



**HAL**  
open science

## The Motivated Memory of Noise

Jeanne Hagenbach, Nicolas Jacquemet, Philipp Sternal

► **To cite this version:**

Jeanne Hagenbach, Nicolas Jacquemet, Philipp Sternal. The Motivated Memory of Noise. 2024.  
halshs-04744490

**HAL Id: halshs-04744490**

**<https://shs.hal.science/halshs-04744490v1>**

Preprint submitted on 18 Oct 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

CES

Centre d'Économie de la Sorbonne  
UMR 8174

**The motivated memory of noise**

Jeanne HAGENBACH, Nicolas JACQUEMET, Philipp STERNAL

2024.10



# The motivated memory of noise\*

Jeanne Hagenbach<sup>†</sup>

Nicolas Jacquemet<sup>‡</sup>

Philipp Sternal<sup>§</sup>

October 2024

## Abstract

We propose a two-stage experiment in which people receive feedback about their relative intelligence. This feedback is a noisy message reminded at every stage, so that subjects cannot forget this ego-relevant information. Instead, we exogenously vary whether the informativeness of the message is reminded in the second stage. We investigate how this treatment variation affects the informativeness reported by subjects, and their posterior beliefs about their intelligence. We show that subjects report informativeness in a self-serving way: subjects with negative messages report that these messages are significantly less informative in the absence of reminder than with it. We also show that the lack of reminder about message informativeness allows subjects to keep a better image of themselves. These results are confirmed by complementary treatments in which we decrease messages informativeness: subjects tend to inflate the informativeness of positive messages that should now be interpreted as bad news.

**JEL codes:** C91, D83, D63.

**Keywords:** Controlled experiment; Motivated beliefs; Overconfidence; Noisy feedback.

---

\*We thank conference and seminar participants at ASFEE 2023 (Lyon), Christian-Albrechts-Universität zu Kiel, Sc Po, U. Ottawa, U. Laval and PSE, as well as Jana Friedrichsen for their valuable comments and feedback. Financial support from the European Research Council (grant agreement n°850996 – MOREV) and from the National Research Agency (program Investissements d’Avenir, ANR-10-LABX-93-0 and ANR-17-EURE-0001) are gratefully acknowledged. The experiment received approval from the PSE-IRB #2021-21. Unless explicitly mentioned in the text, the analysis presented in this paper has been pre-registered as AsPredicted #106055 report ([https://aspredicted.org/F52\\_52G](https://aspredicted.org/F52_52G)).

<sup>†</sup>CNRS and Department of Economics, Sciences Po, Paris. [jeanne.hagenbach@sciencespo.fr](mailto:jeanne.hagenbach@sciencespo.fr).

<sup>‡</sup>Paris School of Economics and Université Paris 1 Panthéon-Sorbonne. Centre d’Economie de la Sorbonne (CES), Maison des Sciences Economiques, 106-112 boulevard de l’Hôpital 75013 Paris. [Nicolas.Jacquemet@univ-paris1.fr](mailto:Nicolas.Jacquemet@univ-paris1.fr).

<sup>§</sup>University of Zürich. Department of Economics, Blümlisalpstr. 10, 8006 Zürich. [philipp.sternal@econ.uzh.ch](mailto:philipp.sternal@econ.uzh.ch).

# 1 Introduction

Selective memory is now a well-known channel by which individuals manage to reach and maintain pleasant beliefs about themselves (see Amelio and Zimmermann, 2023, for a survey).<sup>1</sup> Individuals selectively recall their past actions to think of themselves as being generous or able to take good financial decisions (Saucet and Villeval, 2019; Gödker, Jiao, and Smeets, 2024). They also recall positive feedback about their relative intelligence with a higher accuracy than negative feedback (Zimmermann, 2020) and tend to forget past poor performance (Chew, Huang, and Zhao, 2020). In many situations however, information received about one’s ability cannot easily be forgotten; think for instance about failure at an exam or at getting a job. Our experiment starts from this observation and makes it impossible for subjects to forget ego-relevant pieces of information. In this setting, we investigate whether individuals manipulate how informative they perceive these pieces to be, thereby finding another path to a positive view of themselves.

Our experiment is made up of two sessions in which subjects get information and form beliefs about their relative intelligence. Our design relies on the common hypothesis that subjects intrinsically prefer to believe they are relatively clever than not. In the first session, subjects perform an IQ test and report their prior belief about their performance relative to a benchmark group of participants. Subjects then receive a noisy, binary message about that relative performance. The message states whether their individual performance places them in the upper or in the lower half of the distribution of performance of the benchmark group. The process by which this message is generated is explained to the subjects and works as follows: the computer randomly draws the message from an urn which contains seven true messages about the subject’s relative intelligence and three false messages. We refer to the number of true messages in the urn as the message *informativeness*.

The second session takes place two days later. Subjects are reminded of the message received in the first session but we vary (between subjects) whether they are, in addition, reminded about the message informativeness — REMINDER treatment — or not — NOREMINDER treatment. Thus, subjects cannot forget the feedback about their relative intelligence but, in the latter treatment, can forget and potentially manipulate how informative they perceive this feedback to be. The main experimental outcomes, elicited in the second session, are subjects’ reports about message informativeness and subjects’ updating of beliefs about their relative performance. Subjects receive monetary rewards for accurate responses.

Our main results are twofold. First, comparisons between treatments show that the reported level of message informativeness is significantly, and asymmetrically, affected by the content of

---

<sup>1</sup>Other channels identified in the economic literature include motivated avoidance/search of information (e.g. Dana, Weber, and Kuang, 2007; Golman, Hagmann, and Loewenstein, 2017; Castagnetti and Schmacker, 2022) and self-serving processing of information (e.g. Eil and Rao, 2011; Sharot, 2011; Möbius, Niederle, Niehaus, and Rosenblat, 2022; Hagenbach and Saucet, 2024; Drobner and Goerg, 2024).

the message received. For subjects who received a positive message, the average report of informativeness with and without reminder is unchanged (reports only become noisier in the latter treatment). A positive message is good news for the receiver, so there is little need for self-serving manipulation of informativeness. In contrast, for subjects who received a negative message, the average report in NOREMINDER is shifted downwards compared to the reports in REMINDER: when news are bad but cannot be forgotten, subjects self-servingly deflate the informativeness of these news. Second, while subjects significantly update their beliefs in the direction of the message in REMINDER, the lack of a reminder allows subjects who received a negative message to resist this inconvenient piece of information and update their beliefs upwards.

We complement our findings with additional treatments in which the informativeness of the message is drastically decreased. In the UNRELIABLE treatments, the number of true messages is three out of ten, implying that a negative message should now be interpreted as good news and vice versa. Again, we compare reported informativeness in REMINDER and NOREMINDER. Consistent with the findings in the main experiment, we observe that reported informativeness is distorted only among subjects who received bad news (i.e., a positive message), who now tend to inflate the informativeness of the message (so that the positive message becomes more credible than it actually is). The data on belief updating shows that subjects fail to respond to the information contained in their message when facing this more complex message-generating process. By pooling data from all four treatments, we show this is driven by the tendency of subjects to take into account both the information delivered by the message and the literal content of that message, even with a reminder. With no reminder, subjects tend to form posterior beliefs based mainly on the literal content of the message.

## 2 Design of the experiment

Our experiment is made up of two sessions which are two days apart. In session 1, subjects perform an IQ test. We elicit subjects' prior beliefs about their relative performance in that test before giving them a noisy signal about their true ranking. In session 2, we elicit subjects' recall about the information provided in session 1 and their posterior beliefs. We describe below the content of each session before presenting our treatment variable. Subjects' payment is the sum of a €6 show-up fee (€4 in the first session, €2 in the second session) and the money they make in one (randomly chosen) of the five incentivized tasks they perform in the experiment (two of these tasks being in session 1 and three in session 2).<sup>2</sup> In the invitation, subjects are informed that the experiment involves two sessions but do not learn anything about the content of the second session in the first one.

---

<sup>2</sup>An English translation of the original instructions in French is given in the Supplementary Material, Section D. All further details about payments are given in the Appendix, Section A.

**Overall structure of the experiment.** At the start of the first session, subjects are given 15 minutes to solve the 12-item short form of Raven’s Progressive Matrices Test validated by Arthur and Day (1994). Subjects get €0.5 for every matrix correctly solved. Subjects’ relative performance is then defined in comparison to a benchmark group made up of 127 subjects who performed the same test in a previous experiment.<sup>3</sup> After introducing subjects to the benchmark group, we elicit their prior belief about whether their own performance belongs to the upper half of the performance distribution in the benchmark group. The elicitation of prior beliefs relies on a Binarized Scoring Rule and subjects can gain up to €2 in this task. Next, subjects are told they will receive a message about their relative performance, which is drawn from an urn containing  $p = 7$  true messages and  $10 - p = 3$  false messages. Eventually, subjects see the message which reads either “In the IQ test, you rank in the upper half” – a *positive* message – or “In the IQ test, you rank in the lower half” – a *negative* message. As a manipulation check, each piece of information (number of true messages in the urn and message effectively received) is followed by a (non-incentivized) recall question that asks subjects to report the information they just received.

The same subjects participate in the second session two days later.<sup>4</sup> They are first reminded about the message they received during the first session – along with a reminder, for some subjects, of how this message was generated. Using the same elicitation method as in the first session, subjects are then asked to report their posterior belief about whether their performance in the IQ test belongs to the upper half. At the last stage of the session, as in the first session, we ask subjects (in random order) to report their recall of the message they received and the number of true messages which were in the urn from which their message was drawn. Unlike in session 1, these recall questions are incentivized in session 2: subjects get €1 for accurately reporting the message they received; the report of the number of true messages in the urn is incentivized using a Binarized Scoring Rule and subjects can earn up to €2 in this task. The second session ends with a socio-demographic questionnaire.

**Treatment variable.** The reminders provided at the beginning of the second session aim at controlling precisely what subjects remember in this session. In particular, our experiment ensures that, in session 2, subjects always remember the message they received in session 1. In REMINDER, we remind subjects both about the message received in the first session and about the informativeness of this message, that is, how many true messages were in the urn from which this messages was drawn. In NOREMINDER, we only remind subjects about the message received

---

<sup>3</sup>The data for the benchmark experiment comes from a pilot of a previous version of the experiment, from which we extract only the information about Raven’s PMT. The monetary incentives attached to the PMT are the same as in the main experiment.

<sup>4</sup>We shortened the delay between the two sessions as compared to Zimmermann (2020) to ensure it was possible for at least some subjects to actually remember the number of true messages in the urn from one session to the next, which is confirmed empirically as shown in Figure 2 below.

in the first session but not about its informativeness. The two experimental treatments are implemented between-subjects. Overall, 473 subjects participated in both sessions online, with 248 being assigned to the REMINDER treatment.<sup>5</sup>

### 3 Do subjects self-servingly recall the informativeness of messages?

The experiment aims at investigating two questions: do subjects selectively remember how informative an ego-relevant, noisy piece of information is? If yes, do they act on the selectively recalled information? Our design accordingly delivers two kinds of outcomes in the second session: (i) the answers to the recall question about the number of true messages in the urn; we refer to these answers as the *reports about informativeness of messages*; (ii) the posterior beliefs about the likelihood to belong to the upper half of the performance distribution. We investigate the two questions in turn, relying on comparisons between subjects who have been reminded about the informativeness of messages and those who have not.

#### 3.1 Manipulation check: Recall of reminded information

In both experimental sessions, we elicit subjects' recall of two pieces of information: the message about their relative performance and the informativeness of the messages. In session 1, they receive both pieces of information just before answering the recall questions about these pieces. In session 2, they are reminded about both pieces of information in REMINDER, and solely about the message received in NOREMINDER.

The (non-incentivized) recall questions in session 1 unambiguously show that subjects are able to understand the recall questions and the structure of the experiment, and that they are willing to carefully answer the survey: 98.3% of subjects accurately remember the message they received; and 74.2% accurately report the number of true messages in the urn. The same applies to the (incentivized) recall question in session 2 about the message received, which is accurately reported by 94.5% of subjects overall.

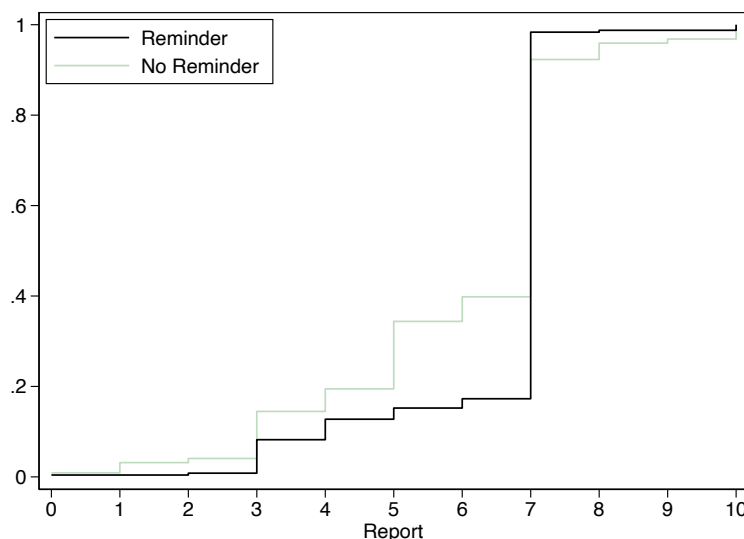
The cumulative distribution (CDF) of answers to the recall question about the number of true messages in session 2 is reported in Figure 1. In REMINDER, the vast majority of subjects (79.4%) report exactly 7, the true number.<sup>6</sup> This distribution stands in sharp contrast with the

---

<sup>5</sup>The subject pool was made up of French speakers registered at Full Factory (<https://www.foulefactory.com/en/>). More details on the experimental procedures, along with descriptive statistics on the sample, are provided in the Appendix, Section A.

<sup>6</sup>In total, about three in four participants (74.4%) return for session 2. We do not find any relationship between the message received and attrition: among subjects who received a negative message in session 1 (317 in total), 77 (i.e., 24%) drop-out, while 86 out of 319 (27%) do so among subjects who received a positive message in session 1. The two attrition rates are not statistically different according to a  $\chi^2$  test,  $p = .441$ . Importantly, the randomization between REMINDER and NOREMINDER is performed at the start of session 2, among subjects who return, so that

Figure 1: Informativeness reports, by treatment



**Note.** Cumulative Distribution Functions (CDF) of the reported number of true messages depending on whether subjects have been exposed to a REMINDER or not (NOREMINDER).

answers to recall questions in NOREMINDER. When subjects were only exposed to the message informativeness in session 1, the share of subjects who accurately report the number of true messages in session 2 goes down to 51.6% (the difference in accuracy with REMINDER is highly significant;  $p = .000$ , conditional t-tests).<sup>7</sup> As a result of these two shifts, the distribution of reports is significantly different between the two treatments ( $p = .000$ , Epps-Singleton (ES) test).<sup>8</sup> These comparisons confirm the efficacy of our treatment manipulation: subjects do report accurately the information they are reminded about, and are more likely to forget about the informativeness of messages when they are not.

### 3.2 Self-serving recall of the informativeness of messages

We now ask whether the recall of the informativeness of messages is self-serving in the sense that it helps subjects keep a good view of their relative intelligence. To achieve this goal, when subjects

---

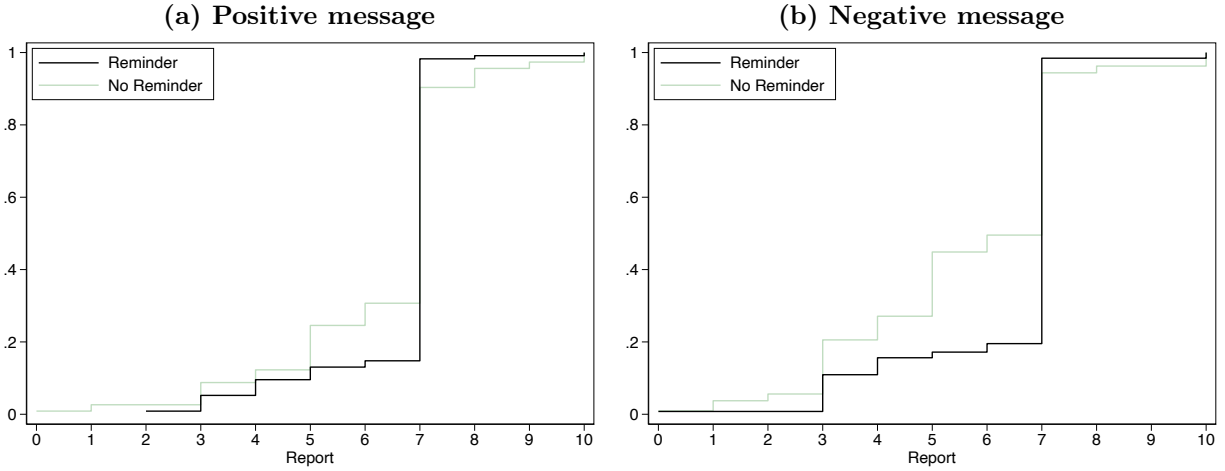
treatment assignment is orthogonal to individual characteristics conditional on returning.

<sup>7</sup>Conditional statistical tests come from linear regressions of the outcome on treatment variables that control for the true rank and the covariates that are unbalanced across treatments (age and gender, see the Appendix, Section A for detailed evidence). The corresponding estimation results are provided in the Appendix, Section E.

<sup>8</sup>We use the ES test rather than the more usual Kolmogorov-Smirnov (KS) test whenever the outcome variable is discrete, as the KS test tends to be particularly under-powered in this case (see, e.g., Goerg and Kaiser, 2009). When outcomes are continuous, we switch back to the KS test. We did not find any discrepancy in statistical significance between the two procedures in these cases; the KS test replication of the corresponding ES tests are available from the authors upon request.



Figure 2: Informativeness reports, conditional on the message received



**Note.** Cumulative Distribution Functions (CDF) of the self-reported number of true messages in REMINDER vs NOREMINDER, depending on whether subjects received a positive (left panel) or a negative (right panel) message.

have room to manipulate informativeness, they should inflate it if the message they received is positive and deflate it if this message is negative.

Figure 2 presents the distributions of informativeness reports conditional on the message received (hence holding statistically constant the distribution of IQ scores). The distribution of reports in REMINDER is significantly different from the one observed in NOREMINDER both for subjects who received a negative ( $p = .000$ , ES test) and a positive ( $p = .089$ , ES test) message. The gap between the two treatments is however much larger after a negative message than after a positive message. This difference in distributions is driven by two different forces: a drop in the share of subjects who accurately report the true number of messages (a kind of extensive margin), and a shift in the report from subjects who do not (the intensive margin).

Although we observe a larger drop of the share of subjects who accurately report the number of true messages after receiving a negative message (a 42.9% decrease, from 77.1% in REMINDER to 44.0% in NOREMINDER,  $p = .000$ , conditional t-test) as compared to those who received a positive message (a 28.6% decrease, from 82.1% in REMINDER to 58.6% in NOREMINDER,  $p = .000$ , conditional t-test), this difference is not statistically significant ( $p = .0344$ , conditional t-test). This suggests that subjects do not selectively forget depending on the message they received.

On the intensive margin, the average report of informativeness remains largely unchanged (from 6.60 in REMINDER to 6.34 in NOREMINDER  $p = .316$ , conditional t-test) when the message is positive despite the significant drop in the share of subjects who correctly recall the number of true messages in NOREMINDER. As shown in Figure 2.a, for subjects with positive messages, the errors in reports in NOREMINDER lead to a flatter distribution — i.e., errors increase to a similar extent on both sides of the distribution. The absence of reminder about informativeness increases the noise in the distribution of reports but subjects do not seem to manipulate their estimates of informativeness in any given direction. One interpretation is that, when the message is positive, it is already good news for the subject’s relative intelligence, so there is not much need to manipulate the reported informativeness of this message (in particular no need to make the news even better by reporting even higher informativeness than 7). The picture looks different for subjects who received a negative message which, if interpreted as informative, is bad news for subjects. Figure 2.b shows that NOREMINDER leads to a leftward shift in the distribution of reports of informativeness. This results in a 12.2% drop in average report after a negative message (6.39 in REMINDER vs 5.61 in NOREMINDER,  $p = .005$ , conditional t-test). When subjects receive a negative message and cannot forget it, they tend to deflate the informativeness of this message.

Given the structure of the message-generating process, extreme manipulations of informativeness could even ‘invert’ the literal meaning of messages: a negative (positive) message would no longer deliver bad (good) news if it were drawn from an urn in which the number of true messages is smaller than the number of false messages (i.e.,  $p \leq 5$ ).<sup>9</sup> In REMINDER, the share of subjects who distort the informativeness of the message to such an extent, that is, who report numbers 5 or less, is similar for subjects with a positive (13.0%) or a negative message (17.2%,  $p = .226$ , conditional t-test). Among subjects who receive a positive message, we only observe a slight (but statistically significant) increase in such reports (to 24.6%,  $p = .050$ , conditional t-test) in NOREMINDER as compared to REMINDER. After a negative message, by contrast, we observe a large and highly significant jump in these low reports, with 44.9% of subjects who report a number of true messages equal to 5 or less ( $p = .000$ , conditional t-test). This increase in NOREMINDER is significantly larger for subjects who received a negative message ( $p = .047$ , conditional t-test).

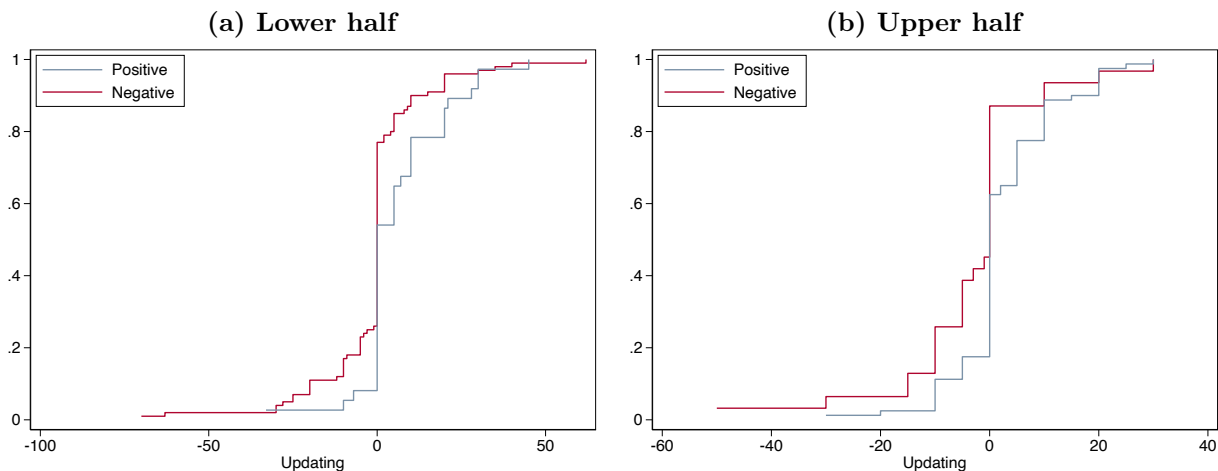
Overall, these results point to self-serving reports of informativeness with an asymmetry between positive and negative messages. A priori, increasing informativeness of positive messages and decreasing informativeness of negative messages both could help subjects to feel better about their relative intelligence. We show that subjects manipulate informativeness of messages to make bad news – negative messages – less bad but not to make good news – positive messages – better.<sup>10</sup> The next question is whether the manipulation of informativeness in front of negative messages

---

<sup>9</sup>This part of the analysis is exploratory.

<sup>10</sup>This result is in line with several results from experiments on the formation of motivated beliefs which also report asymmetries in the way individuals process good and bad news (see, e.g., Zimmermann, 2020; Eil and Rao, 2011; Hagenbach and Saucet, 2024).

Figure 3: Updating as a function of the message, by rank



**Note.** Cumulative Distribution Functions (CDF) of the individual updating (i.e. difference between the posterior, in session 2, and the prior, in session 1, beliefs about the probability of belonging to the upper half of the benchmark distribution) in REMINDER, depending on whether subjects actually belong to the lower (left panel) or the upper (right panel) part of the distribution.

effectively help resist the ego-depleting consequences of these message.

## 4 Do subjects update their beliefs according to reported informativeness?

In this section, we focus on our measures of subjects' beliefs regarding whether their performance belongs to the upper half of the benchmark group.

### 4.1 Manipulation check: Information matters

As a manipulation check, we first ascertain that subjects do take into account the information they are reminded about in REMINDER when forming posterior beliefs about their relative ranking.<sup>11</sup> Figure 3 displays the distribution of updates (the difference between posterior and prior beliefs at the individual level) in this treatment depending on whether subjects received a positive (in blue) or a negative (in red) message. Given that subjects who receive positive and negative messages differ in IQ scores, we compare the effect of messages on beliefs by splitting subjects who belong to the lower (left-hand side) and upper (right-hand side) part of the IQ score distribution.

<sup>11</sup>The Appendix, Section B, provides descriptive statistics on the distribution of priors in the data, which are perfectly balanced across all treatment groups. We did not explicitly pre-register this manipulation check, but provide it as additional evidence showing that subjects understand the experimental setting.

In both groups, receiving a positive message induces a significant shift to the right of the updating: in the bottom half of the IQ score distribution (Figure 3.a), the average updating is  $-0.9$  (resulting in a 46.7% posterior probability on average) among subjects who received a negative message, and jumps to 6.2 (with a posterior of 58.5%) when a positive message has been received ( $p = .015$ , conditional t-tests). Among subjects who belong to the upper half of the distribution (Figure 3.b), the average updating equals  $-3.4$  among subjects who received a negative message (as expected, the average posterior is higher among subjects who belong to the upper half, and equals 61.3% for subjects who received a negative message) and increase to 2.4 (with a posterior of 73.1%) after a positive one ( $p = .009$ , conditional t-tests). The shift in the overall distribution of updates according to the message is significant for both groups of subjects ( $p = .026$  in the lower half,  $p = .033$  in the upper half; KS tests).

## 4.2 Motivated updating of beliefs

The distribution of updates in REMINDER unambiguously confirms that subjects do respond to the feedback they received in the experiment. We now analyze the consequences of the absence of reminder about informativeness on the belief updating process using a normalized measure of updating that rescales the difference between the posterior and the prior to ease comparison between subjects who receive a positive or negative message.<sup>12</sup>

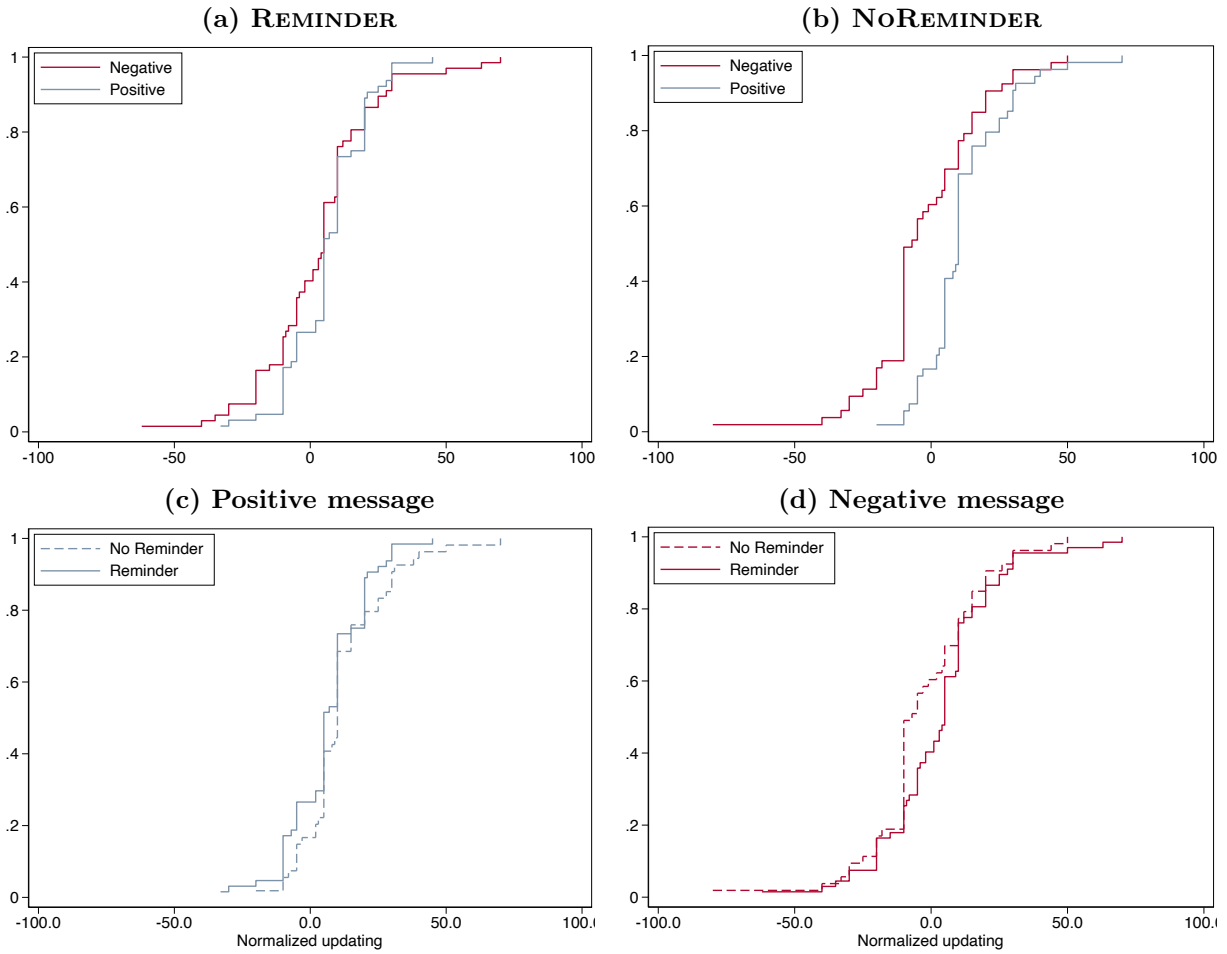
Figure 4 displays the cumulative distributions of this normalized updating variable in the two treatments.<sup>13</sup> The top panel focuses on the difference in updating between subjects who received a positive or a negative message, separately by treatment. In REMINDER (Figure 4.a), the distributions of normalized updating for each message almost exactly coincide (and are not statistically different,  $p = .361$ , KS test). In front of a negative and a positive message, beliefs updating is similar in terms of magnitude but goes in opposite directions. In NOREMINDER, by contrast, there is a clear discrepancy between how subjects update beliefs when receiving a positive or a negative message (Figure 4.b). As expected, subjects who received a positive message tend to update their beliefs upwards. As compared to these subjects, the distribution of normalized updating is shifted to the left for subjects who received a negative message, showing that these subjects also tend to update their beliefs upwards despite the negative message: subjects manage

---

<sup>12</sup>We apply the same normalization as Zimmermann (2020), weighting the difference by  $-1$  for subjects who received a negative message, so that normalized updating is positive if subjects update in the expected direction — upwards/downwards after a positive/negative message.

<sup>13</sup>As is usually the case in this kind of experiment, we observe a very large share of subjects (49.7% on average) whose posterior is exactly equal to their prior. There are very little difference in this share across groups (51.4% among subjects who receive a negative message and 53.4% among subjects who receive a positive message in NOREMINDER; 48.9% and 45.3%, respectively, in REMINDER, also see Figure 3) and none of these differences are statistically significant (the regression results are available from the authors upon request). To ease comparison, the figure displays only subjects for whom the updating is non-zero. All statistical comparisons are performed on the full sample, including subjects whose updating is zero.

Figure 4: Normalized updating by treatment and the message received



**Note.** Cumulative Distribution Functions (CDF) of the normalized updating (= posterior - prior,  $\times -1$  if the message is negative) split according to whether subjects received a positive (in blue) or a negative (in red) message, by treatment in the top panel and within treatments by message in the bottom panel. To facilitate the visualization of the data, the figure only reports the distribution among subjects whose updating is non-zero.

to resist the bad news when they have received a negative message but are not reminded that it is informative (the difference between the two distributions is statistically significant,  $p = .011$ , KS test).<sup>14</sup>

The bottom part of the figure reorganizes the same data according to the message received, which allows to perform treatment comparisons while holding constant the underlying distribution of IQ scores. Subjects who received a positive message (Figure 4.c) tend to update their belief slightly more upwards in the absence of a reminder: the average normalized updating increases from

<sup>14</sup>This comparison between messages for a given treatment are exploratory, but follow the logic of Zimmermann (2020).

3.6 in REMINDER to 5.4 in NOREMINDER. Subjects who received a negative message (Figure 4.d), by contrast, tend to go more against the information they received by updating less downwards in the absence of reminder: the normalized updating drops from 1.5 in REMINDER to  $-1.3$  in NOREMINDER. Although neither of these two differences is statistically significant,<sup>15</sup> these two shifts in opposite direction lead to a statistically significant difference of the response to the reminder between subjects who received a positive or a negative message ( $p = .073$ , conditional t-tests).

## 5 Do subjects self-servingly overturn bad news? Evidence from complementary treatments

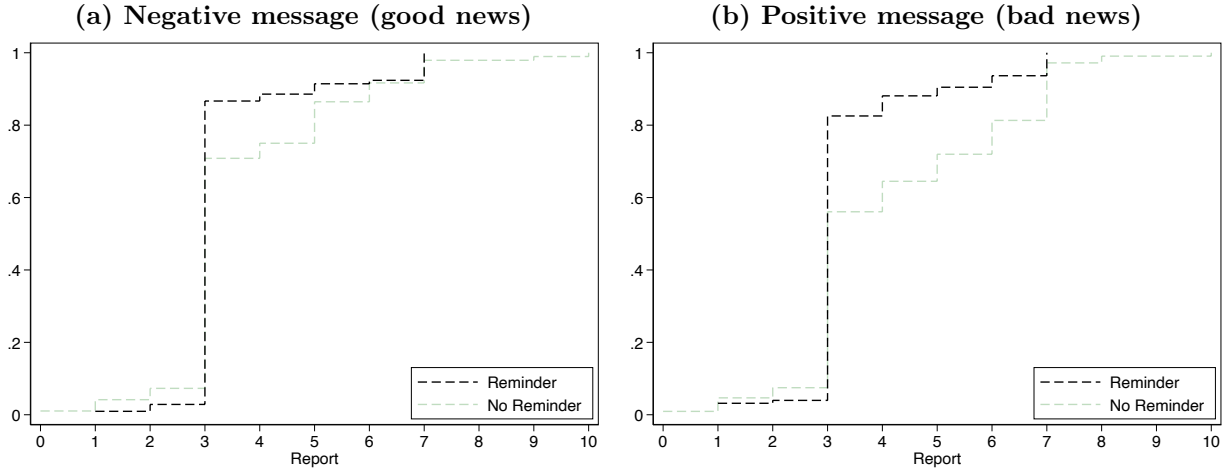
In our main experiment, messages received by subjects are drawn from an urn which contains  $p = 7$  true and  $10 - p = 3$  false messages. Since messages seen by subjects are more likely to be true than false, positive messages can directly be interpreted as good news about relative performance, and negative messages as bad news. In this section, we aim to assess the robustness of our results to situations in which the message should not be taken at face value, i.e., when the message is drawn from an urn containing more false messages than true ones.

We run two complementary between-subjects treatments in which the number of true messages in the urn is  $p = 3$  and the number of false messages is  $10 - p = 7$ . We refer to these treatments as UNRELIABLE (as opposed to the main treatments, in which messages are RELIABLE). We again vary (between subjects) the reminders given to subjects in session 2, and contrast both the recall of the message informativeness and the posterior beliefs between UNRELIABLE-REMINDER and UNRELIABLE-NOREMINDER.<sup>16</sup> As in the main experiment, maintaining a positive self-image implies to inflate the informativeness of a positive message and to deflate it when this message is negative. In Section 3, we have shown that only subjects who received negative messages (or, equivalently when messages are reliable, bad news) manipulate informativeness, suggesting that

<sup>15</sup>Neither the change in the average normalized updating (among subjects who received either a negative,  $p = .227$ , or a positive,  $p = .307$ , message ; conditional t-tests) nor the shifts in the distributions ( $p = .886$  and  $p = .437$ , respectively, KS tests) are statistically significant.

<sup>16</sup> $N = 445$  additional subjects participated to these treatments, among whom 237 were exposed to a REMINDER. Subjects typically report accurately the information they are reminded about in UNRELIABLE: in session 1, 98% of subjects accurately remember the message they received and 75.3% remember (exactly) the number of true messages in the urn. In session 2, 95.7% of subjects accurately remember the message they received. As shown in the Appendix, Section C, REMINDER moreover induces significant variations in session 2 regarding the recall of the number of true messages in the urn: 78.9% of subjects remember (exactly) the number of true messages in the urn in REMINDER, but it is the case of only 54.3% in NOREMINDER ( $p = .000$ , conditional t-test). Similar to the main experiment, about three in four participants (71.9%) return for session 2 in UNRELIABLE. We fail to find any relationship between the message received and attrition: among subjects who received a negative message in session 1 (292 in total), 84 (i.e., 29%) drop-out, while 90 out of 327 (28%) do so among subjects who received a positive message in session 1. The two attrition rates are not statistically different according to  $\chi^2$  test,  $p = .731$ .

Figure 5: Informativeness reports conditional on the message received in UNRELIABLE



**Note.** Cumulative Distribution Functions (CDF) of the self-reported number of true messages in UNRELIABLE-REMINDER vs UNRELIABLE-NOREMINDER, depending on whether subjects received a negative (left panel, which should be interpreted as good news) or a positive (right panel, bad news) message.

no manipulation is needed when news are already good. When information is unreliable, a positive message delivers bad news and a negative message delivers good news. This additional complexity may make bad news more complex to identify, and to go against, for subjects.

### 5.1 Self-serving recall of the informativeness of unreliable messages

Figure 5 displays the distribution of informativeness reports in the UNRELIABLE treatments, split according to the content of the message received. Figure 5.a describes the answers given in REMINDER and NOREMINDER by subjects who received a negative message, which should be interpreted as good news. As in the main experiment, we observe a more noisy distribution of reports in NOREMINDER, but the drop in the share of accurate reports remains limited and amounts to 24% (from 80.7% in REMINDER to 61.6% in NOREMINDER,  $p = .007$ , conditional t-test). The drop is much larger, equal to 38% (77.3% vs 47.7%,  $p = .000$ , conditional t-test), for subjects displayed in Figure 5.b who received positive messages that should be interpreted as bad news (in contrast with the main experiment, this message-dependent shift in accuracy is large enough to achieve statistical significance,  $p = .086$ , conditional t-test).

Looking at average reports of informativeness, we observe a slight (from 3.37 in REMINDER to 3.69 in NOREMINDER), but insignificant ( $p = .190$ , conditional t-test), increase among subjects who received a negative message, that is, good news. In contrast, we observe a large (3.38 vs 4.18), and highly significant ( $p = .000$ , conditional t-test), increase in average guesses after a positive message. This result mirrors the result presented for RELIABLE messages in Section 3.2. Only

subjects who received bad news, here in the form of positive messages, manipulate informativeness. When messages are UNRELIABLE, subjects inflate the informativeness of positive messages (relative to the truth) to protect themselves against the bad news. When messages are RELIABLE, subjects deflate the informativeness of negative messages (relative to the truth) to protect themselves against the bad news.

These interpretations are further substantiated by the observed variations in the share of subjects whose report of informativeness is higher than 5.<sup>17</sup> We remind that, if the informativeness level is higher than 5, the message are reliable in the sense that they can be taken at face value. After negative messages (good news), we only observe a small and insignificant increase in the share of subjects whose reported informativeness is higher than 5 (8.6% in REMINDER vs 13.5% in NOREMINDER,  $p = .378$ , conditional t-test). After a positive message (bad news), by contrast, this share strongly, and significantly, increases from 9.5% in REMINDER to 28% in NOREMINDER ( $p = .000$ , conditional t-test), and is significantly larger than the one observed among subjects receiving a negative message ( $p = .024$ , conditional t-test). That is, after bad news under the form of an unreliable positive message, subjects in NOREMINDER inflate informativeness — which makes the positive message reliable.

Regarding recall of informativeness, the complementary experiments deliver results which are qualitatively similar to the ones of the main experiments: subjects who received bad news are significantly more likely than subjects who received good news to distort the informativeness of the message in a self-serving way. Importantly, the asymmetry between subjects who receive good news and bad news is still present despite good news and bad news being more complex to access at first in UNRELIABLE.

## 5.2 How do subjects update their beliefs in front of unreliable messages?

Applied to the UNRELIABLE setting, the normalized measure of updating used in Section 4.2 amounts to assume that subjects accurately invert the message they received. As shown in the previous section, the response of subjects who receive such an unreliable positive message rather seems to contradict this assumption: they tend to manipulate informativeness so as to take the content of the message at face value. In Figure 6, we accordingly report the updating variable itself rather than the normalized measure.<sup>18</sup>

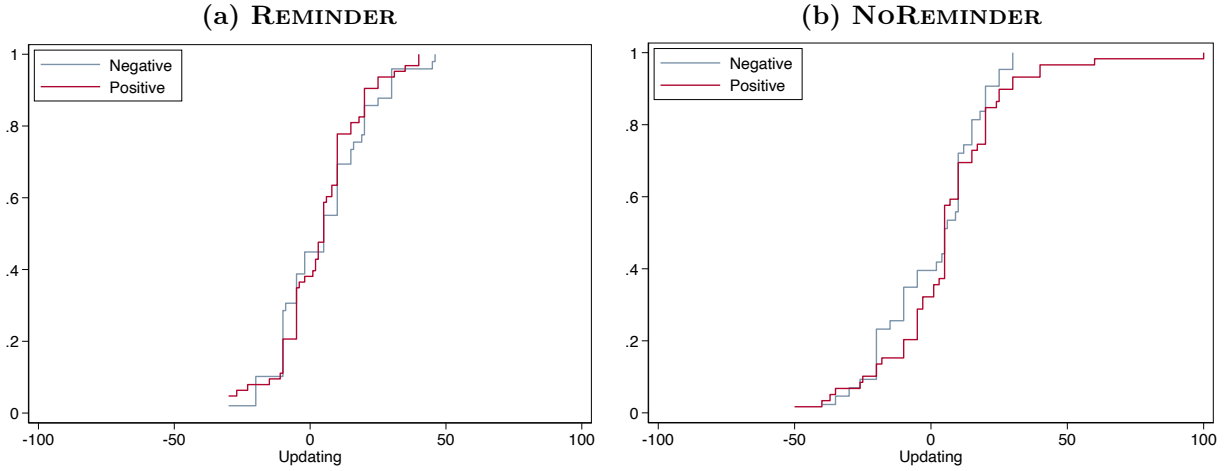
---

<sup>17</sup>This part of the analysis is exploratory.

<sup>18</sup>The distribution of priors is balanced across REMINDER and NOREMINDER in UNRELIABLE, as shown in the Appendix, Section B. This analysis of updating in UNRELIABLE is exploratory.



Figure 6: Updating by treatment and the message received in UNRELIABLE



**Note.** Cumulative Distribution Functions (CDF) of the individual updating ( $= \text{posterior} - \text{prior}$ ) in UNRELIABLE-REMINDER (left panel) and UNRELIABLE-NOREMINDER (right panel) among subjects who received a negative (which should be interpreted as good news, in blue) or a positive (bad news, in red) message. To facilitate the visualization of the data, the figure only reports the distribution among subjects whose updating is non-zero.

The updating observed in front of a REMINDER, reported in Figure 6.a, clearly illustrates how difficult it is for subjects to make use of the information delivered by the message in the UNRELIABLE treatments. In contrast with the main experiment, we fail to observe a clear difference according to the message received: when exposed to a REMINDER, subjects update in a similar way following either a positive or a negative message ( $p = .991$ , KS test). While we do observe a slight shift to the right of the distribution of updates after a positive message in NOREMINDER (see Figure 6.b), the updating is too noisy to achieve statistical significance ( $p = .625$ , KS test).<sup>19</sup> This suggests that the NOREMINDER condition gives more room to subjects to ignore the message-generating process and follow the content of their message.

Pooling the data from all four treatments, Table 1 provides empirical support to this interpretation. We investigate the trade-off between relying on the literal content of the message (positive or negative) and relying on the information it contains (good news if a positive message is drawn in RELIABLE, or if a negative one is drawn in UNRELIABLE) when subjects update their beliefs. In front of a REMINDER, subjects' updating does respond to the informational content of the message, although they are also influenced by the content of the message itself. With NOREMINDER, the observed updating behavior is consistent with subjects focusing solely on the content of the message they received, and ignoring the information it is supposed to deliver.

<sup>19</sup>The average updating among subjects who received a negative message tend to decrease between treatments (from 2.21 in REMINDER to 0.35 in NOREMINDER) and to increase after a positive message (1.78 vs 3.13). None of these differences are statistically significant based on conditional t-tests, though.

Table 1: Regression of updating on pooled data

	(1)	(2)
	REMINDER	NOREMINDER
Positive message	2.329** (0.040)	3.571** (0.010)
Good news	4.561*** (0.000)	1.096 (0.475)
<i>N</i>	485	433

**Note.** OLS regressions on pooled data (from both the main experiment and UNRELIABLE), of the individual updating (= *posterior - prior*) on whether (i) a positive message has been received and (ii) the message should be interpreted as good news according to the message-generating process (equals 1 if a positive message has been received in the main experiment or if a negative message has been received in UNRELIABLE). Model (1) focuses on data from treatments featuring a REMINDER, model (2) on data from NOREMINDER conditions. All models include controls for age, gender, and the true ranking in the IQ distribution. Legend: \*\*\* 1%; \*\*, 5%; \* 10%..

## 6 Conclusion

In an experiment in which subjects cannot forget the feedback they receive about their relative intelligence, we vary whether the reliability of this feedback is reminded to them or not. We show that subjects who receive bad news – both in the form of negative messages that should be taken at face value (in the RELIABLE treatment) and in the form of positive messages that should be inverted (in the UNRELIABLE treatment) – tend to deflate the reliability of this news. When the news is good, we see no such self-serving manipulation of the informativeness of messages. We also find that the absence of a reminder about the informativeness of messages allows subjects to update beliefs about their relative intelligence upwards despite having received negative feedback.

Our work directly contributes to the experimental literature that specifically studies selective recall about performance in IQ tests. In Chew, Huang, and Zhao (2020), participants have to assess, after several months, how the answers they gave in a previous IQ test compare to the correct answers. The authors document that individuals make more recall errors for incorrect than for correct answers. In Zimmermann (2020), participants receive noisy information about their relative performance in an IQ test in the form of positive or negative messages. After a month, subjects are asked to recall these signals and recall negative feedback with lower accuracy than positive feedback. In our experiment, we also provide subjects with noisy information about their relative performance but make sure they cannot forget this signal. Instead, we allow subjects to forget how noisy the feedback was. Our results show this setting gives subjects enough room to use feedback in a self-serving way.

This result provides further behavioral support to the strand of works showing that individuals' trust in news sources is endogenous, with important consequences on political polarization.<sup>20</sup> In that vein, Thaler (2024) experimentally studies motivated trust in the reliability of news. In this experiment, news is provided by either reliable or unreliable sources. Although the news itself is uninformative about the reliability of the source from a Bayesian perspective, subjects tend to assess the veracity of the source in a motivated way. They consider that news which favors their own political party is significantly more likely to come from reliable sources than news which supports the opposing party. In a different setting, our work shows that a source providing unpleasant ego-relevant feedback is considered as less reliable than it truly is, even though the true reliability is initially known. This allows subjects to stay confident about their ability despite negative evidence.

## References

- AMELIO, A., AND F. ZIMMERMANN (2023): "Motivated Memory in Economics—A Review," *Games*, 14(1), 1–15.
- ARTHUR, W., AND D. V. DAY (1994): "Development of a Short Form for the Raven Advanced Progressive Matrices Test," *Educational and Psychological Measurement*, 54(2), 394–403.
- CASTAGNETTI, A., AND R. SCHMACKER (2022): "Protecting the Ego: Motivated Information Selection and Updating," *European Economic Review*, 142, 104007.
- CHEW, S. H., W. HUANG, AND X. ZHAO (2020): "Motivated False Memory," *Journal of Political Economy*, 128(10), 3913–3939.
- DANA, J., R. A. WEBER, AND J. X. KUANG (2007): "Exploiting Moral Wiggle Room: Experiments Demonstrating an Illusory Preference for Fairness," *Economic Theory*, 33(1), 67–80.
- DANZ, D., L. VESTERLUND, AND A. J. WILSON (2022): "Belief Elicitation and Behavioral Incentive Compatibility," *American Economic Review*, 112(9), 2851–2883.
- DROBNER, C., AND S. J. GOERG (2024): "Motivated Belief Updating and Rationalization of Information," *Management Science*, 70(7), 4583–4592.
- EIL, D., AND J. M. RAO (2011): "The Good News-Bad News Effect: Asymmetric Processing of Objective Information about Yourself," *American Economic Journal: Microeconomics*, 3(2), 114–38.
- GENTZKOW, M., M. B. WONG, AND A. T. ZHANG (forthcoming): "Ideological Bias and Trust in Information Sources," *American Economic Journal: Microeconomics*.
- GÖDKER, K., P. JIAO, AND P. SMEETS (2024): "Investor Memory," *Review of Financial Studies*, Forthcoming.

---

<sup>20</sup>See Gentzkow, Wong, and Zhang (forthcoming) for a recent work on that topic. Experimental works in this area include Levendusky (2013); Nisbet, Cooper, and Garrett (2015) and Thaler (2024).

- GOERG, S. J., AND J. KAISER (2009): “Nonparametric Testing of Distributions—the Epps–Singleton Two-Sample Test Using the Empirical Characteristic Function,” *Stata Journal*, 9(3), 454–465.
- GOLMAN, R., D. HAGMANN, AND G. LOEWENSTEIN (2017): “Information Avoidance,” *Journal of Economic Literature*, 55(1), 96–135.
- HAGENBACH, J., AND C. SAUCET (2024): “Motivated Skepticism,” *Review of Economic Studies*, Forthcoming.
- HOSSAIN, T., AND R. OKUI (2013): “The Binarized Scoring Rule,” *Review of Economic Studies*, 80(3), 984–1001.
- KANE, J. V., AND J. BARABAS (2019): “No Harm in Checking: Using Factual Manipulation Checks to Assess Attentiveness in Experiments,” *American Journal of Political Science*, 63(1), 234–249.
- LEVENDUSKY, M. S. (2013): “Why Do Partisan Media Polarize Viewers?,” *American Journal of Political Science*, 57(3), 611–623.
- MÖBIUS, M. M., M. NIEDERLE, P. NIEHAUS, AND T. S. ROSENBLAT (2022): “Managing Self-Confidence: Theory and Experimental Evidence,” *Management Science*, 68(11), 7793–7817.
- NISBET, E. C., K. E. COOPER, AND R. K. GARRETT (2015): “The Partisan Brain: How Dissonant Science Messages Lead Conservatives and Liberals to (Dis) Trust Science,” *The ANNALS of the American Academy of Political and Social Science*, 658(1), 36–66.
- SAUCET, C., AND M. C. VILLEVAL (2019): “Motivated Memory in Dictator Games,” *Games and Economic Behavior*, 117, 250–275.
- SHAROT, T. (2011): “The Optimism Bias,” *Current Biology*, 21(23), R941–R945.
- THALER, M. (2024): “The Fake News Effect: Experimentally Identifying Motivated Reasoning Using Trust in News,” *American Economic Journal: Microeconomics*, 16(2), 1–38.
- WILSON, A., AND E. VESPA (2018): “Paired-Uniform Scoring: Implementing a Binarized Scoring Rule with Non-Mathematical Language,” *Working Paper*.
- ZIMMERMANN, F. (2020): “The Dynamics of Motivated Beliefs,” *American Economic Review*, 110(2), 337–361.

# Appendix

## A Experimental procedures

Since the experiment is self-paced online, we control for the attention of participants by allowing participants to move forward only after a 10-20s delay on all information screens (we do not include such a constraint on decision screens so as to reveal decision times).<sup>21</sup> The sample heterogeneity and its distribution across treatments is described in Table A.

Table A: Sample characteristics by treatment

Noise reminder	RELIABLE			UNRELIABLE			<i>p</i> -value ( <i>F</i> -test)
	NO	YES	<i>p</i> -value	NO	YES	<i>p</i> -value	
<b>IQ Score</b>	6.9	7.1	.488 <sup>a</sup>	6.5	7.1	.038 <sup>a</sup>	.093
<b>Prior</b>	55.8	57.9	.326 <sup>a</sup>	56.0	55.3	.767 <sup>a</sup>	.628
<b>Male (%)</b>	40.0	48.8	.055 <sup>a</sup>	47.6	43.9	.434 <sup>a</sup>	.219
<b>Age</b>	40.9	38.8	.068 <sup>a</sup>	40.1	39.4	.512 <sup>a</sup>	.274
<b>Education (%)</b>							
Less than high school	10.7	9.7		8.7	8.9		
High school	25.3	21.8		24.5	27.8		
2yr college	26.2	23.8		24.0	16.5		
Higher education	37.8	44.8		42.8	46.8		
<i>p</i> -value ( $\chi^2$ test)			.490			.255	.315
<b><i>N</i></b>	225	248		208	237		

<sup>a</sup> *p*-values from t-test comparisons.

Subjects receive a €6 show-up fee for their participation in both sessions of the experiment. Due to constraints specific to the panel provider, two-third of the show-up fee is paid to subjects after their participation to the first session. For the performance-based payment of subjects, one or the five incentivized tasks of the two sessions is randomly selected. This part of the subjects' payment is conditional on participation in both sessions.

Two of the five tasks happen in session 1: the Raven's test (12 items, €0.5 per correct answer) and the prior belief elicitation task. All beliefs in both sessions are elicited using the Wilson and Vespa (2018) non-mathematical implementation of the Binarized Scoring Rule (Hossain and Okui, 2013), whose belief revelation properties have been studied by Danz, Vesterlund, and Wilson (2022, from which our instructions follow the “*Description*” treatment). Subjects are asked for a guess about the likelihood of a statement (probability they belong to the top half of the benchmark group when eliciting beliefs about relative

<sup>21</sup>We initially included two attention checks (one in each session), and borrowed from Kane and Barabas (2019): subjects are asked to answer “5” (“7”) to the question “What is 2 + 2 (3+3) equal to?” before the belief elicitation stage in session 1, and before the provision of reminders in session 2. Ex-post, these attention checks turned out to be too demanding (failure rate of over 25%), so we do not enforce them in our main analysis.

performance, number of true messages in the recall questions). They are explicitly told they maximize their chance of winning the €2 incentives offered for this part if they report their true belief. They are additionally given access to a detailed description of the incentive scheme, based on the draw of the two random numbers (see the instructions in Section D below, for more details).

The three remaining incentivized tasks happen in session 2: the posterior belief (€2 based on the BSR described above), the recall question about the message received (€1 if the report is accurate) and the recall question about the number of true messages (€2 based on the BSR described above). The average duration of both sessions was 25mn and the average payoff of participants is equal to €7.77.

In the four recall questions of the experiment — two non-incentivized recall questions in the first session and two incentivized recall questions in the second session — subjects are offered the option to answer ‘*I don’t know*’. In the first session, none of the subjects used this option when recalling the message seen and 47 subjects out of 918 use the ‘*I don’t know*’ answer when having to state the number of true messages. In the second session, only 4 subjects out of 918 use this answer when asked to recall the message they had just seen, and only 20 subjects use this answer when asked to recall the number of true messages.

## B Distribution of prior beliefs

Table B provides the empirical distribution of subjects’ prior as a function of their true ranking and the message they received in both RELIABLE (main experiment) and UNRELIABLE. The perceived rank (lower / middle / upper half, in row) is deduced from the reported probability of belonging to the upper half. Subjects in the experiment are typically over-confident. Overall, 62.3% of them are accurate on their relative score (assuming people who report 50 believe they belong to the upper half). Priors are balanced across the four treatments according to a Fisher test ( $p = .628$ ).

Table B: Distribution of prior beliefs in the sample

<b>RELIABLE</b>						
Prior	Positive message			Negative message		
	True rank			True rank		
	Lower (%)	Upper (%)	Overall (%)	Lower (%)	Upper (%)	Overall (%)
Lower half	41.2	7.2	18.9	40.4	8.5	34.2
Middle	21.2	13.7	16.3	16.6	14.9	16.2
Upper half	37.5	79.1	64.8	43.0	76.6	49.6

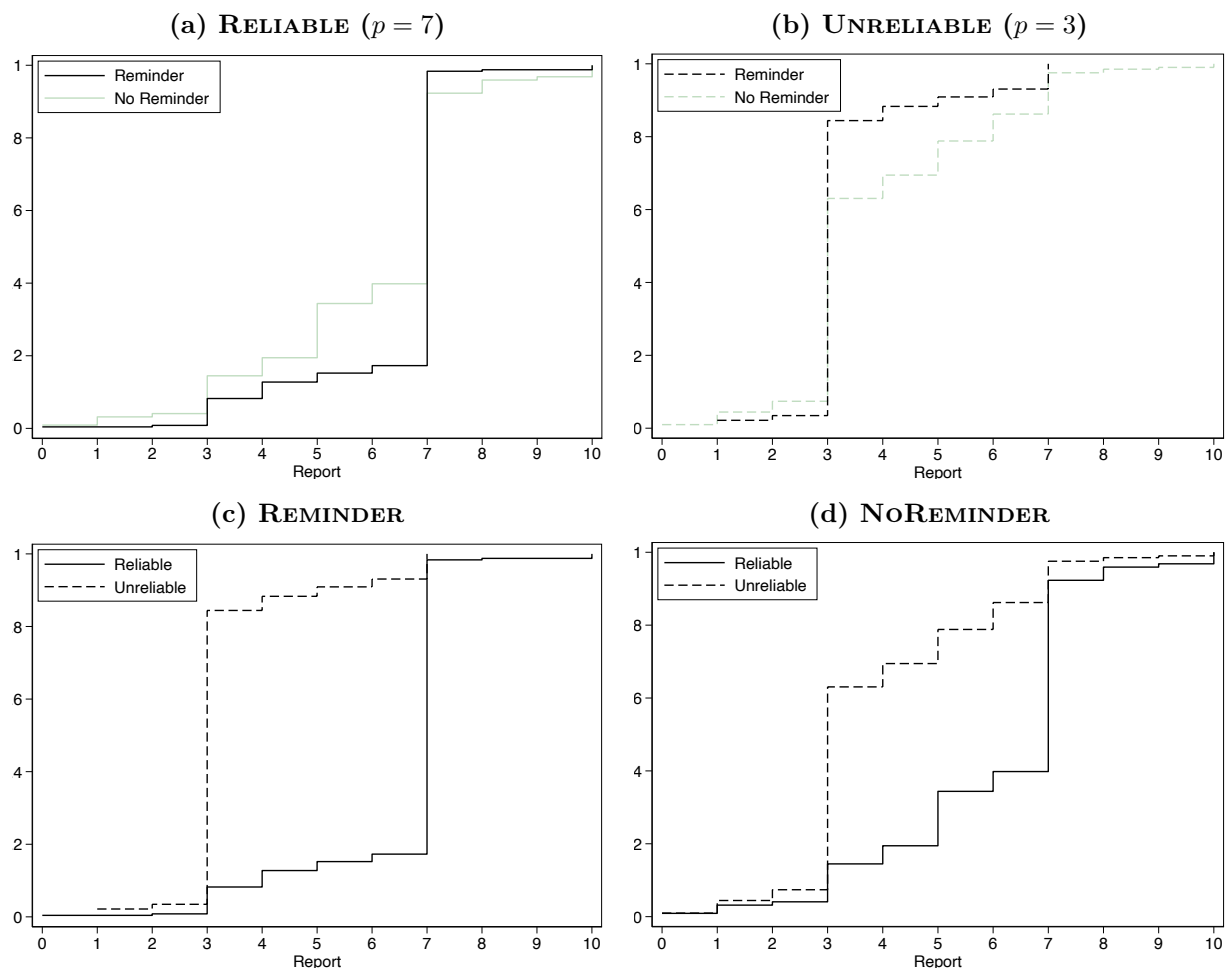
  

<b>UNRELIABLE</b>						
Prior	Positive message			Negative message		
	True rank			True rank		
	Lower (%)	Upper (%)	Overall (%)	Lower (%)	Upper (%)	Overall (%)
Lower half	38.6	5.7	31.2	52.7	13.4	27.4
Middle	23.4	7.5	19.8	12.2	17.2	15.4
Upper half	38.0	86.8	48.9	35.1	69.4	57.2

## C Manipulation check: Recall of reminded information in UNRELIABLE

Figure A displays the distribution of the reported number of true messages in the urn in session 2. The top panel provides comparisons between REMINDER and NOREMINDER separately for the main experiment (labeled RELIABLE, Figure A.a, already shown in Figure 1 in the main text) and for UNRELIABLE (Figure A.b.). The bottom panel reorganizes the same data to show the shift in reported numbers between RELIABLE and UNRELIABLE in both REMINDER (Figure A.c) and NOREMINDER (Figure A.d).

Figure A: Reported number of true messages in the urn, by treatment



**Note.** Cumulative Distribution Functions (CDF) of the reported number of true messages in both the main experiment (RELIABLE) and UNRELIABLE, depending on whether subjects have been exposed to a REMINDER or not (NoREMINDER).

## D Written instructions used in the experiment

*Subjects did not see the sentences which appear in italic below, they are indications for the reader. The sign — indicates a new screen and block letters correspond to the title which appeared on the top of that screen.*

### D.1 First Experimental Session

We thank you for participating in this experiment.

It is financed by a grant from the European Research Council and has received ethical approval by the Institutional Review Board of the Paris School of Economics (decision n° 2021-21). The next screens describe the proceeding of the experiment exactly as it takes place.

#### — INFORMATION AND CONSENT

##### 1) Consent

The study involves a short test designed to measure IQ, a few questions and a socio-demographic survey.

**The study involves 2 sessions:** one taking place today and another one taking place in exactly two days. It is important for our research that you participate in both sessions. **Please, consent to participate in this experiment only if you intend to participate in both sessions.**

You will receive a separate invitation for the second session by e-mail.

The estimated time to complete today's session is 17 minutes. The estimated time to complete the session in two days is 8 minutes.

##### 2) Risks

There is no particular physical or psychological risk associated with participation in this study.

##### 3) Earnings

You will earn a **fixed payment** of 4 euros for having completed the first session. You will also received a **fixed payment** of 2 euros for having completed the second session.

During the two sessions, you can additionally earn a **bonus payment** in ways we will describe later. The minimum bonus is 0 euros and the maximum bonus is 6 euros.

No payments will be made to participants who do not complete both sessions. **The whole bonus payment will be made at the end of the second session.**

##### 4) Data Processing

Data will be collected anonymously and will never be used to identify you as an individual.

##### 5) Data Access, Data Protection and Data Use

The data will be analyzed for scientific purposes. The analysis may result in publications and presentations. Researchers may share the collected anonymous data with other researchers.

##### 6) Withdrawal and Competent Authority



You are free to withdraw from the experiment at any time and for any reason. In this case, the answers you already gave will be deleted. According to the rules of CrowdPanel, you however need to have completed the study to obtain the completion code necessary for payment.

## 7) Contacts

In case of questions about the study, you can contact the researcher responsible of this study: Jeanne Hagenbach (jeanne.hagenbach@sciencespo.fr)

Please click on the button below if you are ready to start:

*Subjects decided whether or not to click on the following sentence:* I consent to participate to the two sessions of this experiment.

### — GENERAL INFORMATION

On some screens of this experiment, the 'Next' button (to proceed to the next page) only appears on the screen after a small delay (typically between 10 and 20 seconds).

**We ask you to read all the instructions carefully. It is important for our research.**

### — PAYMENT IN THIS EXPERIMENT

You will receive a **fixed payment** for your participation to this first session.

Today, the experimental session is made up of **2 parts**. The next session, taking place in two days, will be made up of **3 parts**. **In each of the five parts, you can additionally gain bonus payment depending on your decisions and answers.** At the end of the experiment, the computer will randomly select one of the 5 parts and you will receive the bonus payment earned in this part. **It is therefore in your interest to try to maximize your bonus payment in each of the 5 parts.**

### — PART 1 - IQ TEST

In this part, you will complete a test frequently used to measure Intelligence Quotient (IQ). **This measure of intelligence evaluates ability to reason fluently and to understand complex problems.** Several research articles in social sciences have demonstrated that scores obtained in this test are correlated to educational and professional success.

The test is made up of **12 questions**. For each question, you are presented with a sequence of pieces and you must find the missing piece. You will have to choose it among several proposed pieces. **You will have 15 minutes to answer all the questions.**

Your **IQ score** corresponds to the number of correct answers you give in the test. In this part, you earn 0.50 euros of bonus payment for each correct answer given in the test.

Before you start the test, **we will present you two examples** so that you can familiarize yourself with the questions.

*We then present to the subjects two examples with solutions. Then they do the test for 15 minutes.*

### — END OF PART 1

Part 1 is now over.

—  
Please ignore the text of the question below and answer 5. What is  $2 + 2$  equal to?

*Subjects type in a number.*

— PART 2

In 2022, we administered the same IQ test you just did in Part 1 to a group of more than 100 participants. Like you, these participants completed the test online with a time constraint of 15 minutes. For each participant, the **score** at the IQ test is the total number of correct answers.

**We will compare your score at the IQ test to the scores of these participants.**

- **In the IQ test, you rank in the upper half** if your score is better than at least half of these participants.
- **In the IQ test, you rank in the lower half** if your score is not better than at least half of these participants.

Please note that, in this experiment, you will never learn for sure whether you rank in the upper or lower half.

On the next screen, we will ask you to **estimate the probability that, in the IQ test, you rank in the upper half.**

The **bonus payment** that you can gain in this part is either 0 euros or 2 euros. **The answer you will give on the next screen will determine the probability with which the bonus payment will be 2 euros for this part.** To maximize your chances to get a bonus payment of 2 euros, it is in **your interest to report the estimate of the probability which is closest to what you believe.**

If you want more details about the way we will compute this payoff, you can click here.

*By clicking, subjects could access the details which appear on the following screen.*

—  
The computer will randomly and independently draw two numbers between 0 and 100 (these numbers can have decimals and all have equal probabilities to be drawn). The bonus payment for this part of the experiment will be 2 euros if:

- At the IQ test, you actually rank in the upper half and your estimated probability is higher than the smaller of the two randomly-drawn numbers, or
- At the IQ test, you actually rank in the lower half and your estimated probability is smaller than the larger of the two randomly-drawn numbers.

— PART 2

According to you, what is the probability in percent (between 0 and 100) that you rank in the **upper half** at the IQ test?

*Subjects enter a number.*

— PART 2

Later in this part, you will receive **a message about your ranking in the IQ test** (upper or lower half). **This message is either true or false.**

Precisely, we have generated 10 messages about your ranking:

- **7 messages are true**, that is, **correspond to your true ranking**, and
- **3 messages that are false**, that is, **correspond to the opposite of your true ranking**.

The message you will see has been randomly chosen among these 10 messages.

— PART 2

Among the 10 messages that we have generated about your ranking in the IQ test, how many are **true**, that is, correspond to your true ranking in the IQ test?

*Subjects choose using a pull-down menu going from 0 to 10.*

— PART 2

The message about your ranking in the IQ test is:

*The message is generated and displayed as explained to the subjects.*

— PART 2

Please indicate **the message** you have just seen:

- In the IQ test, you rank in the upper half.
- In the IQ test, you rank in the lower half.
- I don't know which message I have just seen.

—

The second session will take place in exactly two days. You will receive a separate invitation by email. Only participants who complete both sessions will be paid. The payment will be given at the end of the second session.

Thank you for your participation in the first session of our experiment.

## D.2 Second session

— WELCOME

We thank you for participating in the second session of this experiment. It is financed by a grant from the European Research Council and has received ethical approval by the Institutional Review Board of the Paris School of Economics (decision n° 2021-21). The next screens describe the proceeding of the experiment exactly as it takes place.

— GENERAL INFORMATION

On some screens of this experiment, the 'Next' button (to proceed to the next page) only appears on the screen after a small delay (typically between 10 and 20 seconds).

**We ask you to read all the instructions carefully. It is important for our research.**

— GENERAL INFORMATION

Reminder of the payment in the experiment: You will receive a **fixed payment** for your participation to this first session.

The first experimental session was made up of **2 parts**. The session you participate in today will be made up of **3 parts**. **In each of the five parts, you can additionally gain bonus payment depending on your decisions and answers.** At the end of the experiment, the computer will randomly select one of the 5 parts and you will receive the bonus earned in this part. **It is therefore in your interest to try to maximize your bonus payment in each of the 5 parts.**

—

Please ignore the text of the question below and answer 7. What is 3+3 equal to?

*Subjects type in a number.*

— PART 3

Reminder: As explained in the first session, in 2022, we administered the same IQ test you did in the first session to a group of more than 100 participants. Like you, these participants completed the test online with a time constraint of 15 minutes. For each participant, the **score** at the IQ test is the total number of correct answers.

**In the first session, we have compared your score in the IQ test to the scores of these participants.**

- **In the IQ test, you rank in the upper half** if your score was better than at least half of these participants.
- **In the IQ test, you rank in the lower half** if your score was not better than at least half of these participants.

Please note that, in this experiment, you will never learn for sure whether you rank in the upper or lower half.

— PART 3 - *This screen is used in the REMINDER treatment only*

In the first session, you received a **message about your ranking in the IQ test** (upper or lower half). **This message was either true or false.**

Precisely, we had generated 10 messages about your ranking:

- **7 messages were true**, that is, **corresponded to your true ranking**, and
- **3 messages that were false**, that is, **corresponded to the opposite of your true ranking**

The message you saw had been randomly chosen among these 10 messages. We remind you that this message was:

*Subjects see the message they had seen in the first session.*

— PART 3 *This screen is used in the NOREMINDER treatment only*

In the first session, you received a **message about your ranking in the IQ test** (upper or lower half). **This message was either true or false.**

Precisely, we had generated 10 messages about your ranking:

- **some messages were true**, that is, **corresponded to your true ranking**, and
- **some messages that were false**, that is, **corresponded to the opposite of your true ranking**

The message you saw had been randomly chosen among these 10 messages. We remind you that this message was:

*Subjects see the message they had seen in the first session.*

— PART 3

On the next screen, we will ask you to **estimate the probability that, in the IQ test, you rank in the upper half.**

The **bonus payment** that you can gain in this part is either 0 euros or 2 euros. **The answer you will give on the next screen will determine the probability with which the bonus payment will be 2 euros for this part.** To maximize your chances to get a bonus payment of 2 euros, it is in **your interest to report the estimate of the probability which is closest to what you believe.**

If you want more details about the way we will compute this payoff, you can click here.

*By clicking, subjects could access the details which appear on the following screen.*

—

The computer will randomly and independently draw two numbers between 0 and 100 (these numbers can have decimals and all have equal probabilities to be drawn). The bonus payment for this part of the experiment will be 2 euros if

- At the IQ test, you actually rank in the upper half and your estimated probability is higher than the smaller of the two randomly-drawn numbers, or
- At the IQ test, you actually rank in the lower half and your estimated probability is smaller than the larger of the two randomly-drawn numbers.

— PART 3

According to you, what is the probability in percent (between 0 and 100) that you rank in the **upper half** at the IQ test?

*Subjects enter a number.*

*The next Parts 4 and 5 are presented in a random order.*

— PART 4

On the next screen, we ask you to **indicate, among the 10 messages we had generated in the first session, how many were true**, that is, corresponded to your true ranking in the IQ test.

The **bonus payment** that you can gain in this part is either 0 euros or 2 euros. **The answer you will give on the next screen will determine the probability with which the bonus payment will be 2 euros for this part.** To maximize your chances to get a bonus payment of 2 euros, it is in your interest to report the estimate of the number of true messages which is closest to what you believe. If you click on the 'I don't know' option, the bonus will automatically be 0 euro for this part.

If you want more details about the way we will compute this payoff, you can click here.

*By clicking, subjects could access the details which appear on the following screen.*

—

To determine the bonus for this part (provided that you have not chosen the 'I don't know option' which leads to a 0 bonus), the computer will randomly and independently draw two numbers between 0 and 10 (these numbers can have decimals and all have equal probabilities to be drawn). We will also draw randomly a message among the 10 messages we had generated in the first session. The bonus for this part of the experiment will be 2 euros if

- The message drawn is true and your answer is higher than the smaller of the two randomly-drawn numbers, or
- The message drawn is false and your answer is smaller than the larger of the two randomly-drawn numbers.

— PART 4

Among the 10 messages that we had generated in the first session, how many were **true**, that is, corresponded to your true ranking in the IQ test?

*Subjects choose using a pull-down menu going from 0 to 10 and could choose 'I don't know'.*

— PART 5

On the next screen, we ask you to **indicate the message you had seen in the first experimental session** and that we reminded to you today.

The **bonus payment** for this part is either 0 or 1 euro. The bonus is 1 euro if your answer is correct and 0 otherwise. If you click on the option 'I don't know', the bonus will be 0.

— PART 5

Please indicate **the message** you saw in the first session:

- In the IQ test, you rank in the upper half.

- In the IQ test, you rank in the lower half.
- I don't know which message I saw in the first experimental session.

— FINAL QUESTIONNAIRE

*Subjects fill the Rosenberg Self-Esteem questionnaire and give their age, gender, level of education and occupation.*

— PAYMENT SUMMARY

Total payment in this experiment

*Subjects see the overall payment they earned in the experiment (fixed payments for both session and bonus payment) (without seeing which part was drawn for the payment of the bonus).*

Thank you for your participation in our experiment.

## Supplementary Material (not intended for publication)

### E Conditional tests

This section provides the full regression results on which conditional tests provided in the main text are based. Coefficients in bold are the ones commented on in the text.

#### E.1 RELIABLE treatment, guesses (Section 3)

Table C: Conditional tests on the accuracy of reported informativeness, RELIABLE

	(1)	(2)	(3)	(4)
	Pooled	Positive	Negative	Pooled
NOREMINDER	<b>-24.632***</b> (0.000)	<b>-21.113***</b> (0.000)	<b>-28.803***</b> (0.000)	-20.706*** (0.000)
Negative message				4.969 (0.354)
NOREMINDER × Negative message				<b>-7.563</b> (0.344)
True rank	17.831*** (0.000)	24.848*** (0.000)	9.525 (0.165)	18.404*** (0.000)
Age	-0.893*** (0.000)	-0.825*** (0.002)	-0.936*** (0.000)	-0.888*** (0.000)
Male	5.095 (0.203)	1.007 (0.852)	9.978* (0.090)	5.382 (0.178)
Constant	103.643*** (0.000)	95.982*** (0.000)	107.361*** (0.000)	100.421*** (0.000)
Observations	473	233	240	473

**Note.** OLS regressions with robust standard errors on accuracy (i.e., the dummy variable recording whether the reported number of true messages (perfectly) matches the true one). For each model, the top row indicates the sample on which the model is estimated. Coefficients in bold are the ones commented on in the text. *Legend:* \*\*\* 1%; \*\* 5%; \* 10%.



Table D: Conditional tests on reported informativeness, RELIABLE

	Average guesses			Guesses 5 or less		
	(1) Negative	(2) Positive	(3) REMINDER	(4) Negative	(5) Positive	(6) Pooled
NOREMINDER	<b>-0.665***</b> (0.005)	<b>-0.188</b> (0.316)		<b>0.246***</b> (0.000)	<b>0.094*</b> (0.050)	0.091* (0.059)
Negative message			<b>-0.061</b> (0.226)			-0.061 (0.215)
NOREMINDER × Negative message						<b>0.151**</b> (0.047)
True rank	0.467 (0.108)	0.853*** (0.000)	-0.213*** (0.000)	-0.102 (0.104)	-0.269*** (0.000)	-0.198*** (0.000)
Age	-0.019** (0.045)	-0.014* (0.057)	0.003 (0.151)	0.005** (0.020)	0.004 (0.108)	0.005*** (0.005)
Male	0.411* (0.077)	0.056 (0.764)	-0.014 (0.753)	-0.100* (0.080)	-0.015 (0.751)	-0.055 (0.136)
Constant	6.850*** (0.000)	6.522*** (0.000)	0.166* (0.072)	0.029 (0.774)	0.180 (0.105)	0.120 (0.138)
Observations	235	229	243	235	229	464

**Note.** The top row indicates the outcome variable on which regressions are performed: OLS on average guesses in models (1) and (2); OLS with robust standard errors on the dummy variable recording whether guesses are lower or equal to 5 in the remaining columns. For each model, the last part of the top row indicates the sample on which the model is estimated. Coefficients in bold are the ones commented on in the text. *Legend:* \*\*\* 1%; \*\* 5%; \* 10%.

## E.2 RELIABLE treatment, normalized updating (Section 4)

Table E: Conditional tests on (normalized) updating, RELIABLE

	(1)	(2)	(3)	(4)	(5)
	REMINDER Lower	REMINDER Upper	Negative	Positive	Pooled
Positive message	<b>7.410**</b> (0.015)	<b>5.990***</b> (0.009)			
Negative message					-2.662 (0.149)
NOREMINDER			<b>-2.436</b> (0.227)	<b>1.559</b> (0.307)	1.741 (0.331)
NOREMINDER × Negative message					<b>-4.501*</b> (0.073)
True rank			2.365 (0.347)	-3.988** (0.013)	-1.314 (0.362)
Age	-0.085 (0.416)	0.015 (0.867)	-0.054 (0.508)	-0.074 (0.223)	-0.057 (0.259)
Male	-0.962 (0.722)	-3.747* (0.068)	-0.351 (0.861)	-0.863 (0.573)	-0.531 (0.675)
Constant	2.865 (0.544)	-2.123 (0.607)	3.228 (0.375)	9.607*** (0.002)	6.971*** (0.009)
Observations	137	111	240	233	473

**Note.** OLS regressions on the updating variable (in (1) and (2)) or on the normalized updating variable (in (3), (4) and (5)). For each model, the top row indicates the sample on which the model is estimated. Coefficients in bold are the ones commented on in the text. *Legend:* \*\*\* 1%; \*\* 5%; \* 10%.

## E.3 UNRELIABLE treatment (Section 5)

Table F: Conditional tests on guesses, UNRELIABLE (Section 5.1)

	Accuracy			Average guess			Guesses higher than 5		
	(1) Pooled	(2) Positive	(3) Negative	(4) Pooled	(5) Negative	(6) Positive	(7) Negative	(8) Positive	(9) Pooled
NOREMINDER	<b>-24.308***</b> (0.000)	<b>-31.475***</b> (0.000)	<b>-16.754***</b> (0.007)	<b>-16.607***</b> (0.006)	<b>0.261</b> (0.190)	<b>0.838***</b> (0.000)	<b>0.040</b> (0.378)	<b>0.194***</b> (0.000)	0.039 (0.397)
Positive message				8.171 (0.135)					-0.039 (0.338)
NOREMINDER × Positive message				<b>-14.455*</b> (0.086)					<b>0.153**</b> (0.024)
True rank	17.703*** (0.000)	19.022*** (0.002)	17.571*** (0.007)	18.478*** (0.000)	-0.087 (0.676)	-0.519** (0.036)	-0.029 (0.521)	-0.133*** (0.007)	-0.079** (0.018)
Age	-0.602*** (0.001)	-0.616*** (0.009)	-0.632** (0.020)	-0.617*** (0.001)	0.018** (0.035)	-0.001 (0.937)	0.003 (0.168)	0.001 (0.719)	0.002 (0.209)
Male	6.695 (0.105)	11.628** (0.047)	2.364 (0.691)	7.206* (0.083)	-0.254 (0.200)	-0.326 (0.121)	-0.026 (0.560)	-0.057 (0.238)	-0.045 (0.166)
Constant	92.198*** (0.000)	94.272*** (0.000)	91.591*** (0.000)	87.838*** (0.000)	2.877*** (0.000)	3.638*** (0.000)	-0.013 (0.892)	0.118 (0.162)	0.090 (0.187)
Observations	445	237	208	445	201	233	201	233	434

**Note.** The top row indicates the outcome variable on which regressions are performed: OLS with robust standard errors on accuracy (i.e., the dummy variable recording whether the reported number of true messages (perfectly) matches the true one) in models (1)-(4); OLS on average guesses in models (5) and (6); OLS with robust standard errors on the dummy variable recording whether guesses are higher than 5 in the remaining columns. For each model, the last part of the top row indicates the sample on which the model is estimated. Coefficients in bold are the ones commented on in the text. *Legend:* \*\*\* 1%, \*\* 5%, \* 10%.

Table G: Conditional tests on updating, UNRELIABLE (Section 5.2)

	(1)	(2)	(3)
	Negative	Positive	DD
NOREMINDER	-1.855 (0.265)	1.256 (0.505)	-1.961 (0.288)
Positive message			-0.855 (0.645)
NOREMINDER × Positive message			3.261 (0.198)
True rank	-0.558 (0.749)	-1.834 (0.425)	-1.093 (0.445)
Age	-0.014 (0.846)	0.068 (0.383)	0.034 (0.527)
Male	0.097 (0.953)	1.889 (0.329)	1.022 (0.424)
Constant	3.063 (0.370)	-1.272 (0.718)	1.124 (0.680)
Observations	208	237	445

**Note.** OLS regressions on the updating variable. For each model, the top row indicates the sample on which the model is estimated. Coefficients in bold are the ones commented on in the text. The results from the Difference-in-Difference estimation are provided in the last column, but not commented on in the text. These results are not explicitly reported in the text, but referred to in Footnote 19 *Legend*: \*\*\* 1%; \*\* 5%; \* 10%.