Protecting the Ego: Motivated Information Selection and Updating*

Alessandro Castagnetti†

Renke Schmacker‡

May 24, 2020

Abstract

We investigate how individuals search for ego-relevant information and then update their beliefs. In our lab experiment, subjects can select the information structure that gives them feedback regarding their rank in the IQ distribution (ego-relevant treatment) or regarding a random number (control treatment). We find that individuals in the ego-relevant treatment select information structures, in which negative feedback is less salient. When receiving such negative feedback with lower salience they update their beliefs less, but only when feedback is ego-relevant. Hence, subjects select information structures that allow them to misinterpret negative feedback in a self-serving way. Moreover, individuals in the ego-relevant feedback choose less informative feedback.

JEL-codes: D83, D91, C91

Keywords: motivated beliefs, updating, information acquisition

*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 benefitted from discussions with conference participants at the Bogotá Experimental Economics Conference 2020, Nordic Conference on Behavioral and Experimental Economics (Kiel), European Meeting of the ESA (Dijon), Dr@w Forum (Warwick), and 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 Leibniz Competition through the project SAW-2015-DIW-4 is gratefully acknowledged.

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

‡DIW and Free University Berlin, `rschmacker[at]diw[dot]de`

1 Introduction

In many situations individuals can select what kind of feedback they receive. Standard models predict that individuals have a strict preference for choosing the most informative feedback source because this leads to more accurate beliefs and, thus, better-informed decisions. However, if individuals have motivated beliefs, i.e. if they derive utility from 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. Consequently, motivated information selection can facilitate overconfident beliefs.[1]

Information structures in the real world differ in the way information is transmitted. In many instances, the desire to maintain a positive self-view may affect preferences over information structures.For example, individuals choose between mentors and supervisors with different feedback styles. Some supervisors always give explicit positive and negative feedback; others may give positive feedback but withhold negative feedback. In the latter example, the implicit negative feedback is less salient, so messages can be interpreted in a self-serving way. Similarly, students select into majors with different grading policies. They may prefer majors in which it is easier to receive a high absolute grade even though compressed grading distributions render grades less informative (Sabot and Wakeman-Linn, 1991; Ahn *et al.*, 2019).

In this paper, we report results from a lab experiment showing that individuals select information structures in order to protect their belief that they have high ability. We provide evidence for two mechanisms of ego protection. First, individuals choose information structures, in which negative feedback is less salient. Whereas the salience of feedback should not matter for a pure Bayesian updater, our results suggest that individuals choose less salient negative feedback because it facilitates misinterpreting negative feedback in a self-serving way. Second, when feedback is ego-relevant, individuals prefer less informative feedback structures in general. This is to avoid the risk of having to adjust their beliefs downwards.

In the experiment, subjects are asked to form beliefs about the probability of being in the top half of the distribution of subjects. The distribution is either based on (1) performance in an IQ test, or (2) a randomly drawn number. This creates exogenous variation in a between-subject design in whether the rank in the distribution is ego-relevant or not. First, we elicit participants' prior beliefs about whether they are in the top half of the distribution. We then give subjects three consecutive (noisy) signals informing them whether they are in the top

[1] Von Hippel and Trivers (2011) argue that self-deception and overconfidence evolved as an interpersonal strategy to gain a strategic advantage in persuading others. This conjecture is experimentally supported by Schwardmann and Van der Weele (2019), Solda *et al.* (2019), and Smith *et al.* (2017). However, holding inaccurate beliefs about ability is also costly in many domains, e.g., it leads to suboptimal management decisions (Malmendier and Tate, 2005) or over-entry into competition (Camerer and Lovallo, 1999).

half of the distribution. After each signal, we elicit the corresponding posterior belief in an incentive compatible way.

The key feature of our design is that we vary whether subjects receive signals from information structures, which they have selected themselves (endogenous treatment), or from information structures they are exogenously allocated to (exogenous treatment). While the endogenous treatment allows us to study subjects' information preferences, the exogenous treatment enables us to investigate updating behavior absent potential selection. 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 or the other urn. By varying the composition of signals in the urns, we vary the informativeness, salience, and skewness of feedback. *Informativeness* describes by how much beliefs are shifted by a given signal. *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. For example, an information structure is positively skewed if a potential positive signal shifts the posterior more than a negative signal.

In the endogenous treatment, subjects make five pairwise choices between information structures that vary in informativeness, salience, and skewness. Since beliefs are incentivized, we expect subjects to select the most informative feedback structure if beliefs are not ego-relevant. In contrast, subjects who derive utility from believing that they rank high in the IQ distribution may choose an information structure that is less informative, positively skewed, and makes negative feedback less salient, even if it means forming less accurate posterior beliefs.

We find stark differences in the way individuals seek information when the rank is ego-relevant and when it is not. When the ego is at stake, subjects are more likely to choose information structures that are less informative and that make negative feedback less salient. The results do not support the idea that individuals choose information structures that are positively skewed to protect their ego. Our findings are based on the analysis of information structure choices separately, and reinforced by looking at the within-individual choice patterns. Furthermore, we find that subjects who are classified as information avoiding according to the Information Preference Scale by Ho *et al.* (forthcoming) are more likely to choose an information structure that is less informative and features less salient negative feedback.[2]

Moreover, we find that the subsequent belief updating process is heavily influenced by the information structure selected. Individuals in the IQ treatment update less to negative

[2]We do not find evidence for heterogenous information preferences by gender, cognitive ability, or prior belief.

feedback, but only when negative feedback is less salient. We find a first indication for this in the endogenous treatment, where subjects receive signals from the information structure that they select into. The results are corroborated by our exogenous treatment in which subjects are placed into information structures to eliminate potential selection issues. Subjects update asymmetrically given ego-relevant signals, but only when the framing of the signals allows them to do so. When signals are not ego-relevant, subjects update to positive and negative feedback irrespective of its salience. Therefore we can reject the hypothesis that subjects not understanding less salient signals is the explanation for asymmetric updating in the ego-relevant treatment.

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 that it can be optimal for individuals to avoid or distort feedback when believing that one has high ability has a (concave) consumption value. However, since belief distortion in their two-selves model is costly, manipulation of beliefs is only beneficial within the realms of the "reality constraints." In our experiment, we show that reducing the salience of negative feedback can be one way to relax these reality constraints, leading individuals to hold overconfident beliefs about their intelligence. In fact, our results show that subjects who receive feedback that is less informative and in which negative feedback is less salient maintain overconfident beliefs about their intelligence. In contrast, subjects who receive balanced feedback are, on average, no longer overconfident about their rank at the end of the experiment.[3]

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 – general systematic errors regarding how people search and process new information (e.g. confirmation bias) – from motivated biases – biases that are driven by a desire to hold positive views of oneself. In the discussion section, we provide evidence that the treatment differences cannot be explained by cognitive biases like confirmation- or contradiction-seeking behavior, differences in cognitive ability, or confusion about the experimental design.

Our findings on motivated information selection contribute to the burgeoning literature on the production and maintenance of self-serving beliefs about oneself. Bénabou and Tirole (2016) claim that when self-relevant beliefs are involved, people tend to process information differently depending not just on its valence, but also in terms of attention, interpretation, and memory.[4] For instance, people tend to ignore or discount negative news, while more readily incorporating good news into their (posterior) beliefs. However, the resulting experimental

[3]Balanced feedback in our setting describes an information structure that is neither positively nor negatively skewed and in which both positive and negative signals are framed explicitly.

[4]A related literature on motivated reasoning pertains to 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.* (2020) show in a series of online experiments that people's demand for political news goes beyond the desire for acquiring more informative news.

evidence on this mechanism – asymmetric updating – is mixed.[5] On the one hand, Eil and Rao (2011), Möbius *et al.* (2014), and Charness and Dave (2017) find positive asymmetry in updating. On the other hand, some studies either find no asymmetry (Grossman and Owens, 2012; Schwardmann and Van der Weele, 2019; Barron, 2020; Gotthard-Real, 2017; Buser *et al.*, 2018) or even the opposite asymmetry (Ertac, 2011; Kuhnen, 2015; Coutts, 2019). 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 qualifies the existing results by showing that asymmetric updating is only observed when the information structure enables subjects to interpret the signals in a self-serving way.

Our results also contribute to the literature on information avoidance.[6] Eil and Rao (2011) and Möbius *et al.* (2014) present experimental evidence that a significant proportion of subjects who have received prior noisy information regarding their relative rank in an ego-relevant task (i.e., intelligence and attractiveness) have a negative willingness to pay to have their rank fully revealed. However, our study goes beyond pure information avoidance. First, in our study, subjects "choose" the signals that they would like to receive before any feedback is given. We identify preferences about information where information is open to interpretation and subjects can still remain in denial, while in the previous experiments information fully reveals the state. Second, we not only look at preferences for information avoidance, but we also seek to learn how individuals' preferences for information structures depend on the skewness and salience of feedback. Finally, we aim to understand whether there are systematic interactions between information source selection and updating behavior. In particular, our goal is to learn if and how motivated feedback selection interacts with belief formation leading to biased updating and overconfidence.

Our results also relate to an emerging literature on how complexity in the environment influences belief updating. While Epstein and Halevy (2019) and Fryer *et al.* (2019) find that ambiguous signals increase deviations from Bayes rule, Enke (forthcoming) and Jin *et al.* (2018) illustrate that individuals find it difficult to make inferences from the absence of signals. In contrast to this literature, we look at how the informativeness and the framing of the signals affect updating. Moreover, by varying the ego-relevance of the state in a between-subject design, we can analyze the connection between cognitive and motivated biases in updating.

[5] 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.

[6] For a review of the information avoidance literature, see the survey of Golman *et al.* (2017). In the health domain, Oster *et al.* (2013) and Ganguly and Tasoff (2016) 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.

Finally, our research contributes 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 do not have any instrumental value (Falk and Zimmermann, 2016; Zimmermann, 2014; Nielsen, 2018). For example, Masatlioglu *et al.* (2017) find that individuals have a preference for positively skewed information sources; i.e., information structures that resolve more uncertainty regarding the desired outcome than the undesired one. Closer to our setting, some experimental papers study preferences towards information structures in settings where information has instrumental value. Charness *et al.* (2018) and Montanari and Nunnari (2019) study how people seek information from biased information structures. The findings of both papers show that a significant fraction of individuals make suboptimal choices. However, unlike these papers, our goal is to understand how the sub-optimality of information acquisition is driven by ego-relevant motives.

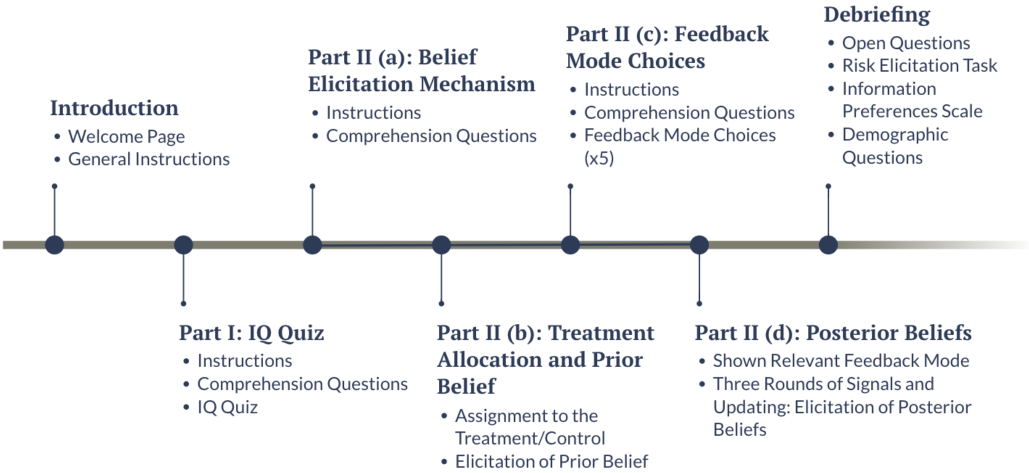
The remainder of this paper is organized as follows. In Section 2, we describe our experimental design, which comprises two treatment variations: ego-relevance of the rank and endogenous/exogenous information structure allocation. In Section 3, we present our experimental results. First, we study how 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 particular, we rule out cognitive biases as an alternative explanation for our main results. Finally, 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 ego-relevance of beliefs; (2) choices between different information structures; and (3) elicitation of updating behavior within different information structures. In a between-subject design we vary whether subjects receive feedback about their relative rank in IQ test performance (IQ treatment) or about a random number (random treatment). Separately, 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).

Figure 1 is an overview of the experiment. It has two parts and only one is randomly selected for payout. In Part I, subjects are paid for their performance on an 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 in Part II the IQ performance is only relevant for subjects in the IQ treatment, all subjects solve the IQ quiz in Part I. This ensures that there are no systematic differences in fatigue, timing, or average earnings between treatments.

Figure 1: Timeline of the Experiment



In Part II, subjects express their belief 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, in Part II (a), we explain to the subjects the matching probabilities method (Karni, 2009), so that subjects maximize their chance to win a prize of £6.00 by stating their true belief (see Appendix B). In Part II (b), subjects give their prior belief about either their IQ performance rank or their random number depending on the treatment they are in (see Figure G.7 for a screenshot). Afterwards, in Part II (c), subjects in the endogenous treatment choose the information structure (feedback mode) they would like to receive signals from. Finally, in Part II (d), there are three rounds of feedback and we elicit posterior beliefs. One out of the four belief elicitations is randomly selected for payout.

Subjects are incentivized to give their true belief, so a payoff-maximizing subject would always choose the most informative information structure and update according to Bayes rule. However, in the IQ treatment the motive to maximize payout can conflict with the desire to protect one's own beliefs about one's (relative) ability. For instance, subjects may forego the expected payoff in order to not impair their belief that they have high intelligence. 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 or not. In line with the experimental literature on motivated beliefs, we assume that further deviations from the rational benchmark are constant between the IQ and random treatments (this assumption is discussed in Section 4).

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. Depending on the treatment, we tell subjects at the beginning of Part II whether we will ask for their beliefs about the IQ performance or the random number. Consistent with previous research, we argue that rank in the IQ treatment is ego-relevant (e.g. Eil and Rao, 2011). To increase the ego-relevance of the IQ treatment, we explicitly told 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).

2.1.1 IQ Treatment

In the IQ treatment, we informed subjects that the second part of the experiment was related to their relative performance on the IQ test they completed in Part I. We told them that the computer divided participants in their session into two groups: one group of subjects whose score was in the top half of the score distribution and the other with scores in the bottom half. The task was to assess whether their IQ performance was in the top or bottom half of the distribution, compared to all other participants in their session.

2.1.2 Random Treatment

In the random treatment, subjects were shown a randomly drawn number between 1 and 100. We told subjects that three other numbers between 1 and 100 (with replacement) had been drawn. They were not shown these numbers. Their task was to assess whether the number they saw was in the top or bottom half of the distribution among these four numbers.[7] The four numbers were randomized at the individual level. The task was deliberately designed to have variation in prior beliefs.[8] If subjects were not given the drawn number or if we compared their number against many numbers, we would expect a degenerate prior distribution.

2.2 Information Structure Selection

2.2.1 Feedback Modes

Table 1 shows the information structures in the experiment. Information structures consist of two urns with ten balls each. A ball drawn from an urn in the selected information structure

[7] In both conditions, subjects were told that ties would be broken randomly.

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

constitutes a signal. If an individuals' IQ score or random number is in the top (bottom) half of the distribution, balls are drawn from the upper (lower) urn with replacement. Every subject receives three independently drawn signals from the urn.

Depending on the information structure, subjects can receive up to three different types of (noisy) signals. Figure G.1 displays how the signals are introduced in the instructions. 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 "...". On the same page, we explain that the informational content of the respective signal depends on the feedback mode and the state (see Figure G.3 for a screenshot). 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.[9] 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. The informativeness of signals in our experiment can be described, first, by their likelihood ratio (LR) and, second, by the probability of getting a non-informative signal. Both of these properties are given in the bottom panel of Table 1. The further away from one the likelihood ratio is, the more informative is 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).[10] The probability of receiving a non-informative signal only applies to Modes A and D, where grey signals are not informative (hence, Mode A is more informative than D).

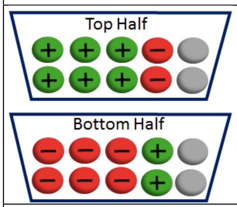
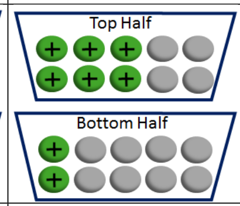
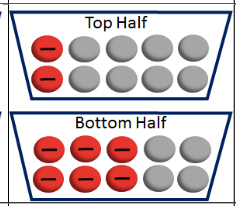
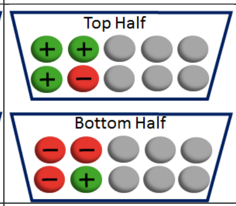
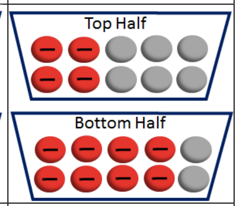
We call an information structure positively skewed if the positive signals are more informative than the negative signals (as in Modes B and E) and negatively skewed if the negative signals are more informative (as in Mode C). An information structure is symmetric if positive and negative signals are equally informative (as in Modes A and D).

Finally, information structures differ in the salience of feedback. Since the color of the grey signal is not associated with positive or negative states and the description is not explicit, we call positive or negative feedback in the form of grey signals *less salient*. For example, in Mode B negative feedback is less salient, while in Modes C and E positive feedback is less salient. An information structure has a balanced framing if positive signals are green (+)

[9] 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 G.3 to G.5, and check if the presented example matters for choices in Appendix D. In the exogenous treatment, we explain the signals using the feedback mode that the participant is assigned to: see the screenshot in Figure G.6.

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

Table 1: Feedback Modes

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

Notes: 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 the informativeness of the signal. E.g., 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 being in the bottom half. Prob(No Info) describes the probability of receiving a non-informative signal.

balls, negative signals are red (-) balls, and non-informative signals are grey balls (as in Mode A and D).

2.2.2 Feedback Mode Choices

We let subjects make five pairwise choices (which we call “scenarios”) 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. Every subject makes all five choices. To control for order effects, we vary the order of information structures (cf. Appendix D). Subjects are told that one of these five pairwise choices will be randomly selected and that the information structure chosen will be used to provide feedback.

Baseline Choice: Mode A vs Mode B In the baseline choice, we make 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., the 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 subject 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 scenarios, we aim to disentangle the underlying preferences for informativeness, salience, and skewness.

Informativeness Choice: Mode A vs Mode D First, individuals might have a preference for less informative feedback structures if information is ego-relevant. The choice between Modes A and D isolates a preference for informativeness. In both 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, e.g., reducing the salience of negative feedback if information is ego-relevant. This could be simple aversion to explicit negative feedback, or anticipation of differential updating behavior (cf. results on 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, individuals could prefer to receive positively skewed information, i.e. positive signals that are more informative than negative ones. We investigate the relative importance of preferences for positive skewness over preferences for less salient negative feedback by letting subjects choose between Mode A and Mode E. While the positive (grey) signals in Mode E have a higher likelihood ratio than the negative (red (-)) signals, positive signals in Mode E are less salient (positive signals are grey and negative signals red). Modes A and E do not just vary in skewness but also in salience and informativeness, so to get at a preference for positive skewness, in the analysis we look at this choice together with our baseline choice. Thus, if subjects have a stronger preference for positive skewness than aversion against salient negative signals, they would choose Mode E over Mode A and Mode B over A in our baseline choice.

Baseline Reversed Choice: Mode A vs Mode C Finally, we also check if individuals have a preference for or against a negatively skewed information structure with less salient positive feedback (Mode C). The signals in this information structure are equally informative about being in the bottom half of the distribution but less informative about being in the top half as compared to Mode A.

2.3 Updating Behavior: Endogenous and Exogenous Treatment

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 received three consecutive signals from one of the feedback modes. After each signal received, subjects were asked to report their (posterior) beliefs.[11] Further, each time they received a signal and

[11] Depending on the condition they were assigned to, they were asked to report their beliefs regarding their IQ scores (IQ treatment) or their number (random treatment) being in the top half of the distribution.

were asked about their beliefs, subjects could view a picture of the feedback mode urns from which they were receiving information by clicking a button (see Figure G.2 for a screenshot of the choice situation).

The feedback mode allocated to each subject depends on whether they are in the endogenous or the exogenous treatment.

2.3.1 Endogenous Treatment

In the endogenous treatment, one out of the five feedback mode choices explained above was randomly selected. The information structure that the subject selected when making their choice becomes relevant for updating. Before receiving signals, each subject was shown the choice they made in that scenario and the feedback mode from which they would be receiving the signals.

2.3.2 Exogenous Treatment

In the exogenous treatment, subjects are not asked to choose a feedback mode – assignments are exogenous. In particular, following the IQ test, subjects are randomly allocated to receive ego or non-ego relevant feedback from Mode A or Mode B.

The reason why we need the exogenous treatment in addition to the endogenous treatment to analyze updating behavior is twofold. First, subjects are randomly allocated into the feedback modes, so there are no systematic differences across groups, whereas in the endogenous treatment subjects self-select into feedback modes. Due to this self-selection, subjects in different feedback modes have, on average, different preferences over information structures. This could drive differences in updating behavior. Second, the exogenous treatment allows us to allocate subjects evenly into the feedback modes and across ego-relevance of the rank. Hence, in the exogenous treatment, we have more statistical power to analyze differences in updating between feedback modes.

Our main interest is in understanding deviations from Bayes rule across feedback modes and according to ego-relevance of the task. Our aim is to disentangle cognitive biases from motivated biases in updating. For this reason, we specifically focus on feedback modes A and B and their interaction with the ego-relevance of the task. On the one hand, a comparison in updating behavior across feedback modes A and B in the random treatment will allow us to understand if differences in the information structure drive cognitive biases. On the other hand, a comparison between updating across ego-relevant conditions will allow us to get at motivated biases in updating.

Importantly, as far as the updating stage is concerned, except for the random assignment into the feedback mode, there are neither differences in the experimental design nor in implementation between the endogenous and exogenous treatments. In this way, analyzing treatment differences in belief formation will permit us to also study whether and how selection affects updating.

2.4 Debriefing

In the last part of the experiment, we asked subjects a battery of questions. First, we asked subjects to complete the Information Preferences Scale by Ho *et al.* (forthcoming), 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 asked subjects to complete the Gneezy and Potters (1997) risk elicitation task. Specifically, each subject received £1.00 and had to decide how much of this endowment to invest in a risky project with a known probability of success. The risky project returned 2.5 times the amount invested with a probability of one-half and nothing with the same probability. Third, we asked subjects to answer two questions in free-form text and they received £0.50 for their answers. We asked them to advice a hypothetical subject who would be performing the feedback mode choices and updating task. In the endogenous treatment, we additionally asked them to explain their motives for choosing the feedback modes across the five scenarios. Finally, we asked subjects a series of demographic questions including age, gender, and nationality. We also asked them a non-incentivized general willingness to take risks question (Dohmen *et al.*, 2011).

2.5 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 feedback mode A or B.

3 Results

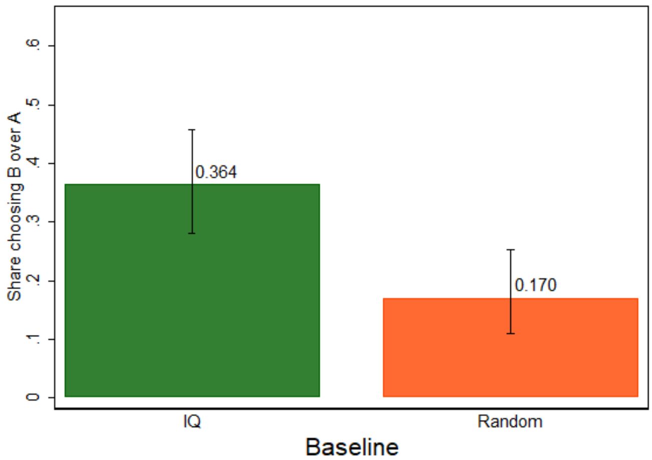
Our analysis proceeds in two steps: First, we investigate treatment differences between IQ and random in feedback mode choices. Second, we analyze how subjects update in response to signals from the corresponding feedback mode. We analyze updating in both the endogenous treatment, where subjects select into feedback modes, and in the exogenous treatment, where subjects are assigned to a feedback mode.

3.1 Information Selection

Information structures in our experiment differ in informativeness, salience, and skewness. The baseline choice is the choice between Mode A and Mode B, which varies in all three of these dimensions. While Mode A gives balanced feedback, Mode B produces less informative, positively skewed signals with less salient negative feedback. Hence, the choice of Mode B is costly because subjects are paid based on the accuracy of their posterior beliefs and Mode B provides less information.

Figure **??** 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 control 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 ($t(214) = 3.278, p = 0.001$).

Figure 2: Share choosing feedback mode B over A in baseline choice



Notes: 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 bar. In the IQ treatment there are N=110 subjects and in control N=106.

In order to disentangle preferences for informativeness, salience, and skewness, we elicit subjects' preferences over information structures in four additional choices. 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 significant ($t(214) = 3.149, p = 0.002$). Hence, the results suggest that subjects in the IQ treatment have indeed a preference for less information compared to subjects in the random treatment.

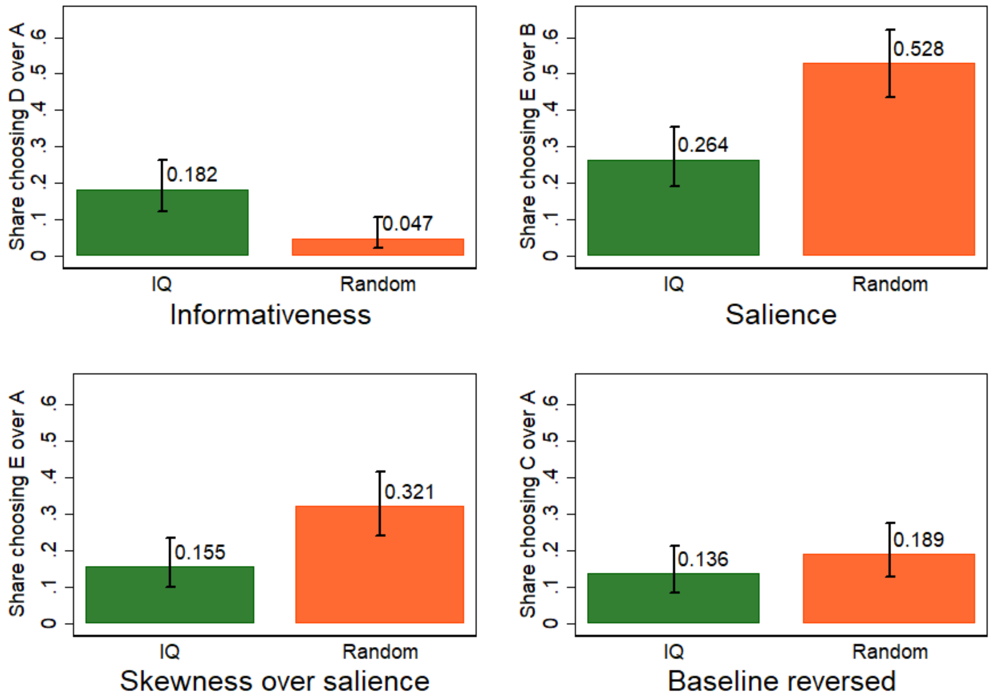
In the salience choice, subjects choose between Mode B, in which negative feedback is less salient, and Mode E, in which positive feedback is less salient. We find that, in the IQ treatment, significantly more subjects exhibit a preference for less salient negative feedback, compared to the random treatment (difference of 26.5 percentage points, $t(214) = 4.116, p < 0.001$). 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 random is not significantly different from 50 percent ($t(105) = 0.581, p = 0.563$). In contrast, in IQ only 26.4 percent choose Mode E, which is significantly lower than the 50 percent predicted by indifference ($t(109) = 5.601, p < 0.001$). 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 gives – just like Mode B – positively skewed information but gives explicit negative feedback and less salient positive feedback. We find that in the IQ treatment, fewer subjects prefer Mode E over Mode A than in the random treatment. Taken together with our finding from the baseline choice, we conclude that subjects have a stronger preference against explicit negative feedback as they have a preference for positive skewness when information is ego-relevant. While in Mode B, positive feedback comes in the form of green (+) signals and negative feedback in the form of grey signals, in Mode E positive feedback is given in the form of grey signals and negative feedback as red (-) signals. This difference in framing 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.[12]

Finally, in the baseline reversed choice, we let subjects decide between feedback mode A and feedback 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 significant difference between treatments with fewer subjects in IQ choosing feedback mode C (difference of 5.2 percentage points, $t(214) = 1.041, p = 0.299$). This result suggests that subjects do not

[12] Although the informativeness of Mode B and E is 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 the reason could be due to some subjects not understanding which feedback mode is more informative. When we explain that either Mode B or Mode E is less informative than Mode A, the levels in random are very similar.

Figure 3: Selection of feedback mode under choice situation



Notes: 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 bar. In the IQ treatment there are N=110 subjects and in control N=106.

prefer less informative feedback modes in the IQ treatment if the feedback mode is negatively skewed and makes positive feedback less salient.

In Table A.2 in the Appendix, we regress the five feedback mode choices on the IQ treatment dummy and different control variables. Controlling for demographics, prior beliefs, score on the IQ test and risk preferences does not alter the results in terms of treatment differences in feedback mode choices.

3.1.1 Information Selection - Within Individual

So far we have looked at each choice separately and analyzed what we can learn from these individual choices. We now address 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 salience of negative feedback/increase salience of positive feedback. We estimate the fraction of subjects who consistently choose information structures that conform with these preferences.[13]

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 with 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 to one of the preferences. This 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 column, 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 who consistently maximize the informativeness in the IQ treatment 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 random, suggesting that many subjects aim to maximize the informativeness of signals but make mistakes.[14] The treatment difference of 23.8 percent is significant ($p<0.01$).

In both treatments, there are only a few subjects who consistently choose feedback modes that are positively skewed. In particular, there are no subjects in IQ and 3.8 percent of subjects in random. In the finite mixture model, the share in random 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). Hence, the results do not suggest that subjects choose positively skewed feedback to protect their ego.

Finally, there are significantly more subjects in IQ than in random who exhibit a preference for reducing the salience of negative feedback but not reducing the salience of positive feedback. While more than 30 percent of subjects follow such a preference in IQ, there are

[13] "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.

[14] In Appendix D, we exploit the order in which feedback modes are presented and find 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***
			(0.053)	(0.042)	(0.069)
Positive skewness	0.000	0.038	0.000	0.107**	-0.107**
			(0.012)	(0.042)	(0.045)
Salience of feedback	0.327	0.057	0.345***	0.000	0.345***
			(0.056)	(0.000)	(0.056)
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: Table shows the share of subjects who choose feedback modes consistent with the respective preference. The 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 choosing accordingly without allowing for implementation mistakes. In "Maximum Likelihood" we estimate the share allowing for implementation noise γ. Standard errors are bootstrapped with 300 replications. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

few to none subjects categorized as such in random. For the salience of feedback, we find the largest treatment difference, at 34.5 percentage points (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 and feedback 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 random than in IQ.

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 on 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.* (forthcoming).[15] The scale

[15] Ho *et al.* (forthcoming) 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.

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). 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 significant interaction term implies that being information seeking according to the IPS scale is associated with a lower probability of choosing a feedback mode that is less informative and in which negative feedback is less salient. Moreover, the IPS scale is not associated with information structure choice in the random treatment, as illustrated by the small and non-significant main coefficient of the IPS scale.

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 significant, we conclude that there is no evidence for heterogeneous treatment effects by gender in the experiment.

Prior Belief In the third Column of Table 3, we investigate whether the treatment effect of the ego-relevant treatment is different depending on the reported prior belief. "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 differently affected by the treatment compared to individuals with priors above 50%.

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 non-significant 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 Beliefs

After subjects made the five information structure choices, one of these choices was randomly selected and the corresponding decision of the subject was implemented.

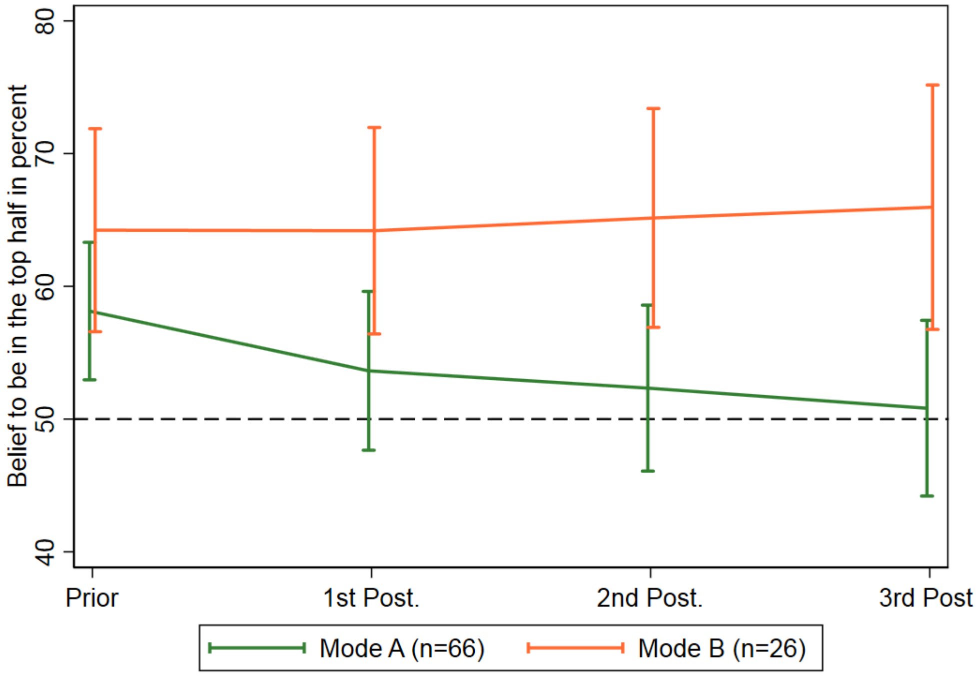
In Figure 4, we plot how the average beliefs in the IQ treatment evolve, depending on the feedback mode from which subjects receive signals. First, we observe that in the IQ treatment,

Table 3: Heterogeneity in baseline choice

	(1) IPS Scale	(2) Gender	(3) Prior Belief	(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: Table shows heterogenous treatment effects of the IQ treatment on the choice of mode B in the baseline choice by IPS Scale, Gender, Prior belief, 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.*, forthcoming). "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. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

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



Notes: Plot shows the average prior and posterior beliefs from treatment endogenous/IQ for the selected feedback mode. The whiskers represent 95% confidence intervals.

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 ($t(65) = 3.138, p = 0.003$) and in Mode B ($t(25) = 3.834, p = 0.001$).[16] Second, there is no significant difference in priors between individuals who end up in Mode A and Mode B ($t(90) = 1.284, p = 0.202$). Third, while the average beliefs of subjects in Mode A seem to converge toward 50% after receiving signals from Mode A, the beliefs in Mode B remain constant. In fact, after three rounds of feedback, we observe a significant difference in posterior beliefs between subjects in Modes A and B ($t(90) = 2.531, p = 0.013$).

These results suggest that selecting an information structure that is less informative and makes negative feedback less salient indeed leads to maintaining high beliefs in the IQ treatment. Moreover, in Appendix F, we show that this pattern is not observed in the control treatment. In the following section, we investigate more closely how individuals update their beliefs depending on the feedback mode from which they receive signals.

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 treatment and the 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:

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

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:

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

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

where δ^{prior} captures the weight put on 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 $\delta^{\text{prior}} = \beta^{\text{pos}} = \beta^{\text{neg}} = 1$. However, in this paper, we do not focus on the comparison with the Bayesian benchmark, rather we are mainly interested in differences in updating depending on the ego-relevance of the underlying state and across information structures. Thus, our analysis focuses on studying the β coefficients and their estimated differences across treatments and information structures.

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

In doing so, we assume that cognitive biases in updating are held constant across the underlying valence of the state. Importantly, however, we do not assume that cognitive or motivated biases are constant across information structures. Indeed, the features of an information structure may have implications for both cognitive and motivated biases in updating.

3.2.2 Updating in the Endogenous Treatment

First, we focus on belief updating in the endogenous treatment. Here, subjects receive signals from a feedback mode, which they selected in (at least) one of the five scenarios. We restrict the analysis to feedback modes A and B because for these feedback modes we have the highest number of subjects: In the IQ treatment, 66 (26) subjects update according to Mode A (B), and in the random treatment 61 (19) subjects update according to Mode A (B).

Table 4 shows estimation results of Equation (2) separately by treatment and feedback mode. We present results for Mode A in Columns (1) and (2) and for Mode B in Columns (3) and (4). We observe conservative updating ($\beta < 1$) in all feedback modes and treatments, which is consistent with previous evidence: subjects update less compared to the Bayesian benchmark (Möbius *et al.*, 2014; Coutts, 2019).

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

	(1) IQ Mode A	(2) Random Mode A	(3) IQ Mode B	(4) Random Mode B
δ^{prior}	0.776***	0.680***	0.828***	0.832***
	(0.086)	(0.075)	(0.107)	(0.163)
β^{Pos}	0.596***	0.762***	0.541**	0.408**
	(0.140)	(0.156)	(0.198)	(0.178)
β^{Neg}	0.527***	0.750***	0.191**	0.205
	(0.111)	(0.129)	(0.084)	(0.272)
p-value (δ^{Prior}=1)	0.011	0.000	0.119	0.316
p-value (β^{Pos}=1)	0.005	0.132	0.028	0.004
p-value (β^{Neg}=1)	0.001	0.057	0.000	0.009
p-value (β^{Pos}=β^{Neg})	0.658	0.952	0.114	0.372
R2	0.692	0.700	0.816	0.664
N	155	144	78	57

Notes: Table shows regression results of Equation (2) in the endogenous treatment, separately by IQ and random treatment and feedback mode 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. Standard errors clustered on subject level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

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 (p-value = 0.658). In Mode B, in contrast, we observe that β^{Pos} is larger than β^{Neg}, both in IQ and random treatments. However, we do not have the statistical power to reject the null hypothesis that they are equal, since relatively few subjects end up in Mode B.

To have a sufficient number of subjects in Mode B and to exclude self-selection into feedback modes, in the next section we investigate updating in the exogenous treatment.

3.2.3 Updating in the Exogenous Treatment

In the exogenous treatment, subjects are randomly assigned into feedback mode 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 5, we display the estimation results for the exogenous treatment. As before, we do not find evidence for asymmetric updating in Mode A. Both in IQ and in random, the coefficients β^{Pos} and β^{Neg} are of similar magnitude and not significantly different from each other (p-value = 0.659 in IQ and p-value = 0.676 in random).

Table 5: Updating across feedback modes and treatments (exogenous treatment)

	(1) IQ Mode A	(2) Random Mode A	(3) IQ Mode B	(4) Random Mode B
δ^{prior}	0.799***	0.726***	0.896***	0.779***
	(0.096)	(0.065)	(0.046)	(0.061)
β^{Pos}	0.631***	0.835***	0.482***	0.758***
	(0.122)	(0.124)	(0.077)	(0.174)
β^{Neg}	0.559***	0.761***	0.244***	0.938***
	(0.128)	(0.124)	(0.088)	(0.142)
p-value (δ^{Prior}=1)	0.042	0.000	0.026	0.001
p-value (β^{Pos}=1)	0.004	0.191	0.000	0.171
p-value (β^{Neg}=1)	0.001	0.059	0.000	0.662
p-value (β^{Pos}=β^{Neg})	0.659	0.676	0.028	0.325
R2	0.633	0.729	0.677	0.735
N	132	139	165	186

Notes: Table shows regression results of Equation (2) in the exogenous treatment, separately by IQ and random treatment and feedback mode 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. Standard errors clustered on subject level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

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 update less to negative than to positive signals. The updating coefficient for positive signals is about twice as large as for negative signals and the coefficients differ significantly (p-value=0.028). However, this is not the case when feedback is not ego-relevant: in the random treatment, subjects update, if anything, more to negative feedback that is less salient but the difference in coefficients is not significant (p-value=0.735).

In Table 6, we interact the feedback mode with the signal received, separately by treatment. In both endogenous and exogenous treatments, subjects update significantly less to negative feedback in Mode B when signals are ego-relevant (Columns 1 and 3). However, when information is not ego-relevant, the interaction is only marginally significant in the endogenous treatment (Column 2). The interaction is not significant and even slightly positive in the exogenous treatment (Column 4), suggesting that the negative coefficient in the endogenous treatment is due to selection. Namely, a number of subjects who end up in Mode B in the endogenous/random treatment, chose Mode B in the baseline choice – in random this decision can neither be explained by payout maximization nor by motivated beliefs. Hence, this is a selected group of subjects, whose updating behavior should be interpreted with caution.

Taken together these results suggest that subjects' belief formation is driven by motivated reasoning, as subjects asymmetrically update only in the IQ treatment. However, while individuals might have a preference for forming high beliefs about themselves, our results

Table 6: Updating interacted with feedback modes by treatments

	Endogenous		Exogenous	
	(1)	(2)	(3)	(4)
	IQ	Random	IQ	Random
δ^{prior}	0.770***	0.671***	0.814***	0.716***
	(0.068)	(0.058)	(0.084)	(0.066)
β^{Pos}	0.594***	0.661***	0.623***	0.839***
	(0.108)	(0.126)	(0.119)	(0.127)
β^{Neg}	0.634***	0.751***	0.568***	0.767***
	(0.107)	(0.123)	(0.124)	(0.125)
δ^{prior} x Mode B	0.057	0.161	0.081	0.063
	(0.125)	(0.168)	(0.096)	(0.090)
β^{Pos} x Mode B	-0.053	-0.253	-0.142	-0.080
	(0.222)	(0.213)	(0.142)	(0.215)
β^{Neg} x Mode B	-0.443***	-0.546*	-0.324**	0.171
	(0.135)	(0.291)	(0.151)	(0.189)
R2	0.705	0.615	0.715	0.720
N	330	318	330	357

Notes: Table shows regression results of Equation (2), fully interacted with the feedback mode. The regressions are estimated separately by IQ and random treatment as well as by endogenous and exogenous treatment. 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 as well as whether the signal comes from Mode A or B. The model does not include a constant. Standard errors clustered on subject level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

show that this does not seem to be always possible. In fact, our results show 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

In our experimental findings there is a striking difference across treatments in the way subjects choose between different feedback modes. We interpret these results as evidence for differential preferences over information structures driven by motivated reasoning – biases that are driven by specific individuals' goals (e.g., of having high opinions of oneself 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 (e.g., the belief that one has high intelligence). In contrast, general failures in cognitive reasoning that depend on one's prior beliefs, like confirmation-seeking and contradiction-seeking behavior, usually go in either direction. Here, we discuss whether these tendencies explain our results.

4.1.1 Confirmation-Seeking Behavior

Confirmation-seeking behavior, also known as confirmation bias, is the tendency to search for, interpret, favor, and recall information in a way that confirms one's preexisting beliefs.[17] In our experiment, this bias could explain treatment differences in information structure selection if participants with different priors have different beliefs over the informational content of our feedback modes and/or have a preference for receiving signals that confirm their priors. This could (partly) explain treatment differences as subjects in the treatment group have (slightly) higher prior beliefs than those in the control (60.1% vs. 54.2%).[18] For instance, it could be the case that participants with high priors believe that, in our baseline choice, Mode B is more informative than Mode A. However, even when controlling for prior beliefs, we see that subjects in the ego-relevant treatment are more likely to choose feedback mode B compared to those in the control. Similarly, as shown in Table A.2 in the Appendix, our treatment differences in all feedback mode choices hold when controlling for prior beliefs. Finally, it is relevant to note that, in the informativeness choice, there is no

[17]See Nickerson (1998) for a review of the psychological literature on confirmation bias.
[18]However, the difference is not statistically significant at conventional levels (p-value=0.101).

role for confirmation-seeking behavior. Thus, this difference cannot be explained through confirmation bias. Taken together, these results show that confirmation bias falls short of explaining the treatment differences in the information selection stage of our experiment.

4.1.2 Contradiction-Seeking Behavior

Contrary to confirmation bias, contradiction-seeking behavior can be defined as the tendency to search for, interpret, favor, and recall information that goes against one's prior beliefs. As above, this bias could explain treatment differences in information structure selection if participants with different priors have different beliefs over the informational content of the feedback modes and/or have a preference for receiving signals that do not confirm their priors. However, by the same arguments as for confirmation-seeking behavior, we can rule out contradiction-seeking behavior as an alternative explanation for our treatment differences.

4.2 Bounded Rationality

4.2.1 Cognitive Ability

Cognitive errors in processing and interpreting information do vary by 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.

By taking into account participants' abstract reasoning ability, measured by their scores in the IQ test, 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.[19] Second, our treatment differences in feedback mode selection are robust to controlling for individuals' cognitive ability (see Column 4 in Table A.2 in the appendix).

4.2.2 Confusion

Our experimental design features some fairly complex elements and, thus, might have affected participants' understanding. To tackle this possibility, we paid close attention to the way we presented the experimental instructions to our participants. We also ensured participants' understanding by letting them answer comprehension questions. Any participant confusion is unlikely to be a driving force of our results as it is held constant across our treatment conditions.

[19] Mean IQ scores in the IQ treatment is 11.34, while it is 11.41 in the control treatment. A Mann-Whitney test fails to reject the null that the difference in distributions is significantly different from zero (p-value=0.78).

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 control make the suboptimal choice. This finding is reassuring as it implies that subjects understood our experimental instructions and were sufficiently incentivized to choose the optimal decision.

4.3 Emotional Involvement: Heat vs. Light

Finally, 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, there is suggestive evidence in favor of emotions arising in the treatment group. This supports our claim that motivated reasoning drives our results and not cognitive failures. First, in the open-text question where we asked subjects to describe how they chose between different information structures, participants in the IQ treatment were more likely to report answers that stated their willingness to avoid explicit negative (red) signals; this is not the case in the control treatment. This differential response by treatment suggests that participants in the IQ treatment chose specific feedback modes to avoid feeling negative emotions. Second, the informativeness choice result clearly shows that individuals in the IQ treatment are more inclined to protect their beliefs (and, presumably their emotions) since a Bayesian, or even boundedly rational thinker, without motivated beliefs would welcome more information.

In summary, in this section we argue that our results cannot be accounted for by cognitive biases and, in particular, by endogenous directionality and bounded rationality, including cognitive ability and confusion. On the other hand, there is evidence that participants' behavior could be driven by emotional reactions due to the ego-relevance of the underlying state.

5 Conclusion

We run an experiment to study individuals' preferences towards information structures and subsequent belief updating if information is ego-relevant or neutral. 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. These findings suggest that individuals selectively choose information structures that allow them to protect their ego. Moreover, 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 results are also informative to the literature on asymmetric updating, and their mixed findings. Indeed, the different ways in which the signals are framed across studies could partly explain the differential results in the literature.

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 show that the framing of feedback is one dimension of "reality constraints" that allows individuals to maintain and nurture motivated beliefs. Zimmermann (2020) shows that raising the incentives to recall negative feedback can constitute another "reality constraint." This raises a question for future research over which other dimensions in the environment may constrain individuals from holding motivated beliefs and how consciously people engage in motivated thinking.

Taken together, our findings suggest that motivated information acquisition might play a key 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 choice can enable us to protect ourselves from receiving "bad" news about our abilities. Future research should pay more attention to the different ways in which individuals can and do self-select into environments to receive (avoid) more flattering (damaging) feedback to the self. Moreover, as found in our experimental data, information source selection also interacts with the subsequent belief formation process as some signals allow one to interpret the underlying information in a self-serving manner. These results have important implications. Namely, feedback procedures at the workplace that are intended to disclose unbiased information should not allow employees discretion over the information sources, to avoid biased information transmission. This could be implemented by having an objective third-party assessing workers' performance. Similarly, the anticipation of different feedback cultures (e.g., at the workplace, at the industry-level, or profession) may be an important factor that discourages people from undertaking certain career paths.

References

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.

Barron, K. (2020). Belief updating: Does the 'good-news, bad-news' asymmetry extend to purely financial domains? *Experimental Economics*.

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.

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. (2018). How Do People Choose Between Biased Information Sources? Evidence from a Laboratory Experiment, `http://econ.ucsb.edu/~sevgi/CharnessOpreaYuksel_Dec2018.pdf` (retrieved 07/26/2019).

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, `https://ssrn.com/abstract=2127795` (retrieved 03/02/2020).

Chopra, F., Haaland, I. and Roth, C. (2020). Do people value more informative news?, `https://drive.google.com/file/d/1D4OAyAXM3gFWDjDwk9b8SycRKkKMzBwh` (retrieved 03/26/2020).

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.

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. (forthcoming). What you see is all there is.

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. (forthcoming). Measuring information preferences. *Management Science*.

Jin, G. Z., Luca, M. and Martin, D. (2018). Is no news (perceived as) bad news? An experimental investigation of information disclosure, NBER Working Paper Series 21099.

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.

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).

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. (2018). Preferences for the resolution of uncertainty and the timing of information, `https://kirbyknielsen.com/wp-content/uploads/kirby/RoU.pdf` (retrieved 07/26/2019).

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

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

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.

Smith, M. K., Trivers, R. and von Hippel, W. (2017). Self-deception facilitates interpersonal persuasion. *Journal of Economic Psychology*, **63**, 93–101.

Solda, A., Ke, C., Page, L. and von Hippel, B. (2019). Strategically delusional. *Experimental Economics*.

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.

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: 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***	0.192***	0.178***	0.193***	0.189***	0.161***
	(0.059)	(0.060)	(0.059)	(0.059)	(0.058)	(0.061)
Informativeness	0.135***	0.136***	0.132***	0.134***	0.132***	0.132***
	(0.042)	(0.044)	(0.043)	(0.042)	(0.042)	(0.042)
Framing	-0.264***	-0.266***	-0.243***	-0.265***	-0.263***	-0.263***
	(0.064)	(0.067)	(0.065)	(0.065)	(0.065)	(0.065)
Skewness over framing	-0.166***	-0.129**	-0.164***	-0.165***	-0.167***	-0.122**
	(0.057)	(0.059)	(0.057)	(0.057)	(0.057)	(0.058)
Baseline reversed	-0.052	-0.036	-0.049	-0.053	-0.052	-0.037
	(0.050)	(0.052)	(0.050)	(0.050)	(0.051)	(0.058)
Demographics		✓				✓
Prior			✓			✓
IQ score				✓		✓
Risk					✓	✓
N	216	216	216	216	216	216

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

B Belief Elicitation Mechanism

After Part I was completed but before we explained Part II, we told subjects that for the following part of the experiment we would ask them their beliefs regarding some events. In particular, they were told that they would be asked four belief questions and that one question would be chosen at random to count for payments.

Then, we explained our belief elicitation procedure to them. We used the belief elicitation mechanism proposed by Karni (2009) called the matching probabilities method.[20] Under this method, 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 to the belief question) indicate 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 the chances of them betting on the event or the lottery and their chances of winning the prize. We also emphasized that truthful reporting was the answer that maximized the chances of earning the prize. To ensure that subjects understood the main features of this elicitation procedure, we asked subjects to answer comprehension questions about the belief elicitation procedure.

[20]This method is also referred to as the "crossover mechanism," "reservation probabilities," and "lottery method." This belief elicitation mechanism is first introduced in an experiment by Möbius *et al.* (2014); since then it is extensively applied to other experiments in the asymmetric updating literature, including 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 the choices prescribed by the respective preference, the smaller will be the estimated implementation noise.

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

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

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 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, e.g., 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 G.3 to G.5). 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. 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: 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 insignificant interaction effects, we do not find much 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 scenario (i.e., in the third order). 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 G.5). Hence, in Order 3 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 control treatment when abstracting from the skewness over framing Choice.

Table D.2: Feedback mode choice by presented order

	(1) Baseline	(2) Informativeness	(3) Framing	(4) Skewness over framing	(5) Baseline 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: 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.*, forthcoming)

- 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]

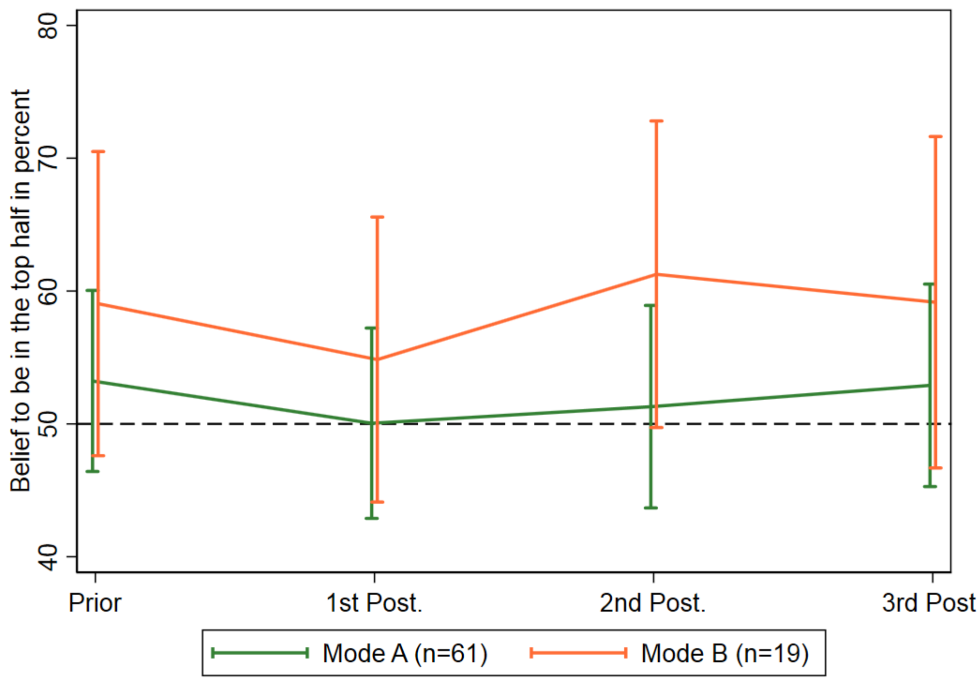
- You gave a close friend one of your favorite books for her birthday. Visiting her apartment a couple of months later, you notice the book on her shelf. She never said anything about it; do you want to know if she liked the book? [1: Definitely don't want to know; 4: Definitely want to know]

- 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 Information Selection and Beliefs in the Control Treatment

In Figure F.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 ($t(78) = 0.853, p = 0.396$). 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 posterior beliefs after three signals are not significantly different ($t(78) = 0.824, p = 0.413$).

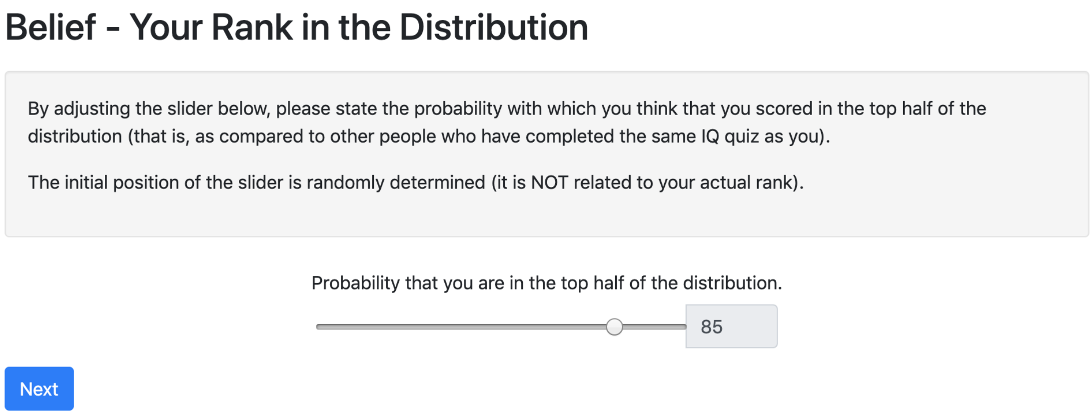
Figure F.1: Beliefs before and after signals by feedback mode (endogenous/random treatment)



Notes: 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.

G Screenshots of the Experiment

Figure G.1: 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.

Figure G.2: 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

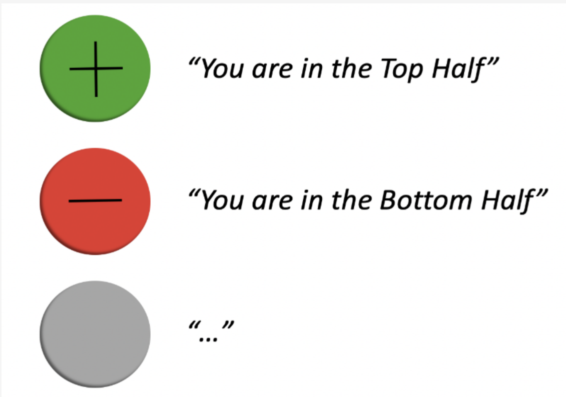
Now, you will receive additional information (feedback) about your performance in the IQ quiz to help you assess whether or not you are in the top half of the distribution.

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



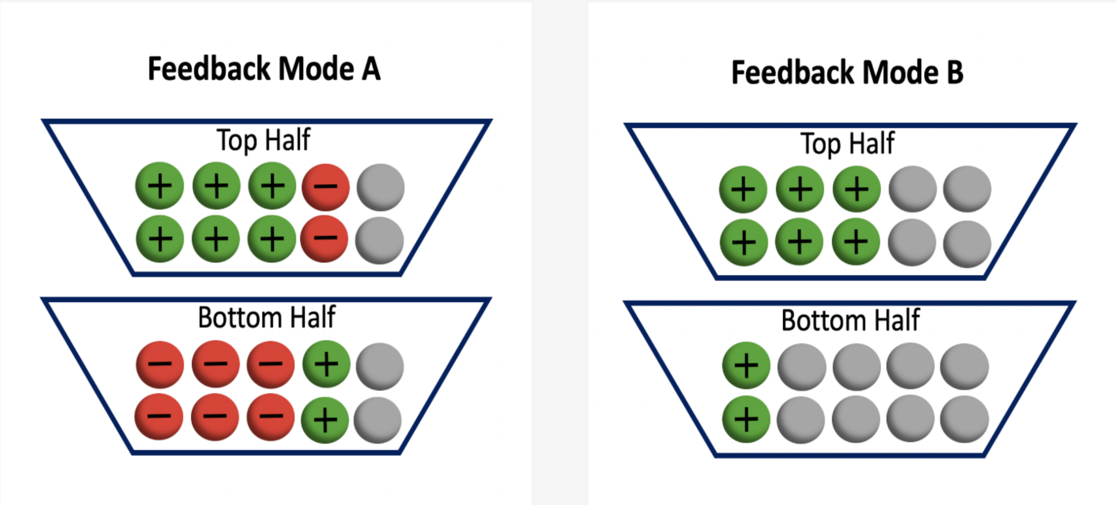
Notes: The figure displays a screenshot of the template in which we explain the participant, in the IQ/endogenous treatment, the possible signals he/she can receive about his/her performance.

Figure G.3: 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.

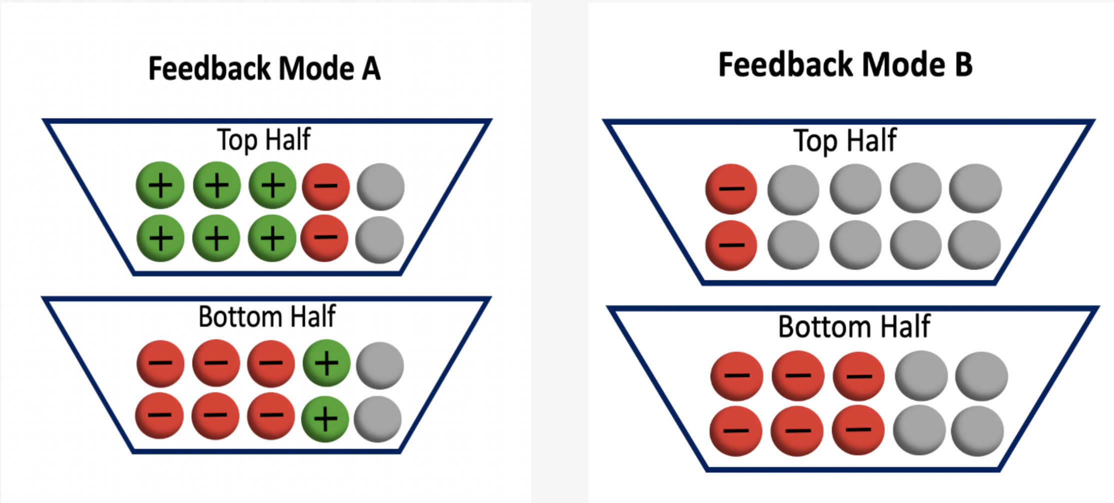
Notes: The figure displays a screenshot of the template in which we explain to the participant, in the IQ/endogenous treatment, the selection of feedback modes and how feedback modes differ. In Order 1, we use the baseline choice as an example.

Figure G.4: 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:

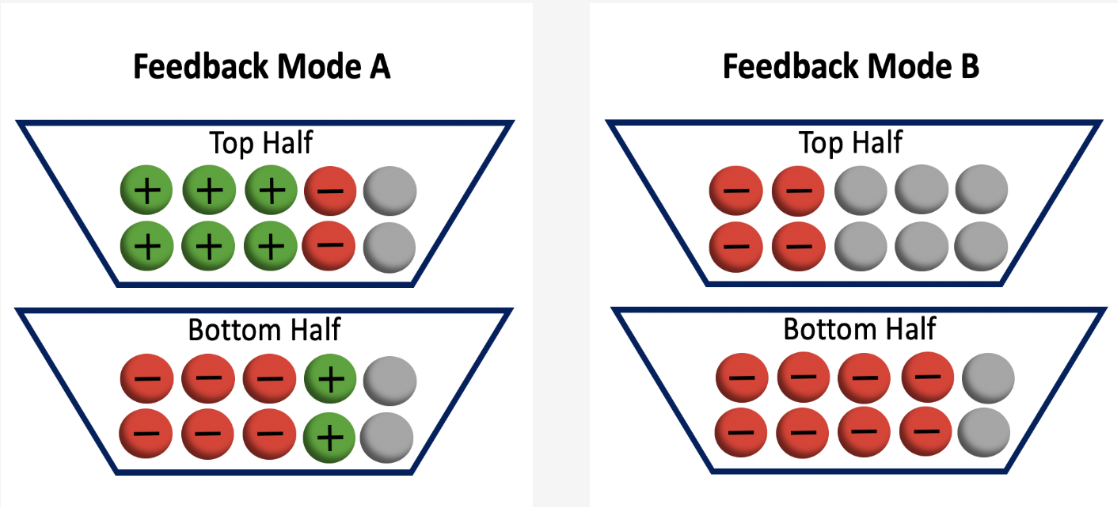
- Irrespective of the feedback mode you choose, if you are in the top half of the distribution your feedback will be determined by the urn at the top of each figure. If you are in the bottom half, your feedback will be determined by the urn at the bottom.
- Consider the example above. In both feedback 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 G.5: 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:

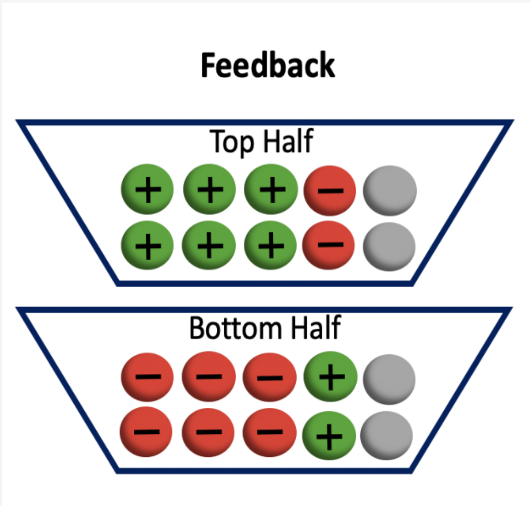
- Irrespective of the feedback mode you choose, if you are in the top half of the distribution your feedback will be determined by the urn at the top of each figure. If you are in the bottom half, your feedback will be determined by the urn at the bottom.
- Consider the example above. In both feedback 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 G.6: Screenshot of the instructions' template about the feedback mode (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.

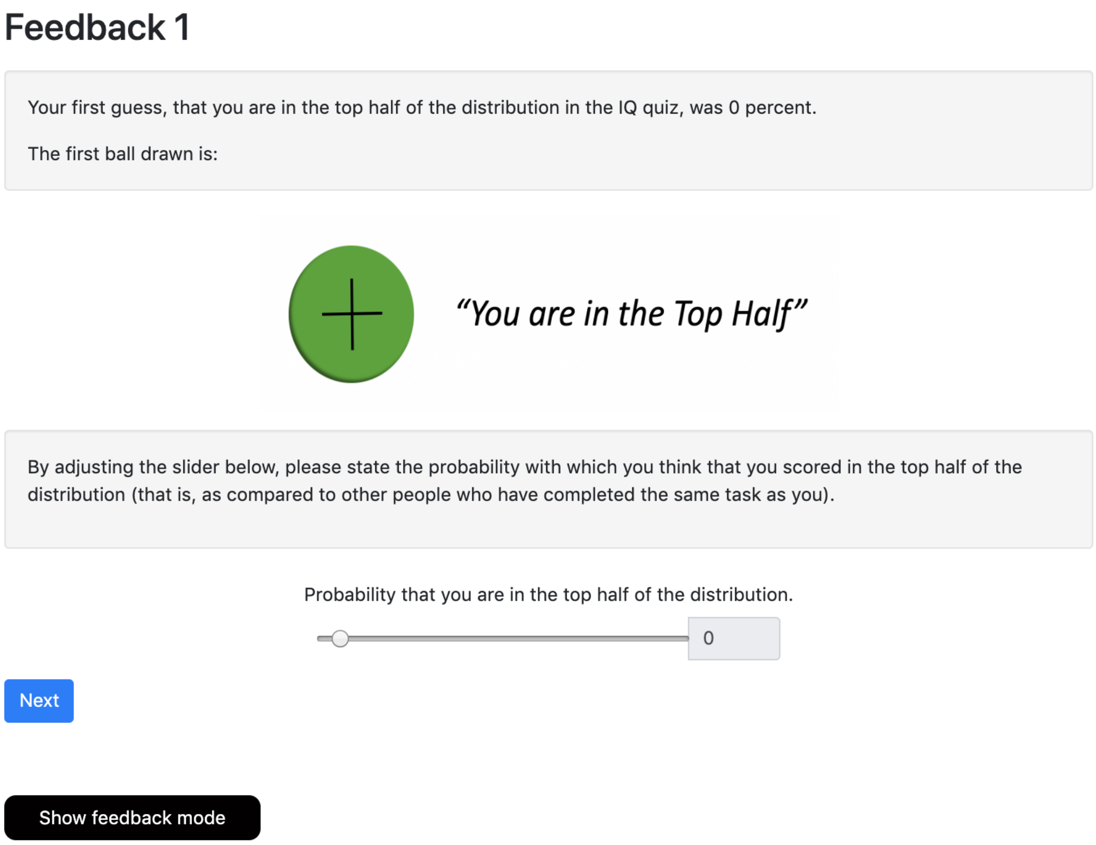


Notice that:

- 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 G.7: 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.