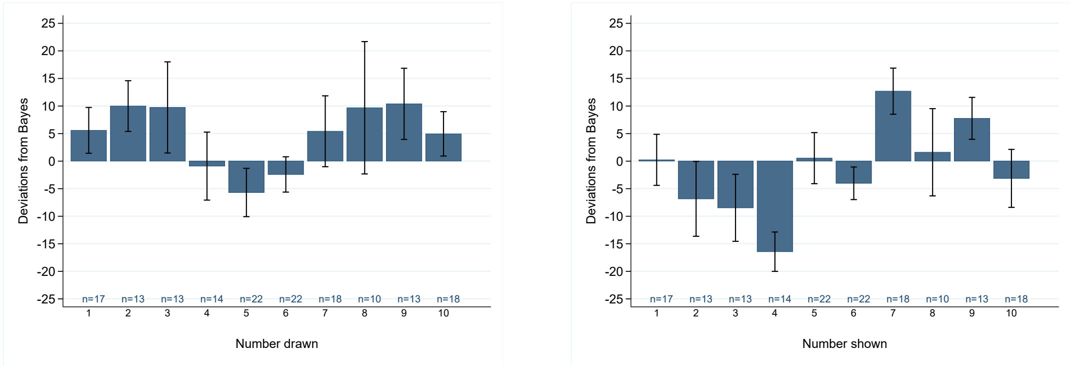
$$\hat{Y}_i^N = \sum_{j=1}^{J} w_j^i Y_j^C, \tag{1}$$

where $\hat{Y}_i^N$ denotes beliefs of subject i from the treatment condition if he had not received the signal (the counterfactual outcome), Y_j^C denotes beliefs of subject j in the control condition (it includes a correction for potential bias as in Abadie and Imbens, 2011), $j = \{1, \ldots, J\}$, and w_j^i is the weight assigned to j in the counterfactual outcome of subject i. The weights are normalized such that $0 \leq w_j \leq 1$ and $\sum_{j=1}^{J} w_j = 1$. I estimate the weights using a kernel regression for each participant in the treatment condition. I describe the estimation procedure in detail in Appendix B.2.

Intuitively, I construct the counterfactual to the participant i in the following way: I take the decisions of *all* participants in the control condition regarding the number that the participant i saw. However, not all observations in the control condition receive the same weight. Those participants whose true ranks and prior beliefs were closer to that of the participant i, receive a higher weight.

In the baseline specification, I match subjects using their true rank and prior beliefs. In Appendix D, I report the results based on two alternative specifications: in Specification 2, I match participants based on their true rank and prior beliefs about the number under consideration, and in Specification 3, I use only the distribution of prior beliefs (in theory, it subsumes information that a subject has about his performance). The results are very similar to the baseline specification.

Figure 9: Mean deviation from Bayes in the treatment condition and counterfactual.



(a) In the treatment condition.

(b) In the counterfactual.

On Panel a) and b) in Figure 9, I present the average deviations from the Bayesian benchmark in the treatment and the new control. The two graphs differ significantly for numbers "2", "3", and "4". Participants in the treatment tend to report significantly higher beliefs about signal informativeness after observing the signal "2", "3", or "4". However, the presented graphs do not say anything about the relation between signals and subjects' expectations. For this reason, I turn to the regression analysis.

4.4.2 Regressions Analysis

To further investigate the relation between the signal realization and its perceived informativeness, I conduct regression analysis using the constructed counterfactual and report its results in Table 3. The dependent variable is the difference between points allocated to Box 2 (indicative of one's rank) in the treatment and points that would have been allocated if the subject was in the control condition (the counterfactual).

Firstly, to test for the overall treatment effect, I regress the dependent variable on a constant. The coefficient is significant and equal to 4.95, a value similar to the one obtained in the regression based on all observations from the control condition.

In the second specification, I add an indicator variable "Good Signal", which takes the value 1 if a signal received by the subject was a "good" signal – it indicated a rank lower or equal to the subject's median belief. The coefficient is positive and highly

Table 3: The effect of the signal's valence.

	(1)	(2)	(3)	(4)
Good Signal		10.55***	7.72*	15.68***
		(3.87)	(3.94)	(5.05)
Outside Priors			-10.59***	-2.91
			(3.92)	(4.96)
Outside Priors × Good				-19.40**
				(7.89)
Constant	4.95**	0.40	6.45*	2.06
	(1.96)	(2.54)	(3.35)	(3.75)
Observations	160	160	160	160

Standard errors in parentheses

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

Note: The dependent variable is the difference between numbers of points allocated to Box 2 in the treatment and in the counterfactual (kernel-based matching). "Good Signal" indicator variable takes value 1 if the signal was below or equal to the median of subject's belief distribution, and 0 otherwise. "Outside priors" indicator variable takes value 1 if the subject attached a zero prior probability to the signal being his rank, and 0 otherwise.

significant, meaning that participants allocated on average 10.55 points more to the Box 2 after good news. The interpretation of the results is that, after a favorable signal, participants assign a higher probability to the signal being informative than they would have if they were in the control condition and the signal hasn't affected their utility.

The Good Signal coefficient changes if we control for the signal being outside of the support of prior belief distribution (Specification 3). Adding an interaction of the Outside Prior variable and the good signal indicator increases its value to 15.68, but a large and negative coefficient at the interaction term makes the overall effect negative.

The results suggest that the good-news effect is not universal across signals. Getting a signal "too good to be true" makes the agent skeptical and leads him to assign a lower probability than he would in the control condition. It does not contradict the theory, as for the subject to experience the belief-based utility it is necessary that the signal affects his beliefs, and it may not be the case if the signal is outside of the subject's priors.

The standard errors reported in Table 3 do not account for the matching procedure. Abadie and Imbens (2006) derive analytical formulas for a consistent estimator for the large-sample variance of the nearest-neighbor matching estimator. However, large-sample techniques may not be well suited when the number of units in the comparison group is small (Abadie et al., 2010). For a robust inference in the finite sample at hand, I employ the inferential techniques proposed in Abadie et al. (2010).

4.4.3 Placebo Studies

Although the sample size of 222 subjects would not be considered small for an experimental economist, it is a small sample given our set-up. If we divide participants based on their prior belief distribution, observed signal, and its relation to the subject's priors, we end up with much smaller groups.

For a robust inference in a finite sample, Abadie et al. (2010) propose an inferential technique based on "placebo studies". The idea behind it is to compare the actual treatment effect to the distribution of "placebo" treatment effect. The latter is calculated by assigning the treatment status to a random sample of participants in the control condition and estimating the same regression as in Specification 4 in Table 3. I provide more details about the procedure in Appendix B.3.

Figures 10 and 11 summarize the results of the placebo studies. In Figure 10, I present a histogram of coefficients at the Good Signal variable. Figure 11 shows the coefficients at the interaction of the Good Signal and the Outside Prior variable. The vertical lines denote the actual treatment effects. One can notice that their magnitudes are extreme relative to the distributions of coefficients in the placebo studies, indicating the statistical significance of the actual treatment effects. The empirical distribution of the placebo effects allows me to calculate the p-value of a two-sided test to assess the statistical significance of the actual treatment effect. Formally, I test a hypothesis that there is no difference between the actual treatment effect and the placebo treatment effect. The corresponding p-values are 0.003 in the case of the Good Signal variable and 0.039 for the interaction term.

Figure 10: Distribution of coefficients at the Good Signal variable (Specification 4).

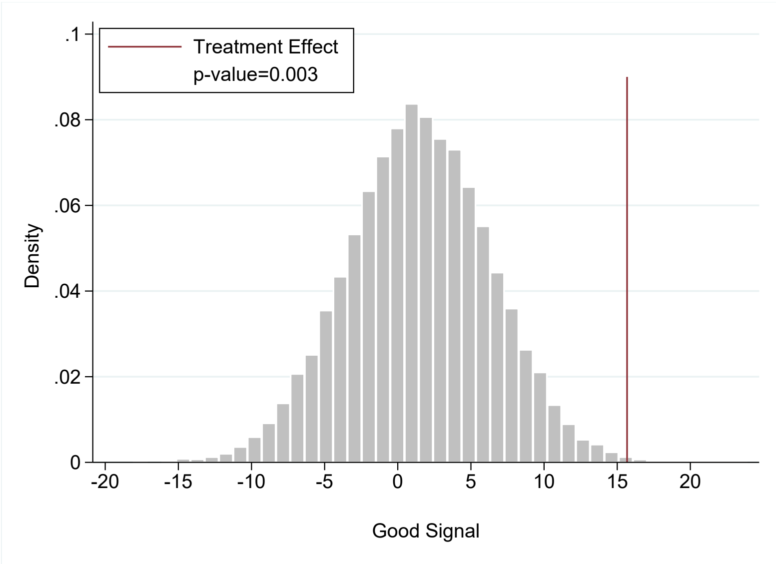
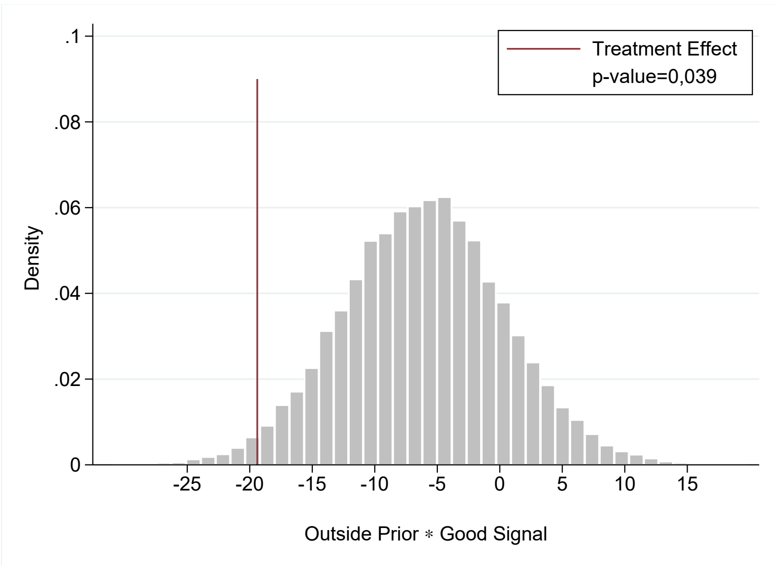


Figure 11: Distribution of coefficients at the interaction term (Specification 4).



4.5 Do subjects in the treatment condition earn more?

In the following section, I look at the payoffs from the main task in the treatment and the control conditions. In both conditions, subjects were remunerated with "lottery tickets": a higher probability of receiving a large reward of 12 Euro.

Firstly, I compare the average probabilities in the treatment and in the control condition, as well as in the treatment and in the counterfactual scenario. There are significant differences in the average probabilities in both cases.

Participants in the treatment condition obtained, on average, 65.5% probability of receiving a large reward, a probability that was 12.6% lower (p-value = 0.000) than the probability earned by the subjects in the control condition. If we compare the treated subjects to the control constructed in Section 4, participants in the treatment condition earned 8.7% less than they would have in the counterfactual (p-value = 0.006).

One can conclude that, on average, subjects in the treatment condition are worse-off than they would be if they were not given an opportunity to observe a signal and learn. The result gives rise to several questions: to what extend subjects are aware of their propensity to interpret realized signals differently? If they could commit to their decisions ex ante, would they do that? The result and its implications suggest interesting directions for future research.

Interestingly, there is no significant difference in the foregone earnings between over-confident and underconfident (p-value = 0.863). Moreover, there is no significant difference between subjects whose IQ test score was in the bottom of the test score distribution and those who scored in the top half of the distribution (p-value = 0.883).[14]

However, when we look closer at probabilities forgone by participants who differ with respect to their rank and the extend to which they are over- or under-confident, we see much more variation. The next step would be to exploit this variation to empirically investigate the belief-based component of the utility function.

[14]I compared subjects in the aforementioned groups with the respective counterfactuals, and then the average differences in earned probabilities were compared between the groups. The average losses in probability were very similar, ranging from 8.2% to 9.0%.

5 Additional Evidence

In this section, I examine a complementary data set of subjects' answers to questionnaires described in Section 2. Firstly, I look at the subjects' personality traits, anxiety levels, as well as habitual use of emotion regulation strategies, and report their correlations with subjects' decisions in the second task.

5.1 Emotion Regulation Questionnaire

In this section, I examine subjects' answers to the emotion regulation questionnaire, BIG-5 and STAI. In Table 4, I report correlations between subjects' decisions in the treatment condition (relative to the Bayesian benchmark) and the above-mentioned measures. The absolute deviations from Bayesian updating are correlated with the habitual use of reappraisal. The coefficient value of −0.18 indicates a weak, negative correlation significant at the 0.05 level.

Table 4: Deviations from rationality and agents' characteristics in the treatment condition.

	DevB	Extr	Cons	Open	Neur	Agre	Trait	State	Reapp	Supr
DevB	1.00									
Extr	0.00	1.00								
Cons	0.05	−0.01	1.00							
Open	−0.09	0.22*	0.10	1.00						
Neur	0.12	−0.24*	−0.26*	0.16*	1.00					
Agre	−0.03	0.07	0.07	0.07	−0.13	1.00				
Trait	−0.07	0.29*	0.35*	−0.09	−0.71*	0.23*	1.00			
State	−0.15	0.28*	0.17*	−0.03	−0.58*	0.24*	0.70*	1.00		
Reapp	−0.18*	0.09	0.15	0.18*	−0.17*	0.22*	0.13	0.17*	1.00	
Supr	−0.04	−0.19*	0.05	−0.17*	−0.04	0.03	−0.13	−0.14	0.38*	1.00

* $p < 0.05$

Note: "DevB" stands for deviations from Bayesian update. I use the labels: "Extr", "Cons", "Open", "Neur", and "Agre" for BIG-5 personality traits: extraversion, conscientiousness, openness to experience, agreeableness and neuroticism, respectively. I denote Anxiety trait and state with "Trait" and "State" (the two measures are defined such that a higher score indicates less anxious individual). "Reapp" and "Supr" stands for emotion regulation strategies: reappraisal and suppression.

In Table 5, I present the estimates of regressions based on decisions made by participants in the treatment condition. I regress the independent variable, the absolute deviations from Bayesian update, on the independent variable "Reappraisal" that measures subject's habitual use of reappraisal.

The coefficient at the Reappraisal variable is negative and significant at the 0.05 level. Reporting one point higher response on the 7-point Likert scale in questions about one's habitual use of reappraisal leads to a 3-point decrease in the distance from Bayesian update. The value doesn't change much if I control for subject's rank, median belief or whether the signal he received was below or above his median belief or not within the prior belief distribution.

Table 5: The effect of reappraisal on deviations from Bayes.

	(1)	(2)
Reappraisal	-2.96**	-2.82**
	(1.29)	(1.29)
Constant	26.61***	27.33***
	(5.76)	(7.50)
Controls	No	Yes
Observations	160	160

Standard errors in parentheses

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

Note: The dependent variable is absolute deviations from the Bayesian update. Controls include the subject's rank and median prior belief, a dummy variable equal 1 if the signal was below or equal to the median prior belief, and a dummy variable equal 1 if the signal was outside of the subject's prior beliefs.

The results show that subjects' decisions correlate with the way they handle positive and negative emotions in their daily life. The more used they are to regulate their emotions by thinking differently about the situation they found themselves in, the more they adhere to rational decision-making. To investigate this further, I take a closer look at emotion regulation strategies together with self-reported emotions experienced before the task.

5.2 Test-related Emotions

In addition to the data presented so far, I collected survey data about test-related emotions experienced by participants before receiving the signal[15]

Out of eight test-related emotions, anxiety and hopelessness significantly correlate with absolute deviations from Bayesian updating in the treatment condition. However, when I regressed absolute deviations from Bayesian updating on all test-related emotions, only helplessness was highly statistically significant (p-value = 0.02) and remained so, even after adding additional controls on subjects' rank, median belief, and signal's value or its relation to the subject's beliefs.

Hopelessness was measured by agreement with the statement "I felt that I would rather not do this part because I've lost all hope.". As reported in the first column in Table 6, stating a 1-point higher answer to the question translates to an increase of 4.3 points in absolute deviation from Bayesian updating (controlling for all remaining test-related emotions).

The coefficient at the Hopelessness variable remains unchanged if I control for the emotion regulation strategies: suppression and reappraisal (Specification 2) in Table 6. Of the two strategies, only reappraisal is different from zero and significant. Moreover, it has the expected negative sign and value similar to that reported in Table 5.

I hypothesize that the use of reappraisal counteracts the negative impact of Hopelessness. To test this hypothesis, I add to the regression the interaction of Hopelessness and Reappraisal. I report the estimation results in the last column of Table 6.

[15]In the instructions displayed on the screen, I highlighted that questions refer to the particular moment in time: *after* learning the nature of the task, but *before* seeing the number.

Table 6: The effect of self-reported emotions on deviations from rationality.

	(1)	(2)	(3)
Hopelessness	4.31^{**}	4.30^{**}	17.23^{***}
	(1.83)	(1.82)	(4.62)
Reappraisal		-2.82^{**}	2.21
		(1.42)	(2.16)
Hopelessness × Reappraisal			-3.10^{***}
			(1.02)
Constant	10.00	20.18^{**}	2.73
	(8.28)	(10.11)	(11.40)
Controls 1	Yes	Yes	Yes
Controls 2	No	Yes	Yes
Observations	160	160	160

Standard errors in parentheses

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

Note: The dependent variable is absolute deviations from the Bayesian update. The independent variable "Hopelessness" was measured by the extent to which a subject agreed with the statement "I felt that I would rather not do this part because I've lost all hope.". "Reappraisal" refers to self-reported habitual use of reappraisal. Controls 1 include all other emotions reported by subjects; Controls 2 include the measure of habitual use of suppression.

The coefficient at the interaction term is negative and highly significant, whereas the coefficient at the Reappraisal variable loses its significance. At the same time, the coefficient at Hopelessness increases fourfold and gains significance, suggesting that its impact is much larger without the offsetting effect of reappraisal.

While only suggestive, the evidence presented in this section supports the view that the treatment effect is stemming from the visceral, emotion-based reaction to signals. That reaction lies at the heart of what economists call "the belief-based utility" and is the driving force behind asymmetric updating.

6 Conclusions

In this paper, I examine how the belief-based utility influences the agent's interpretation of a signal he receives. To this end, I designed a simple experiment with two experimental conditions. In the treatment condition, participants observe a signal about their intelligence and decide whether the signal is informative or not. In the control condition, participants make the same choice but without receiving the factual signal: they have to make decisions ex ante, conditioning on possible signal realizations. I find a strong and significant effect, with a positive asymmetry: subjects tend to interpret favorable signals as more informative. I argue that the difference stems from subjects' reactions to changes in the belief-based utility.

One may wonder whether there is any evidence that the signal received in the treatment condition actually affected the subject's utility. To properly address this question, I would conduct a similar experiment but with learning about an ego-neutral parameter. A less satisfactory but immediate answer is that as economists, we rely on choices to reveal underlying preferences represented by a utility function. In the case of belief-based utility, if a decision change in response to a received signal, as compared to a signal considered but not actually received, the signal must have affected the agent's utility.

There is mounting evidence that people derive utility not only from physical outcomes but also from their beliefs about the current or future state. At the same time, there are important reasons why we should care about the belief-based utility. First and foremost, it is related to widely discussed behavioral phenomena such as belief polarization, demand for (and avoidance of) information, and persistent overconfidence. Although the belief-based utility is likely to be a driving force behind all above, the precise mechanism is not fully understood. My study takes the next step towards unraveling the underlying mechanism by investigating how belief-based utility shapes the way we interpret reality.

References

Abadie, Alberto, Alexis Diamond, and Jens Hainmueller (2010). "Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program". In: *Journal of the American statistical Association* 105.490, pp. 493–505.

Abadie, Alberto and Guido W. Imbens (2006). "Large sample properties of matching estimators for average treatment effects". In: *econometrica* 74.1, pp. 235–267.

— (2011). "Bias-corrected matching estimators for average treatment effects". In: *Journal of Business & Economic Statistics* 29.1, pp. 1–11.

Ambuehl, Sandro and Shengwu Li (2018). "Belief updating and the demand for information". In: *Games and Economic Behavior* 109, pp. 21–39.

Barber, Brad M and Terrance Odean (2001). "Boys will be boys: Gender, overconfidence, and common stock investment". In: *The quarterly journal of economics* 116.1, pp. 261–292.

Bénabou, Roland and Jean Tirole (2016). "Mindful Economics: The Production, Consumption, and Value of Beliefs". In: *Journal of Economic Perspectives* 30.3, pp. 141–164.

Benoît, Jean-Pierre and Juan Dubra (2011). "Apparent overconfidence". In: *Econometrica* 79.5, pp. 1591–1625.

Brandts, Jordi and Gary Charness (2009). *The strategy method: A survey of experimental evidence*. Tech. rep. mimeo, Department of Business Economics U. Autònoma de Barcelona and ...

Brunnermeier, Markus K and Jonathan A Parker (2005). "Optimal expectations". In: *American Economic Review* 95.4, pp. 1092–1118.

Buser, Thomas, Leonie Gerhards, and Joël Van Der Weele (2018). "Responsiveness to feedback as a personal trait". In: *Journal of Risk and Uncertainty* 56.2, pp. 165–192.

Camerer, Colin and Dan Lovallo (1999). "Overconfidence and Excess Entry: An Experimental Approach". In: *The American Economic Review* 89.1, pp. 306–318.

Caplin, Andrew and John V Leahy (2019). *Wishful Thinking*. Tech. rep. National Bureau of Economic Research.

Chew, Soo Hong, Wei Huang, and Xiaojian Zhao (2019). "Motivated false memory". In: *Available at SSRN 2127795*.

Coutts, Alexander (2019). "Good news and bad news are still news: Experimental evidence on belief updating". In: *Experimental Economics* 22.2, pp. 369–395.

Coutts, Alexander, Leonie Gerhards, and Zahra Murad (2020). "What to blame? Self-serving attribution bias with multi-dimensional uncertainty". In:

Eil, David and Justin M. Rao (2011). “The good news-bad news effect: Asymmetric processing of objective information about yourself”. In: *American Economic Journal: Microeconomics* 3.2, pp. 114–138.

Engelmann, Jan, Maël Lebreton, Peter Schwardmann, Joel J van der Weele, and Li-Ang Chang (2019). “Anticipatory anxiety and wishful thinking”. In:

Enke, Benjamin and Thomas Graeber (2019). *Cognitive uncertainty.* Tech. rep. National Bureau of Economic Research.

Ertac, Seda (2011). “Does self-relevance affect information processing? Experimental evidence on the response to performance and non-performance feedback”. In: *Journal of Economic Behavior & Organization* 80.3, pp. 532–545.

Gerlitz, Jean-Yves and Jürgen Schupp (2005). “Zur Erhebung der Big-Five-basierten persoenlichkeitsmerkmale im SOEP”. In: *DIW Research Notes* 4, p. 2005.

Gilboa, Itzhak and Massimo Marinacci (2016). “Ambiguity and the Bayesian paradigm”. In: *Readings in formal epistemology.* Springer, pp. 385–439.

Golman, Russell, David Hagmann, and George Loewenstein (2017). “Information avoidance”. In: *Journal of Economic Literature* 55.1, pp. 96–135.

Gross, James J and Oliver P John (2003). “Individual differences in two emotion regulation processes: implications for affect, relationships, and well-being.” In: *Journal of personality and social psychology* 85.2, p. 348.

Grossman, Zachary and David Owens (2012). “An unlucky feeling: Overconfidence and noisy feedback”. In: *Journal of Economic Behavior & Organization* 84.2, pp. 510–524.

Hahn, Jinyong (1998). “On the role of the propensity score in efficient semiparametric estimation of average treatment effects”. In: *Econometrica*, pp. 315–331.

Hastie, Trevor, Robert Tibshirani, and Jerome Friedman (2009). *The elements of statistical learning: data mining, inference, and prediction.* Springer Science & Business Media.

Heckman, James J, Hidehiko Ichimura, and Petra Todd (1998). “Matching as an econometric evaluation estimator”. In: *The review of economic studies* 65.2, pp. 261–294.

Hestermann, Nina and Yves Le Yaouanq (2020). “Experimentation with Self-Serving Attribution Biases”. In: *American Economic Journal: Microeconomics.*

Hossain, Tanjim and Ryo Okui (2013). “The binarized scoring rule”. In: *Review of Economic Studies* 80.3, pp. 984–1001.

Huffman, David, Collin Raymond, and Julia Shvets (2019). “Persistent overconfidence and biased memory: Evidence from managers”. In: *Pittsburgh: University of Pittsburgh.*

Kőszegi, Botond (2006). “Ego Utility, Overconfidence, and Task Choice”. In: *Journal of the European Economic Association* 4.June, pp. 673–707.

Lerner, Jennifer S, Ye Li, Piercarlo Valdesolo, and Karim S Kassam (2015). "Emotion and decision making". In: *Annual review of psychology* 66.

Malmendier, Ulrike and Geoffrey Tate (2005). "Does overconfidence affect corporate investment? CEO overconfidence measures revisited". In: *European Financial Management* 11.5, pp. 649–659.

— (2008). "Who makes acquisitions? CEO overconfidence and the market's reaction". In: *Journal of financial Economics* 89.1, pp. 20–43.

Mezulis, Amy H, Lyn Y Abramson, Janet S Hyde, and Benjamin L Hankin (2004). "Is there a universal positivity bias in attributions? A meta-analytic review of individual, developmental, and cultural differences in the self-serving attributional bias." In: *Psychological bulletin* 130.5, p. 711.

Möbius, Markus M, Muriel Niederle, Paul Niehaus, and Tanja S Rosenblat (2014). "Managing Self-Confidence". Working Paper.

Niederle, Muriel and Lise Vesterlund (2007). "Do women shy away from competition? Do men compete too much?" In: *The Quarterly Journal of Economics* 122.3, pp. 1067–1101.

Ortoleva, Pietro and Erik Snowberg (2015). "Overconfidence in political behavior". In: *American Economic Review* 105.2, pp. 504–35.

Pekrun, Reinhard, Thomas Goetz, Anne C Frenzel, Petra Barchfeld, and Raymond P Perry (2011). "Measuring emotions in students' learning and performance: The Achievement Emotions Questionnaire (AEQ)". In: *Contemporary educational psychology* 36.1, pp. 36–48.

Rosenbaum, Paul R and Donald B Rubin (1983). "The central role of the propensity score in observational studies for causal effects". In: *Biometrika* 70.1, pp. 41–55.

Schwardmann, Peter and Joel Van der Weele (2019). "Deception and self-deception". In: *Nature human behaviour* 3.10, pp. 1055–1061.

Spielberger, Charles D (1983). "State-trait anxiety inventory for adults". In:

Uusberg, Andero, Jamie L Taxer, Jennifer Yih, Helen Uusberg, and James J Gross (2019). "Reappraising reappraisal". In: *Emotion Review* 11.4, pp. 267–282.

Van den Steen, Eric (2004). "Rational overoptimism (and other biases)". In: *American Economic Review* 94.4, pp. 1141–1151.

Zimmermann, Florian (2020). "The dynamics of motivated beliefs". In: *American Economic Review* 110.2, pp. 337–61.

A Regression Analysis

In the following section, I describe the regression analysis of subjects' decisions in the second task. The dependent variable is the number of points allocated to Box 2, which reflects the probability that a subject assigns to the signal being his rank. Since the decision depends on the subject's prior beliefs about his rank, I regress it on Bayesian updated belief about the box that accounts for subjects' priors.

In Table 7, I present the estimation results of different specifications. Firstly, I regress the dependent variable on the Bayesian belief (the independent variable "Bayes") and the treatment dummy (I refer to it as the "Treatment" variable). As reported in the first column, both coefficients are positive and highly significant.

Table 7: The effect of receiving a signal on decisions.

	(1)	(2)	(3)	(4)	(5)	(6)
Bayes	0.69***	0.83***	0.67***	0.79***	0.67***	0.79***
	(0.04)	(0.09)	(0.04)	(0.09)	(0.04)	(0.09)
Treatment	4.55**	4.45**	4.84**	4.74**	1.62	1.58
	(1.95)	(1.93)	(1.91)	(1.89)	(2.60)	(2.59)
Good Signal			5.08**	4.82**	3.29	3.07
			(2.14)	(2.18)	(2.59)	(2.63)
Treatment × Good					7.36*	7.22*
					(4.27)	(4.27)
Outside Priors		9.69*		8.88		8.74
		(5.23)		(5.38)		(5.34)
Constant	9.49***	0.56	7.85***	-0.25	8.69***	0.70
	(1.29)	(5.30)	(1.46)	(5.34)	(1.64)	(5.40)
Observations	650	650	650	650	650	650

Standard errors in parentheses

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

Note: The dependent variable is the number of points allocated to Box 2. The independent variable "Bayes" refers to the Bayesian prediction based on the prior belief distribution. "Treatment" is a dummy variable indicating the Treatment condition. "Good Signal" dummy variable takes value 1 if the signal was below or equal to the median of the subject's belief distribution and 0 otherwise. "Outside priors" dummy variable indicates that the subject attached a prior probability of 0 to the signal being his rank.

The coefficient at the Bayes variable is smaller than 1 (p-value = 0.000), implying that, on average, subjects allocate fewer points to Box 2 (with numbers equal to their rank) than a Bayesian decision-maker would do. However, subjects in the treatment condition tend to allocate 4.5 points more to Box 2 than in the control condition.

The coefficient at the Bayes variable significantly increases when I add a dummy "Outside Priors" that takes the value of 1 if a signal was outside of the subject's priors (Specification 2). On average, subjects allocated 9.69 points to Box 2 after a signal to which they assigned a prior probability of zero (the coefficient is significant at 0.1 level).

In the third specification, I add a dummy variable "Good", which takes the value of 1 if a number displayed on the individual screen was lower or equal to the subject's median belief – a favorable signal about one's rank. The coefficient is positive and significant, implying that participants allocated on average more points to the box indicative of their rank after good news. Its value remains unchanged if I control for the signal being outside of the subject's prior distribution (Specification 4).

Finally, I test whether subjects treated good signals differently in the treatment and control conditions. In Specification 5), I add a variable "Treatment × Good" that is an interaction of the Treatment and Good Signal variables. As I see in column 5 in Table 7, the coefficient at the interaction term is high and significant at the 0.1 significance level, while the coefficients at the Treatment and Good Signal variables drop and lose their significance.

B Estimation Procedure

In this section, I provide more details on the estimation procedure. I begin by motivating the use of a matching estimator and contrasting it with a linear regression. Next, I describe the procedure that I use to determine the parameters of the matching estimator. Finally, I discuss the inference.

I adopt the standard model of treatment effects (Abadie and Imbens, 2006; Heckman et al., 1998; Rosenbaum and Rubin, 1983). Let $Y_i(W_i)$ denote the outcome of interest: the number of points that a participant i allocated to Box 2. W_i is a binary variable indicating whether the subject was assigned to the treatment ($W_i = 1$) or the control condition ($W_i = 0$). The average treatment effect is defined as

$$\tau = \mathbb{E}\left[Y_i(1) - Y_i(0)\right]. \tag{2}$$

Let subjects' decisions be described by an additive model

$$\begin{aligned} Y_i(1) &= \mu_1(\mathbf{X}_i) + \varepsilon_1 \\ Y_i(0) &= \mu_0(\mathbf{X}_i) + \varepsilon_0, \quad \varepsilon_1, \varepsilon_0 \sim i.i.d \end{aligned} \tag{3}$$

where μ_1 and μ_0 are unknown functions of a k–dimensional vector of individual characteristics $\mathbf{X}$.

Because of a random assignment to the two conditions, the sample average of outcomes recorded in the control condition is an unbiased estimator of the counterfactual outcome $Y_i(0)$. Therefore, a consistent estimation of the treatment effect entails a simple comparison of mean outcomes in both groups of participants

$$\hat{\tau} = \frac{1}{|N_T|}\sum_{i \in N_T} Y_i - \frac{1}{|N_C|}\sum_{i \in N_C} Y_i, \tag{4}$$

where N_T and N_C denote the set of participants in the treatment and the control condition respectively, and $|A|$ is the cardinality of a set A.

However, I am interested in heterogeneous treatment effects, defined as

$$\tau(\mathbf{x}) = \mathbb{E}\left[Y_i(1) - Y_i(0)|\mathbf{X}_i = \mathbf{x}\right]. \tag{5}$$

In case of a random assignment, one can simply compare means

$$\tau(\mathbf{x}) = \mathbb{E}\left[Y_i|W_i = 1, \mathbf{X}_i = \mathbf{x}\right] - \mathbb{E}\left[Y_i|W_i = 0, \mathbf{X}_i = \mathbf{x}\right]. \tag{6}$$

Note that the conditional expectation satisfy $\mathbb{E}\left[Y_i|W_i = w, \mathbf{X}_i = \mathbf{x}\right] = \mu_w(\mathbf{x})$.

B.1 Potential Problem: Selection

Estimation of $\mu_w(\mathbf{x}) = \mathbb{E}\left[Y_i|W_i = 0, \mathbf{X}_i = \mathbf{x}\right]$ turns out to be quite challenging. To illustrate why the OLS estimate may not be consistent, let's consider the task of estimating the treatment effect among participants who received a good signal. Consequently, X_i consists of a dummy variable equal to 1 for people who received a signal above or equal to their median belief, and 0 otherwise. The treatment effects are defined as

$$\tau(1) = \mathbb{E}\left[Y_i|W_i = 1, X_i = 1\right] - \mathbb{E}\left[Y_i|W_i = 0, X_i = 1\right]$$

and

$$\tau(0) = \mathbb{E}\left[Y_i|W_i = 1, X_i = 0\right] - \mathbb{E}\left[Y_i|W_i = 0, X_i = 0\right].$$

We are interested in the difference $\tau(1) - \tau(0)$.

Although the assignment to the treatment and control conditions is random, the experimental design may lead to the covariance between the treatment status and signals received by participants. This correlation may arise because participants in the treatment condition are presented one number, which is their rank true rank with probability $\frac{1}{2}$. Therefore, underconfident agents will see a good signal more often than overconfident agents. In contrast, in the control conditions subjects see all 10 signals.

Formally, let U_i denote a binary variable indicating whether an agent is underconfident ($U_i = 1$) or not ($U_i = 0$). The measured treatment effect of a good signal can be decomposed into

$$\underbrace{\hat{\mathbb{E}}\left[Y_i|W_i=1,X_i=1,U_i\right]-\hat{\mathbb{E}}\left[Y_i|W_i=0,X_i=1,U_i\right]}_{\text{observed difference between treatment and control}}=\underbrace{\mathbb{E}\left[Y_i(1)-Y_i(0)|W_i=1,X_i=1,U_i\right]}_{\text{treatment effect}}$$
$$+\underbrace{\mathbb{E}\left[Y_i(0)|W_i=1,X_i=1,U_i\right]-\mathbb{E}\left[Y_i(0)|W_i=0,X_i=1,U_i\right]}_{\text{selection}}$$

The selection arises if participants in the treatment and control conditions behave differently (on average) even absent any intervention. Let's decompose the selection term further. Let $Pr(X_i=1|W_i=1,U_i=1)$ denote the probability that a subject i observes a good signal, while receiving the treatment and being underconfident. We can expand the selection term as follows

$$\begin{aligned}\mathbb{E}\left[Y_i(0)|W_i=1,X_i=1,U_i\right]&=\mathbb{E}\left[Y_i(0)|W_i=1,X_i=1,U_i=1\right]Pr(X_i=1|W_i=1,U_i=1)\\&+\mathbb{E}\left[Y_i(0)|W_i=1,X_i=1,U_i=0\right](1-Pr(X_i=1|W_i=1,U_i=1))\end{aligned}$$

and

$$\begin{aligned}\mathbb{E}\left[Y_i(0)|W_i=0,X_i=1,U_i\right]&=\mathbb{E}\left[Y_i(0)|W_i=0,X_i=1,U_i=1\right]Pr(X_i=1|W_i=0,U_i=1)\\&+\mathbb{E}\left[Y_i(0)|W_i=0,X_i=1,U_i=0\right](1-Pr(X_i=1|W_i=0,U_i=1))\end{aligned}$$

Due to the random assignment to the treatment and control, it follows that

$$\gamma_1\equiv\mathbb{E}\left[Y_i(0)|W_i=1,X_i=1,U_i=1\right]=\mathbb{E}\left[Y_i(0)|W_i=0,X_i=1,U_i=1\right]$$

and

$$\gamma_0\equiv\mathbb{E}\left[Y_i(0)|W_i=1,X_i=1,U_i=0\right]=\mathbb{E}\left[Y_i(0)|W_i=0,X_i=1,U_i=0\right].$$

In other words, conditional on X, U, the assignment is as good as random. This simplifies the selection term

$$\begin{aligned}
&\mathbb{E}\left[Y_i(0)|W_i = 1, X_i = 1, U_i\right] - \mathbb{E}\left[Y_i(0)|W_i = 0, X_i = 1, U_i\right] \\
&\qquad = \gamma_1 * Pr(X_i = 1|W_i = 1, U_i = 1) + \gamma_0 \left(1 - Pr(X_i = 1|W_i = 1, U_i = 1)\right) \\
&\qquad\quad - \gamma_1 * Pr(X_i = 1|W_i = 0, U_i = 1) - \gamma_0 \left(1 - Pr(X_i = 1|W_i = 0, U_i = 1)\right)
\end{aligned}$$

This means that the selection term is zero if and only if participants in the treatment condition are as likely to receive good signal as participants in control condition,

$$Pr(U_i = 1|W_i = 1, X_i = 1) = Pr(U_i = 1|W_i = 0, X_i = 1).$$

However, participants the condition condition see all signals, while participants in the treatment see their own rank with probability $\frac{1}{2}$. Although we make sure that the participants are randomly allocated to the two groups, we cannot ensure that the signals relative to prior beliefs are randomly allocated to participants.

B.2 Solution: Matching Estimator

To deal with the potential selection issues (as well as to deal with potentially complex non-linearity of $\mu_w(\mathbf{x})$), I follow Heckman et al. (1998) and construct a matching estimator. For every participant in the treatment condition, I construct a counterfactual outcome based on decisions of participants in the control conditions with similar characteristics. In the example described in the previous section, we would match the underconfident agent who saw a good signal in the treatment condition with participants in the control condition, who were also underconfident and considered the same signal.

Following Heckman et al. (1998), I use a kernel regression to estimate the counterfactual outcomes $Y_i(0)$ for every participant in the treatment condition. The key identification assumption is that, conditional on all observables included in the matching procedure, the assignment of signals to participants is as good as random.

Formally, the treatment effect can be written as

$$\hat{\tau}(\mathbf{x}) = |N_T|^{-1} \sum_{i \in N_T} \left(Y_i - \sum_{j \in N_C} w_j^i \left(Y_j + \hat{\mu}_1(\mathbf{X}_i) - \hat{\mu}_1(\mathbf{X}_j) \right) \right), \tag{7}$$

where for each participant i in the treatment condition, w_j^i is a weight that I assign to a subject j from the control condition. The more similar participants i and j are (in terms of characteristics in $\mathbf{X}$), the higher the weight w_j^i. The weights are normalized such that $\sum_j w_j^i = 1$ and $w_j^i > 0$, $\forall_{i,j}$. The correction term $\hat{\mu}_1(\mathbf{X}_i) - \hat{\mu}_1(\mathbf{X}_j)$ removes the potential asymptotic bias of the matching estimator, as suggested by Abadie and Imbens (2011), where $\hat{\mu}_w(\mathbf{X})$ is a consistent regression estimator of $\mu_w(\mathbf{X})$. The intuition behind the bias correction is as follows. For a good match, the distance between $\mathbf{X}_i$ and $\mathbf{X}_j$ is small and the correction term $\hat{\mu}_1(\mathbf{X}_i) - \hat{\mu}_1(\mathbf{X}_j)$ vanishes. At the same time, the bias correction provides insurance in case of an imprecise match. In this case, the decision that a subject i would have made in the control condition becomes

$$\hat{Y}_i(1) = \hat{\mathbb{E}}\left[Y_i | W_i = 1, \mathbf{X}_i\right] + \sum_{j \in N_C} w_j^i \left(Y_j - \hat{\mathbb{E}}\left[Y_i | W_i = 1, \mathbf{X}_j\right] \right).$$

That is, the counterfactual outcome is equal to the regression prediction augmented with the matching term. The latter makes the whole estimator robust to a potential misspecification of the regression function stemming from the non-linearity of $\mu_w(x)$.

The weights w_j^i were constructed using the Epanechnikov kernel $K_{\mathbf{h}}(x)$ with a vector of parameters $\mathbf{h} > 0$

$$K_{\mathbf{h}}\left(\left|\left| \mathbf{X}_i - \mathbf{X}_j \right|\right| \right) = \begin{cases} \frac{3}{4}\left(1 - \left(\left|\left| \mathbf{h}\left(\mathbf{X}_i - \mathbf{X}_j\right) \right|\right| \right)^2 \right) & \text{if } \left|\left| \mathbf{h}\left(\mathbf{X}_i - \mathbf{X}_j\right) \right|\right| \leq 1 \\ 0 & \text{otherwise,} \end{cases} \tag{8}$$

where the length of $\mathbf{h}$ is equal to the number of characteristics included in $\mathbf{X}$.

Thereby, the weights are given by

$$w_j^i(\mathbf{h}) = \begin{cases} \frac{K_{\mathbf{h}}\left(\left|\left| \mathbf{X}_i - \mathbf{X}_j \right|\right| \right)}{\sum_{j \in \mathcal{J}_i} K_{\mathbf{h}}\left(\left|\left| \mathbf{X}_i - \mathbf{X}_j \right|\right| \right)} & \text{if } j \in \mathcal{J}_i \\ 0 & \text{otherwise ,} \end{cases} \tag{9}$$

where $\mathcal{J}_i \subset N_C$ is the set of individuals in the control condition who considered the same number as subject i. Epanechnikov kernel has similar interpretation to the nearest-neighbor matching often employed in the applied literature. It assigns positive weights only to a compact subset of neighbors whose characteristics are the closest to those of the target point. However, in contrast to nearest-neighbor matching which gives all the points in the neighborhood equal weight, Epanechnikov kernel assigns weights that decline smoothly with distance from the target point. This ensures that the resulting approximation of the conditional expectations $\mu_w(\mathbf{x})$ is smooth.

Parameter vector $\mathbf{h}$ controls the support of the kernel – how many closest neighbors to include and how quickly the weights decay with the distance. I follow a standard practice in the literature and estimate $\mathbf{h}$ using leave-one-out cross validation on the sample of participants in the control condition (see e.g. Hastie et al., 2009, Chapter 6). For a given $\mathbf{h}$, I estimate, using kernel regression, the probability each individual $k \in N_C$ assigned to Box 2

$$\hat{Y}_k(\mathbf{h}) = \sum_{j \in N_C \setminus \{k\}} w_j^k(\mathbf{h}) \left(Y_j + \hat{\mu}_1(\mathbf{X}_i) - \hat{\mu}_1(\mathbf{X}_j)\right).$$

I choose $\mathbf{h}$ to minimize the mean squared prediction error

$$|N_C|^{-1} \sum_{k \in N_C} \left(Y_k - \hat{Y}_k(\mathbf{h})\right)^2.$$

As for the choice of observables $\mathbf{X}$, I match agents in the treatment to those in the control who make decision regarding *the same* signal. Given this initial selection, I consider three specifications. In the baseline specification, I match participants based on their rank and prior belief distribution.

In the second specification, I match participants based on their rank and prior beliefs with respect to the number displayed on the computer screen. For example, if a participant observed the number "3" displayed on his screen, he would be matched based on his rank and how many points he allocated to rank 3 in the prior beliefs elicitation. In the third specification, I use only the prior belief distribution.

I estimate a consistent regression estimate $\hat{\mu}_w(x)$ using the estimator proposed in Hahn (1998)

$$\hat{\mu}_1(\mathbf{x}) = \frac{\hat{E}\left[YW \middle| X = \mathbf{x}\right]}{\hat{E}\left[W \middle| X = \mathbf{x}\right]} \tag{10}$$

To obtain consistent estimators for the conditional expectations on the right-hand side of (10), I use an OLS for the nominator and a logit regression for the denominator.

B.3 Inference

For a robust inference in a small sample, I employ the inferential techniques proposed in Abadie et al. (2010). The idea behind these techniques is to test whether the estimated treatment effect is large relative to the distribution of so-called "placebo effects". The placebo effect is estimated by assigning the treatment status to a random sample of all participants and conducting the same regression analysis.

To this end, I draw a random sample of 160 observations (i.e. equal to the number of subjects in the treatment condition) from all observations in the experiment and assign them the treatment status. I follow the matching procedure to create a counterfactual for each of those 160 observations. Next, I use the observations and their counterfactuals to estimate a placebo treatment effect. I run the same regression as those described in Table 3 and store the resulting coefficients.

I repeat this procedure 20 000 times to obtain a distribution of the placebo effects. Those coefficients are presented in the histograms in Figures 10 and 11. The empirical distribution of the placebo effects allows me to calculate the p-values of a two-sided test to assess the statistical significance of the actual treatment effect.

C Additional Results

Table 8: The effect of the signal's valence if we control for how far the signal was from the subject's median belief.

	(1)	(2)	(3)	(4)	(5)
Good Signal		10.55***	9.23**	18.34***	18.42***
		(3.87)	(4.16)	(6.30)	(6.14)
Distance			-0.88	0.60	3.78**
			(1.00)	(1.26)	(1.61)
Distance × Good				-3.92*	-3.88*
				(2.05)	(2.00)
Outside Priors					-17.37***
					(5.70)
Constant	4.95**	0.40	3.26	-1.55	-2.02
	(1.96)	(2.54)	(4.14)	(4.82)	(4.70)
Observations	160	160	160	160	160

Standard errors in parentheses

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

Note: The dependent variable is the difference between numbers of points allocated to Box 2 in the Treatment and the Control (Kernel-Based Matching). "Distance" indicates the absolute difference between the signal and the subject's median belief.

Table 9: The effect of the signal's valence if we exclude underconfident participants.

	(1)	(2)	(3)	(4)
Good Signal		8.21*	5.58	16.18***
		(4.31)	(4.30)	(5.17)
Outside Priors			-10.91***	-2.65
			(4.03)	(4.56)
Outside Priors × Good				-28.96***
				(8.54)
Constant	2.17	-0.47	5.41*	0.96
	(2.04)	(2.44)	(3.22)	(3.35)
Observations	115	115	115	115

Standard errors in parentheses

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

Note: The dependent variable is the difference between numbers of points allocated to Box 2 in the treatment and in the counterfactual (kernel-based matching).

C.1 Random Assignment to Signals

The results from Appendix A and Section 4.4 based on a subset of participants who were guessing a random number in the treatment condition.

Table 10: The effect of receiving an actual signal on decisions (no matching).

	(1)	(2)	(3)	(4)	(5)	(6)
Bayes	0.69***	0.80***	0.66***	0.77***	0.66***	0.76***
	(0.04)	(0.10)	(0.04)	(0.10)	(0.04)	(0.10)
Treatment	7.21***	7.00***	7.39***	7.20***	5.13	5.06
	(2.60)	(2.58)	(2.57)	(2.55)	(3.63)	(3.60)
Good Signal			4.20*	3.96*	3.41	3.22
			(2.32)	(2.37)	(2.58)	(2.62)
Treatment × Good					5.37	5.09
					(5.47)	(5.48)
Outside Priors		7.95		7.17		7.00
		(5.77)		(5.92)		(5.89)
Constant	9.77***	2.42	8.42***	1.87	8.79***	2.38
	(1.37)	(5.85)	(1.58)	(5.88)	(1.69)	(5.93)
Observations	577	577	577	577	577	577

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 11: The effect of the signal's valence (with matching, specification 1).

	(1)	(2)	(3)	(4)
Good Signal		7.14	3.64	15.38*
		(5.66)	(5.68)	(8.10)
Outside Priors			-13.99**	-4.26
			(5.64)	(7.38)
Outside Priors × Good				-22.34**
				(11.17)
Constant	7.27**	4.32	13.65***	7.16
	(2.80)	(3.64)	(5.16)	(6.02)
Observations	87	87	87	87

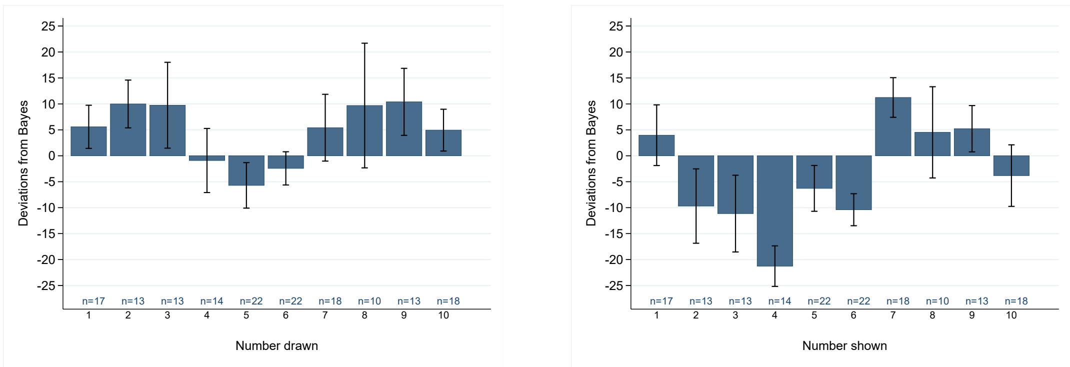
Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

D Matching

In this section, I present the results of my analysis described in Section 4.4 with the counterfactual based on different matching criteria. In Specification 2, I match participants based on their true rank and prior beliefs about the number under consideration (as opposed to the entire belief distribution in Specification 1). In Specification 3, I use only the prior belief distribution.

D.1 Matching Specification 2

Figure 12: Mean deviation from Bayes in the treatment condition and counterfactual.



(a) In the treatment condition. (b) In the counterfactual.

Table 12: The effect of the signal's valence.

	(1)	(2)	(3)	(4)
Good Signal		11.44***	7.38*	14.69***
		(4.13)	(4.12)	(5.31)
Outside Priors			-15.18***	-8.13
			(4.10)	(5.22)
Outside Priors × Good				-17.82**
				(8.29)
Constant	7.49***	2.56	11.24***	7.21*
	(2.09)	(2.71)	(3.51)	(3.94)
Observations	160	160	160	160

Standard errors in parentheses

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

Note: The dependent variable is the difference between numbers of points allocated to Box 2 in the treatment and in the counterfactual.

Figure 13: Distribution of coefficients at the Good Signal variable (Specification 4).

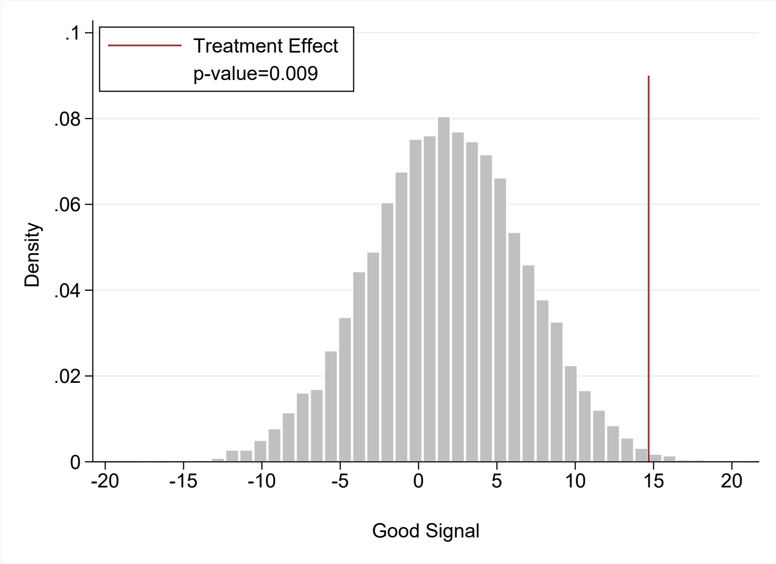
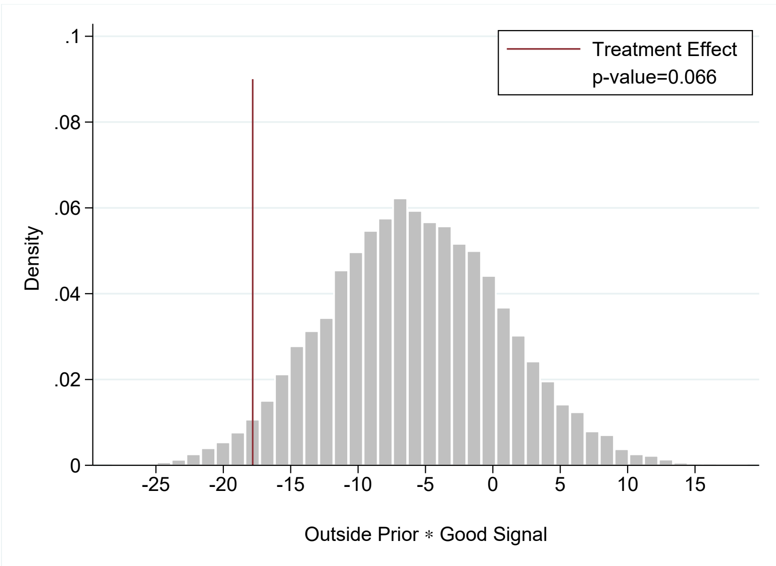
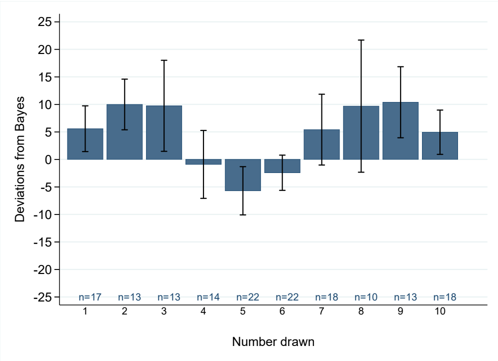


Figure 14: Distribution of coefficients at the interaction term (Specification 4).

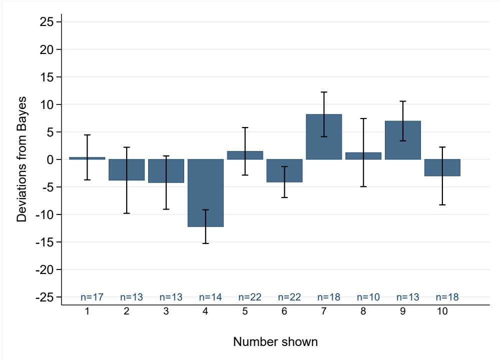


D.2 Matching Specification 3

Figure 15: Mean deviation from Bayes in the treatment condition and counterfactual.



(a) In the treatment condition.



(b) In the counterfactual.

Table 13: The effect of the signal's valence.

	(1)	(2)	(3)	(4)
Good Signal		10.57***	8.88**	15.88***
		(3.77)	(3.89)	(5.01)
Outside Priors			-6.36	0.41
			(3.87)	(4.92)
Outside Priors × Good				-17.09**
				(7.82)
Constant	4.43**	-0.13	3.51	-0.36
	(1.91)	(2.48)	(3.31)	(3.72)
Observations	160	160	160	160

Standard errors in parentheses

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

Note: The dependent variable is the difference between numbers of points allocated to Box 2 in the treatment and in the counterfactual (kernel-based matching). "Good Signal" indicator variable takes value 1 if the signal was below or equal to the median of subject's belief distribution, and 0 otherwise. "Outside priors" indicator variable takes value 1 if the subject attached a zero prior probability to the signal being his rank, and 0 otherwise.

Figure 16: Distribution of coefficients at the Good Signal variable (Specification 4).

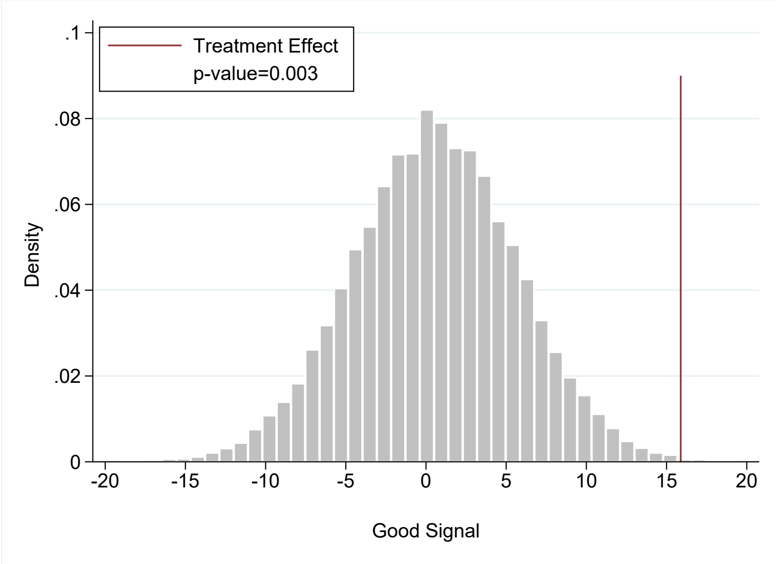
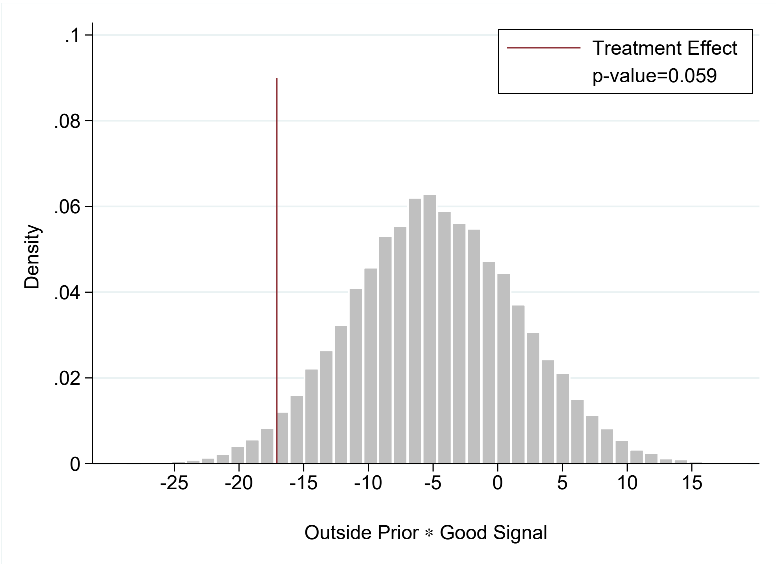


Figure 17: Distribution of coefficients at the interaction term (Specification 4).



E Overconfidence

Table 14: Belief distribution and Confidence Type.

	All	Overconfident	Unbiased	Underconfident
Rank	5.65	7.13	4.38	2.93
	(2.69)	(1.97)	(1.44)	(1.98)
Mean	4.47	4.17	4.43	5.13
	(1.75)	(1.53)	(1.35)	(2.16)
1^{st} **Quartile**	3.71	3.40	3.71	4.41
	(1.74)	(1.47)	(1.26)	(2.21)
Median	4.45	4.14	4.38	5.16
	(1.79)	(1.54)	(1.44)	(2.20)
3^{rd} **Quartile**	5.16	4.84	5.21	5.84
	(1.87)	(1.70)	(1.44)	(2.19)
Range	4.89	4.82	5.04	5.00
	(1.57)	(1.64)	(1.44)	(1.60)
Bias	2.44	2.99	0	2.23
	(1.84)	(1.74)	(.)	(1.58)
N	209	127	24	58

E.1 The Benoit-Dubra Critique

(soon to be updated)

F RTC Registration details

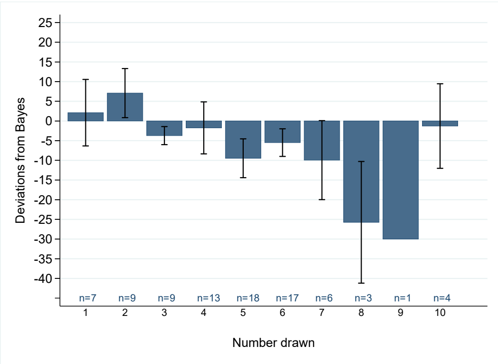
(soon to be updated)

G Instructions

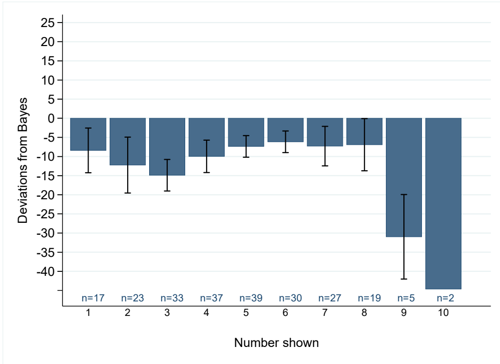
(soon to be updated)

H Deviations from Bayes

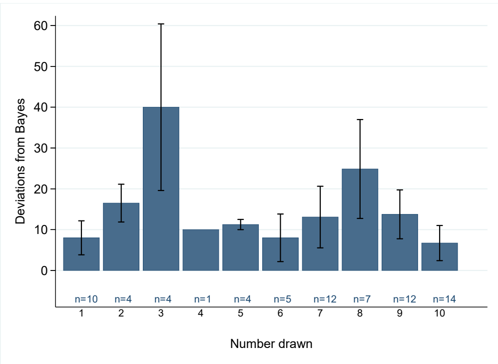
Figure 18: Mean deviation from Bayes for different type of signals (Part I).



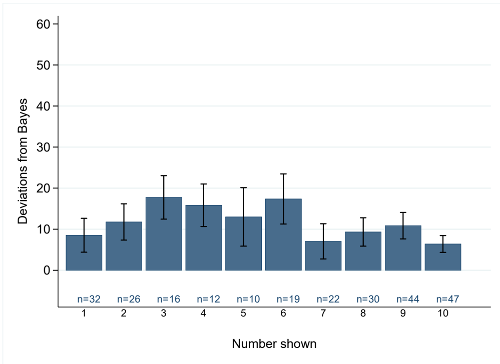
(a) treatment condition, signal within priors.



(b) control condition, signal within priors.

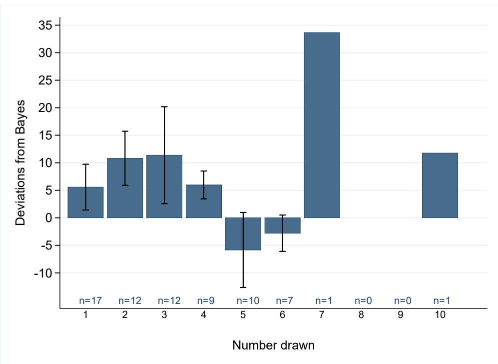


(c) treatment condition, signals outside priors.

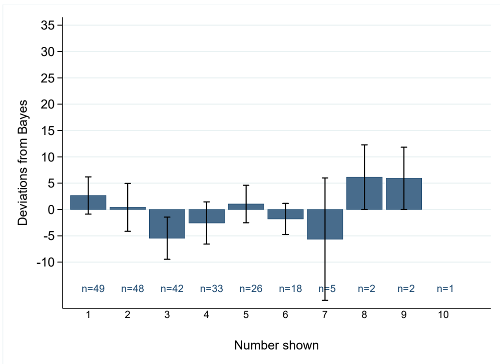


(d) control condition, signals outside priors.

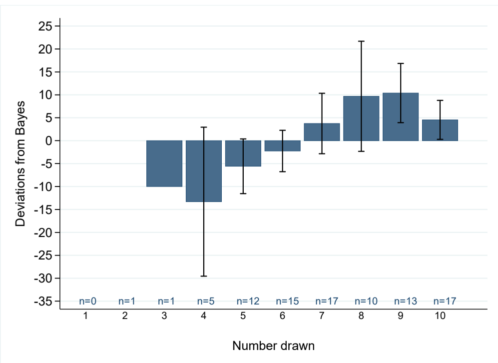
Figure 19: Mean deviation from Bayes for different type of signals (Part II).



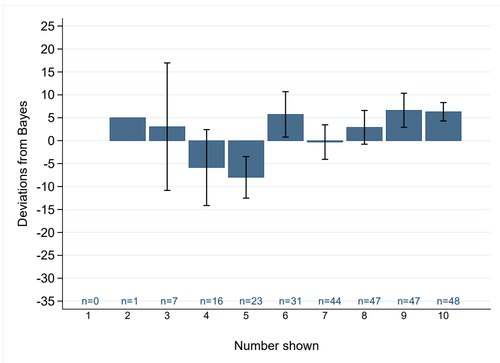
(a) treatment condition, after a "good" signal.



(b) control condition, after "good" signals.



(c) treatment condition, after a "bad" signal.



(d) control condition, after "bad" signals.