Protecting the Ego: Motivated Information Selection and Updating[*]

Alessandro Castagnetti[†]

Renke Schmacker[‡]

This draft: December 13, 2021

First draft: December 16, 2019

Abstract

We investigate whether individuals self-select feedback that allows them to maintain their motivated beliefs. In our lab experiment, subjects can choose the information structure that gives them feedback regarding their rank in the IQ distribution (ego-relevant treatment) or regarding a random number (control). Although beliefs are incentivized, individuals are less likely to select the most informative feedback in the ego-relevant treatment. Instead, many individuals select information structures in which negative feedback is less salient. When receiving negative feedback with lower salience subjects update their beliefs less, but only in the ego-relevant treatment and not in the control. Hence, our results suggest that individuals sort themselves into information structures that allow them to misinterpret negative feedback in a self-serving way. Consequently, subjects in the IQ treatment remain on average overconfident despite receiving feedback.

JEL-codes: D83, D91, C91

Keywords: motivated beliefs, updating, information acquisition, overconfidence

---

[*]We thank Robert Akerlof, Kai Barron, Roland Bénabou, Christine Exley, Guillaume Fréchette, Tobias König, Dorothea Kübler, Muriel Niederle, Kirill Pogorelskiy, Jan Potters, Eugenio Proto, Andy Schotter, Daniel Sgroi, Bertil Tungodden, Yves Le Yaouanq, Sevgi Yuksel, and seminar participants in Berlin, Vienna, and Warwick for helpful comments. The paper also benefited from discussions with conference participants at the EEA Annual Congress 2020, the Bogotá Experimental Economics Conference 2020, the Nordic Conference on Behavioral and Experimental Economics (Kiel), the European Meeting of the ESA (Dijon), the Dr@w Forum (Warwick), and the PhD Conference (Warwick). We started working on this project during our visit at the Center for Experimental Social Science at New York University, and we thank the Center for its hospitality. Financial support from the Warwick University PhD Conference Grant, the Deutsche Forschungsgemeinschaft through CRC TRR 190 (project number 280092119), and the Leibniz Competition through the project SAW-2015-DIW-4 is gratefully acknowledged.

[†]University of Warwick, `S.Castagnetti[at]warwick[dot]ac[dot]uk`

[‡]University of Lausanne, `renke.schmacker[at]unil[dot]ch`

1 Introduction

In many instances, individuals can choose the source from which they receive feedback about their ability. For example, politicians and executives can surround themselves with subordinates who give their honest opinion or loyalists who habitually praise the capabilities of their superior. In educational contexts, individuals select supervisors and mentors that differ in their feedback style. Similarly, college students select majors with more or less compressed grading distributions and thus differing informativeness of grades (Ahn *et al.*, 2019; Sabot and Wakeman-Linn, 1991).

Standard models predict that individuals have a strict preference for choosing the most informative feedback source because this leads to more accurate beliefs and, in turn, better-informed decisions. However, if individuals have motivated beliefs, for example the desire to hold the belief that they have high ability (Bénabou and Tirole, 2002; Köszegi, 2006), they may prefer feedback structures that allow them to preserve this positive self-view.[1]

In this paper, we study individuals' preferences over feedback concerning their ability. Our experiment resembles the situation where a student can choose between two supervisors: supervisor A and supervisor B. The student knows that both supervisors aim to give clear, positive feedback when the student's performance is good. However, the student also knows that supervisor A usually tells the student clearly when she considers the work poor, while supervisor B abstains from giving any comment when she is critical about the student's work. Hence, not receiving feedback from supervisor B is in fact negative feedback, but it is framed in a way that is not explicit. If the student anticipates that this framing will allow to disregard potential negative feedback, the student may prefer supervisor B, which can lead to the formation and maintenance of overconfident beliefs.

We conduct the first lab experiment in which individuals can select the information structure that provides noisy feedback about their ability. We compare information selection when feedback is related to relative IQ test performance to a control condition, in which feedback relates to the draw of a random number. Feedback is instrumental since individuals are incentivized for the accuracy of their beliefs. Nonetheless, we find that many subjects in the IQ treatment do not choose the most informative feedback structure. Instead, our results suggest that they choose feedback that allows them to maintain their belief that they have high ability.

We provide evidence for a mechanism of ego protection so far unrecognized in the literature: When feedback is ego-relevant, individuals choose information structures in which

[1] We are agnostic about whether individuals derive utility from holding themselves in high regard (Köszegi, 2006), whether they do so to gain a strategic advantage in persuading others (Von Hippel and Trivers, 2011), or to motivate themselves (Bénabou and Tirole, 2002). However, we note that holding inaccurate beliefs about ability is costly in many domains, for example, it leads to suboptimal management decisions (Malmendier and Tate, 2005) or over-entry into competition (Camerer and Lovallo, 1999).

negative feedback is less salient and, thus, potentially easier to misinterpret. We consider a signal to be less salient when it is framed in a neutral way (i.e., not linked to a positive or negative cue), although it carries either positive or negative information when its context is taken into account. In the experiment, a negative signal with high salience is presented as a red sign with the description "You are in the bottom half", while a negative signal with low salience is a grey sign with the description "...". Whereas the lower salience of feedback should not matter for a rational Bayesian updater, our results show that it leads individuals to update less in response to unpleasant news. When information is not ego-relevant, we do not find evidence for such asymmetric updating. Thus, the results are consistent with the explanation that individuals select feedback that allows them to distort their beliefs in a self-serving way.

The key feature of our design is that subjects can choose the information structure they receive feedback from. Information structures are presented in the form of two urns with varying compositions of positive and negative signals. Depending on whether the subject is in the top or bottom half of the distribution, signals are drawn from one urn or the other. This design allows us to vary the properties of the information structures and cleanly identify preferences for ego-relevant information. In particular, by varying the composition of signals in the urns, we vary the informativeness, salience, and skewness of feedback. *Informativeness* describes the noisiness of the signals. *Salience* refers to the way in which positive and negative signals are framed, holding informativeness constant. We vary the salience of feedback by framing it as either green/red signals with an explicit description (high salience) or grey signals without description (low salience). An information structure is *skewed* if positive signals are more or less informative than negative signals (Masatlioglu *et al.*, 2017; Nielsen, 2020). For example, an information structure is positively skewed if a potential positive signal is more precise than a negative signal.

The incentivization of beliefs gives us a clear prediction for subjects in the control condition: we expect that subjects select the most informative feedback structure in order to maximize their expected payoff. In contrast, subjects who derive utility from the belief that they rank high in the IQ distribution may prefer an information structure that allows to maintain this belief—for example, by selecting feedback that is less informative, positively skewed, or makes negative feedback less salient (which potentially facilitates future belief distortion).

We find that individuals seek different information when the rank is ego-relevant versus when it is not. When the ego is at stake, subjects are more likely to choose information structures that are less informative (treatment difference of 13.5 percentage points) and that make negative feedback less salient (treatment difference of 26.4 percentage points). However, we find no evidence that subjects in the ego-relevant treatment are more likely to choose information structures that are positively skewed. Our findings are based on aggregate treatment differences in information structure choices, and reinforced when looking at within-individual choice patterns (in the control, we classify 89.3 percent of subjects as information maximizers, and in IQ only 65.5 percent). Furthermore, we find that subjects who are classified as avoiding information according to the Information Preference Scale by Ho *et al.* (2021) are more likely to choose an information structure that is less informative and features less salient negative feedback. We do not find evidence for heterogenous information preferences by gender, cognitive ability, or prior beliefs.

Moreover, we find that the subsequent belief updating process is influenced by the information structure. Subjects in the IQ treatment react less to negative feedback than to positive feedback, but only when negative feedback is less salient. We find the first indication of this when subjects receive signals from the information structure that they self-select (endogenous treatment). The results are corroborated by a treatment in which subjects are exogenously placed into information structures to eliminate potential selection issues (exogenous treatment). Furthermore, in the control condition, subjects react similarly to positive and negative feedback irrespective of its salience. Therefore, asymmetric updating in the ego-relevant treatment is unlikely to be explained by difficulty in understanding less salient signals.

We assess explanations other than motivated reasoning for the treatment effects. The treatment variation in the ego-relevance of the state allows us to distinguish cognitive biases (i.e., general systematic errors regarding how people search and process new information, such as confirmation bias) from motivated biases (i.e., biases that are driven by a desire to hold positive views of oneself). We provide evidence that the treatment differences cannot be explained by cognitive biases such as confirmation-seeking or contradiction-seeking behavior, differences in cognitive ability, or confusion about the experimental design. Subjects' free-text explanations for their choices indicate that they make a "gut-level" decision to prefer positive-looking signals over explicit negative signals when feedback is ego-relevant.

Our results shed light on the conditions under which individuals with a desire to protect their ego can distort their beliefs. Bénabou and Tirole (2002) show in a two-selves model that it can be optimal for individuals to avoid or distort feedback when the belief in high ability has consumption or motivation value. However, since belief distortion in the model by Bénabou and Tirole (2002) is costly, manipulation of beliefs is only beneficial within the realms of the "reality constraints," implying that individuals cannot simply choose the beliefs they like. Our results suggest that selecting an information structures in which unpleasant feedback is easier to misinterpret can be one way to relax these reality constraints.

The idea that an individual selects an information structure that allows negative feedback to be interpreted in a self-serving way can be seen as a form of self-deception. It is a long-standing philosophical puzzle whether self-deception is possible at all, given that the same mind has to be the deceiver and the deceived (e.g., Mele, 2001). In the moral domain, Saccardo and Serra-Garcia (2020) find direct evidence for anticipated belief distortion. In their experiment, many advisors postpone acquiring the information whether the product they ought to recommend is beneficial for the advisee or not, thereby reducing the salience of the information. These subjects are also more likely to underrespond to the information when it is in conflict with their incentives and to behave ultimately in a self-serving way, showing that individuals are somewhat sophisticated about their future belief distortion. We find suggestive evidence that anticipated belief distortion also contributes to the maintenance of overconfident beliefs about intelligence. Our results show that subjects who select feedback that is less informative and in which negative feedback is less salient remain overconfident about their IQ rank over the course of the experiment. In contrast, subjects who receive explicit negative feedback are, on average, no longer overconfident at the end of the experiment.

Our findings contribute to the literature on how individuals process positive and negative feedback regarding their ego-relevant characteristics. Bénabou and Tirole (2016) argue that when ego-relevant beliefs are involved, people tend to process information differently depending on whether the information is more or less desirable.[2] For instance, people might tend to ignore or discount negative news, while more readily incorporating good news into their (posterior) beliefs.[3] However, the resulting experimental evidence on this mechanism—asymmetric updating—is mixed. On the one hand, Charness and Dave (2017), Eil and Rao (2011), and Möbius *et al.* (2014) find positive asymmetry in updating. On the other hand, a number of studies either find no asymmetry (Barron, 2021; Buser *et al.*, 2018; Gotthard-Real, 2017; Grossman and Owens, 2012; Schwardmann and Van der Weele, 2019) or even the opposite asymmetry (Coutts, 2019; Ertac, 2011; Kuhnen, 2015). Moreover, in a paper related to how errors in updating can be driven by motivated beliefs, Exley and Kessler (2019) find that people update their beliefs based on completely uninformative signals, but only when the signals carry positive information and the updating state is ego-relevant. Our paper helps explain the existing results by studying belief updating in different information structures. We show that asymmetric updating is only observed when the framing of the information structure allows subjects to interpret the signals in a self-serving way. In contrast, we do not observe asymmetric updating when positive and negative feedback is highly salient.

We complement the literature on information avoidance by giving subjects more control over the amount and type of feedback they receive.[4] Eil and Rao (2011) and Möbius *et al.*

[2] Related literature on motivated reasoning studies the demand for and consumption of political news. This literature shows evidence that consumers prefer like-minded news (Garz *et al.*, 2020; Gentzkow and Shapiro, 2010). Moreover, Chopra *et al.* (forthcoming) show in a series of online experiments that people's demand for political news goes beyond the desire for acquiring more informative news.

[3] Selective recall of ego-relevant feedback is documented in the experiments of Chew *et al.* (2020) and Zimmermann (2020). Both papers find that negative feedback on IQ test performance is more likely to be forgotten, compared to positive feedback.

[4] For a review of the information avoidance literature, see the survey of Golman *et al.* (2017). In the health domain, Ganguly and Tasoff (2016) and Oster *et al.* (2013) provide empirical evidence that people avoid medical testing. In a financial context, Karlsson *et al.* (2009) and Sicherman *et al.* (2015) show that investors check their portfolios less often when the market is falling.

(2014) present experimental evidence that a considerable proportion of subjects who have received prior noisy information regarding their relative rank in ego-relevant domains (intelligence and attractiveness) have a negative willingness to pay to have their rank fully revealed. In contrast, subjects in our study *choose* the noisy signals that they would like to receive before any feedback is given. On the one hand, we find that subjects indeed choose less informative feedback when it is ego-relevant. On the other hand, they disregard unpleasant information using a mechanism that is so far unexplored in the literature: subjects in the ego-relevant treatment choose information structures that makes unpleasant feedback less salient.

Our results add novel insights about motivated cognition to the literature on how signal structures affect belief updating. While Epstein and Halevy (2019) and Fryer *et al.* (2019) find that ambiguous signals increase deviations from Bayes rule, Enke (2020) and Jin *et al.* (2021) illustrate that individuals find it difficult to interpret the lack of a signal. Enke *et al.* (2020) provide evidence that individuals update more in response to signals, which are framed with valenced cues that activate associative memory. In contrast to these experiments, we vary the ego-relevance of the state and show that the framing of signals affects whether subjects update asymmetrically.

Finally, our research adds a motivated beliefs perspective to the literature on information structure selection. Within this literature, several papers study preferences about the timing and skewness of information disclosure in settings where information structures have—in contrast to our experiment—no instrumental value (Falk and Zimmermann, 2016; Nielsen, 2020; Zimmermann, 2014). For example, Masatlioglu *et al.* (2017) find that individuals have a preference for positively skewed information sources; that is, information structures that resolve more uncertainty regarding the desired outcome than the undesired one. Some experimental papers study preferences over information structures in settings where information has instrumental value (but is not ego-relevant). Charness *et al.* (2021) and Montanari and Nunnari (2019) study how people seek information from biased information structures and show that many individuals do not maximize the informativeness. Hoffman (2016) finds that businesspeople are on average too confident in their expert knowledge and underpay for instrumental information, possibly to avoid negative feedback. We show that the tendency to select less informative feedback is more pronounced in ego-relevant domains and can be one driver for the prevalence of overconfident beliefs.

The remainder of this paper is organized as follows. In Section 2, we describe our experimental design, which comprises two treatment variations: ego-relevance of the rank and endogenous/exogenous information structure allocation. In Section 3, we present our experimental results. First, we study how participants select their preferred information structures depending on the ego-relevance of the rank. Second, we study subsequent belief updating. In Section 4, we discuss our findings, and in Section 5 we conclude.

2 Experimental Design

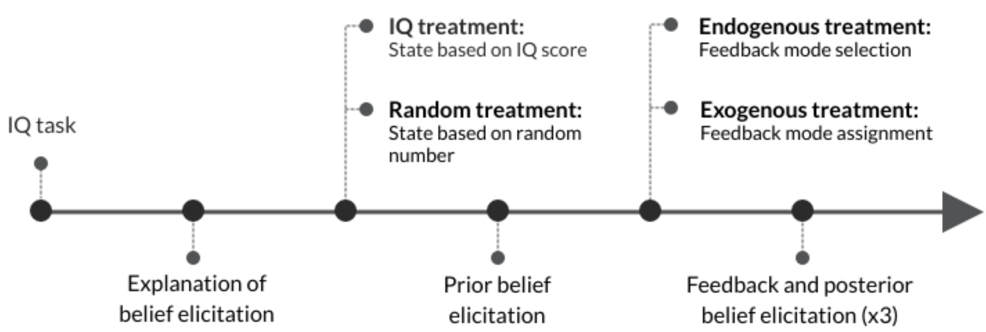
To investigate whether individuals choose information structures that protect their ego, we design an experiment that contains (1) exogenous variation in the ego-relevance of beliefs; (2) choices between different information structures; and (3) elicitation of updating behavior within different information structures.

In a 2x2 between-subject design, we vary, on the one hand, whether subjects receive feedback about their relative rank in IQ test performance (IQ treatment) or about a random number (random treatment). This treatment variation allows us to study the impact of motivated beliefs on preferences over feedback. On the other hand, we vary whether subjects receive signals from the information structure they selected into (endogenous treatment) or from an information structure they are exogenously assigned to (exogenous treatment). The latter treatment variation allows us to study how subjects update their beliefs with and without self-selection into feedback.

2.1 Overview

Figure 1 presents an overview of the experiment. In the beginning, all subjects perform an incentivized IQ test. They have 10 minutes to solve 20 matrices from the Raven Advanced Progressive Matrices (APM) test. They can earn £2.00 per correct answer out of three randomly chosen matrices. Although the IQ performance is only relevant for subjects in the IQ treatment in the later stages of the experiment, all subjects solve the IQ quiz. This ensures that there are no systematic differences in fatigue, timing, or average earnings between treatments.

Figure 1: Timeline of the Experiment



In the remainder of the experiment, subjects are asked to express their beliefs about the state of the world, which is either related to their rank in the IQ test (IQ treatment) or to the draw of a random number (random treatment). First, we explain the matching probabilities method to the subjects (Karni, 2009), so that they know that they maximize their chance of winning a £6.00 prize by stating their true beliefs (see Appendix B). Next, subjects are informed whether the belief questions refer to their IQ score or their random number and are asked for their prior beliefs. Afterwards, subjects in the endogenous treatment choose the feedback mode they would like to receive signals from, while subjects in the exogenous treatment are shown the feedback mode assigned to them. Feedback modes are presented in the form of urns with a varying composition of signals, as explained below. Finally, three rounds of feedback are drawn from the respective mode. We elicit subjects' posterior beliefs, allowing us to study how subjects update their beliefs in response to receiving signals.

At the end of the experiment, either the IQ task or the belief elicitation is randomly selected for payout. If the belief elicitation is chosen, one out of the four beliefs (one prior and three posteriors) is randomly selected for payout. Thereby, we aim to exclude any hedging motives that could lead to a reported belief about IQ performance that differs from the true belief (cf. Blanco *et al.*, 2010; Azrieli *et al.*, 2018). Screenshots of the experimental instructions are provided in Appendix H.

Subjects are incentivized to give their true beliefs, so a payoff-maximizing subject would always choose the most informative feedback mode and update according to Bayes rule. However, in the IQ treatment, subjects' motives to maximize payout can conflict with the desire to protect their beliefs about their (relative) ability. For instance, subjects may be willing to accept a smaller expected payoff in order to not impair their belief that they are in the top of the intelligence distribution. If that is the case, we expect a treatment difference in information structure selection and/or updating behavior depending on whether the beliefs are ego-relevant.

2.2 Treatments

2.2.1 IQ and Random Treatment

We vary the ego-relevance of beliefs by randomizing subjects into an IQ treatment and a random treatment at the session level. Consistent with previous research, we argue that rank in the IQ treatment is ego-relevant (e.g., Eil and Rao, 2011), and that subjects care more for their IQ rank than their random number. We assume that further deviations from the rational benchmark are constant between the IQ and random treatments (this assumption is discussed in Section 4).

IQ Treatment In the IQ treatment, we inform subjects that the second part of the experiment is related to their relative performance in the IQ test they completed in the beginning. We tell them that the computer divides participants in their session into two groups: one group of subjects whose score is in the top half of the score distribution and the other with scores in the bottom half. The task is to assess whether their IQ performance is in the top or bottom half of the distribution, compared to all other participants in their session. To increase the ego-relevance of the IQ treatment (Drobner and Goerg, 2021), we explicitly tell subjects that the APM test is commonly used to measure fluid intelligence and that high scores in this test are regarded as a good predictor for academic and professional success, occupation, income, health, and longevity (Sternberg *et al.*, 2001; Gottfredson and Deary, 2004).

Random Treatment In the random treatment, subjects are shown a randomly drawn number between 1 and 100. We tell subjects that three other numbers between 1 and 100 (with replacement) were drawn. These numbers are not shown to them. Their task is to assess whether the number they are given is in the top or bottom half of the distribution among these four numbers. The four numbers are randomized at the individual level and ties are broken randomly. The task is deliberately designed to generate variation in prior beliefs.[5] If subjects were not given the drawn number or if we compared their number against many numbers, we would have expected a degenerate prior distribution instead.

In Figure A.1, we plot the prior distribution by treatment. As expected, the prior distribution in the random treatment is centered around 50 percent, while priors in the IQ treatment are left-skewed with more mass in the top half of the distribution. Hence, in the analysis, we include priors as a covariate and use a matching strategy to ensure that the differences in priors do not drive differences in feedback choice.

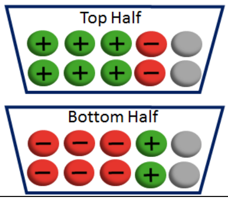
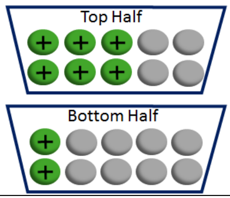
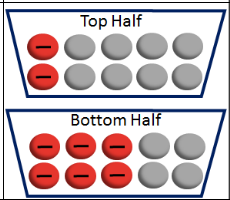
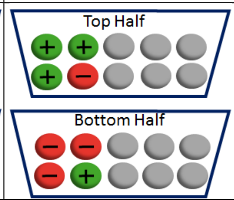
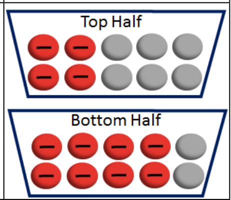
2.2.2 Endogenous and Exogenous Treatment

Moreover, we vary whether subjects endogenously select or are exogenously assigned a feedback mode. The rationale for the exogenous treatment is that it allows us to study updating behavior absent self-selection. In contrast, in the endogenous treatment, subjects in different feedback modes have, on average, different preferences over information structures, which could affect updating behavior.

Endogenous Treatment In the endogenous treatment, subjects make five pairwise choices between feedback modes. Figures H.16 to H.18 illustrate the choice situations. In each of the five choice situations, we vary the properties of the feedback modes, as explained in detail in the next section. Afterwards, one out of the five feedback mode choices is randomly selected

[5] In the endogenous treatment, the standard deviation of prior beliefs in the random treatment turns out to be 25.065, compared to 19.993 in the IQ treatment. Hence, the random treatment generates similar variance compared to the IQ treatment.

Table 1: Feedback Modes

Mode A	Mode B	Mode C	Mode D	Mode E
				
LR(Top|Green)=3	LR(Top|Green)=3	LR(Top|Grey)=2	LR(Top|Green)=3	LR(Top|Grey)=3
LR(Top|Red)=1/3	LR(Top|Grey)=1/2	LR(Top|Red)=1/3	LR(Top|Red)=1/3	LR(Top|Red)=1/2
Prob(No Info)=1/5	Prob(No Info)=0	Prob(No Info)=0	Prob(No Info)=3/5	Prob(No Info)=0

Notes: The table shows the feedback modes that can be selected in the experiment. Depending on the state (top or bottom half), a signal is drawn from the upper or lower urn. LR(State|Signal) describes the likelihood ratio of the signal concerning the state and is a measure for its informativeness. For example, LR(Top|Green(+)) is the likelihood of receiving a green (+) signal when being in the top half divided by the likelihood of receiving a green (+) signal when in the bottom half. Prob(No Info) describes the probability of receiving a non-informative signal.

and the choice of the subject is implemented. Before receiving signals, each subject is shown the selected feedback mode from which they would receive the signals.

Exogenous Treatment In the exogenous treatment, subjects are not asked to choose a feedback mode—instead, assignments are exogenous. In particular, following the IQ test, subjects are randomly allocated to receive ego or non-ego relevant feedback from one of the feedback modes.

2.3 Feedback Modes

Table 1 shows the information structures in the experiment. Information structures consist of two urns with 10 balls each. A ball drawn from an urn in the selected information structure constitutes a signal. If an individual's IQ score or random number is in the top (bottom) half of the distribution, balls are drawn from the upper (lower) urn with replacement. This design allows us to cleanly vary the properties of the information structure in a way that is transparent to the subjects.

Depending on the information structure, subjects can receive up to three different types of (noisy) signals. Subjects in the IQ (random) treatment can either receive a green signal with a plus (+) sign and the description "You are in the top half" ("Your number is in the top half"), a red signal with a minus (-) sign and the description "You are in the bottom half" ("Your number is in the bottom half"), or a grey signal with the description "...". Figure H.8 shows how the signals are introduced in the instructions.

On the same page, we explain that the informational content of the respective signal depends on the feedback mode and the state. For example, subjects are told that in feedback mode A, they are more likely to get the green (+) signal if they are in the top half of the distribution and that they are more likely to get the red (-) signal if they are in the bottom half of the distribution.[6] Subjects have to correctly answer comprehension questions about these features before they can proceed. In all feedback modes, the green (+) signal increases the posterior that one is in the top half and the red (-) signal increases the posterior that one is in the bottom half. Depending on the feedback mode, the grey signal can be positive, negative, or non-informative feedback.

Information structures differ in their informativeness, skewness, and framing. Informativeness describes how noisy a feedback mode is. Note that a perfectly informative feedback structure could feature only green (+) signals in the top urn and only red (-) signal in the bottom urn. We introduce noise by, for example, adding red (-) signals to the top urn or uninformative grey signals to either urn. The informativeness of signals in our experiment can be described first by their likelihood ratio (LR) and second by the probability of receiving a non-informative signal. Both of these properties are given in the bottom panel of Table 1. The further away the likelihood ratio is from unity, the more informative the signal and the more it shifts the posterior belief of a Bayesian updater (e.g., the negative signal in Mode A is more informative than the negative signal in Mode B).[7] The probability of receiving a non-informative signal only applies to Modes A and D, in which grey signals are not informative (hence, Mode A is more informative than D).

We call an information structure positively skewed if the positive signals are more informative in terms of their likelihood ratio than the negative signals (as in Modes B and E) and negatively skewed if the negative signals are more informative (as in Mode C).

Finally, information structures differ in their feedback salience. Since the color grey is typically not associated with positive or negative states and the description is not explicit, we say that positive or negative feedback in the form of grey signals is *less salient*. For example, in Mode B negative feedback is less salient, while in Modes C and E positive feedback is less salient.

2.4 Feedback Mode Choices

Subjects make five pairwise choices between information structures. By carefully varying the information structures they can choose from, we are able to elicit whether subjects have preferences for informativeness, salience, or skewness of feedback depending on its ego-relevance.

[6] In the endogenous treatment, we explain these characteristics by always using one feedback mode choice as an example. We use three different examples, as illustrated in Figures H.9 to H.11, and check if the presented example matters for choices in Appendix D. In the exogenous treatment, we explain the signals using the participant's assigned feedback mode: see the screenshot in Figure H.12.

[7] In fact, a likelihood ratio of one implies that the signal is fully uninformative about the underlying state.

Every subject makes all five choices. To control for order effects, we vary the order of information structure choices (cf. Appendix D).

Baseline Choice: Mode A vs Mode B In the baseline choice, subjects choose between two feedback modes that vary in informativeness, skewness, and salience. First, Mode A is more informative than Mode B. Second, while Mode A gives balanced positive and negative feedback depending on the state, Mode B is positively skewed and negative feedback is less salient (i.e., negative feedback is framed as grey signals). Hence, if more subjects choose Mode B in the IQ treatment compared to the random treatment, we can interpret this as evidence that subjects protect their ego by choosing an information structure that gives less informative and positively skewed signals and where negative feedback is less salient.

Using the remaining feedback mode choices, we aim to disentangle the underlying preferences for informativeness, salience, and skewness.

Informativeness Choice: Mode A vs Mode D First, individuals might want to avoid information if it is ego-relevant. The choice between Modes A and D isolates a preference for informativeness. In both feedback modes, the skewness and salience of signals is held constant, only the probability of receiving a completely uninformative (grey) signal varies. Hence, subjects who have a preference for avoiding information prefer Mode D over Mode A.

Salience Choice: Mode B vs Mode E Second, individuals could have a preference for the salience of feedback, for example, for reducing the salience of negative feedback if information is ego-relevant. This could be due to an aversion to explicit negative feedback, or due to anticipation of differential updating behavior (cf. results of updating in Section 3.2). To test for salience preferences, we let subjects choose between Modes B and E, which have the same informativeness and skewness but only differ in the salience of positive and negative feedback (in Mode B negative feedback is less salient and in Mode E positive feedback is less salient).

Skewness over Salience Choice: Mode A vs Mode E Third, we investigate whether individuals have a preference for positive skewness. We test individuals' prefererences for positive skewness relative to their preference for feedback salience. Therefore, we consider the choice between Modes A and E together with the baseline choice between Modes A and B. If subjects' choices are driven by a preference for positive skewness, they would prefer both Mode E and Mode B over Mode A.

Baseline Reversed Choice: Mode A vs Mode C Finally, we also check if individuals have a preference for a negatively skewed information structure with less salient positive feedback (Mode C).

2.5 Updating Behavior

Besides information structure selection, we also analyze the updating behavior of subjects and how it interacts with the feedback mode. During the updating stage, subjects receive three consecutive signals from one of the feedback modes. After each signal, subjects are asked to report their posterior beliefs. Further, each time they receive a signal and are asked about their beliefs, subjects can view a picture of the feedback mode urns from which they receive information by clicking a button (see Figure H.19 for a screenshot of the choice situation).

Our main interest is to understand differences in updating across feedback modes and according to the ego-relevance of the task. For this reason, we specifically focus on Modes A and B and their interaction with the ego-relevance of the task. On the one hand, a comparison in updating behavior across Modes A and B in the random treatment will allow us to understand whether differences in the information structure drive cognitive biases (i.e., general deviations from Bayes rule). On the other hand, a comparison of updating across IQ and random treatment will allow us to study the extent of motivated biases in updating (i.e., deviations from Bayes rule specific to ego-relevant feedback).

2.6 Debriefing

In the last part of the experiment, we ask subjects a battery of questions. First, we ask subjects to complete the Information Preferences Scale by Ho *et al.* (2021), which is a 13-item questionnaire that measures an individual's desire to obtain or avoid information that has an instrumental value but is also unpleasant. The scale measures information preferences in three domains: consumer finance, personal characteristics, and health. Second, we ask subjects to complete the Gneezy and Potters (1997) risk elicitation task. Specifically, each subject receives £1.00 and has to decide how much of this endowment to invest in a risky project with a known probability of success. The risky project returns 2.5 times the amount invested with a probability of one-half and nothing with the same probability. We also ask them a non-incentivized general willingness to take risks question (Dohmen *et al.*, 2011). Third, we ask subjects to answer two questions in free-form text and they receive £0.50 for their answers. We ask them to advise a hypothetical subject who would be performing the feedback mode choices and updating task. In the endogenous treatment, we additionally ask them to explain their motives for choosing the feedback modes across the five choice situations. Finally, we ask subjects a series of demographic questions including age, gender, and nationality.

2.7 Experimental Procedure

The experimental sessions were conducted from June to October 2019 in the Economics Laboratory of Warwick University, United Kingdom. Overall, we recruited 445 subjects through the Sona recruitment system to take part in the experiment. We conducted 14 sessions (216 subjects) for the endogenous treatment and 15 sessions (229 subjects) for the exogenous treatment. Sessions lasted an average of 60 minutes. Participants earned an average payment of £11.00, including the show-up fee of £5.00. We conducted the experiment using oTree (Chen *et al.*, 2016). Descriptive statistics of the sample are provided in Table A.1.

In each session, subjects were randomly assigned a cubicle and general instructions were read aloud. The remaining instructions were provided onscreen. In both the endogenous and exogenous sessions, it was randomly determined whether the cubicle belonged to the IQ or random treatment. Moreover, in the exogenous treatment, it was randomly determined if the cubicle was allocated to Modes A or B.

3 Results

Our analysis proceeds in two steps: First, we investigate treatment differences between IQ and random in feedback mode choices. Second, we analyze how subjects update in response to signals from the feedback modes.

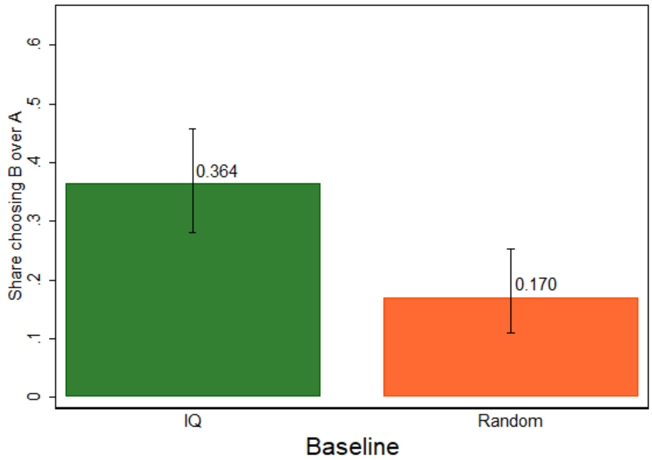
3.1 Information Selection

We first focus on the baseline choice between Modes A and B that vary in informativeness, salience, and skewness. While Mode A gives balanced feedback, Mode B gives less informative, positively skewed signals with less salient negative feedback. Hence, choosing Mode B is costly because subjects are paid based on the accuracy of their posterior beliefs and Mode B provides less information.

Figure 2 illustrates the percentage of subjects who prefer to receive signals from Mode B over Mode A. While only 17.0 percent of subjects in the random treatment choose Mode B over Mode A, 36.4 percent in the IQ treatment prefer Mode B. The difference of 19.4 percentage points is statistically significant (t-test, $p = 0.001$; Wilcoxon rank-sum test, $p = 0.001$).

In order to disentangle preferences for informativeness, salience, and skewness, we elicit subjects' preferences over information structures in four additional choice situations. Figure 3 plots the results. In the informativeness choice, subjects can choose between Mode A and Mode D, where both give balanced feedback but where Mode D is less informative than Mode A. The top left panel of Figure 3 shows that a higher proportion of subjects in IQ choose the less informative Mode D over Mode A. The difference of 13.5 percentage points is statistically significant (t-test, $p = 0.002$; Wilcoxon rank-sum test, $p = 0.002$). Hence, the results suggest

Figure 2: Share choosing Mode B over A in baseline choice



Notes: The plot shows the fraction of subjects who prefer Mode A over B in the baseline choice by treatment. In contrast to Mode A, Mode B is less informative, positively skewed, and makes negative feedback less explicit. The 95% confidence intervals (Wilson) are shown by the bar. In the IQ treatment, there are N=110 subjects and in random treatment N=106.

that subjects in the IQ treatment do indeed have a preference for less information compared to subjects in the random treatment.

In the salience choice, subjects choose between Mode B, in which negative feedback is less salient, and Mode E, in which positive feedback is less salient. We find that, in the IQ treatment, significantly more subjects exhibit a preference for less salient negative feedback, compared to the random treatment (difference of 26.5 percentage points, t-test, $p < 0.001$; Wilcoxon rank-sum test, $p < 0.001$). Since the informativeness and skewness of the feedback modes are held constant, we expect subjects in the random treatment to be indifferent. Indeed, the share of 52.8 percent choosing Mode E in the random treatment is not significantly different from 50 percent (t-test, $p = 0.563$). In contrast, in the IQ treatment only 26.4 percent choose Mode E, which is significantly lower than the 50 percent predicted by indifference (t-test, $p < 0.001$). Hence, we infer that people care about the salience of signals when it concerns ego-relevant information.

In the skewness over salience choice, we give subjects the choice between Mode A and Mode E. Mode E— just like Mode B—gives positively skewed information, but gives explicit negative feedback and less salient positive feedback. We find that fewer subjects prefer Mode E over Mode A in the IQ treatment than in the random treatment. Taken together with our findings from the baseline choice, we conclude that subjects' preference for Mode B in the IQ treatment is not driven by a preference for positive skewness. The difference in framing of Mode E is enough to overturn the treatment difference from the baseline choice, which suggests that the preference for positive skewness is not as strong as the preference against explicit negative signals.[8]

Finally, in the baseline reversed choice, we let subjects decide between Mode A and Mode C. Mode C gives less informative, negatively skewed signals with less salient positive feedback. In contrast to the baseline choice, we do not find a statistically significant difference between treatments with fewer subjects in IQ choosing Mode C (difference of 5.2 percentage points, t-test, $p = 0.299$; Wilcoxon rank-sum test, $p = 0.298$). This result suggests that subjects do not prefer less informative feedback modes in the IQ treatment if the feedback mode is negatively skewed and makes positive feedback less salient.

Robustness In Table A.2 in the appendix, we regress the five feedback mode choices on the IQ treatment dummy and various control variables. Controlling for demographics, prior beliefs, IQ scores and risk preferences does not alter the results in terms of treatment differences in feedback mode choices. In Table A.3, we perform a nearest-neighbor matching (with replacement) with respect to priors. By matching subjects in IQ to subjects with similar priors in the random treatment, we control for differences in the prior distributions non-parametrically. However, the treatment effects turn out to be very similar compared to the main analysis.

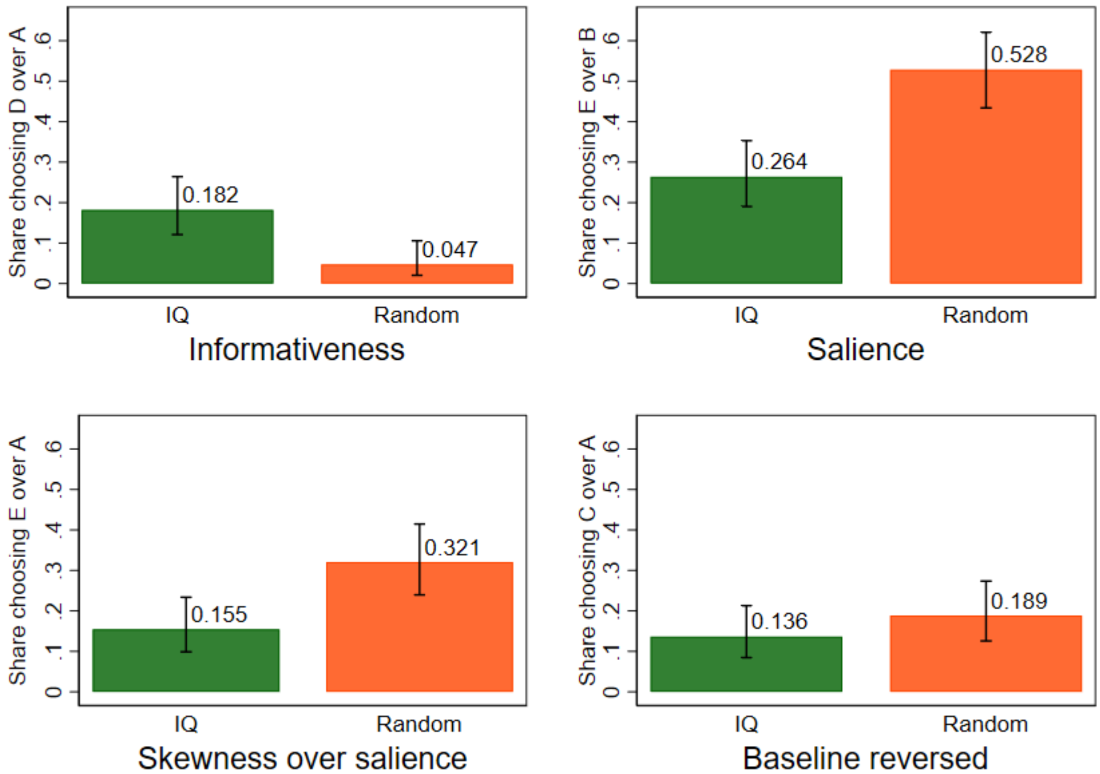
In the main analysis, we test treatment differences in five information structure choices. We account for multiple hypothesis testing in Table A.4 in the appendix. To control for the family-wise error rate, we calculate p-values based on the procedure by List *et al.* (2019) and report p-values using Holm (1979) and Bonferroni adjustment. None of the adjustments change our assessment of the effects' statistical significance.

3.1.1 Information Selection—Within Individual

So far we have looked at aggregate treatment differences in separate choices and analyzed what we can learn from them. We now examine the within-subject choice patterns. To perform this within-analysis, we suppose that each subject has a fixed preference over information structures (conditional on the treatment) and chooses information structures according to this preference. In line with our experimental design, we focus on three preferences for information structures: maximize the informativeness, seek positive skewness, and reduce

[8] Although the informativeness of Modes B and E are the same, we find different levels in the random treatment for the baseline choice and the skewness over salience choice. In Appendix D, we find evidence that this could be because some subjects do not understand without further explanation which feedback mode is more informative. When we explain that Mode B and Mode E are less informative than Mode A, the levels in the random treatment are very similar.

Figure 3: Selection of feedback mode by choice situation



Notes: The plot shows the fraction of subjects who prefer one feedback mode over the other in the respective choice by treatment. Informativeness: Mode D is less informative than Mode A. Salience: Mode E makes negative feedback less salient than Mode B. Skewness over salience: Mode E is positively skewed and less informative than Mode A, but makes positive feedback less salient. Baseline reversed: Mode C is negatively skewed, less informative than Mode A, and makes positive feedback less salient. The 95% confidence intervals (Wilson) are shown by the bar. In the IQ treatment, there are N=110 subjects and in random treatment N=106.

salience of negative feedback/increase salience of positive feedback. We estimate the fraction of subjects who consistently choose information structures that conform to these preferences.[9]

First, we calculate the fraction of subjects who make choices according to each of these preferences and the fraction of subjects who make choices that do not conform to one of these preferences. Second, we allow subjects to make mistakes and estimate which of the preferences can best explain subjects' choice patterns using a finite mixture model. Hence, we estimate the share of preference types and the amount of implementation noise (γ) necessary to classify subjects. The estimation strategy is described in detail in Appendix C.

In Table 2, we compare the relative prevalence of the implied preferences by treatment. In the first two columns, we present the empirically observed fraction of subjects who adhere to a given preference when they are not allowed to make mistakes. In the third and fourth columns, we show the estimated fractions using the finite mixture model. Note that 26.4 percent of subjects in the IQ and 36.8 percent in the random treatment are not classified if we do not allow for mistakes. In contrast, in the finite mixture model we use maximum likelihood to assign to every subject the preference that describes her choice pattern best.

First, consider the strategy to maximize the informativeness of information structures. There are fewer subjects in the IQ treatment who consistently maximize the informativeness than in the random treatment. When using the finite mixture model, the share increases from 40.9 to 65.5 percent in IQ and from 53.8 to 89.3 percent in the random treatment, suggesting that many subjects aim to maximize the informativeness of signals but make mistakes.[10] The treatment difference of 23.8 percent is statistically significant.

Second, there are significantly more subjects in the IQ than in the random treatment who exhibit a preference for reducing the salience of negative feedback but not of positive feedback. While more than 30 percent of subjects follow such a preference in IQ, there are few to none who are categorized as such in the random treatment. For the salience of feedback, we find the largest treatment difference, at a statistically significant 34.5 percentage points.

Finally, there are only a few subjects who consistently choose feedback modes that are positively skewed. In particular, there are no subjects in the IQ and 3.8 percent of subjects in the random treatment. In the finite mixture model, the share in the random treatment increases to 10.7 percent. However, note that it only requires three consistent choices to be attributed to this preference (in contrast to four for the other preferences). Overall, the results do not suggest that subjects choose positively skewed feedback to protect their ego.

[9] "Maximum information" predicts that subjects make choices according to $A \succ B, C, D, E$, "Positive skewness" predicts $B \succ A; E \succ A; A \succ C$, and "Salience of feedback" predicts $B \succ A; A \succ C, E; B \succ E$. Note that subjects can follow more than one preference but we assume that they have one dominant preference when these preferences conflict.

[10] In Appendix D, we exploit the order in which feedback modes are presented and find suggestive evidence that the difference is, to a large degree, driven by subjects who do not understand that Mode E reveals less information than Mode A.

Table 2: Share of subjects revealing a consistent preference by treatment

Preferences	No Mistakes		Maximum Likelihood		
	IQ	Random	IQ	Random	Difference
Maximum information	0.409	0.538	0.655***	0.893***	-0.238***
			(0.059)	(0.043)	(0.072)
Salience of feedback	0.327	0.057	0.345***	0.000	0.345***
			(0.059)	(0.001)	(0.059)
Positive skewness	0.000	0.038	0.000	0.107	-0.107**
					(0.045)
Variance of error term (γ)			0.492***	0.585***	
			(0.046)	(0.050)	
Not classified	0.264	0.368			
N	110	106	110	106	

Notes: The table shows the share of subjects who choose feedback modes consistent with the respective preference. Maximum information prescribes $A \succ B, C, D, E$. Positive skewness prescribes $B \succ A; A \succ C; E \succ A$. Salience of feedback means to seek explicit positive feedback but avoid explicit negative feedback and prescribes $B \succ A; A \succ C, E; B \succ E$. In No Mistakes, we calculate the share without allowing for implementation mistakes. In Maximum Likelihood, we estimate the share allowing for implementation noise γ. The share of "Positive skewness" is implied since the shares have to sum to 1. Standard errors in parentheses are bootstrapped with 1,000 replications. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

To sum up, the within-individual choices support our findings from the individual information structure choices. When looking at internally consistent choice patterns, subjects in IQ prefer less informative feedback modes as well as modes in which positive feedback is explicit but negative feedback less salient. Moreover, we find few subjects who have a preference for positive skewness but, if anything, the share is higher in the random than in the IQ treatment.

3.1.2 Heterogeneity in Information Structure Selection

We investigate heterogeneity in information structure selection based on self-reported information preferences, gender, prior beliefs, and performance in the IQ quiz. We focus on the baseline choice as it combines all three channels of ego protection: informativeness, skewness, and salience.

Information Preference Scale In the post-experimental questionnaire, subjects are asked to answer the Information Preference Scale (IPS) by Ho *et al.* (2021).[11] The scale consists of 13 scenarios from different domains (health, consumer finance, personal life) in which an individual can receive potentially unpleasant information (the items are shown in Appendix E).

[11] Ho *et al.* (2021) design and validate the Information Preference Scale in order to measure an individual's trait to obtain or avoid information. They show that it correlates strongly with related scales and that it even predicts information avoidance in the political domain, a domain not represented in the scale itself.

Table 3: Heterogeneity in baseline choice

	(1) IPS Scale	(2) Gender	(3) Prior Beliefs	(4) IQ Score
IQ treatment	0.323***	0.243**	0.196**	0.182**
	(0.083)	(0.107)	(0.073)	(0.091)
IPS (Info seeking)	0.064			
	(0.074)			
IPS (Info seeking) × IQ treatment	-0.259**			
	(0.117)			
Female		-0.121		
		(0.079)		
IQ treatment × Female		-0.064		
		(0.127)		
Low prior			-0.060	
			(0.073)	
IQ treatment × Low prior			-0.028	
			(0.125)	
Low IQ				0.018
				(0.074)
IQ treatment × Low IQ				0.020
				(0.120)
Constant	0.140***	0.244***	0.191***	0.159***
	(0.046)	(0.068)	(0.048)	(0.056)
R2	0.075	0.076	0.054	0.049
N	216	216	216	216

Notes: The table shows heterogenous treatment effects of the IQ treatment on the choice of Mode B in the baseline choice by IPS Scale, Gender, Prior beliefs, and IQ score. IPS (Info seeking) is an indicator for subjects who score in the top half of the Information Preference Scale (Ho *et al.*, 2021). Low prior indicates if a subject reports a prior below 50 and Low IQ indicates if a subject has not scored more than the median in the IQ quiz. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The respondent has to indicate her preference on a 4-point scale from "Definitely don't want to know" to "Definitely want to know."

In the first column of Table 3, we regress the choice of Mode B in the baseline choice on the IQ treatment indicator, an indicator if a subject scores above the median in the IPS scale, and an interaction of the two variables. The statistically significant interaction term implies that individuals, who are information seeking according to the IPS scale, are less likely to avoid information and choose less salient negative feedback in the IQ treatment. Moreover, the IPS scale is not associated with information structure choice in the random treatment, as illustrated by the small and statistically insignificant main coefficient of the IPS variable.

Gender In the second column of Table 3, we regress the choice of Mode B in the baseline choice on the IQ treatment indicator interacted with a dummy variable that indicates whether a subject is female. However, since the interaction term is small and far from statistically significant, we conclude that there is no evidence for heterogeneous treatment effects by gender in the experiment.

Prior Beliefs In the third Column of Table 3, we investigate whether the effect of the ego-relevant treatment is different depending on the reported prior beliefs. "Low Prior" indicates that an individual reports a prior that is lower than 50%, while the reference group reports a prior above or equal to 50%. We do not observe that subjects with priors below 50% are affected differently by the treatment compared to individuals with priors above 50%.[12]

IQ Score Finally, in the last column of Table 3, we analyze whether there is a differential treatment effect for subjects who performed better or worse in the IQ quiz. "Low IQ" indicates that a subject has correctly solved the same number or fewer of the Raven matrices than the median (12). The statistically insignificant and small interaction term suggests that there is no differential treatment effect depending on the performance in the IQ task (i.e., their measured cognitive ability).

3.1.3 Information Selection and Overconfidence

After subjects made the five information structure choices, one of these choices was randomly selected and the corresponding decision of the subject was implemented. Before analyzing belief updating in detail, we check descriptively how the selected information structure relates to the development of overconfident beliefs about the IQ rank.

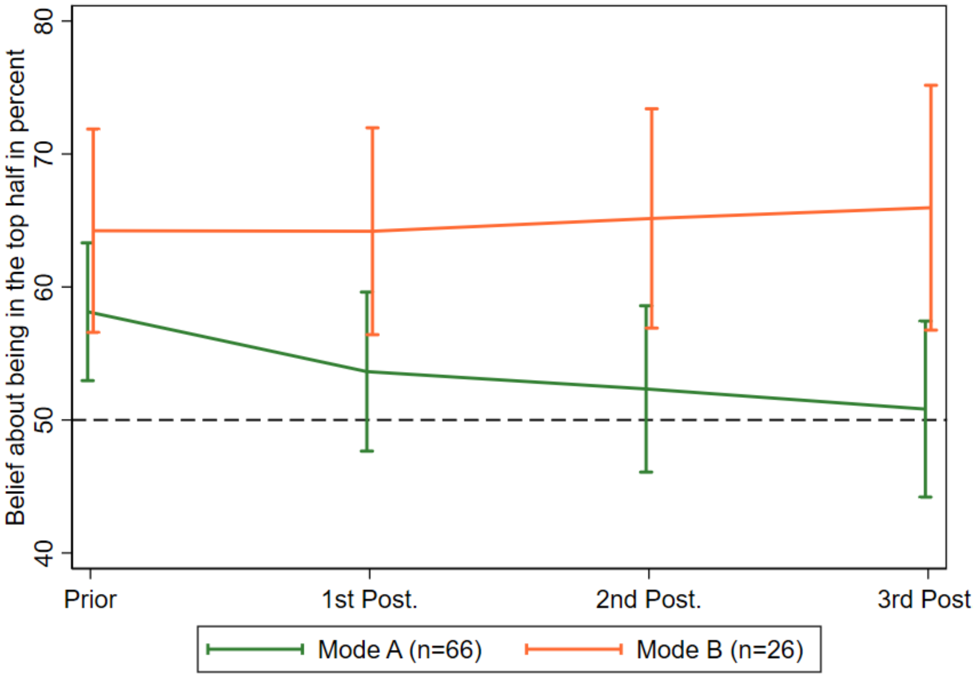
In Figure 4, we plot how the average beliefs in the IQ treatment evolve, depending on the feedback mode from which subjects receive signals. Although the sample becomes quite small when conditioning on the selected feedback mode (n=66 in Mode A and n=26 in Mode B), some interesting patterns emerge. We observe that in the IQ treatment, subjects are *overconfident* in their prior beliefs: on average, they report a likelihood higher than 50% of being in the top half of the IQ distribution, both in Mode A (t-test, $p = 0.003$) and in Mode B (t-test, $p = 0.001$).[13] Subjects in Mode A and Mode B start out with similar priors, but while the average beliefs of subjects in Mode A seem to converge toward 50% after receiving signals, the beliefs in Mode B remain constant.[14]

[12] Although not statistically significant, the main coefficient of the prior is negative, which is consistent with confirmation-seeking behavior. In Section 4, we discuss confirmation-seeking behavior in more detail.

[13] Benoît *et al.* (2015) show that true overconfidence is observed when the average stated probability to be in the top 50% of the distribution is significantly larger than 50%.

[14] In fact, after three rounds of feedback, subjects in Mode A have an average belief that is statistically indistinguishable from 50% (t-test, $p = 0.806$), while subjects in Mode B have an average belief that is still

Figure 4: Beliefs before and after signals by feedback mode (endogenous/IQ treatment)



Notes: The plot shows the average prior and posterior beliefs from the endogenous/IQ treatment for the selected feedback mode. Subjects are overconfident, if the average belief about being in the top half is larger than 50 percent. The whiskers represent 95% confidence intervals.

These results suggest that selecting an information structure that is less informative and makes negative feedback less salient indeed leads to maintaining overconfident beliefs in the IQ treatment. Moreover, in Appendix G, we show that this pattern is not observed in the random treatment. These descriptive patterns motivate us to investigate belief updating by information structure in more detail in the next section.

3.2 Belief Updating

We aim to investigate how subjects process the signals they receive from different feedback modes. First, we introduce the estimation framework for analyzing potential deviations from Bayesian updating. Then, we analyze updating in both the endogenous and exogenous treatment. While subjects in the endogenous treatment receive signals from the self-selected feedback mode, subjects in the exogenous treatment are allocated to a feedback mode.

3.2.1 Estimation Framework

We follow the approach developed by Grether (1980) and Möbius *et al.* (2014) to estimate updating behavior. The framework allows individuals to put different weights on the prior and the positive or negative signals they may receive, nesting the Bayesian benchmark as a special case. In the case of binary signals, Bayes rule can be written in the following form:

$$(1) \qquad logit(\mu_t) = logit(\mu_{t-1}) + \mathbb{1}(s_t = pos)ln(LR_{pos}) + \mathbb{1}(s_t = neg)ln(LR_{neg})$$

where μ_t is the belief at time t and LR_k is the likelihood ratio of the signal $s_t = k \in \{pos, neg\}$.

To estimate the model, we add an error term and attach coefficients to the prior and to the positive and negative signals an individual receives:

$$(2) \quad logit(\mu_{it}) = \delta^{\text{prior}} logit(\mu_{i,t-1}) + \beta^{\text{pos}}\mathbb{1}(s_{it} = pos)ln(LR_{pos}) + \beta^{\text{neg}}\mathbb{1}(s_{it} = neg)ln(LR_{neg}) + \epsilon_{it}$$

where δ^{prior} captures the weight of the prior while β^{pos} and β^{neg} measure the responsiveness to positive and negative signals, respectively. ϵ_{it} captures non-systematic errors in updating. A Bayesian updater would exhibit $\delta^{\text{prior}} = \beta^{\text{pos}} = \beta^{\text{neg}} = 1$.

significantly larger than 50% (t-test, $p = 0.002$). The difference between Mode A and B in final posterior minus prior is statistically significant (Welch's unequal variances t-test, $p = 0.022$; Wilcoxon exact rank-sum test, $p = 0.007$).

The estimated β coefficients in the random treatment (i.e., where the state is not ego-relevant) inform us how updating behavior deviates from the Bayesian benchmark. Following the literature on belief updating, we interpret these deviations as being driven by "cognitive" biases, while the differential updating across ego-relevant and non-ego relevant states allows us to identify "motivated" biases in processing information. In particular, we test whether there is asymmetric updating ($\beta^{pos} \neq \beta^{neg}$). For example, subjects in the ego-relevant treatment might have a desire to put more weight on positive rather than negative signals when forming their posteriors.

Moreover, we investigate whether there are cognitive or motivated biases across information structures. Hence, we investigate whether the properties of an information structure have implications for updating behavior.

3.2.2 Updating in the Endogenous Treatment

First, we briefly consider belief updating in the endogenous treatment. Here, subjects receive signals from the feedback mode they chose in the randomly selected choice situation. We restrict the analysis to Modes A and B to allow comparisons with the exogenous treatment.

A word of caution may be necessary here as the analysis of belief updating in the endogenous treatment can only be seen as suggestive for two main reasons: First, subjects self-select into feedback modes, that is, subjects in Mode A and B may not be comparable. Second, there are relatively few subjects who end up receiving signals from Mode B. In the IQ treatment, 26 subjects update according to Mode B, and in the random treatment there are only 19 subjects (in contrast, 66 subjects in the IQ treatment and 61 subjects in the random treatment end up in Mode A).[15]

Table A.5 in the appendix presents the estimation results of Equation (2) separately by treatment and feedback mode. In Mode A, we do not find evidence for asymmetric updating in either treatment since the coefficients β^{Pos} and β^{Neg} are of similar magnitude and we cannot reject the null hypothesis that they are equal (Wald tests, $p = 0.620$ in IQ and $p = 0.928$ in random). In Mode B, in contrast, we observe that β^{Pos} is larger than β^{Neg}, both in the IQ and the random treatment. However, as noted above, due to the small number of subjects in Mode B, we lack the statistical power to reject the null hypothesis that the coefficients are equal. The minimum detectable difference in coefficients is 0.598 for the IQ treatment and 0.622 in the random treatment (for $p < 0.05$ at 80% power), which is well above the differences that we find (0.349 in IQ and 0.204 in random).[16]

[15] Part of the reason for this imbalance is the fact that Mode A features in four choice situations and Mode B only in two, such that the probability is twice as high that a choice is drawn, in which Mode A could have been selected.

[16] To calculate the minimum detectable difference (MDD), we recode the independent variables such that one coefficient indicates the difference between the positive and negative updating coefficients. The p-value on this variable is equivalent to the Wald test p-value reported in the bottom of Table A.5. The standard

Table 4: Updating across feedback modes and treatments (exogenous treatment)

	(1) IQ Mode A	(2) IQ Mode B	(3) Random Mode A	(4) Random Mode B
δ^{prior}	0.814**	0.896***	0.716***	0.779***
	(0.085)	(0.046)	(0.066)	(0.061)
β^{Pos}	0.623***	0.482***	0.839	0.758
	(0.120)	(0.077)	(0.127)	(0.174)
β^{Neg}	0.568***	0.244***	0.767*	0.938
	(0.124)	(0.088)	(0.125)	(0.142)
Wald p-value (β^{Pos}=β^{Neg})	0.715	0.028	0.685	0.325
R2	0.640	0.781	0.697	0.730
N	165	165	171	186
Chow p-value ($\delta_{\text{A}}^{\text{prior}}$=$\delta_{\text{B}}^{\text{prior}}$)	0.398		0.480	
Chow p-value ($\beta_{\text{A}}^{\text{Pos}}$=$\beta_{\text{B}}^{\text{Pos}}$)	0.321		0.710	
Chow p-value ($\beta_{\text{A}}^{\text{Neg}}$=$\beta_{\text{B}}^{\text{Neg}}$)	0.035		0.367	

Notes: The table shows regression results of Equation (2) in the exogenous treatment, separately by IQ and random treatment and Modes A and B. We regress the posterior belief on the prior belief and the signal's likelihood ratio, interacted with an indicator if the signal is positive or negative. The model does not include a constant. Stated beliefs of 100 are replaced with 99 and beliefs of 0 with 1, respectively. Standard errors clustered on subject level in parentheses. Stars indicate whether the coefficient is statistically different from 1. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

To have more statistical power and to exclude self-selection into feedback modes, we now turn our focus to belief updating in the exogenous treatment.

3.2.3 Updating in the Exogenous Treatment

In the exogenous treatment, subjects are randomly assigned to Modes A or B. In the IQ treatment, 55 (55) subjects update according to Mode A (B), and in the random treatment 57 (62) subjects update according to Mode A (B).

In Table 4, we display the estimation results for the exogenous treatment. As before, we do not find evidence for asymmetric updating in Mode A. In both IQ and random treatment the coefficients β^{Pos} and β^{Neg} are of similar magnitude and are not significantly different from each other (Wald tests, $p = 0.715$ in IQ and $p = 0.685$ in random).

However, we find asymmetric updating in Mode B in the IQ treatment: when information is ego-relevant and negative signals are less salient (i.e., framed as grey signals), subjects

error of this coefficient times 2.8 gives us the minimum detectable difference that would be necessary to find a statistically significant coefficient at the 0.05 level with 80% power. The idea behind the factor 2.8 is the following: For the t-test to be significant, the coefficient has to be 1.96 standard errors away from zero. To have a 80% probability of drawing a t-value greater than 1.96, the "true" t-statistic has to be 1.96+0.84=2.8, with 0.84 being the inverse normal at the 80th percentile (see, e.g., Ioannidis *et al.*, 2017).

update less in response to negative than in response to positive signals. The updating coefficient for positive signals is about twice as large as for negative signals and the coefficients differ significantly (Wald test, $p = 0.028$). However, this is not the case when feedback is not ego-relevant: in the random treatment, subjects update, if anything, more in response to negative, less salient feedback, but the difference in coefficients is not statistically significant (Wald test, $p = 0.325$).

In the exogenous treatment, we have more statistical power compared to the endogenous treatment. However, to detect more subtle differences in updating, we would need a larger sample size. The minimum detectable difference in coefficients at 80% power is 0.422 (IQ) and 0.494 (random) in Mode A, while it is 0.295 (IQ) and 0.506 (random) in Mode B. It is notable that we are relatively well powered for detecting a difference in the IQ treatment in Mode B, but that we have less statistical power to detect a difference in the random treatment. However, since our subjects in the random treatment update, if anything, *more* in response to negative feedback, we still interpret it as suggestive evidence in favor of a motivated bias in updating (instead of a cognitive bias).

This interpretation is also supported by results from Chow tests in the bottom of Table 4. Subjects update significantly less in response to negative feedback in Mode B when signals are ego-relevant. However, when information is not ego-relevant, the difference is not statistically significant.

The previous analysis studies updating in a Bayesian framework that takes prior beliefs and signal informativeness into account. However, since it is derived from Bayes' rule it imposes a lot of structure. To control for differences in prior beliefs between treatments in a more flexible way, we follow a nearest-neighbor matching strategy. We compute the treatment difference in final posterior beliefs after matching subjects in IQ and control based on their prior beliefs. First, we trim the prior belief distribution to ensure common support between treatments.[17] Next, individuals are matched to at least 1, 2, or 5 neighbors in the other treatment by minimizing the distance in priors. The average treatment effect is calculated by comparing the average final posterior in one treatment to the posteriors of the neighbors in the other treatment.

Table 5 presents the results of the matching exercise. In Mode A, we observe that after three rounds of signals there are no significant differences in posterior beliefs between individuals in the IQ and random treatment. However, in Mode B, in which negative feedback is less salient, subjects in the IQ treatment are 11.33 to 11.97 percentage points more likely to believe that they are in the top half at the end of the experiment. Put differently, subjects in Mode A who start out with similar priors end up with similar posteriors in the IQ and ran-

[17] For example, in Mode B the lowest prior in the IQ treatment is 20, while there are 14 observations in the random treatment with a prior below 20 (see Figure A.1b). We drop these 14 observations as they have no counterpart in the IQ treatment. Trimming is described as an important pre-processing step, e.g., in Imbens (2015).

Table 5: Differences in final posterior beliefs between IQ and random treatment using nearest-neighbor matching (exogenous treatment)

	(1) Mode A	(2) Mode B
1 Neighbor	-2.609	11.746**
	(5.976)	(5.743)
2 Neighbors	-0.490	11.971**
	(5.615)	(5.668)
5 Neighbors	0.813	11.334**
	(5.385)	(5.684)
N	107	103

Notes: The table shows average treatment effects for final posterior beliefs between IQ and random treatment. The analysis is based on nearest neighbor matching on priors with replacement. We require subjects to be matched to at least 1, 2, or 5 neighbors by minimizing the absolute distance in priors. To ensure common support, we trimmed 5 observations in Mode A (1 from IQ and 4 from random) and 14 observations in Mode B (14 from random), which were above the maximum prior or below the minimum prior in the respective other treatment. Standard errors in parentheses are based on Abadie and Imbens (2006). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

dom treatment. Only in Mode B do they arrive at different posteriors, depending on whether feedback is ego-relevant.

Taken together, the results suggest that subjects' belief formation is driven by motivated reasoning, as subjects update asymmetrically only in the IQ treatment. However, while individuals might have a preference for forming confident beliefs about themselves, our results show that this does not always seem to be possible. In fact, our results suggest that asymmetric updating arises only in the information structure that features negative signals that are less salient and, thus, easier to misperceive. In the next section, we further discuss and interpret our experimental results.

4 Discussion

Our experimental findings show systematic differences across treatments in the way subjects choose between different feedback modes. We interpret these results as evidence for preferences over information structures driven by motivated reasoning (individuals' desire to have high opinions of themselves and self-enhancement motives).

We now discuss whether treatment differences could alternatively be explained by cognitive biases. In doing so, we follow the key features that distinguish motivated thinking from cognitive failures according to Bénabou and Tirole (2016).

4.1 Endogenous Directionality

A distinct feature of motivated reasoning is that it is directed toward some end (for example, the belief that one is highly intelligent). In contrast, general failures in cognitive reasoning that depend on one's prior beliefs, like confirmation-seeking and contradiction-seeking behavior, can go in either direction.

Confirmation-seeking behavior, or confirmation bias, is the tendency to search for, interpret, favor, and recall information in a way that confirms preexisting beliefs.[18] On the contrary, contradiction-seeking behavior is the tendency to favor information that goes against one's prior beliefs. In a non-ego-relevant setting, Charness *et al.* (2021) find that many individuals choose information structures that confirm their prior beliefs, but relatively few subjects can be classified as contradiction-seeking.[19]

In our experiment, confirmation-seeking could potentially explain treatment differences in information structure selection, as individuals in the IQ treatment have slightly higher priors on average. For instance, confirmation-seeking subjects with high priors could prefer Mode B since it is less informative about the negative state. However, even when controlling for prior beliefs, we see that subjects in the ego-relevant treatment are more likely to choose Mode B compared to those in the random treatment. In fact, all treatment differences hold when controlling for prior beliefs in regressions (Table A.2) and when employing a matching strategy on priors (Table A.3).

Finally, it is relevant to note that, in the informativeness choice, the treatment difference cannot be explained by confirmation- or contradiction-seeking behavior. Taken together, these results show that confirmation bias falls short of explaining the treatment differences in the information selection stage of our experiment.

4.2 Bounded Rationality

4.2.1 Cognitive Ability

Cognitive errors in processing and interpreting information depend on individuals' cognitive ability and analytical sophistication. That is, more able and more analytically sophisticated agents are less prone to cognitive biases. On the other hand, motivated reasoning does not necessarily imply a negative correlation.

[18]See Nickerson (1998) for a review of the psychological literature on confirmation bias.

[19]In the random treatment, we find some support for confirmation-seeking as a cognitive bias. In the baseline choice we find that subjects with a high prior are slightly more likely to choose Mode B (although the difference is not statistically significant) and in the baseline reversed choice they are less likely to choose Mode C (t-test, $p = 0.068$). However, in the skewness over framing choice, subjects with high priors are slightly less likely to choose Mode E (t-test, $p = 0.238$), speaking against confirmation-seeking behavior. In general, our experiment is not designed to provide conclusive evidence regarding confirmation-seeking behavior since the information structures vary both in framing and informativeness. It is an interesting question for future research how confirmation-seeking individuals trade off informativeness and salience of feedback.

By taking into account participants' abstract reasoning ability, measured by their IQ scores, our results do not seem to be driven by cognitive ability. Two pieces of evidence support this conclusion. First, individuals across treatment groups do not vary in their cognitive ability.[20] Second, our treatment differences in feedback mode selection are robust to controlling for individuals' cognitive ability (see Table A.2 in the appendix).

4.2.2 Confusion

Our experimental design features some rather complex elements and thus might have affected participants' understanding. To tackle this we paid close attention to the way we presented the experimental instructions. We also ensured understanding by letting participants answer comprehension questions (see screenshots in Figures H.2, H.4, and H.15). Moreover, participant confusion is unlikely to affect our conclusions as it is held constant across treatments.

Furthermore, if we look at the informativeness choice, in which the information structures only differ in the likelihood of receiving an uninformative signal and where we held constant their skewness and framing, we see that less than five percent of subjects in the random treatment make the suboptimal choice. This finding is reassuring as it implies that subjects understood our experimental instructions and were sufficiently incentivized to make well thought-out decisions.

4.3 Emotional Involvement: Heat vs Light

Bénabou and Tirole (2016) argue that motivated beliefs evoke and trigger emotional reactions, whereas cognitively driven biases do not. While we do not measure participants' emotions in the experiment, we find some suggestive evidence for emotions arising in the IQ treatment in the free-text questions.

In the post-experimental questionnaire, we asked subjects to explain how they chose between feedback modes. In the IQ treatment, we find that many subjects report a "gut-level" response to select the more positive-looking feedback and avoid explicit negative (red) signals.[21] In Appendix F, we conduct a quantitative content analysis of the free-text responses. We enlisted three research assistants, who had no information about the treatment variation, to code the free-text responses according to a pre-defined codebook (shown in Appendix F.1). A coding is counted if at least two out of the three research assistants coded a response in the same way.

[20]Mean IQ score in the IQ treatment is 11.34, while it is 11.41 in the random treatment. We cannot reject the H0 that IQ scores are significantly different between IQ and random (t-test, $p = 0.882$; Wilcoxon rank-sum test, $p = 0.780$).

[21]Subjects who chose Mode B in the IQ treatment stated, for example, "I always chose the feedback that looked the more positive. For example, with the most green dots/the less red dots", "My mind told me to avoid red and go for green", "I took a preference for those featuring green, motivational I guess?", etc.

In Table F.1, we find that subjects in the IQ treatment are more likely to state that they preferred the feedback mode that featured more green signals (Fisher's exact test, $p = 0.001$) and less red signals (Fisher's exact test, $p = 0.060$) compared to the random treatment. Subjects in the random treatment are more likely to state that they made their selection according to the feedback's informativeness (Fisher's exact test, $p = 0.013$). We do not find a statistically significant treatment difference for any of the other explanations (e.g., ease of understanding feedback, confirmation of prior belief). In Table F.2, we find that the stated preference for more green/less red signals and for more information is indeed correlated with observed behavior.

5 Conclusion

This paper documents an experiment to study individuals' preferences toward information structures and subsequent belief updating if information is ego-relevant or not. Our results from the information selection stage show that individuals in the ego-relevant treatment are more likely to choose feedback modes that are less informative and that make negative feedback less salient, compared to the control. The results from the belief updating stage indicate that individuals' belief formation is asymmetric (i.e., individuals respond more to positive news than to negative news), but only in the ego-relevant condition and when the negative feedback is less salient, and therefore easier to misperceive. These findings are consistent with individuals selectively choosing information structures that allow them to protect their ego.

Our results suggest that while individuals might have a motivated tendency to process information differently depending on its valence, their ability to do so depends on the "reality constraints" in the environment. We provide evidence that the framing and salience of feedback is one dimension of "reality constraints" that limits individuals' ability to nurture motivated beliefs. Zimmermann (2020) shows that raising the incentives to recall negative feedback can constitute another "reality constraint." Similarly, Drobner (forthcoming) shows that subjects, who are informed that uncertainty about the ego-relevant state will be resolved at the end of the experiment are less likely to update in a motivated way. This raises a question for future research regarding other dimensions in the environment that may constrain individuals from maintaining motivated beliefs and how consciously people engage in motivated thinking.

We find that subjects update symmetrically when feedback is made explicit, but that they update asymmetrically when negative feedback is less salient. In light of the mixed evidence on asymmetric updating in the literature, it would be interesting to study in more detail how the propensity to update asymmetrically depends on the framing and style of feedback. Moreover, future research could investigate how subjects choose between information structures, in which no asymmetric updating was found.

Taken together, our findings suggest that motivated information selection might play an important role in producing overconfident beliefs. Indeed, it is often the case that in our everyday life, we can choose our information sources and exert some control over the type of signals that we receive. This can help us protect ourselves from having to adjust our ego-relevant beliefs downwards. Further research could investigate in real-world settings how preferences over information structures impact how individuals sort into feedback environments, for example, in educational contexts.

While overconfidence is costly in some settings, it may be beneficial in others, for example, for motivational purposes or when trying to persuade others. Therefore, it depends on the context whether institutions should restrict or enhance individuals' ability to select their feedback sources. Future studies could investigate how much discretion over feedback sources is optimal in educational or professional environments.

References

Abadie, A. and Imbens, G. W. (2006). Large sample properties of matching estimators for average treatment effects. *Econometrica*, **74** (1), 235–267.

Ahn, T., Arcidiacono, P., Hopson, A. and Thomas, J. R. (2019). Equilibrium Grade Inflation with Implications for Female Interest in STEM Majors, NBER Working Paper 26556.

Azrieli, Y., Chambers, C. and Healy, P. (2018). Incentives in experiments: A theoretical analysis. *Journal of Political Economy*, **126** (4), 1472–1503.

Barron, K. (2021). Belief updating: Does the 'good-news, bad-news' asymmetry extend to purely financial domains? *Experimental Economics*, **24**, 31—58.

Bénabou, R. and Tirole, J. (2002). Self-confidence and personal motivation. *The Quarterly Journal of Economics*, **117** (3), 871–915.

— and Tirole, J. (2016). Mindful economics: The production, consumption, and value of beliefs. *Journal of Economic Perspectives*, **30** (3), 141–64.

Benoît, J.-P., Dubra, J. and Moore, D. A. (2015). Does the better-than-average effect show that people are overconfident? Two experiments. *Journal of the European Economic Association*, **13** (2), 293–329.

Blanco, M., Engelmann, D., Koch, A. and Normann, H. (2010). Belief elicitation in experiments: Is there a hedging problem? *Experimental Economics*, **13** (4), 412–438.

Buser, T., Gerhards, L. and van der Weele, J. (2018). Responsiveness to feedback as a personal trait. *Journal of Risk and Uncertainty*, **56** (2), 165–192.

Camerer, C. and Lovallo, D. (1999). Overconfidence and excess entry: An experimental approach. *American Economic Review*, **89** (1), 306–318.

Charness, G. and Dave, C. (2017). Confirmation bias with motivated beliefs. *Games and Economic Behavior*, **104**, 1–23.

—, Oprea, R. and Yuksel, S. (2021). How do people choose between biased information sources? evidence from a laboratory experiment. *Journal of the European Economic Association*, **19** (3), 1656–1691.

Chen, D. L., Schonger, M. and Wickens, C. (2016). oTree—an open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance*, **9**, 88–97.

Chew, S. H., Huang, W. and Zhao, X. (2020). Motivated false memory. *Journal of Political Economy*, **128** (10), 3913–3939.

Chopra, F., Haaland, I. and Roth, C. (forthcoming). Do people value more informative news? *Journal of Public Economics*.

Coutts, A. (2019). Good news and bad news are still news: Experimental evidence on belief updating. *Experimental Economics*, **22** (2), 369–395.

Dal Bó, P. and Fréchette, G. R. (2011). The evolution of cooperation in infinitely repeated games: Experimental evidence. *American Economic Review*, **101** (1), 411–29.

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J. and Wagner, G. G. (2011). Individual risk attitudes: Measurement, determinants, and behavioral consequences. *Journal of the European Economic Association*, **9** (3), 522–550.

Drobner, C. (forthcoming). Motivated beliefs and anticipation of uncertainty resolution. *American Economic Review: Insights*.

— and Goerg, S. J. (2021). Motivated belief updating and rationalization of information, `https://www.econstor.eu/handle/10419/243173` (retrieved 11/22/2021).

Eil, D. and Rao, J. M. (2011). The good news-bad news effect: asymmetric processing of objective information about yourself. *American Economic Journal: Microeconomics*, **3** (2), 114–38.

Enke, B. (2020). What you see is all there is. *Quarterly Journal of Economics*, **153** (3), 1363–1398.

—, Schwerter, F. and Zimmermann, F. (2020). Associative memory and belief formation, `https://benjamin-enke.com/pdf/Memory.pdf` (retrieved 10/29/2020).

Epstein, L. and Halevy, Y. (2019). Hard-to-interpret signals, `https://yoram-halevy.faculty.economics.utoronto.ca/wp-content/uploads/SignalAmbiguity.pdf` (retrieved 07/26/2019).

Ertac, S. (2011). Does self-relevance affect information processing? Experimental evidence on the response to performance and non-performance feedback. *Journal of Economic Behavior & Organization*, **80** (3), 532–545.

Exley, C. L. and Kessler, J. B. (2019). Motivated Errors, `https://www.dropbox.com/s/fqb32hb1g7e6vrm/ExleyKessler_MotivatedErrors.pdf` (retrieved 07/26/2019).

Falk, A. and Zimmermann, F. (2016). Beliefs and utility: Experimental evidence on preferences for information, IZA Discussion Paper No. 10172.

FRYER, R. G., HARMS, P. and JACKSON, M. O. (2019). Updating Beliefs when Evidence is Open to Interpretation: Implications for Bias and Polarization. *Journal of the European Economic Association*, **17** (5), 1470–1501.

GANGULY, A. and TASOFF, J. (2016). Fantasy and dread: the demand for information and the consumption utility of the future. *Management Science*, **63** (12), 4037–4060.

GARZ, M., SOOD, G., STONE, D. F. and WALLACE, J. (2020). The supply of media slant across outlets and demand for slant within outlets: Evidence from us presidential campaign news. *European Journal of Political Economy*, **63** (101877).

GENTZKOW, M. and SHAPIRO, J. M. (2010). What drives media slant? Evidence from us daily newspapers. *Econometrica*, **78** (1), 35–71.

GNEEZY, U. and POTTERS, J. (1997). An experiment on risk taking and evaluation periods. *The Quarterly Journal of Economics*, **112** (2), 631–645.

GOLMAN, R., HAGMANN, D. and LOEWENSTEIN, G. (2017). Information avoidance. *Journal of Economic Literature*, **55** (1), 96–135.

GOTTFREDSON, L. S. and DEARY, I. J. (2004). Intelligence predicts health and longevity, but why? *Current Directions in Psychological Science*, **13** (1), 1–4.

GOTTHARD-REAL, A. (2017). Desirability and information processing: An experimental study. *Economics Letters*, **152**, 96–99.

GRETHER, D. M. (1980). Bayes rule as a descriptive model: The representativeness heuristic. *The Quarterly Journal of Economics*, **95** (3), 537–557.

GROSSMAN, Z. and OWENS, D. (2012). An unlucky feeling: Overconfidence and noisy feedback. *Journal of Economic Behavior & Organization*, **84** (2), 510–524.

HO, E., HAGMANN, D. and LOEWENSTEIN, G. (2021). Measuring information preferences. *Management Science*, **67** (1), 126–145.

HOFFMAN, M. (2016). How is information valued? evidence from framed field experiments. *The Economic Journal*, **126**, 1884–1911.

HOLM, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, **6** (2), 65–70.

IMBENS, G. W. (2015). Matching methods in practice. three examples. *The Journal of Human Resources*, **50** (2), 373–419.

IOANNIDIS, J. P. A., STANLEY, T. D. and DOUCOULIAGOS, H. (2017). The power of bias in economics research. *The Economic Journal*, **127**, F236–F265.

Jin, G. Z., Luca, M. and Martin, D. (2021). Is no news (perceived as) bad news? An experimental investigation of information disclosure. *American Economic Journal: Microeconomics*, **13** (2), 141–173.

Karlsson, N., Loewenstein, G. and Seppi, D. (2009). The ostrich effect: Selective attention to information. *Journal of Risk and Uncertainty*, **38** (2), 95–115.

Karni, E. (2009). A mechanism for eliciting probabilities. *Econometrica*, **77** (2), 603–606.

Köszegi, B. (2006). Ego utility, overconfidence, and task choice. *Journal of the European Economic Association*, **4** (4), 673–707.

Kuhnen, C. M. (2015). Asymmetric learning from financial information. *The Journal of Finance*, **70** (5), 2029–2062.

List, J. A., Shaikh, A. M. and Xu, Y. (2019). Multiple hypothesis testing in experimental economics. *Experimental Economics*, **22**, 773–793.

Malmendier, U. and Tate, G. (2005). CEO overconfidence and corporate investment. *The Journal of Finance*, **60** (6), 2661–2700.

Masatlioglu, Y., Orhun, A. Y. and Raymond, C. (2017). Intrinsic information preferences and skewness, `http://econweb.umd.edu/~masatlioglu/Masatlioglu_Orhun_Raymond.pdf` (retrieved 10/01/2019).

Mele, A. R. (2001). *Self-Deception Unmasked*. Princeton, NJ: Princeton University Press.

Möbius, M., Niederle, M., Niehaus, P. and Rosenblat, T. (2014). Managing self-confidence: Theory and experimental evidence, `https://web.stanford.edu/~niederle/Mobius.Niederle.Niehaus.Rosenblat.paper.pdf` (retrieved 07/26/2019).

Montanari, G. and Nunnari, S. (2019). Audi alteram partem: An experiment on selective exposure to information, `http://www.salvatorenunnari.eu/mn_selectexposure.pdf` (retrieved 12/15/2019).

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, **2** (2), 175–220.

Nielsen, K. (2020). Preferences for the resolution of uncertainty and the timing of information. *Journal of Economic Theory*, **189**, 105090.

Oster, E., Shoulson, I. and Dorsey, E. (2013). Limited life expectancy, human capital and health investments. *American Economic Review*, **103** (5), 1977–2002.

Romano, J. P. and Wolf, M. (2010). Balanced control of generalized error rates. *The Annals of Statistics*, **38**, 598–633.

Sabot, R. and Wakeman-Linn, J. (1991). Grade Inflation and Course Choice. *Journal of Economic Perspectives*, **5** (1), 159–170.

Saccardo, S. and Serra-Garcia, M. (2020). Cognitive flexibility or moral commitment? Evidence of anticipated belief distortion, `https://ssrn.com/abstract=3676711` (retrieved 08/02/2021).

Schwardmann, P. and Van der Weele, J. (2019). Deception and self-deception. *Nature Human Behaviour*, **3** (10), 1055–1061.

Sicherman, N., Loewenstein, G., Seppi, D. J. and Utkus, S. P. (2015). Financial attention. *The Review of Financial Studies*, **29** (4), 863–897.

Sternberg, R. J., Grigorenko, E. L. and Bundy, D. A. (2001). The predictive value of IQ. *Merrill-Palmer Quarterly*, **47** (1), 1–41.

Von Hippel, W. and Trivers, R. (2011). The evolution and psychology of self-deception. *Behavioral and Brain Sciences*, **34** (1), 1.

Zimmermann, F. (2014). Clumped or piecewise? Evidence on preferences for information. *Management Science*, **61** (4), 740–753.

— (2020). The dynamics of motivated beliefs. *American Economic Review*, **110** (2), 337–361.

Online Appendix for "Protecting the Ego: Motivated Information Selection and Updating"

A Additional Tables

Table A.1: Descriptive statistics

	Endogenous		Exogenous	
	IQ	Random	IQ	Random
Age (Mean)	20.973	21.830	20.236	20.168
	(2.960)	(4.095)	(2.933)	(2.304)
Female (Share)	0.664	0.613	0.618	0.529
	(0.475)	(0.489)	(0.488)	(0.501)
Native English speakers (Share)	0.427	0.377	0.364	0.370
	(0.497)	(0.487)	(0.483)	(0.485)
Studying (Share)	0.964	0.972	0.955	0.975
	(0.188)	(0.167)	(0.209)	(0.157)
First year students (Share)	0.455	0.481	0.545	0.504
	(0.500)	(0.502)	(0.500)	(0.502)
IQ puzzles solved (Mean)	11.336	11.406	10.964	10.924
	(3.425)	(3.397)	(3.038)	(3.051)
N	110	106	110	119

Notes: The table shows descriptive statistics of the experimental dataset. Standard deviations are in parentheses.

Table A.2: Information structure choices controlling for covariates

	(1)	(2)	(3)	(4)	(5)	(6)
Baseline	0.194***	0.191***	0.178***	0.193***	0.189***	0.164***
	(0.059)	(0.060)	(0.059)	(0.059)	(0.058)	(0.061)
Informativeness	0.135***	0.138***	0.132***	0.134***	0.132***	0.128***
	(0.042)	(0.045)	(0.043)	(0.042)	(0.042)	(0.045)
Framing	-0.265***	-0.266***	-0.243***	-0.265***	-0.263***	-0.239***
	(0.064)	(0.066)	(0.065)	(0.065)	(0.065)	(0.068)
Skewness over framing	-0.166***	-0.130**	-0.163***	-0.165***	-0.167***	-0.122**
	(0.057)	(0.059)	(0.057)	(0.057)	(0.058)	(0.058)
Baseline reversed	-0.052	-0.038	-0.049	-0.053	-0.052	-0.038
	(0.050)	(0.052)	(0.050)	(0.050)	(0.051)	(0.053)
Demographics		✓				✓
Prior			✓			✓
IQ score				✓		✓
Risk					✓	✓
N	216	216	216	216	216	216

Notes: The table shows the coefficient of the IQ treatment dummy in the regression of the feedback mode choice on the respective covariates. Demographics comprises controls for gender, age, years of study, and whether English is the native language. The risk measure is by Gneezy and Potters (1997). Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.3: Information structure choices using nearest-neighbor matching by prior

	(1)	(2)	(3)
	1 Neighbor	2 Neighbors	5 Neighbors
Baseline	0.227***	0.225***	0.189***
	(0.065)	(0.067)	(0.062)
Informativeness	0.136***	0.142***	0.135***
	(0.050)	(0.049)	(0.050)
Salience	-0.250***	-0.238***	-0.219***
	(0.069)	(0.068)	(0.065)
Skewness over framing	-0.159***	-0.145**	-0.150***
	(0.058)	(0.057)	(0.055)
Baseline reversed	-0.076	-0.070	-0.058
	(0.049)	(0.048)	(0.047)
n	215	215	215

Notes: The table shows average treatment effects for information structure choices between IQ and random treatment. The analysis is based on nearest neighbor matching on priors with replacement. We require subjects to be matched to at least 1, 2, or 5 neighbors by minimizing the absolute distance in priors. To ensure common support, we trimmed 1 observation in the random treatment, which was below the minimum prior in the IQ treatment. Standard errors in parentheses are based on Abadie and Imbens (2006). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.4: Information structure choices applying multiple testing correction

		p-values			
	Difference	Unadjusted	MHT	Holm	Bonferroni
Baseline	0.194	0.001	0.004	0.004	0.006
Informativeness	0.135	0.002	0.005	0.005	0.009
Salience	-0.265	0.000	0.000	0.001	0.001
Skewness over framing	-0.166	0.003	0.007	0.007	0.017
Baseline reversed	-0.052	0.300	0.300	0.300	1.000

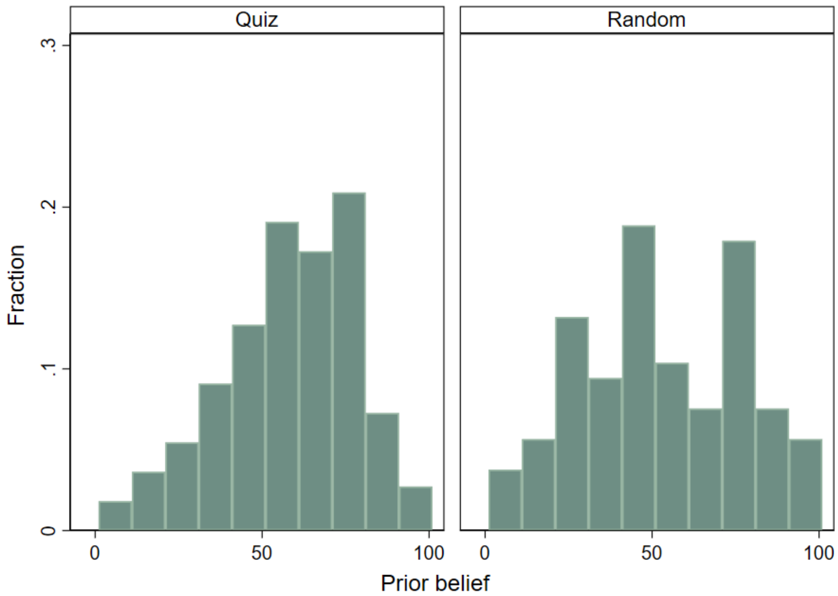
Notes: The table shows treatment differences in information structure choices between IQ and random treatment, together with unadjusted p-values and p-values corrected for multiple hypothesis testing. All p-values are calculated using the Stata package mhtexp with 10,000 bootstrap replications (List *et al.*, 2019). MHT uses the procedure proposed by List *et al.* (2019) that builds on Romano and Wolf (2010) to control the familywise error rate. Holm multiplies the smallest unadjusted p-value by the number of hypotheses, the second smallest by the number of hypotheses minus 1 etc. Bonferroni multiplies all unadjusted p-values by the number of hypotheses.

Table A.5: Updating across feedback modes and treatments (endogenous treatment)

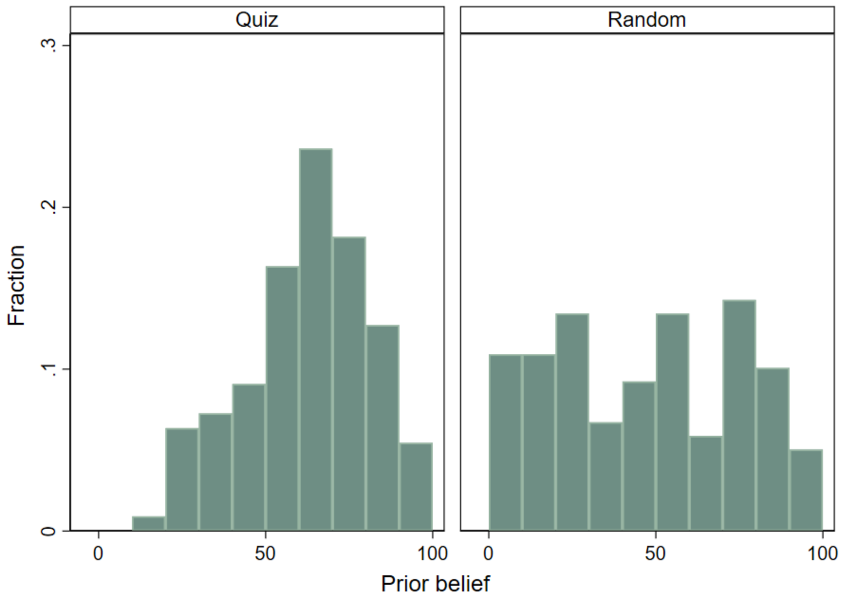
	(1)	(2)	(3)	(4)
	IQ Mode A	Random Mode A	IQ Mode B	Random Mode B
δ^{prior}	0.767***	0.650***	0.828	0.832
	(0.076)	(0.070)	(0.107)	(0.163)
β^{Pos}	0.603***	0.781	0.541**	0.408***
	(0.135)	(0.157)	(0.198)	(0.178)
β^{Neg}	0.527***	0.763*	0.191***	0.205***
	(0.111)	(0.131)	(0.084)	(0.272)
Wald p-value (β^{Pos}=β^{Neg})	0.620	0.928	0.114	0.372
R2	0.684	0.651	0.809	0.646
N	198	183	78	57

Notes: The table shows regression results of Equation (2) in the endogenous treatment, separately by IQ and random treatment and Modes A and B. We regress the posterior belief on the prior belief and the signal's likelihood ratio, interacted with an indicator if the signal is positive or negative. The model does not include a constant. Stated beliefs of 100 are replaced with 99 and beliefs of 0 with 1, respectively. Standard errors clustered on subject level in parentheses. Stars indicate whether the coefficient is statistically different from 1. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure A.1: Prior beliefs in deciles by treatment



(a) Endogenous treatment



(b) Exogenous treatment

Notes: The plot shows the distribution of prior beliefs to be in the top half of the distribution by treatment.

B Belief Elicitation Mechanism

After the IQ task, we tell subjects that for the following part of the experiment we will ask them their beliefs regarding some events. In particular, they are told that they will be asked four belief questions and that one question will be chosen at random to count for payment.

Then, we explain our belief elicitation procedure. We use the belief elicitation mechanism proposed by Karni (2009), called the matching probabilities method.[22] Subjects are presented with two possible bets: the lottery and the event. Each bet either pays a prize p (£6.00 in our experiment) or nothing. More specifically:

- The Event: pays the prize p if the event occurs, 0 otherwise.
- The Lottery: pays the prize p with probability x for $x \in \{0, 1, 2, ..., 100\}$, and 0 otherwise.

Hence, subjects (through their answers) indicate with what probability x makes them indifferent between betting on the event or the lottery. After they indicate the indifference point, one probability $y \in \{0, 1, 2, ..., 100\}$ is drawn. If $x \geq y$, the subject bets on the event and earns the prize p if the event occurs. On the other hand, if $x < y$, the subject bets on the lottery, which has probability y of paying the prize p. Intuitively, by choosing x, the subject affects her chances of betting on the event or the lottery and the chances of earning the prize p in case she ends up betting on the lottery. Under this mechanism, reporting one's subjective probability of the event occurring maximizes the chances of earning the prize, regardless of risk preferences.

Given the complexity of this belief elicitation mechanism, we decided to make instructions intuitive for subjects by walking them through an example and explaining how their answer would affect their chances betting on the event or the lottery, and their chances of winning the prize. We also emphasize that truthful reporting is the answer that maximizes the chances of earning the prize. The exact wording of the instructions is provided as a screenshot in Figure H.3. To ensure that subjects understand the main features of this elicitation procedure, we asked subjects to answer comprehension questions (see Figure H.4).

[22] This method is also referred to as the "crossover mechanism," "reservation probabilities," and "lottery method." It has been extensively applied in experimental economics, including Möbius *et al.* (2014), Coutts (2019), Buser *et al.* (2018), and Schwardmann and Van der Weele (2019).

C Maximum Likelihood Estimation of Within-Subject Choice Patterns

We use a finite mixture model to estimate the share of subjects who exhibit consistent choice patterns that pertain to one of three preferences ("Maximum information," "Positive skewness," or "Salience of feedback"). We allow for a deviation between the observed choice and the choice prescribed by a subject's preference: $y_{ic} = I\left\{s_{ic}(s^p) + \gamma\epsilon_{ic} \geq 0\right\}$, where y_{ic} is the choice by subject i in choice situation c (0 for the first alternative and 1 for the second alternative). s_{ic} is the choice that is prescribed by the preference s^p (coded by -1 for the first alternative and 1 for the second alternative). $I\left\{.\right\}$ is 1 if the term in brackets is positive and 0 otherwise. ϵ_{ic} is an iid error term that is type 1 extreme value distributed. γ scales the variance of the error term and can be interpreted as the amount of implementation noise to be estimated. Thus, the more an individual's choices align with those prescribed by the respective preference, the smaller the estimated implementation noise will be.

The likelihood of subject i to follow preference s^p over all choices c is

$$\pi_i(s^p) = \prod_C \left(\frac{1}{1+exp(-s_{ic}(s^p))/\gamma}\right)^{y_{ic}} \left(\frac{1}{1+exp(s_{ic}(s^p))/\gamma}\right)^{1-y_{ic}}. \tag{3}$$

The resulting log likelihood is $\sum_I \ln(\sum_P \pi(s^p)\pi_i(s^p))$, which is summed over all I subjects by treatment and where P represents the set of preferences we consider. $\pi(s^p)$ is the estimated fraction of the sample with preference p. For identification, one preference is taken as reference category and its share is implied by the fact that shares have to sum to one. For the estimation, we adapt the code by Dal Bó and Fréchette (2011).

D Investigation of Order Effects

The subjects in our experiment make five consecutive choices between information structures. Since only one of these five choices is randomly selected to be implemented, each individual choice can be treated as an independent choice. However, one may be concerned about potential order effects if subjects' subsequent choices are affected by their previous choices, for example, due to a preference for consistency. Moreover, in the experimental instructions, we always use the first feedback mode choice as an example to explain the choice situation (see Figures H.9 to H.11). Hence, it is possible that subjects have a better understanding of the feedback mode choice that is presented first. However, note that potential order effects do not affect our results if they are constant between IQ and random treatments. Nonetheless, to check if there are differential order effects between treatments, we vary the order in which subjects make pairwise choices. The different orders with the respective number of subjects are presented in Table D.1.

Table D.1: Order of feedback mode choices

Feedback choice	Order 1	Order 2	Order 3
Baseline (A vs B)	1st	2nd	3rd
Informativeness (A vs D)	3rd	3rd	4th
Framing (B vs E)	5th	4th	2nd
Skewness over framing (A vs E)	4th	5th	1st
Baseline reversed (A vs C)	2nd	1st	5th
N	116	51	49

Notes: The table shows the order with which the respective feedback mode choice is presented.

In Table D.2, we interact the treatment dummy with a dummy indicating if a subject makes feedback mode choices according to the first, second, or third order. Most importantly, as indicated by the statistically insignificant interaction effects, we do not find support for differential order effects between treatments. This suggests that order effects are of no concern for our conclusions.

Interestingly, in Column (4), we observe that in both treatments significantly fewer subjects select Mode E in the skewness over framing choice when this choice represents the first feedback mode choice (i.e., in Order 3). This could be explained by the fact that in Order 3 this choice is presented first and is used as an example in the instructions (cf. Figure H.11). Hence, subjects may better understand that Mode E is, in fact, less informative than Mode A. This is supported by the observation that the fraction of subjects who follow a strict preference to maximize the informativeness in Table 2 increases substantially from 0.409 to 0.500 in the IQ treatment, and from 0.538 to 0.689 in the random treatment when taking out the skewness over framing choice.

Table D.2: Feedback mode choices regressed on presented order

	(1)	(2)	(3)	(4)	(5)
				Skewness	Baseline
	Baseline	Informativeness	Framing	over framing	reversed
IQ treatment	0.199***	0.151***	-0.324***	-0.200**	-0.057
	(0.076)	(0.057)	(0.087)	(0.083)	(0.070)
Order 2	0.077	0.005	-0.081	-0.066	-0.073
	(0.092)	(0.047)	(0.121)	(0.115)	(0.085)
Order 3	0.127	0.048	-0.061	-0.219**	0.057
	(0.100)	(0.062)	(0.123)	(0.101)	(0.104)
IQ treatment x Order 2	0.024	0.001	0.075	-0.005	0.014
	(0.148)	(0.105)	(0.158)	(0.141)	(0.110)
IQ treatment x Order 3	-0.049	-0.075	0.184	0.153	0.007
	(0.154)	(0.110)	(0.167)	(0.131)	(0.140)
Constant	0.123***	0.035	0.561***	0.386***	0.193***
	(0.044)	(0.025)	(0.067)	(0.065)	(0.053)
R2	0.060	0.047	0.084	0.062	0.019
N	216	216	216	216	216

Notes: The table shows results from regressing the choice of the second alternative in the respective choice situation on the IQ treatment dummy and dummies indicating the order in which choices where presented. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

E Information Preference Scale (Ho *et al.*, 2021)

- As part of a semi-annual medical checkup, your doctor asks you a series of questions. The answers to these questions can be used to estimate your life expectancy (the age you are predicted to live to). Do you want to know how long you can expect to live? [1: Definitely don't want to know; 4: Definitely want to know]
- You provide some genetic material to a testing service to learn more about your ancestors. You are then told that the same test can, at no additional cost, tell you whether you have an elevated risk of developing Alzheimer's. Do you want to know whether you have a high risk of developing Alzheimer's? [1: Definitely don't want to know; 4: Definitely want to know]
- At your annual checkup, you are given the option to see the results of a diagnostic test which can identify, among other things, the extent to which your body has suffered long-term effects from stress. Do you want to know how much lasting damage your body has suffered from stress? [1: Definitely don't want to know; 4: Definitely want to know]
- Ten years ago, you had the opportunity to invest in two retirement funds: Fund A and Fund B. For the past 10 years, you have invested all your retirement savings in Fund A. Do you want to know the balance you would have, if you had invested in Fund B instead? [1: Definitely don't want to know; 4: Definitely want to know]
- You decide to go to the theater for your birthday and give your close friend (or partner) your credit card so they can purchase tickets for the two of you, which they do. You aren't sure, but suspect that the tickets may have been expensive. Do you want to know how much the tickets cost? [1: Definitely don't want to know; 4: Definitely want to know]
- You bought an electronic appliance at a store at what seemed like a reasonable, though not particularly low, price. A month has passed, and the item is no longer returnable. You see the same appliance displayed in another store with a sign announcing 'SALE.' Do you want to know the price you could have bought it for? [1: Definitely don't want to know; 4: Definitely want to know]
- You gave a close friend one of your favorite books for her birthday. Visiting her apartment a couple of months later, you notice the book on her shelf. She never said anything about it; do you want to know if she liked the book? [1: Definitely don't want to know; 4: Definitely want to know]

- Someone has described you as quirky, which could be interpreted in a positive or negative sense. Do you want to know which interpretation he intended? [1: Definitely don't want to know; 4: Definitely want to know]
- You gave a toast at your best friend's wedding. Your best friend says you did a good job, but you aren't sure if he or she meant it. Later, you overhear people discussing the toasts. Do you want to know what people really thought of your toast? [1: Definitely don't want to know; 4: Definitely want to know]
- As part of a fund-raising event, you agree to post a picture of yourself and have people guess your age (the closer they get, the more they win). At the end of the event, you have the option to see people's guesses. Do you want to learn how old people guessed that you are? [1: Definitely don't want to know; 4: Definitely want to know]
- You have just participated in a psychological study in which all the participants rate one-anothers' attractiveness. The experimenter gives you an option to see the results for how people rated you. Do you want to know how attractive other people think you are? [1: Definitely don't want to know; 4: Definitely want to know]
- Some people seek out information even when it might be painful. Others avoid getting information that they suspect might be painful, even if it could be useful. How would you describe yourself? [1: If it could be painful, I don't want to know; 4: Even if it could be painful, I always want to know]
- If people know bad things about my life that I don't know, I would prefer not to be told. [1: Strongly agree; 4: Strongly disagree]

F Content analysis of free-text responses

In the post-experimental questionnaire, we asked subjects to explain how they decided between feedback modes. In particular, we asked: "Please explain, in general, how you decided between feedback modes across the five scenarios. For example, why did you choose one feedback mode over another? What specific characteristics of the feedback modes were you looking at?"

To analyze the free-text responses, we enlisted three research assistants to code the responses according to a pre-defined codebook. The research assistants were not familiar with the experiment, except for the instructions shown in Appendix F.1. The research assistants were only given the randomly sorted list of free-text responses, which did not contain additional variables or information about the treatment variation. A coding is counted if at least two out of three research assistants agreed on the coding.

Table F.1 shows how often an explanation for a feedback mode choice is given by treatment. In the IQ treatment, subjects are more likely to state that they preferred the feedback mode with more green signals ($p = 0.001$) and less red signals ($p = 0.060$) compared to the random treatment. In the random treatment, they more often gave the explanation that they chose the feedback mode according to its informativeness ($p = 0.013$). The frequency of all the other explanations does not differ significantly between treatments.

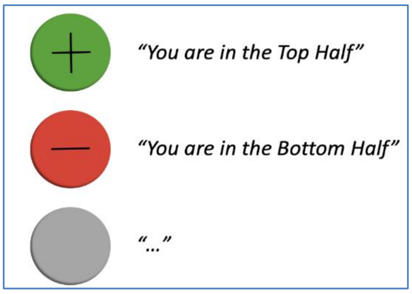
In Table F.2, we explore whether the stated explanations correlate with information structure choices. The desire to receive more green signals and less red signals is indeed highly correlated with the propensity to select the respective feedback mode (e.g., Mode B in the baseline choice). Similarly, individuals who stated that they had maximized the informativeness or minimized the number of grey signals chose the more informative feedback mode in the baseline and informativeness choice.[23]

[23] The only inconsistent association is by subjects who state that they chose "More grey signals", but instead chose the feedback mode with less grey signals. However, these are only two subjects in the IQ and three in the random treatment (see Table F.1).

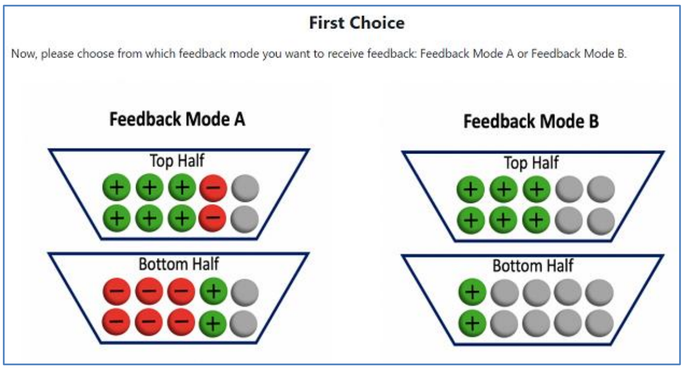
F.1 Coding instructions for research assistants

Text analysis of survey responses

We conducted a lab experiment, in which subjects could select the type of feedback/signals that they would like to receive. The feedback informed them whether they were in the top or bottom half of a distribution. There are three types of possible signals (green, red, or grey balls):



Subjects had to make decisions between two "feedback modes" each, which were presented in the form of urns with a varying composition of balls/signals. Signals would be drawn from the top urn if the subject was in the top half, or from the bottom urn if the subject was in the bottom half of the distribution. In total, there were five choice scenarios. Below you see an example for the choice scenario between Mode A and Mode B:



After the experiment, we asked the subjects:

"Please explain, in general, how you decided between feedback modes across the five scenarios. For example, why did you choose one feedback mode over another? What specific characteristics of the feedback modes were you looking at?"

Your task

We give you the responses given by the 216 subjects to the question above. Your task is to code their responses according to the following codebook. If a response matches a category, type a 1 in the respective column. A response can match with more than one category. If a response does not match any of the categories or is difficult to understand, type a 1 in the column "none". In case of doubt, please classify it as "none" instead of guessing what is meant.

The possible categories are given on the next page.

Categories

- **more_greens:** chose the feedback mode with more green balls
    - Example: “I choose the feedback mode which had more green balls”
- **more_reds:** chose the feedback mode with more red balls
    - Example: “I was looking for the more negative approach (i.e. red balls) over the positive balls”
- **more_greys:** chose the feedback mode with more grey balls
    - Example: “maybe more Grey than red (-) balls”
- **less_greens:** chose the feedback mode with less green balls
    - Example: “I choose the feedback mode which had less green balls”
- **less_reds:** chose the feedback mode with less red balls
    - Example: “I wanted the feedback mode that had less red”
- **less_greys:** chose the feedback mode with less grey balls **(or more green *and* red balls)**
    - Example 1: “If there were less '...' signs I was more likely to choose the feedback mode”
    - Example 2: “I chose one feedback mode over another if the feedback mode showed more red or green counters.”
- **more_information:** chose the feedback more that is more informative/precise or has the largest difference in signals between top and bottom urn
    - Example 1: “The mode I chose looks like the most informative one “
    - Example 2: “I was more concerned about having more coloured balls than grey balls as that would give a clearer picture of whether I was in the top half or bottom half.”
- **easier_to_understand:** chose the feedback more that is easier to understand
    - Example: “I decided to choose the more simpler feedback modes (mostly B and E) as they were easier to read and interpret”
- **confirmation:** chose the feedback mode that is more likely to confirm the initial belief
    - Example: “Assuming I was in the bottom I wanted the feedback modes that would be most likely to confirm that.”
- **random:** chose the feedback mode randomly
    - Example: “i really did not understand the task very well, so I just chose randomly”
- **none:** statement cannot be classified by one of the above statements
    - Example: “Known values of balls preferred.”

Table F.1: Number of times an explanation is mentioned (free-text responses)

	IQ	Random	p-value
More green signals	27	8	0.001
More red signals	5	2	0.446
More grey signals	2	3	0.679
Less green signals	0	0	.
Less red signals	5	0	0.060
Less grey signals	15	23	0.153
More information	20	35	0.013
Easier to understand	7	6	1.000
Confirmation	6	4	0.748
Random	2	4	0.439
None	19	21	0.727
Number of coded answers	108	106	

Notes: The table shows how often a respective response was given to the question how subjects decided between feedback modes by treatment. The coding was conducted by three research assistant using the instructions in Appendix F.1. A coding is counted if at least two out of three research assistants agreed on the coding. Column 3 shows p-values based on Fisher's exact tests (two-sided).

Table F.2: Feedback mode choices regressed on free-text explanations

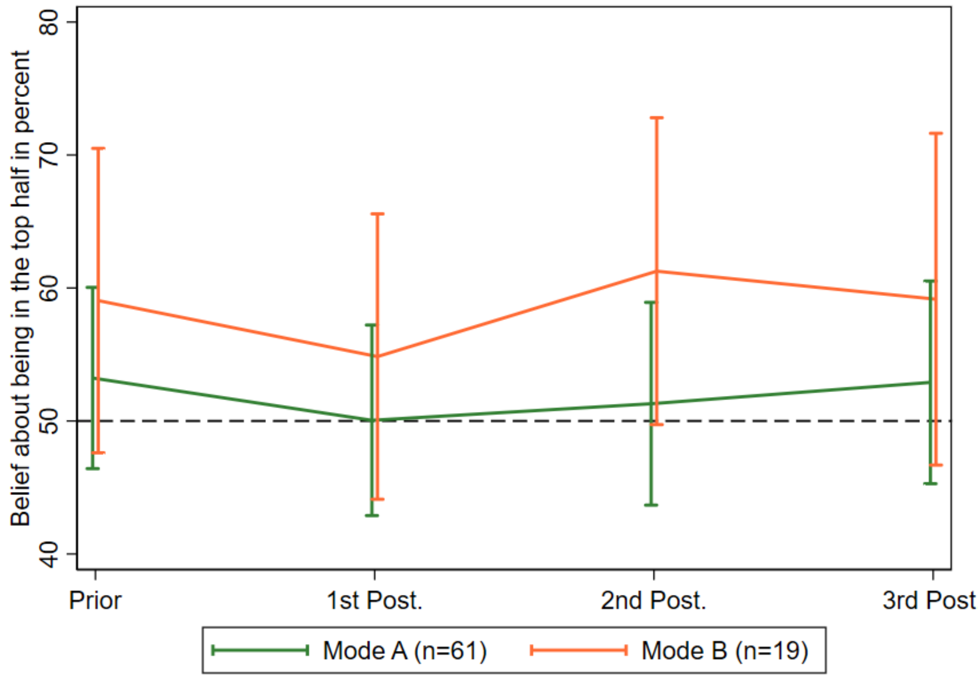
	(1) Baseline	(2) Informativeness	(3) Salience	(4) Skewness over salience	(5) Baseline reversed
More green signals	0.371***	-0.028	-0.326***	-0.237***	-0.116**
	(0.086)	(0.077)	(0.085)	(0.080)	(0.056)
More red signals	-0.305***	0.022	0.204	0.068	0.453**
	(0.089)	(0.145)	(0.200)	(0.214)	(0.179)
More grey signals	-0.262**	-0.213***	-0.092	-0.095	0.115
	(0.126)	(0.072)	(0.222)	(0.216)	(0.136)
Less red signals	0.468***	0.207	-0.280***	-0.172**	-0.058
	(0.100)	(0.230)	(0.092)	(0.067)	(0.044)
Less grey signals	-0.249***	-0.141***	0.078	-0.017	-0.058
	(0.056)	(0.036)	(0.094)	(0.085)	(0.057)
More information	-0.180***	-0.131**	-0.014	-0.083	-0.002
	(0.065)	(0.052)	(0.092)	(0.089)	(0.061)
Easier to understand	0.190	-0.188***	-0.410***	-0.081	0.112
	(0.144)	(0.054)	(0.101)	(0.142)	(0.132)
Confirmation	-0.167	-0.183***	0.118	-0.005	0.179
	(0.108)	(0.052)	(0.169)	(0.168)	(0.156)
Random	-0.310***	0.124	-0.142	0.185	0.706***
	(0.061)	(0.205)	(0.211)	(0.222)	(0.165)
None	-0.010	-0.085	-0.075	-0.065	0.073
	(0.096)	(0.078)	(0.108)	(0.101)	(0.084)
Constant	0.310***	0.210***	0.475***	0.315***	0.127**
	(0.061)	(0.056)	(0.073)	(0.073)	(0.053)
R2	0.310	0.099	0.114	0.053	0.188
n	216	216	216	216	216

Notes: The table shows results from regressing the choice of the second alternative in the respective choice situation on the reasons given in the free-text part of the post-experimental questionnaire. The coding of free-text responses was conducted by three research assistant using the instructions in Appendix F.1. A coding is counted if at least two out of three coders agreed on the coding. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

G Information Selection and Beliefs in the Random Treatment

In Figure G.1, we plot the prior and posterior beliefs after three rounds of feedback in the endogenous/random treatment. First, as in the IQ treatment, we observe that prior beliefs between Modes A and B are not significantly different from each other (t-test: $p = 0.396$). However, unlike in the IQ treatment, we do not observe that beliefs in the feedback mode diverge with the arrival of signals and, in fact, also the final posterior minus prior beliefs after three signals are not significantly different (t-test: $p = 0.950$).

Figure G.1: Beliefs before and after signals by feedback mode (endogenous/random treatment)



Notes: The plot shows the average prior and posterior beliefs (after each of the three signals) from treatment endogenous/random for the selected feedback mode. The whiskers represent 95% confidence intervals.

H Instructions

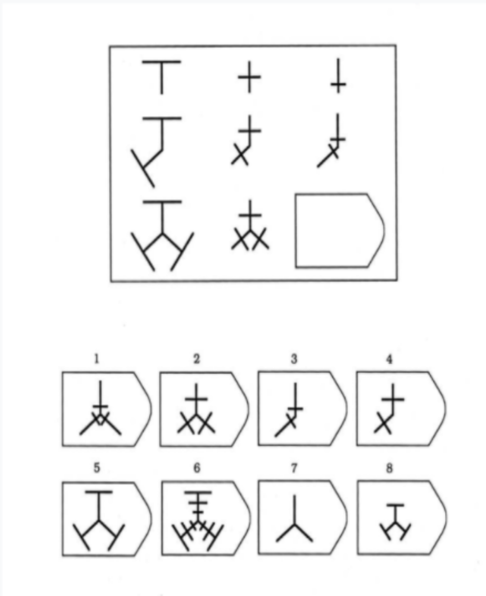
Figure H.1: Screenshot of the instructions template about the IQ task

Instructions Task 1 - The Quiz

In this task you are asked to solve a quiz. The quiz is a non-verbal test that measures abstract reasoning and can be used to estimate fluid intelligence (IQ). High scores in this test are regarded as one of the best predictors for academic and professional success, occupation, income, health, and longevity.

More specifically, in the quiz you are asked to solve 20 puzzles in 10 minutes. This quiz consists of two sets of 10 puzzles. For each set, you will have 5 minutes (300 seconds) to solve them.

Each puzzle consists of a visual geometric design with a missing piece. You are asked to fill in the missing piece from eight possible choices. One example is below (here, the correct answer is piece number 1):



If this task is selected to count for payments, you will be paid for three randomly chosen puzzles. That is, for each of these randomly chosen puzzles, you will be paid £2.00 if your answer is correct.

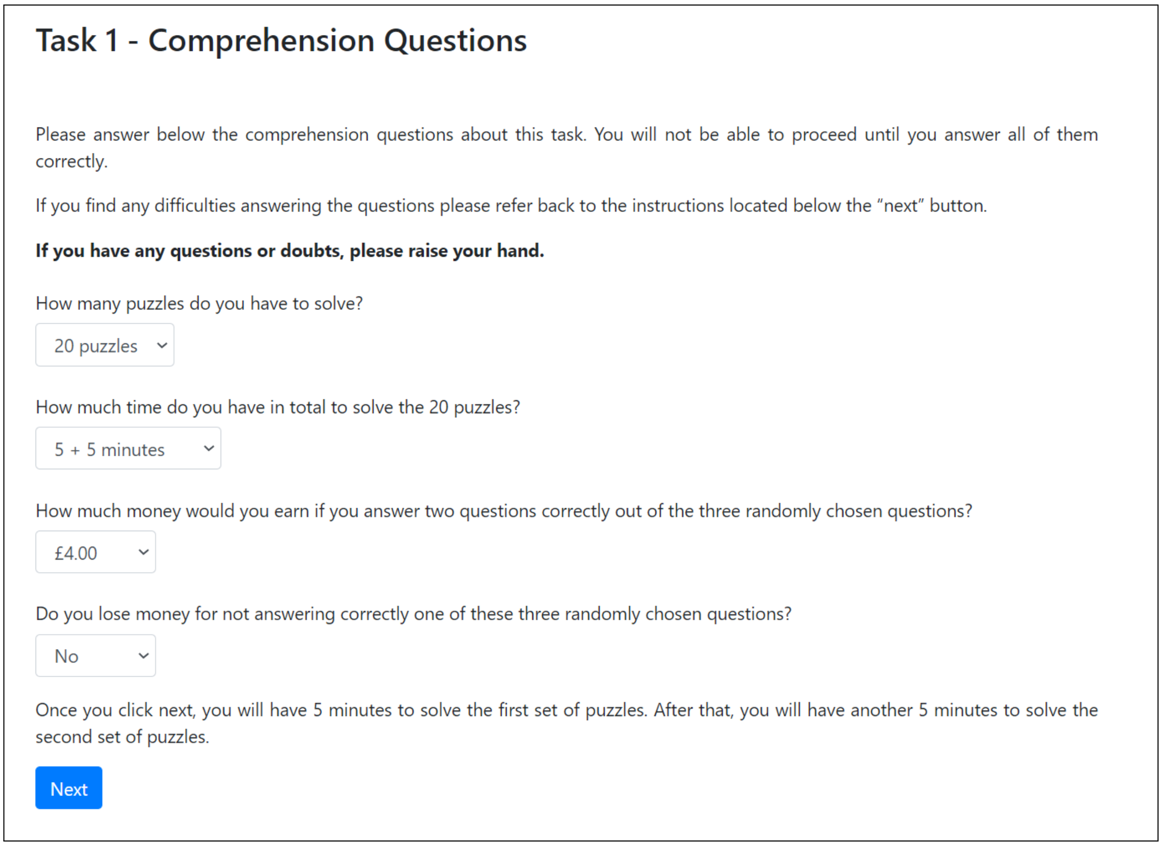
On the next page, you will be asked comprehension questions about the instructions. You can only proceed with the experiment if you have solved them correctly.

To continue with the comprehension questions for this task, please type in the cell below the number "10".

Next

Notes: The figure displays a screenshot of the template in which we explained to participants the IQ task.

Figure H.2: Screenshot of the comprehension quiz about the IQ task



Task 1 - Comprehension Questions

Please answer below the comprehension questions about this task. You will not be able to proceed until you answer all of them correctly.

If you find any difficulties answering the questions please refer back to the instructions located below the “next” button.

If you have any questions or doubts, please raise your hand.

How many puzzles do you have to solve?

20 puzzles

How much time do you have in total to solve the 20 puzzles?

5 + 5 minutes

How much money would you earn if you answer two questions correctly out of the three randomly chosen questions?

£4.00

Do you lose money for not answering correctly one of these three randomly chosen questions?

No

Once you click next, you will have 5 minutes to solve the first set of puzzles. After that, you will have another 5 minutes to solve the second set of puzzles.

Next

Notes: The figure displays a screenshot of the comprehension quiz about the IQ task. Subjects can only proceed once all questions are answered correctly.

Figure H.3: Screenshot of the instructions template about the belief elicitation mechanism

In this task, we will ask your beliefs regarding the probabilities of some events. The probabilities you state will affect your payment in this experiment. In particular, the payment method is such that you have the highest chance of earning (more) money by stating the true probability with which you think the event will occur.

We will now explain you how the payment method works. For ease of understanding, let us consider a specific event: The probability that it rained yesterday in New York. Note that this example is only for illustrative purposes, in the experiment it will be replaced by other events.

We give you £6.00 and you have to decide whether you want to place your bet on the lottery or the event:

- **The lottery**: you earn the £6.00 if a purple ball is drawn from an urn containing 100 purple and orange balls. The computer will randomly determine the composition of purple and orange balls (each possible composition is equally likely);
- **The event**: you earn the £6.00 if the event occurred (it rained yesterday in New York).

Should you place your bet on the lottery or the event (to have the highest chances of earning the £6.00)? This will depend on the number of purple balls in the urn and the probability you think it rained yesterday in New York.

For example, if there are 5 purple balls in the urn, the chance to win the £6.00 by picking the lottery is only 5%. Hence, most people would choose the event since the chance that it rained in New York is probably higher than 5%. However, if there are 90 purple balls in the urn, picking the event will only give you a higher chance to win the £6.00 if you believe the probability that it rained in New York is higher than 90%. Therefore in this case most people would pick the lottery.

How are you going to place your bet?

We will ask you to state your belief regarding the probability with which the event occurred. The mechanism ensures that it is optimal for you to state your true belief. In particular, this is because the number you give determines how many purple balls need to be there (in the lottery) for you to prefer to place your bet in the lottery instead of the event. The computer will then determine the composition of the urn. If there happen to be fewer (or equally many) purple balls than the minimum you chose, you will be betting on the event. Thus, you will earn £6.00 if it rained yesterday in New York and £0.00 otherwise. If there happen to be more purple balls than the minimum you chose, you will be betting on the lottery. Then the computer draws a ball. If the ball drawn is purple, you earn £6.00, if the ball is orange you earn £0.00. This method guarantees that you have the highest chances of earning the £6.00 if you report your true belief of the event occurring (if you wish, below the "next" button you can find a more detailed explanation of why this is the case).

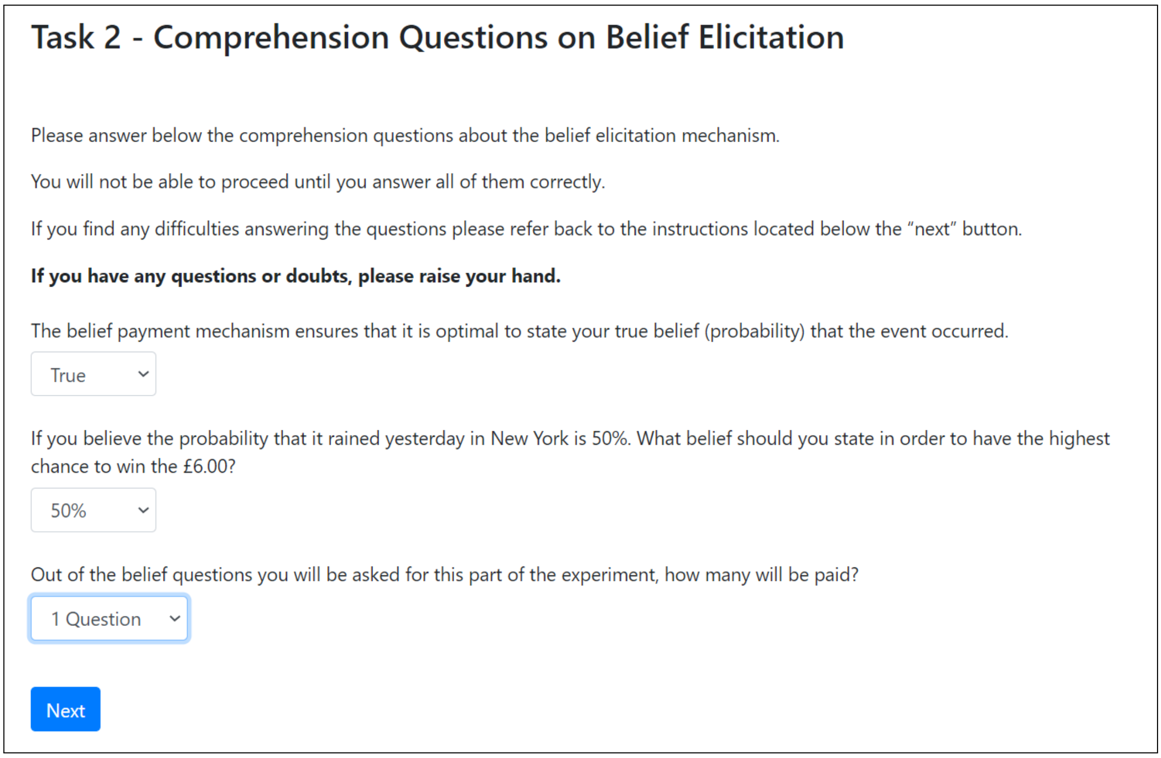
While this method may look complicated, its implications are simple: you have the highest chance of earning (more) money if you honestly report your best guess of the probability of the event occurring. For example, if you believe that it rained yesterday in New York with 80%, you should state 80%.

Task 2: Payment

In this task, we will ask you four belief questions. Out of the four, the computer will randomly draw one. The randomly drawn question will count for your payments (if this task is selected to count for payments). In particular, you will be paid for your answer in that question following the method explained here.

Notes: The figure displays a screenshot of the template in which we explained to participants the belief elicitation mechanism.

Figure H.4: Screenshot of the comprehension quiz about the belief elicitation mechanism



Task 2 - Comprehension Questions on Belief Elicitation

Please answer below the comprehension questions about the belief elicitation mechanism.

You will not be able to proceed until you answer all of them correctly.

If you find any difficulties answering the questions please refer back to the instructions located below the "next" button.

If you have any questions or doubts, please raise your hand.

The belief payment mechanism ensures that it is optimal to state your true belief (probability) that the event occurred.

True

If you believe the probability that it rained yesterday in New York is 50%. What belief should you state in order to have the highest chance to win the £6.00?

50%

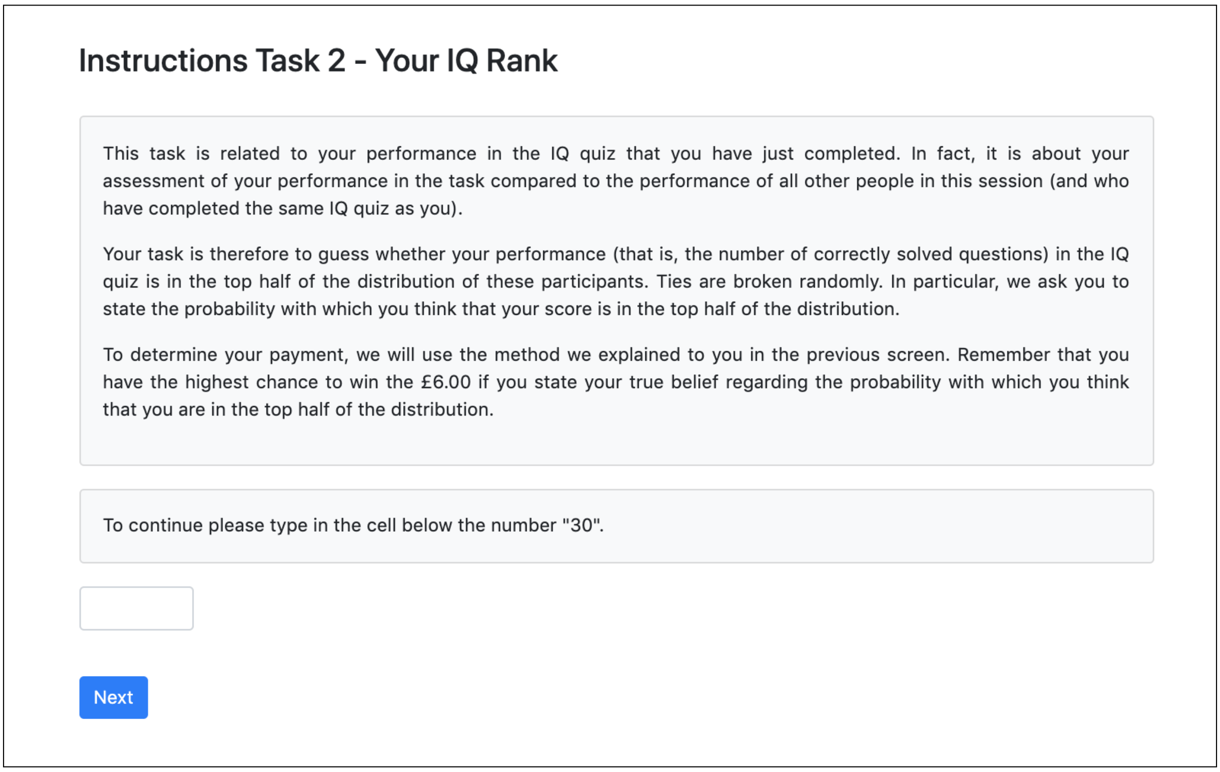
Out of the belief questions you will be asked for this part of the experiment, how many will be paid?

1 Question

Next

Notes: The figure displays a screenshot of the comprehension quiz about the belief elicitation mechanism. Subjects can only proceed once all questions are answered correctly.

Figure H.5: Screenshot of the instructions template about the rank determination in the IQ treatment



Instructions Task 2 - Your IQ Rank

This task is related to your performance in the IQ quiz that you have just completed. In fact, it is about your assessment of your performance in the task compared to the performance of all other people in this session (and who have completed the same IQ quiz as you).

Your task is therefore to guess whether your performance (that is, the number of correctly solved questions) in the IQ quiz is in the top half of the distribution of these participants. Ties are broken randomly. In particular, we ask you to state the probability with which you think that your score is in the top half of the distribution.

To determine your payment, we will use the method we explained to you in the previous screen. Remember that you have the highest chance to win the £6.00 if you state your true belief regarding the probability with which you think that you are in the top half of the distribution.

To continue please type in the cell below the number "30".

Next

Notes: The figure displays a screenshot of the template in which we explain how participant rank in the IQ treatment is determined.

Figure H.6: Screenshot of the instructions template about the rank determination in the random treatment

Instructions Task 2 - Your Number's Rank

From now on all the following tasks will **NOT** be related to the previous IQ quiz that you have just completed.

For this task the computer has drawn for you a random number between 1 and 100, with each number in this interval being equally likely to be drawn. We will call this draw your **number**.

In particular, the computer has drawn the following number:

10

Now, the computer will draw (with replacement) three other numbers (between 1 and 100 with each number in this interval being equally likely to be drawn). Unlike for your number, you will not know the realizations/draws of these three numbers.

Your task is to guess whether your number (10) is in the top half of the distribution of these four randomly drawn numbers (your number and the three other numbers drawn by the computer). Put differently, we ask you whether your number is among the two highest (top half) or among the two lowest numbers (bottom half). Ties are broken randomly. In particular, we ask you to state the probability with which you think that your number is in the top half of the distribution.

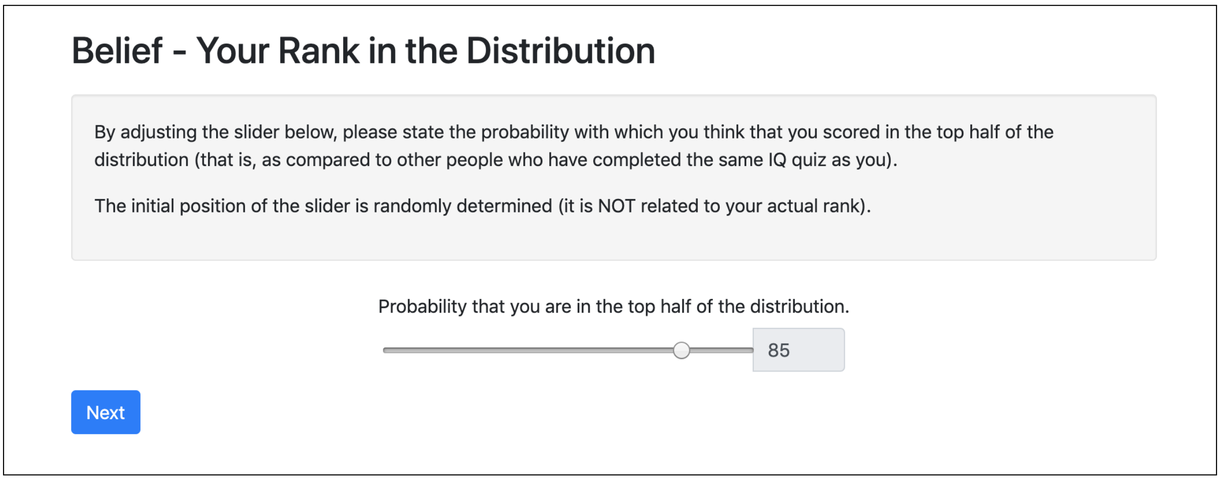
To determine your payment, we will use the method we explained to you in the previous screen. Remember that you have the highest chance to win the £6.00 if you state your true belief regarding the probability with which you think your number is in the top half of the distribution.

To continue please type in the cell below the number "30".

Next

Notes: The figure displays a screenshot of the template in which we explain how participant rank in the random treatment is determined.

Figure H.7: Screenshot of the prior belief elicitation template in the IQ treatment



Notes: The figure displays a screenshot of the template in which we asked the participant, in the IQ treatment, to state his/her prior belief about his/her relative rank. Accordingly, subjects in the random treatment are asked whether their number is in the top half of the distribution of randomly drawn numbers.

Figure H.8: Screenshot of the instructions template about the possible signals the participant can receive in the IQ/endogenous treatment

Instructions Task 2 - Feedback about your IQ Rank

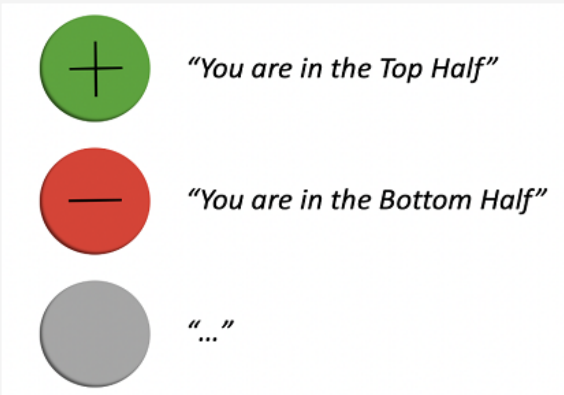
Now, you will receive additional information (feedback) about your performance in the IQ quiz to help you assess whether or not you are in the top half of the distribution.

What is Feedback?
Depending on your rank in the distribution, you will receive feedback about your rank. You can receive three types of feedback in the form of evaluations:

- The green evaluation that tells you: "You are in the Top Half";
- The red evaluation that tells you: "You are in the Bottom Half";
- The grey evaluation that tells you: "...".

Figure 1 shows you the exact three possible evaluations that you can receive.

Figure 1: Feedback



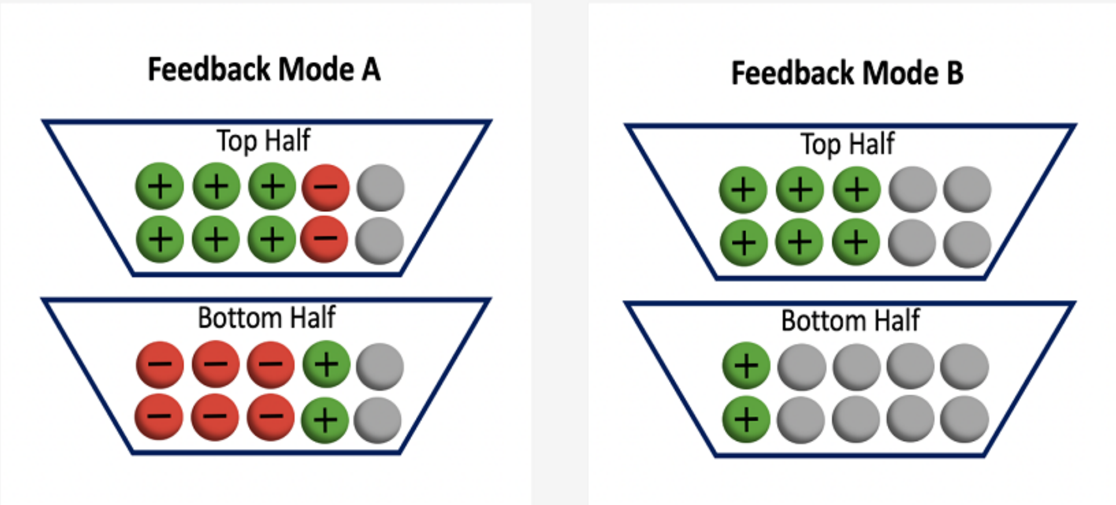
Notes: The figure displays a screenshot of the template in which we explain, in the IQ/endogenous treatment, the possible signals each participant can receive about the relative performance.

Figure H.9: Screenshot of the instruction template regarding the feedback mode selection (Order 1) in the IQ/endogenous treatment

How is the feedback determined?

Which evaluation (feedback) you receive depends on your actual rank in the distribution in the IQ quiz and the "feedback mode" from which the feedback is generated. However, the feedback does not completely reveal your rank in the distribution.

Let's consider an example of two feedback modes to make things clearer:



Notice that:

- Irrespective of the feedback mode you choose, if you are in the top half of the distribution your feedback will be determined by the urn at the top of each figure. If you are in the bottom half, your feedback will be determined by the urn at the bottom.
- Consider the example above. In both feedback modes you are more likely to get the green evaluation if you are in the top half of the distribution. However, if you choose to receive feedback from Feedback Mode A, you are more likely to get the red evaluation if you are in the bottom half of the distribution. In Feedback Mode B, in contrast to Feedback Mode A, you will never get the red evaluation but instead the grey evaluation. Thus, Feedback Mode A is more informative than Mode B in case that you are in the bottom half.

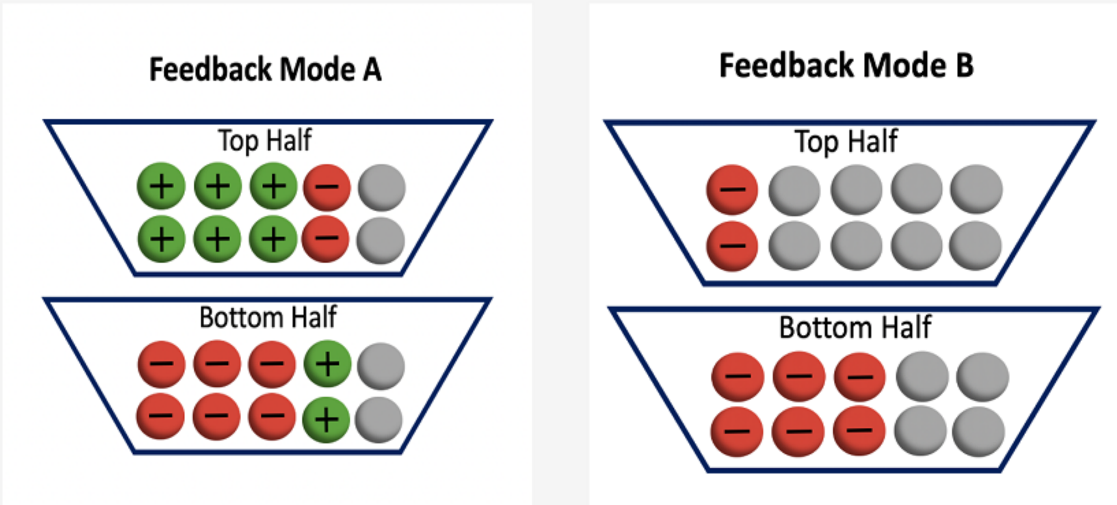
Notes: The figure displays a screenshot of the template in which we explain to the participant, in the IQ/endogenous treatment, the selection of feedback modes and how feedback modes differ. In Order 1, we use the baseline choice as an example.

Figure H.10: Screenshot of the instructions template about the feedback mode selection (Order 2) in the IQ/endogenous treatment

How is the feedback determined?

Which evaluation (feedback) you receive depends on your actual rank in the distribution in the IQ quiz and the "feedback mode" from which the feedback is generated. However, the feedback does not completely reveal your rank in the distribution.

Let's consider an example of two feedback modes to make things clearer:



Notice that:

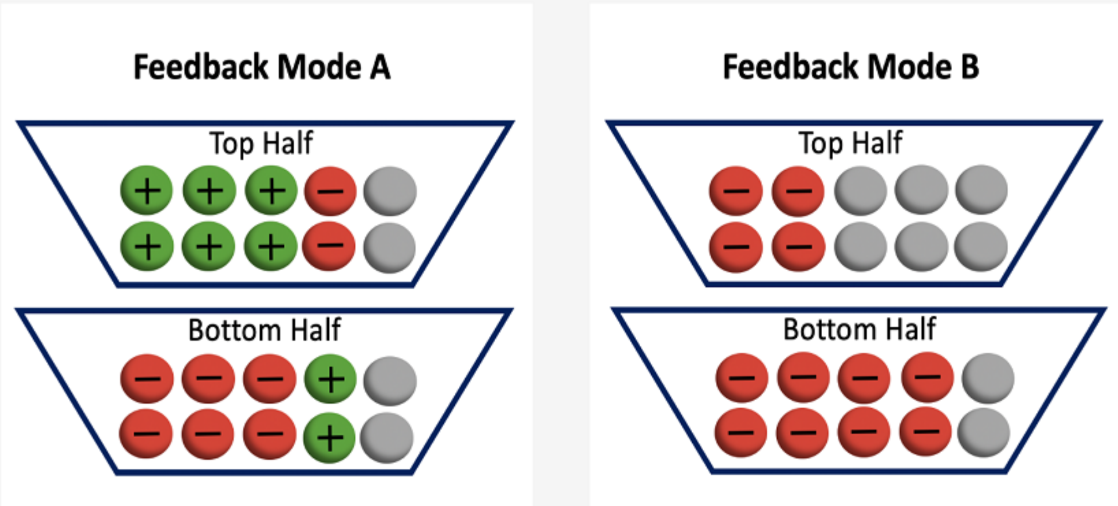
- Irrespective of the feedback mode you choose, if you are in the top half of the distribution your feedback will be determined by the urn at the top of each figure. If you are in the bottom half, your feedback will be determined by the urn at the bottom.
- Consider the example above. In both feedback modes you are more likely to get the red evaluation if you are in the bottom half of the distribution. However, if you choose to receive feedback from Feedback Mode A, you are more likely to get the green evaluation if you are in the top half of the distribution. In Feedback Mode B, in contrast to Feedback Mode A, you will never get the green evaluation but instead the grey evaluation. Thus, Feedback Mode A is more informative than Mode B in case that you are in the top half.

Notes: The figure displays a screenshot of the template in which we explain to the participant, in the IQ/endogenous treatment, the selection of feedback modes and how feedback modes differ. In Order 2, we use the baseline reversed choice as an example. Hence, Mode B in this screenshot is called Mode C in the remainder of the paper.

Figure H.11: Screenshot of the instructions template about the feedback mode selection (Order 3) in the IQ/endogenous treatment

How is the feedback determined?
Which evaluation (feedback) you receive depends on your actual rank in the distribution in the IQ quiz and the "feedback mode" from which the feedback is generated. However, the feedback does not completely reveal your rank in the distribution.

Let's consider an example of two feedback modes to make things clearer:



Notice that:

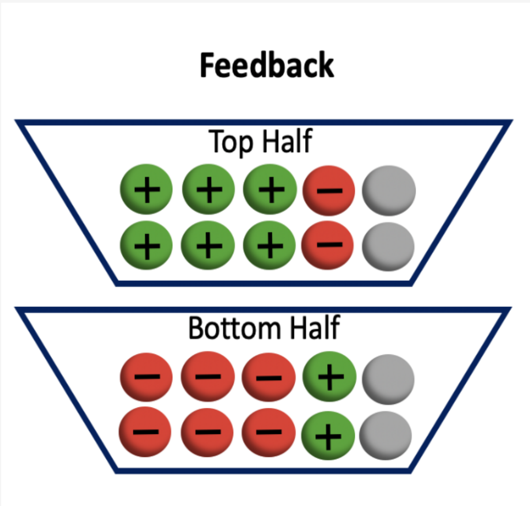
- Irrespective of the feedback mode you choose, if you are in the top half of the distribution your feedback will be determined by the urn at the top of each figure. If you are in the bottom half, your feedback will be determined by the urn at the bottom.
- Consider the example above. In both feedback modes you are more likely to get the red evaluation if you are in the bottom half of the distribution. However, if you choose to receive feedback from Feedback Mode A, you are more likely to get the green evaluation if you are in the top half of the distribution. In Feedback Mode B, in contrast to Feedback Mode A, you will never get the green evaluation. Note that Feedback Mode A is more informative than Mode B in case that you are in the bottom half.

Notes: The figure displays a screenshot of the template in which we explain to the participant, in the IQ/endogenous treatment, the selection of feedback modes and how feedback modes differ. In Order 3, we use the skewness over framing choice as an example. Hence, Mode B in this screenshot is called Mode E in the remainder of the paper.

Figure H.12: Screenshot of the instructions template about feedback mode A in the IQ/exogenous treatment

How is the feedback determined?

Which evaluation you receive depends on your actual rank in the distribution in the IQ quiz. If you are in the top half of the distribution, your feedback will be determined by the urn at the top of the figure. If you are in the bottom half, your feedback will be determined by the urn at the bottom. However, the feedback does not completely reveal your rank in the distribution.



Notice that:

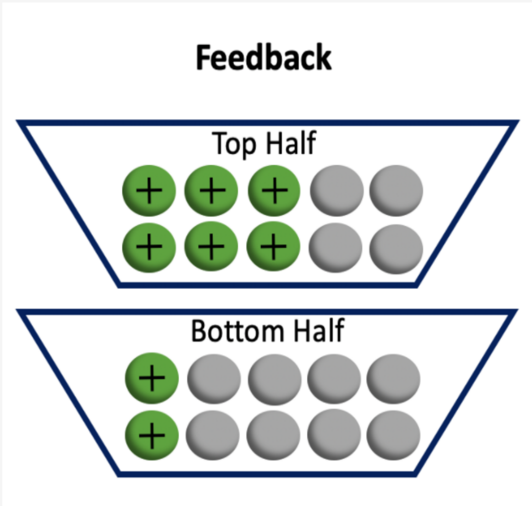
- You are more likely to get the green evaluation if you are in the top half of the distribution.
- You are more likely to get the red evaluation if you are in the bottom half of the distribution.

Notes: The figure displays a screenshot of the template in which we explain to the participant, in the IQ/exogenous treatment, how signals are drawn. In this example, the participant is exogenously assigned to Mode A.

Figure H.13: Screenshot of the instructions template about feedback mode B in the IQ/exogenous treatment

How is the feedback determined?

Which evaluation you receive depends on your actual rank in the distribution in the IQ quiz. If you are in the top half of the distribution, your feedback will be determined by the urn at the top of the figure. If you are in the bottom half, your feedback will be determined by the urn at the bottom. However, the feedback does not completely reveal your rank in the distribution.



Notice that:

- You are more likely to get the green evaluation if you are in the top half of the distribution.
- You are more likely to get the grey evaluation if you are in the bottom half of the distribution.

Notes: The figure displays a screenshot of the template in which we explain to the participant, in the IQ/exogenous treatment, how signals are drawn. In this example, the participant is exogenously assigned to Mode B.

Figure H.14: Screenshot of the instructions template about feedback in the IQ treatment

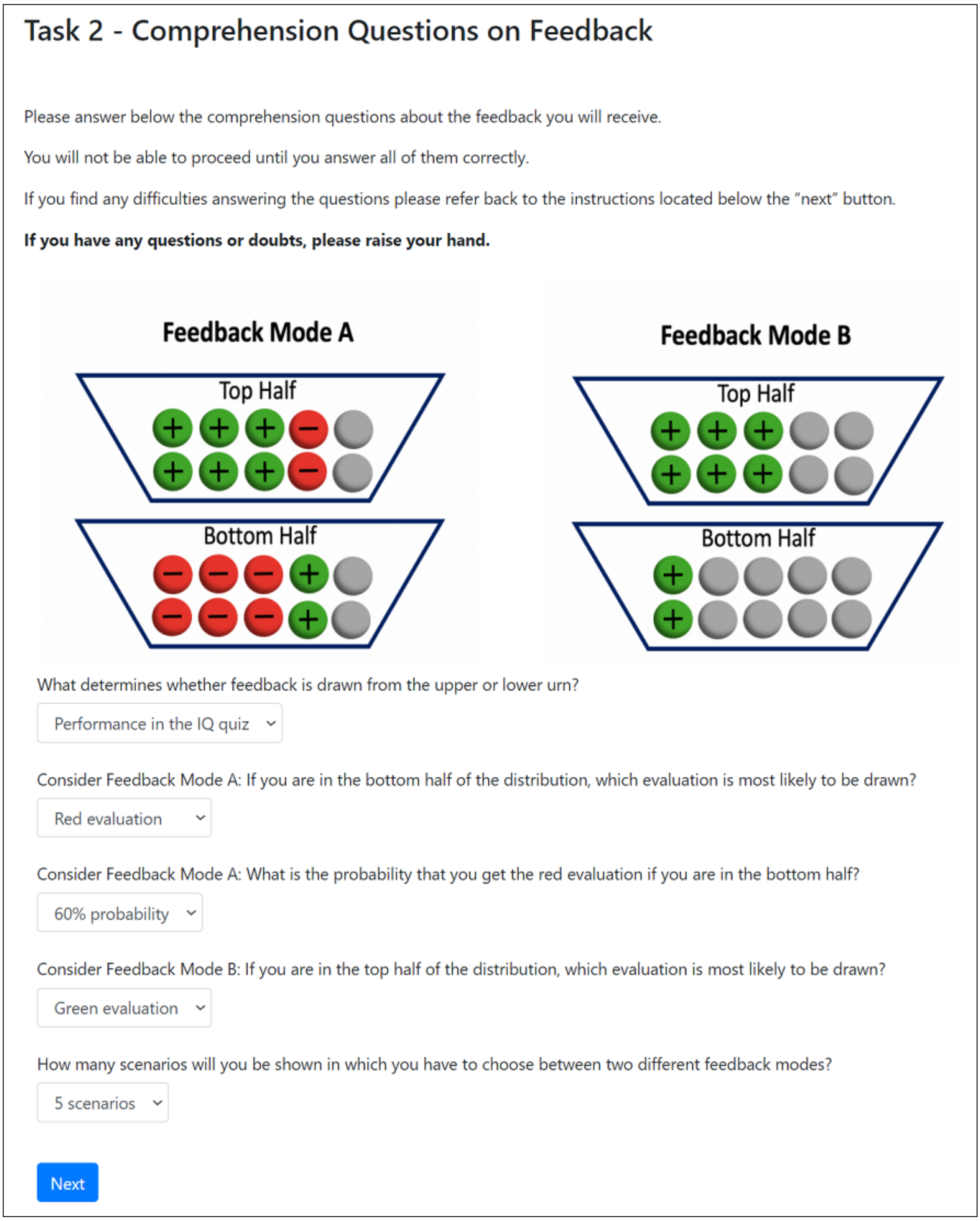
Three Rounds of Feedback

In particular, you will be provided with **three rounds of feedback**. In each round, the computer will randomly draw an evaluation from the urn that corresponds to your IQ rank. Realize that the evaluations will be drawn one at a time and with replacement. That is, after the evaluation is drawn, it will be put again in the urn and only then the following evaluation will be drawn.

After each evaluation that you receive, you will have to state again the probability with which you think you are in the top half of the distribution. One of these questions could be randomly drawn to count for payments. If so, we will pay you following the method described to you earlier: **Depending on the accuracy of your answer you could either earn £6.00 or nothing.**

Notes: The figure displays a screenshot of the template in which we explain to the participant, in the IQ treatment, how feedback is provided.

Figure H.15: Screenshot of the comprehension quiz about the feedback selection in the IQ/endogenous treatment


Task 2 - Comprehension Questions on Feedback

Please answer below the comprehension questions about the feedback you will receive.

You will not be able to proceed until you answer all of them correctly.

If you find any difficulties answering the questions please refer back to the instructions located below the "next" button.

If you have any questions or doubts, please raise your hand.

What determines whether feedback is drawn from the upper or lower urn?

Performance in the IQ quiz

Consider Feedback Mode A: If you are in the bottom half of the distribution, which evaluation is most likely to be drawn?

Red evaluation

Consider Feedback Mode A: What is the probability that you get the red evaluation if you are in the bottom half?

60% probability

Consider Feedback Mode B: If you are in the top half of the distribution, which evaluation is most likely to be drawn?

Green evaluation

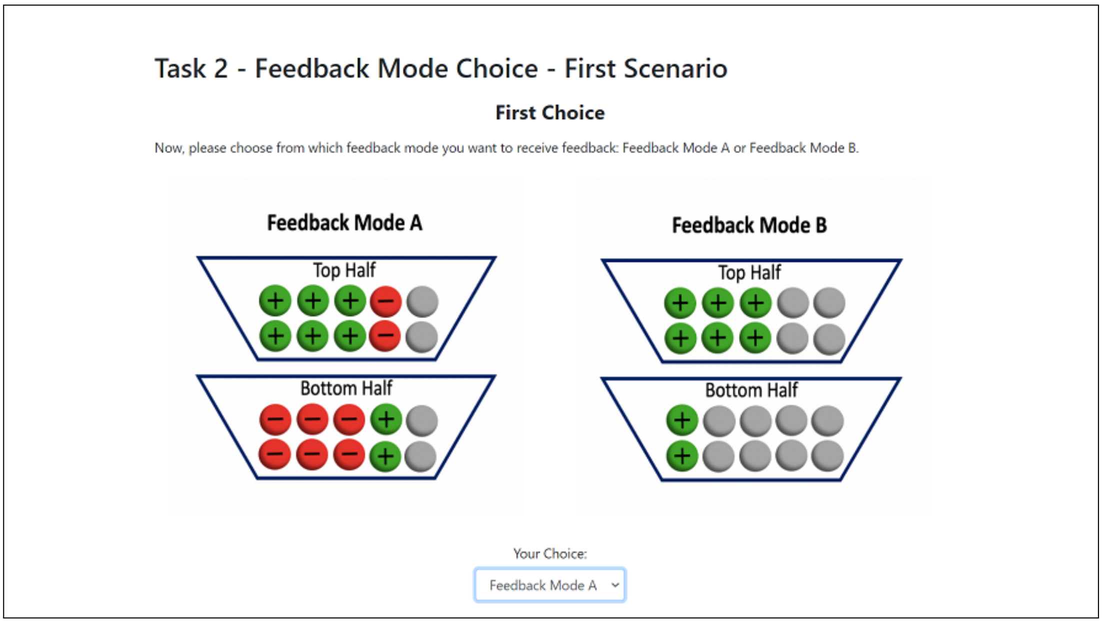
How many scenarios will you be shown in which you have to choose between two different feedback modes?

5 scenarios

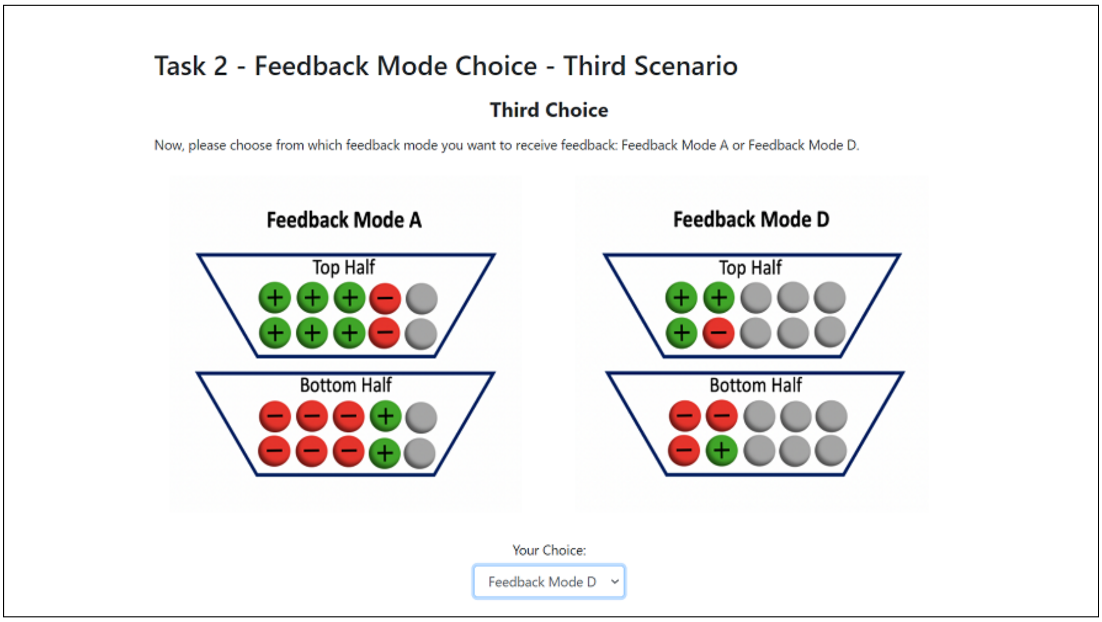
Next

Notes: The figure displays a screenshot of the comprehension quiz about the feedback selection. Subjects can only proceed once all questions are answered correctly.

Figure H.16: Screenshot of feedback selection in the endogenous treatment



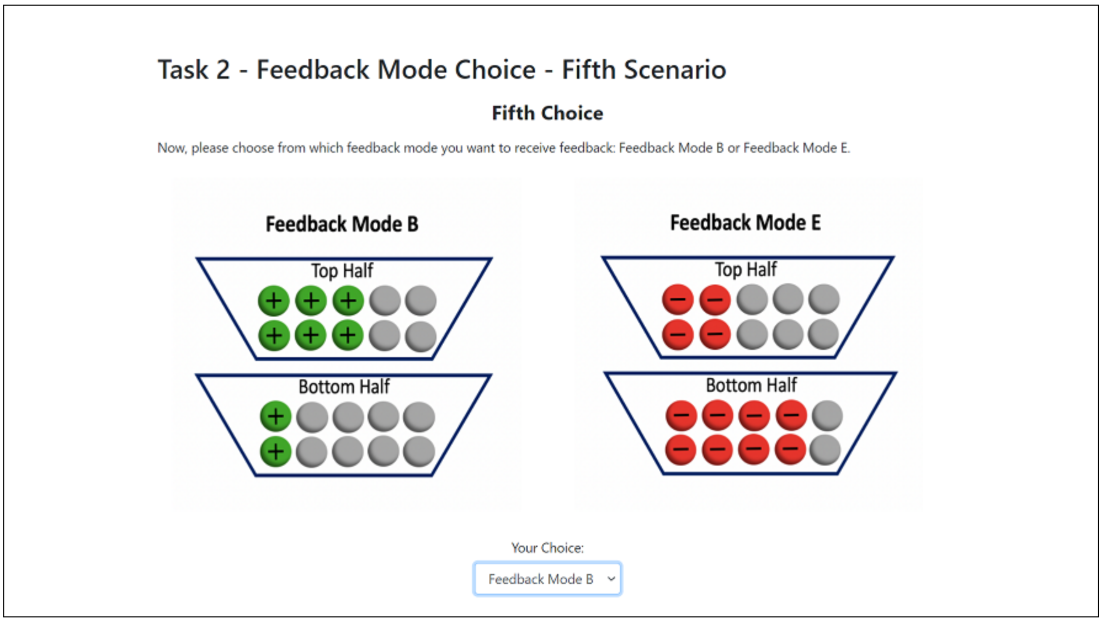
(a) Baseline choice



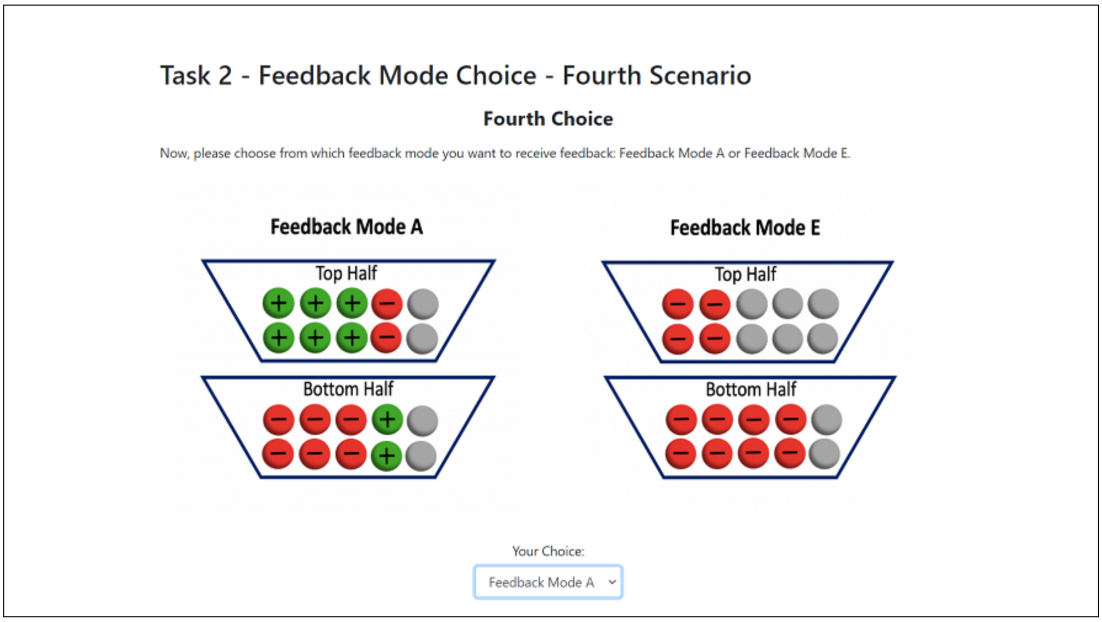
(b) Informativeness choice

Notes: The figure displays a screenshot of the templates in which we asked the participant, in the endogenous treatment, to select the treatment from which they would like to receive feedback. Each information structure choice is presented on a separate screen. In the depicted example, the information structures are presented in Order 1 (cf. Appendix D).

Figure H.17: Screenshot of feedback selection in the endogenous treatment



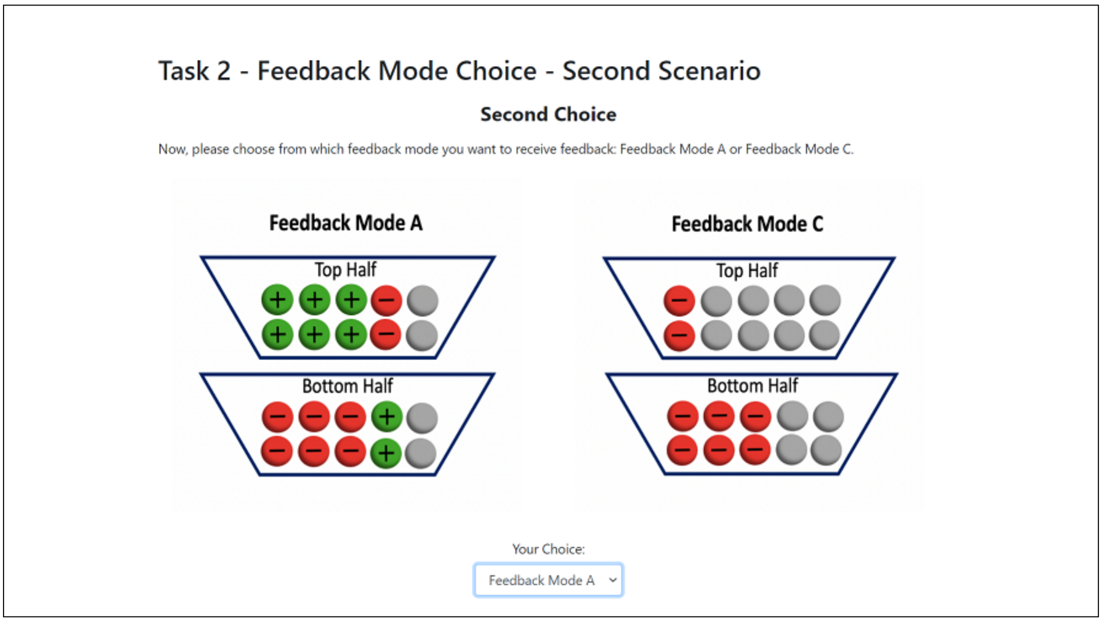
(a) Salience choice



(b) Skewness over Salience choice

Notes: The figure displays a screenshot of the templates in which we asked the participant, in the endogenous treatment, to select the treatment from which they would like to receive feedback. Each information structure choice is presented on a separate screen. In the depicted example, the information structures are presented in Order 1 (cf. Appendix D).

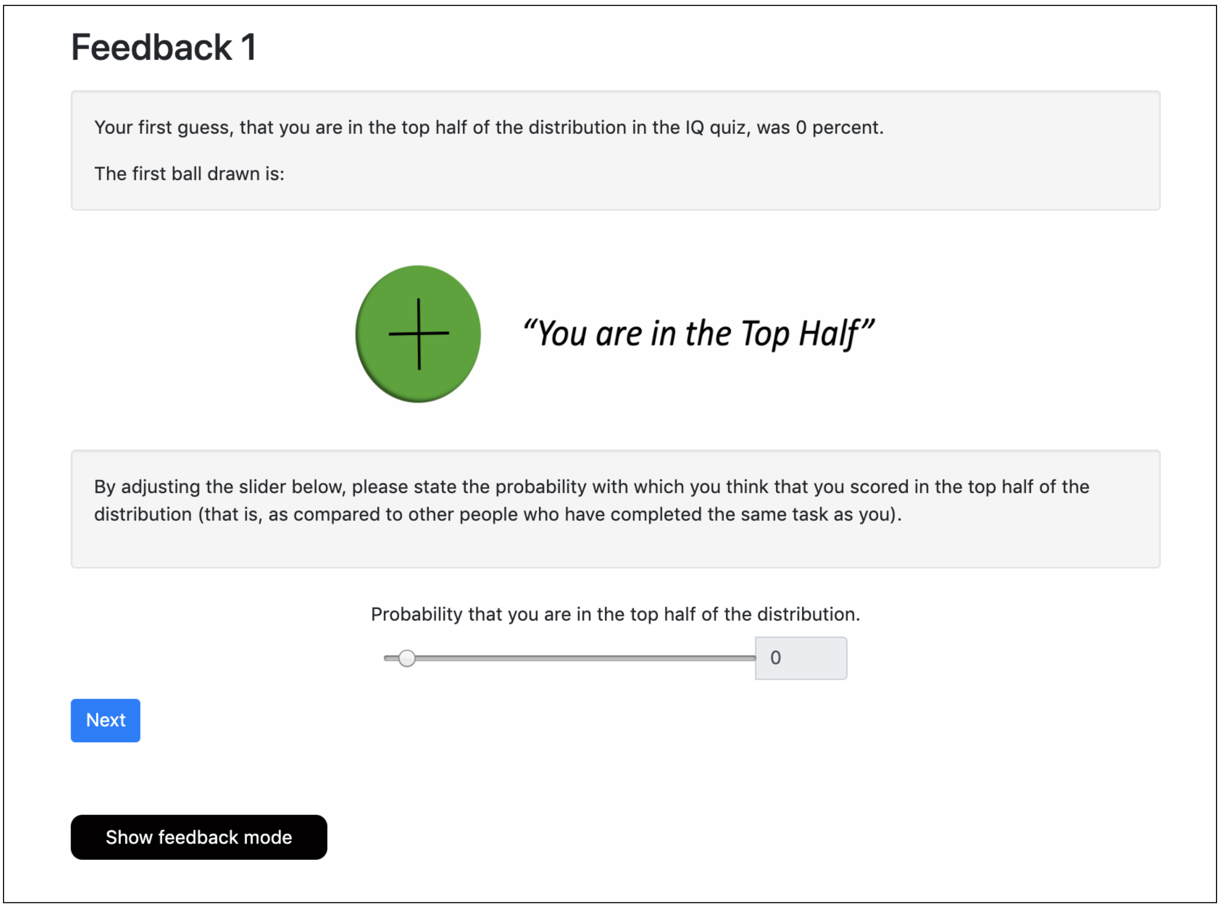
Figure H.18: Screenshot of feedback selection in the endogenous treatment



(a) Baseline Reversed choice

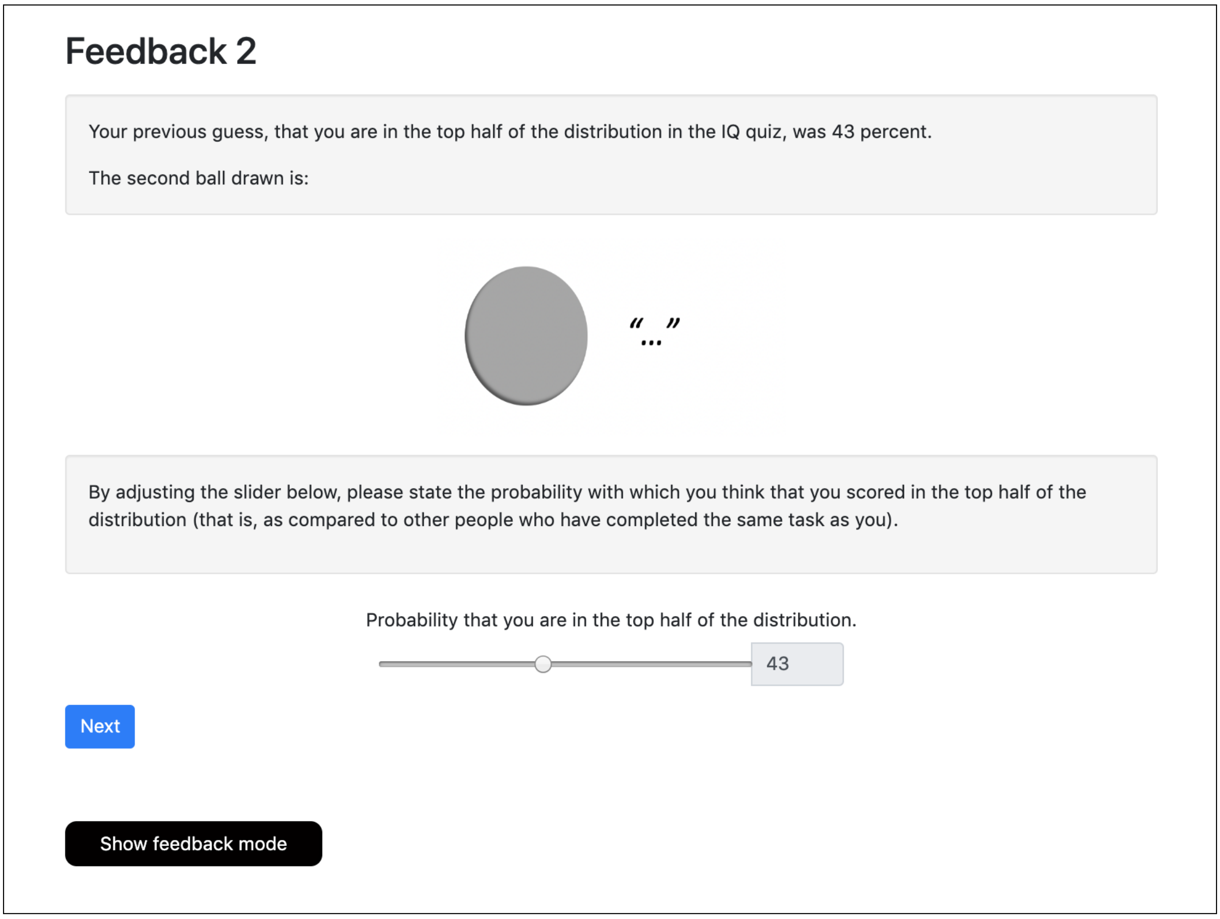
Notes: The figure displays a screenshot of the templates in which we asked the participant, in the endogenous treatment, to select the treatment from which they would like to receive feedback. Each information structure choice is presented on a separate screen. In the depicted example, the information structures are presented in Order 1 (cf. Appendix D).

Figure H.19: Screenshot of the posterior belief elicitation template in the IQ treatment following a green (+) signal



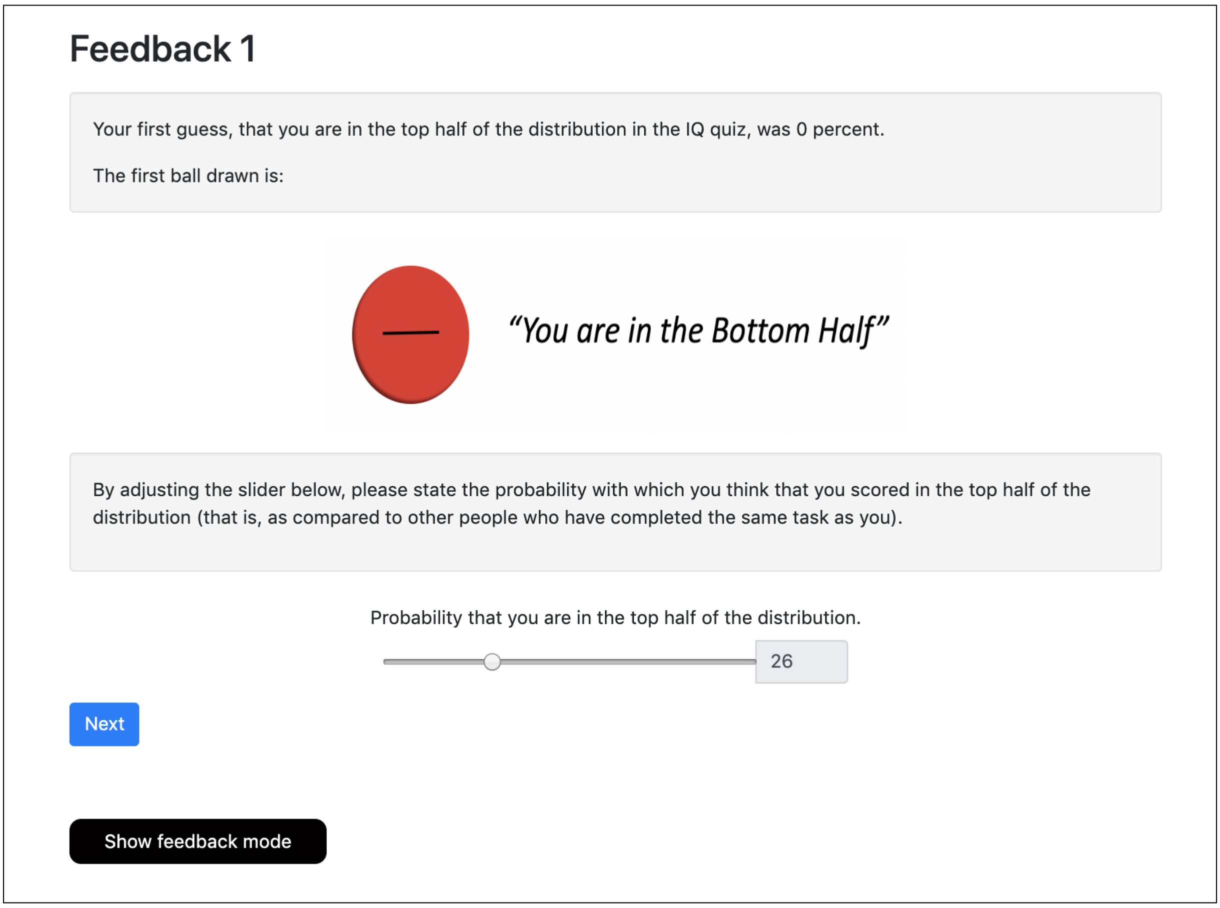
Notes: The figure displays a screenshot of the template in which we asked the participant, in the IQ treatment, to state his/her posterior belief about his/her relative rank following a green (+) signal.

Figure H.20: Screenshot of the posterior belief elicitation template in the IQ treatment following a grey signal



Notes: The figure displays a screenshot of the template in which we asked the participant, in the IQ treatment, to state his/her posterior belief about his/her relative rank following a grey signal.

Figure H.21: Screenshot of the posterior belief elicitation template in the IQ treatment following a red (-) signal



Notes: The figure displays a screenshot of the template in which we asked the participant, in the IQ treatment, to state his/her posterior belief about his/her relative rank following a red (-) signal.