# Motivated Skepticism*

Jeanne Hagenbach[†] Charlotte Saucet[‡]

February 5, 2024

**Abstract**

We experimentally study how individuals read strategically-transmitted information when they have preferences over what they will learn. Subjects play disclosure games in which Receivers should interpret messages skeptically. We vary whether the state that Senders communicate about is ego-relevant or neutral for Receivers, and whether skeptical beliefs are aligned or not with what Receivers prefer believing. Compared to neutral settings, skepticism is significantly lower when it is self-threatening, and not enhanced when it is self-serving. These results shed light on a new channel that individuals can use to protect their beliefs in communication situations: they exercise skepticism in a motivated way, that is, in a way that depends on the desirability of the conclusions that skeptical inferences lead to. We propose two behavioral models that can generate motivated skepticism. In one model, the Receiver freely manipulates his beliefs after having made skeptical inferences. In the other, the Receiver reasons about evidence in steps and the depth of his reasoning is motivated.

**Keywords:** Disclosure games, hard information, unraveling result, skepticism, motivated beliefs, endogenous depth of reasoning.

**JEL Codes:** C72, C91, D82, D91.

*We thank the editor, Andrea Galeotti, and four anonymous referees for their excellent suggestions. We also thank Larbi Alaoui, Victor Augias, Kai Barron, Daniel Barreto, Roland Bénabou, Alessandra Casella, Tilman Fries, Alexis Ghersengorin, Juan Ivars, Frédéric Koessler, Johannes Leutgeb, Yves Le Yaouanq, Franz Ostrizek, Eduardo Perez-Richet, Rémi Suchon, Eugenio Verrina, Marie Claire Villeval, and Sevgi Yuksel for helpful comments. We are grateful to the seminar participants at Sciences Po, PSE, CREST, UCL, WZB, Queen Mary University, EUI, Universities of Aalto, Constance, Cologne, Prague, Maastricht, Kyoto, Osaka, Sinica, Hong Kong (BU), Taiwan (NTU), the ASFEE conference, the Bounded Rationality workshop in Tel-Aviv and the BBU workshop in Amsterdam. We thank Quentin Thevenet for programming the experiment, Nina Bonge for helping at the WZB-TU Lab and Alix Charpentier for excellent research assistance. Jeanne thanks Dorothea Kübler for a fruitful stay at the WZB in 2021. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement n°850996 – MOREV).

[†]CNRS, Sciences Po, CEPR, WZB - jeanne.hagenbach@sciencespo.fr

[‡]University of Paris 1 - charlotte.saucet@univ-paris1.fr

# 1 Introduction

In many economic situations, communication is considered as an effective way to reduce information asymmetry. The less informed parties are assumed to be willing to learn the truth and, knowing the objectives of the communicating parties, to be able to make inferences from disclosed and undisclosed information. At the extreme, the *unraveling result* (Milgrom, 1981, Grossman, 1981) establishes that information can be fully learnt by uninformed parties provided that they read vague disclosure with skepticism. And indeed, it is often natural to interpret the absence of precise evidence as unfavorable for the communicating party. This paper shows that individuals' ability to exercise skepticism, and therefore the relevance of the unraveling result, importantly depends on whether or not individuals *want* to learn the truth in the first place. Using an online-lab experiment, we investigate how subjects interpret hard information when they have preferences over what they want to be true.

Situations of strategic communication often share the feature that agents are not indifferent about what they learn. Think for example of firms revealing hard information to consumers about products attributes. Consumers wish to know these attributes to best-adapt purchasing decisions, but they may also benefit *per se* from believing that products have particular attributes, such as being environmental-friendly or ethically produced. In advising settings, advisors often communicate with advisees about their abilities or chances of success. Advisees have an interest in learning the truth but they may also be directly affected by the beliefs they hold about their abilities or chances. In these situations, the agents who read information trade-off the need to form accurate beliefs and the wish to hold comforting or pleasant beliefs. Our experiment examines whether agents form motivated beliefs when receiving strategically-transmitted information. By doing so, it brings together the literature on *disclosure games* and the literature on *motivated beliefs*.

The theoretical literature on voluntary disclosure games makes especially sharp predictions about the reading of information in equilibrium. In these games, the information transmitted by the privately-informed Sender to the Receiver is hard in the following sense: a message from the Sender consists of a subset of types that must include the true type. The message is considered precise when it is a singleton set – the type is fully disclosed –, and vague otherwise. In the classical version of these games initially proposed by Milgrom (1981), the Sender's payoff simply increases with the decision finally taken by the Receiver. The Receiver's objective is to take an action that matches the type. In equilibrium, information is fully revealed by the Sender because different types never pool on sending the same vague message: the highest type has an interest in separating from the lower types and inducing a higher action; at the next step, it is the second-highest type who should fully disclose, etc. Because of this unraveling mechanism, the Receiver's equilibrium beliefs after any vague message are *skeptical*, that is, assign probability one to the lowest type disclosed. Skepticism captures the intuitive idea that, when facing vague statements such as "the student is in the top ten" or "at least half of the ingredients are organic", the rational reading is that the student actually ranked ten and that no more than half of the ingredients are organic.

Our experiment is designed to study the extent to which Receivers form skeptical beliefs as a function of the desirability of these beliefs. In the lab, Sender-subjects transmit verifiable information about a type to uninformed Receiver-subjects, and Receivers need to guess the type right to maximize their monetary payoffs. We vary whether skeptical beliefs are pleasant or not for Receivers by crossing two exogenous variations, all else equal. The first variation is whether or not the Receiver is directly affected by his beliefs about the type, which we implement by having the Sender communicate about a type which is neutral or ego-relevant for the Receiver.[1] In *Loaded* treatments, the type is a noisy measure of the Receiver's relative intelligence. In *Neutral* treatments, the type is also a rank but it has no particular meaning for the Receiver.[2] The second variation concerns the objective of the Sender. In *High* treatments, the Sender's payoff strictly increases with the Receiver's guess. In that case, the Sender should disclose precisely high types (i.e., low Receiver's ranks) and obfuscate low types (i.e., high Receiver's ranks). In *Low* treatments, the Sender's payoff strictly decreases with the Receiver's guess. The Sender should now conceal high types (i.e., low Receiver's ranks). The *High vs. Low* variation determines whether being skeptical consists in believing the highest or the lowest rank disclosed. This makes no difference in *Neutral* treatments but affects the desirability of the skeptical beliefs in *Loaded* treatments. Specifically, in the *High_Loaded* treatment, the skeptical belief assigns probability one to the highest rank disclosed, a belief we consider *self-serving*. In the *Low_Loaded* treatment, the skeptical belief assigns probability one to the lowest rank disclosed, a belief we consider *self-threatening*.

We implement the four treatments between subjects. Each subject is Sender or Receiver, and plays ten games with random rematch and no feedback. Our data contains 2000 games (around 500 by treatment) for which we record the type seen by the Sender, the message sent and the Receiver's guess. In each game, we measure Receiver's skepticism by evaluating how close his guess is to the *skeptical guess*, i.e., the guess he would have made if he held skeptical beliefs. We also measure skepticism using the frequency of skeptical guesses. We test three main pre-registered hypotheses.[3] First, we hypothesize that Receivers' skepticism is unaffected by the *High vs. Low* variation in *Neutral* treatments. Second, we hypothesize that Receivers' skepticism will be at least as high

[1] The use of ego-relevant information is a commonly-used artefact to create preferences over beliefs in the lab (see Schwardmann and Van der Weele, 2019 and Zimmermann, 2020 for instance).

[2] Across these treatments, we keep the Sender's behavior constant by only telling him that he discloses information about a secret number.

[3] The reference for pre-registration is AEARCTR-0007541.

in *High_ Loaded* as in *High_ Neutral*, that is, when skeptical beliefs are self-serving. In theory, Receivers are fully skeptical in the *Neutral* settings but, if they are not, there is room for increased skepticism when it is self-serving. Third, we hypothesize that Receivers' skepticism is strictly lower in *Low_ Loaded* than in *Low_ Neutral*, that is, when skeptical beliefs are self-threatening. In short, we expect individuals' to interpret vague information skeptically when it is good news to them, and less so when it is bad news to them.

We begin the data analysis by checking that Senders and Receivers understood the basics of the game. Senders' communication strategies have a clear structure that takes into account their reversed objectives: in *High* treatments, Senders of type $t$ most often disclose that their type is *at least* $t$; in *Low* treatments, Senders of type $t$ most often disclose that their type is *at most* $t$ (that is, use the "top ten" kind of messages). On their side, Receivers take the evidence into account in the sense that their guesses are almost always within the set of types disclosed. When comparing Receivers' skepticism in the two *Neutral* treatments that serve as benchmarks, the difference is marginal and goes in the direction of a slightly lower level of skepticism in *Low_ Neutral* than in *High_ Neutral*. We attribute this small difference to the Sender's payoff function being slightly more complex in the *Low* than in the *High* condition.

Our main finding is that skepticism is significantly lower when it is self-threatening (as in the *Low_ Loaded* treatment) than when it is not (as in the *Low_ Neutral* treatment). Said differently, individuals read information less skeptically when it implies believing they ranked low in the IQ test, than in comparable neutral environments. In contrast, we find no significant difference in the level of Receivers' skepticism when skepticism is self-serving (as in the *High_ Loaded* treatment) and when it is not (as in the *High_ Neutral* treatment). These results demonstrate two important things. First, the exercise of skepticism does not depend on the object that individuals reason about: subjects are able to make skeptical inferences about their relative IQ in the *High_ Loaded* treatment. Second, and it is our main message, individuals exercise skepticism in a way that depends on the conclusions that skeptical inferences lead to. In particular, they lack skepticism when it leads to beliefs they dislike.

Because the set of messages we consider is rich, we can observe nuances in the desirability of skeptical guesses that go beyond the ones created by our experimental variations. Again, we observe that the level of skepticism depends finely on its effect on the ego. Precisely, skeptical guesses are significantly less frequent when theses guesses are strongly self-threatening than when they are mildly so; skeptical guesses are significantly more frequent when they are strongly self-serving than when they are mildly so. We also establish that individuals' lack of skepticism when it is self-threatening is not driven by an overall tendency of these individuals to be confident about their relative intelligence and to stick to these prior views. To do so, we check the robustness of our main results to the exclusion of confident subjects. We also run a complementary treatment in which we make Receivers a priori more confident about their neutral ranks in the *Low_Neutral* treatment and show that this change does not affect their subsequent guesses.

Our results point to the following asymmetry: compared to the *Neutral* setting, the *Loaded* setting hinders skepticism in the *Low* treatments more than it enhances it in the *High* treatment. This finding is in line with previous works establishing that individuals process good news and bad news asymmetrically to protect their beliefs (see for instance Eil and Rao, 2011, Möbius et al., 2022, Chew et al., 2020 or Zimmermann, 2020). One important difference however is that, in our strategic setting, individuals first need to reason about the Sender's strategy to uncover whether a message is good or bad news to them. We view motivated skepticism as a new channel by which individuals can protect their beliefs in strategic settings. From a more applied point of view, our work suggests that individuals may read product information differently when it leads to disturbing conclusions than when it does not. Motivated skepticism provides a new, psychological reason for the failure of unraveling, which is often observed in the field and contradicts standard theory.[4]

Finally, we propose two behavioral models that can generate motivated skepticism.[5] The first model, presented in Section 2, is an adaptation to our communication setting of the models of motivated beliefs formation which follow Brunnermeier and Parker (2005). The Receiver, whose beliefs enter directly his utility function, first forms equilibrium skeptical beliefs. Next, he manipulates these beliefs freely by trading-off the material benefit of keeping the accurate beliefs and the psychological benefit of having a good expectation of his relative intelligence. The predictions of this model correspond to the three hypotheses listed above. The alternative model, presented in Section 6, is more original. In this model, we assume, following the work of Schipper and Li (2020), that the Receiver reasons in steps when reading hard information. We next propose that how deep the Receiver goes in the reasoning process depends on how psychologically pleasant or unpleasant his beliefs become along that process. As the Receiver reasons in *High_Loaded*, he converges step by step to believing that his rank is the highest one disclosed. As the Receiver reasons in *Low_Loaded*, he converges step by step to believing his rank is the lowest one disclosed. These treatments provide different psychological incentives to reason while keeping constant the material benefits and the cognitive costs to do so. Following the framework of Alaoui and Penta (2015), we predict that

[4]Understanding when unraveling fails is important to decide whether or not to mandate disclosure. Dranove and Jin (2010) report that unraveling is incomplete in many markets and already provide explanations, complementary to ours, of why this may be the case: disclosure may not be costless, the Sender may not be fully informed, etc.

[5]In both models, we focus on the Receiver side and do not consider that the Sender could, in some contexts, expect motivated skepticism and react to it. In our experiment, Senders are never made aware of the fact that types can be ego-relevant for Receivers.

the reasoning towards skeptical guesses will be less deep in *Low_ Loaded* than in *Low_ Neutral*, and deeper in *High_ Loaded* than in *High_ Neutral*. Looking at individual data, we observe that Receivers make fewer reasoning steps in *Low_ Loaded* than in *Low_ Neutral*, and similar number of steps in *High_ Loaded* and in *High_ Neutral*. Overall, the two models make similar predictions about the effect of our treatments on skepticism, but they differ in *how* motivated skepticism can emerge. In one model, the Receiver distorts his beliefs after having made the skeptical inference. In the other, preferences over beliefs limit the inference process itself.

**Related literature.** Our experiment connects the literature on *disclosure games* and the literature on *motivated beliefs*.

A central result in the literature on disclosure games is the *unraveling result* of Milgrom (1981) and Grossman (1981), which establishes that information is fully disclosed by the informed party in equilibrium.[6] This result crucially relies on Receivers reading information in a specific way, unfavorable to the Sender. While the theory is clear, the empirical literature reports that voluntary disclosure is not always complete (see Mathios, 2000, Jin, 2005, Brown et al., 2012, Luca and Smith, 2015 or Bederson et al., 2018 for concrete examples, and Dranove and Jin, 2010 for a survey). Experiments have offered some elements to understand this partial disconnection between theory and practice. Jin et al. (2021) and Deversi et al. (2021) respectively show that subjects' ability to make skeptical inferences depends on their experience and on the language used by the Sender.[7] King and Wallin (1991) and Dickhaut et al. (2003) consider costs to disclosure or a probability that the Sender is not informed, also inducing failures of unraveling. Benndorf et al. (2015) observe a low amount of disclosure when Senders are framed into being workers disclosing their productivity. Their experiment suggests that psychological aspects can affect communication in important ways, a point also made in Loewenstein et al. (2014). Our experiment is the first to consider that the reading of hard information may depend on the preference Receivers have over beliefs.

The idea that, in some contexts, individuals' well-being directly depends on their beliefs is at the center of the recently-growing literature on *motivated beliefs* (surveyed in Bénabou, 2015). In the last decade, the experimental literature on that topic has identified various channels that individuals use to reach favored conclusions. These channels include asymmetric belief updating, selective recall, or motivated information selection.[8] In the works on asymmetric updating, subjects form beliefs

[6] See Okuno-Fujiwara et al. (1990), Seidmann and Winter (1997), Giovannoni and Seidmann (2007) and Hagenbach et al. (2014) for related theoretical works.

[7] Schipper and Li (2020) analyze stepwise reasoning in experimental disclosure games and link Receivers' levels of reasoning to their IQ. Hagenbach and Perez-Richet (2018) show that Receivers can exercise skepticism also when the Senders' payoffs are not monotonic as in Milgrom (1981).

[8] Asymmetric information processing is documented in various experiments such as Eil and Rao (2011), Sharot et al. (2011), Charness and Dave (2017), Drobner and Goerg (2021), Kogan et al. (2021) or Möbius et al. (2022).

about states they intrinsically care about, such as their relative intelligence or beauty. They do so based on noisy signals provided to them by the experimenter. In our work, subjects update their beliefs using information which is *strategically disclosed* by another player. Subjects therefore need to reason about the other party's strategy to determine whether a given vague message is good or bad news. We shed light on a new channel that individuals can use to protect their beliefs: motivated skepticism.

We propose two behavioral models that can generate motivated skepticism. One model, presented in Section 2, is related to existing models of motivated beliefs formation that followed Bénabou and Tirole (2002) and Brunnermeier and Parker (2005). We share with these models the assumptions that the agent's utility depends directly on beliefs, and that the agent then faces a trade-off between holding accurate beliefs to take optimal decisions and holding comfortable beliefs.[9] The other model, in Section 6, proposes that the agent's favored beliefs are reached through a stepwise reasoning whose depth is motivated. The heuristic for strategic reasoning in disclosure games is borrowed from Schipper and Li (2020), and broadly related to the literature on strategic thinking in games (see Nagel, 1995; Camerer et al., 2004; Costa-Gomes and Crawford, 2006; Crawford et al., 2013; Crawford and Iriberri, 2007). Alaoui and Penta (2015) propose a framework to make players' depth of reasoning in games endogenous to the *material* benefits to reason. We complement their work by making reasoning depth endogenous to the *psychological* benefits to reason.

Only a few papers examine the formation of motivated beliefs in experimental settings which are strategic or social. In Schwardmann and Van der Weele (2019), Schwardmann et al. (2022), and Solda et al. (2020), individuals convince themselves that a state is true to better persuade others. In Thaler (2022) and Burro and Castagnetti (2022), Senders adapt their cheap talk to what Receivers want to hear about political issues or relative intelligence. Using a design similar to ours, Burro and Castagnetti (2023) show that Receivers are not fooled by positive cheap-talk messages about their relative intelligence. Oprea and Yuksel (2021) show that subjects rely more on peers' beliefs when these beliefs are self-serving. We consider hard information transmission and study motivated deviations from skeptical beliefs.

Thaler (2022) shows that subjects assess the veracity of information sources in a motivated way. For evidence of selective memory, see the survey by Amelio and Zimmermann (2023). Motivated information selection is documented in Grossman (2014), Grossman and Van der Weele (2017), Serra-Garcia and Szech (2021), Exley and Kessler (2021) and Chen and Heese (2023). See Golman et al. (2017) for a survey on information avoidance and Thompson and Loewenstein (1992) on biased search of information. A general discussion of the mechanics of motivated reasoning can be found in Epley and Gilovich (2016), and earlier in Kunda (1990).

[9]Beliefs can enter directly the agent's utility function through self-image concerns (as assumed here and in Bénabou and Tirole, 2011), anticipatory emotions (as in Kőszegi, 2006, Caplin and Eliaz, 2003 or Schwardmann, 2019) or motivational concerns (Bénabou and Tirole, 2002). Hagenbach and Koessler (2022) classify some of these functions and provide a discussion of the models of motivated-belief formation.

# 2 Theoretical framework

In this section, we present the well-known theoretical framework that guided our design and modify it to consider a Receiver whose utility is belief-based. We also give definitions and predictions that will help us analyze the experimental data.

## 2.1 Standard disclosure games

We consider a version of the standard Sender-Receiver disclosure game of Milgrom (1981) and Grossman (1981). The Sender is privately informed of a type $t$ and sends a costless message $m$ about it to the Receiver. We assume that $t$ is initially drawn from a finite set of positive real numbers $T$. Upon receiving $m$, the Receiver updates his beliefs about $t$ and chooses an action $a \in \mathbb{R}$ which affects both players. The Sender's utility $u_S(a)$ is type-independent. The Receiver's utility is given by $u_R(a;t) = -(a-t)^2$, capturing the idea that the Receiver wants his action to match the type.

When the type is $t$, the set of messages available to the Sender is $M(t)$. This set contains all the subsets of $T$ which include $t$ and are made up of consecutive numbers.[10] As an example, consider $T = \{1, 2, 3, 4, 5\}$. The set of messages available to a Sender of type $t = 1$ is $M(1) = \{\{1\}, \{1, 2\}, \{1, 2, 3\}, \{1, 2, 3, 4\}, \{1, 2, 3, 4, 5\}\}$. The set of all messages available to the Sender is $M = \cup_{t \in T} M(t)$. With this message structure, information is hard: a message $m \in M$ provides *evidence* that the true type is in $m$. A message is *precise* when it is a singleton, and is *vague* otherwise. The *size* of message $m$ is its cardinality $|m|$. The highest and lowest types in $m$ are respectively denoted $t_{sup}(m)$ and $t_{inf}(m)$.

A (pure) strategy of the Sender is a mapping $\sigma_S(\cdot)$ from $T$ to $M$ such that $\sigma_S(t) \in M(t)$. The Sender's strategy is *fully revealing* when it is separating: $\sigma_S(t) \neq \sigma_S(t')$ for every $t \neq t'$. A (pure) strategy of the Receiver is a mapping $\sigma_R(\cdot)$ from $M$ to the set of actions $\mathbb{R}$. $\beta_m \in \Delta(T)$ denotes the belief of the Receiver following message $m$. We say that a belief $\beta_m$ is *consistent* with $m$ if $\beta_m$ has support in $m$.

We solve this game using Perfect Bayesian equilibrium. In equilibrium, *(i)* $\sigma_S(\cdot)$ is a best-reply to $\sigma_R(\cdot)$, *(ii)* $\beta_m$ is derived from Bayes' rule after any message $m$ sent in equilibrium, and *(iii)* for every $m$, the action that maximizes the Receiver's utility given $\beta_m$ is $\sigma_R(m) = E_{\beta_m}(t)$. *(ii)* implies that Receivers' beliefs are consistent after any equilibrium message $m$. We add the requirement, common in disclosure games, that $\beta_m$ is also consistent after any messages $m$ sent off equilibrium

[10] This message structure corresponds to the *rich language* considered in Hagenbach and Koessler (2017) or Ali et al. (2021). An alternative language that is commonly considered in theoretical and experimental works is the *simple language*: the Sender of type $t$ either discloses $\{t\}$ or $T$. The rich language allows for more nuances in disclosure.

path. By extension of consistent beliefs, we define consistent actions:

**Definition 1** An action $\sigma_R(m)$ is *consistent* with $m$ if it is optimal for a belief that has support in $m$.

In the experimental data, we will examine the consistency of Receivers' actions. It will serve as a check that Receivers have understood that Senders transmit hard evidence.

## 2.2 Receiver's skepticism

The notion of skepticism, introduced by Milgrom (1981) and Milgrom and Roberts (1986), plays a central role in disclosure games as it prescribes how the Receiver should interpret vague messages. This interpretation naturally depends on the commonly-known Sender's objective. In our experiment, we will vary this objective by changing whether the Sender wants to induce a *High* or a *Low* Receiver's action.

In *High* games, the Sender's payoff strictly increases in the Receiver's action $a$.[11] Since the Sender wants to induce a high action, a Sender of high type should disclose this type precisely. In contrast, a Sender of low type may have an interest in shrouding information and sending vague messages. It follows that an intuitive way for the Receiver to read a vague message is to be skeptical and interpret it as coming from a Sender of low type. In *High* games, we define skeptical beliefs as follows:

**Definition 2** In *High* games, the belief $\beta_m$ is *skeptical* if it assigns probability one to the lowest type in $m$. We denote this type $t_{skep}(m)$.

In *Low* games, the Sender's objective if reverted: the Sender's payoff strictly decreases in the Receiver's action $a$. Since the Sender's objective is now to induce a low action, skepticism consists in interpreting vague messages as coming from the highest type disclosed:

**Definition 3** In *Low* games, the belief $\beta_m$ is *skeptical* if it assigns probability one to the highest type in $m$. We denote this type $t_{skep}(m)$.

We now recall the unraveling result which establishes that every equilibrium of *High* and *Low* games is fully-revealing, and that the Receiver's equilibrium beliefs are always skeptical. The proof is given in Appendix A.

[11] As explained in Hagenbach et al. (2014), most of the disclosure literature assumes this specific state-independent and monotonic payoff for the Sender. It also corresponds to the Sender's payoff considered in all the experiments on strategic disclosure mentioned in the literature review. Compared to more complex Sender's payoff functions, this payoff has the advantage of making it relatively simple for the Receiver to identify how to be skeptical. This leaves less room a priori for the Receiver to lack skepticism.

**Proposition 1** *In High and Low games, every equilibrium is fully revealing. In equilibrium, the Receiver's belief is skeptical after every (on and off-path) message.*

In our experiment, we will observe the Receivers' actions but not directly their beliefs. We define an action following $m$ as *skeptical* if it is optimal for the skeptical belief $\beta_m$.

**Definition 4** In *High* games, the skeptical action after message $m$ equals $t_{inf}(m)$. In *Low* games, the skeptical action after message $m$ equals $t_{sup}(m)$.

## 2.3 A psychological Receiver

In our experiment, we will consider settings in which the Receiver's utility is standard in that it depends only on his action and the type, $u_R(a,t)$. We will also consider settings in which the Receiver additionally and intrinsically cares about the beliefs he holds about $t$. For simplicity, we incorporate this fact into the model by considering that the Receiver's utility is given by the sum of his material utility $u_R(a,t)$ and a psychological payoff that depends directly and only on his expectation about $t$. We assume that the psychological payoff decreases in this expectation because, in our experiment, the Receiver will intrinsically prefer to have a lower expectation of an ego-relevant type $t$. For a given action $a$, type $t$ and belief $\beta$ about $t$, we assume that the Receiver's utility is given by

$$\tilde{u}_R(a;t,\beta) = -(a-t)^2 - \alpha E_\beta(t)$$

where $\alpha > 0$ parameterizes self-image concerns.[12]

When the Receiver's utility is given by $\tilde{u}_R(a;t,\beta)$ and he faces a vague message $m$, he may trade-off the material benefit of holding skeptical beliefs after $m$ $-$ these beliefs are accurate in equilibrium $-$ and the psychological benefit of having a lower expectation of $t$. To examine this trade-off, we assume that, after receiving message $m$ in equilibrium, the Receiver chooses his beliefs optimally while considering (i) the information available at that time, that is, that the true type is $t_{skep}(m)$, and (ii) that he will subsequently choose an action which is optimal for the beliefs he will eventually hold.[13] Regarding point (ii) and considering $\tilde{u}_R(a;t,\beta)$, the Receiver's optimal action $a$ is equal to $E_{\beta_m}(t)$ when his belief is $\beta_m$. It follows that, after receiving message $m$, the Receiver

[12]For a categorization of psychological utilities $u(a;t,\beta)$ which received recent attention in economics, see Hagenbach and Koessler (2022). The form we consider fits into the category of *separable utilities*, with the particularity that beliefs affect utility only through the expectation of the type.

[13]There exist at least two categories of belief-formation models. In one approach, which we follow in spirit here, beliefs are chosen freely by trading-off the costs and benefits of distorting beliefs away from the Bayesian framework (see Brunnermeier and Parker, 2005 or Caplin and Leahy, 2019). In another approach, beliefs are formed in equilibrium of a multi-self game (see Bénabou and Tirole, 2002, Hestermann et al., 2020 or Hagenbach and Koessler, 2022). In both categories, a single agent is initially informed of the state of the world but manipulates his beliefs about it for psychological reasons. Point (i) is an adaption of this assumption to our communication setting.

chooses the belief $\beta_m$ that maximizes

$$-(E_{\beta_m}(t) - t_{skep}(m))^2 - \alpha E_{\beta_m}(t)$$

under the constraint, which we add, that $\beta_m$ stays consistent with $m$. Solving this program, we get that the chosen belief $\beta_m$ is such that $E_{\beta_m}(t) = \max\{t_{inf}(m), t_{skep}(m) - \frac{\alpha}{2}\}$.

In *High* games, $t_{skep}(m) = t_{inf}(m)$, so $\beta_m$ is such that $E_{\beta_m}(t) = t_{inf}(m)$. The Receiver does not manipulate his belief away from the skeptical belief. Indeed, the skeptical belief in *High* games is the consistent belief that maximizes the psychological payoff. We say that the skeptical belief is *self-serving*. In *Low* games, $t_{skep}(m) = t_{sup}(m)$, so $\beta_m$ is not such that $E_{\beta_m}(t) = t_{sup}(m)$. The Receiver manipulates his beliefs away from the skeptical beliefs (and more so when self-image concerns are more important). Indeed, the skeptical belief in *Low* games is the consistent belief that minimizes the psychological payoff. We say that the skeptical belief is *self-threatening*. We summarize our findings below, and use the term *psychological Receiver* for a Receiver with utility $\tilde{u}_R(a; t, \beta)$ and who manipulates his belief as described above.

**Proposition 2** *In High games, the psychological Receiver's belief is skeptical after every vague message. In Low games, the psychological Receiver's belief is never skeptical after any vague message.*

According to Propositions 1 and 2, the Receiver with a standard utility function takes the skeptical action in *High* and *Low* games. When the Receiver is psychological, he takes the skeptical action in the *High* but not in the *Low* games.

## 2.4 On the Sender's fully-revealing strategy

Proposition 1 establishes that the Sender uses a fully-revealing strategy in every equilibrium. This strategy can be made up of precise or vague messages. Consider that the Sender thinks that the Receiver could, for any reason including mistakes, act as follows: when a type $t$ is contained in the set of types disclosed, there is always a probability that the Receiver takes action $a = t$. In this case, a specific fully-revealing strategy is optimal for the Sender: in *High* games, every type $t$ discloses that the type is *at least* $t$, that is, sends the message $\{t, \ldots, t_{sup}\}$ with $t_{sup}$ the highest type in $T$; conversely, in *Low* games, every type $t$ discloses that the type is *at most* $t$, that is, sends the message $\{t_{inf}, \ldots, t\}$ with $t_{inf}$ the lowest type in $T$. In each game, we refer to this specific Sender's strategy as *selective disclosure*.[14]

[14] Milgrom and Roberts (1986) already point to this result which is proved in Hagenbach and Koessler (2017).

# 3 Experimental design

We present the overall structure of our experiment before describing the two experimental variations that we consider.

## 3.1 Overall structure

The experiment has two parts. Subjects' payment is the sum of a show-up fee and of the money earned in each part.

**Part 1: IQ test.** Subjects begin by completing a 15-minutes test made up of 15 Raven matrices (Raven, 1936). They earn 0.50 euro per correctly solved matrix. A subject's *performance* is an integer between 0 and 15 that corresponds to the number of correctly-solved matrices. When the IQ test is over, we elicit in an incentivized way the subjects' beliefs about their performance relative to the performance of a benchmark group made up of 99 subjects who did the same IQ test.[15]

**Part 2: Sender-Receiver games.** In the second part, subjects play 10 times the same Sender-Receiver game. Before the 10 games start, subjects learn whether they will play as a Sender or as a Receiver. Subjects keep this role for the 10 games but are randomly matched in Sender-Receiver pairs at the beginning of every game, which is common knowledge. Each of the 10 games has four steps:

*Step 1* The computer generates a type $t$ in $\{1, 2, 3, 4, 5\}$. In Subsection 3.2, we detail how the type is generated, a key element of our experimental manipulation.

*Step 2* The Sender is privately informed about the generated type $t$.

*Step 3* The Sender decides which message $m$ about $t$ to sent to the Receiver. For any given type $t$, the set of messages available to the Sender is restricted to the sets of consecutive types which contain $t$.

*Step 4* The Receiver observes the message $m$ sent in *Step 3* and makes a guess $a \in [1, 5]$ about the type $t$. We allow for guesses with one digit.[16]

When a Sender-Receiver game is over, subjects move to the next game without getting any feedback about the type $t$ that the Sender had effectively seen or about their realized payoff.

[15] We had previously ran 5 sessions with the only objective to gather the performance of these 99 subjects in the IQ test. The beliefs elicitation procedure and payment is detailed in Subsection 5.3.

[16] The complete instructions are given in Online Appendix 8. Online Appendix 9 provides an example of the screens seen by the Sender and by the Receiver.

**Payoffs.** Players' monetary payoff functions are common knowledge. The Sender's payoff function only depends on the Receiver's guess $a$ and not on the type generated in *Step 1* of the game. The Sender's payoff function is part of our experimental manipulation and is detailed in Subsection 3.2. The Receiver's payoff function is the same in all treatments and depends both on his guess $a$ and on the type $t$. It is given by the following formula: $5 - |a - t|$.[17] When the Receiver believes a type $t'$ for sure, his optimal guess equals $t'$. We give subjects the exact formula, and explain them that their payoff is higher when their guess is closer to the true type. In each game, both players' payoffs can range from 1 to 5 euros. One of the 10 games is randomly selected for payment of Part 2.

## 3.2 Treatments

Our experiment has four treatments that result from the crossing of the two variations introduced in Section 2: the *High* or *Low* Sender's objective, and whether or not the Receiver intrinsically cares about the type $t$. We now explain how we implement these variations in the lab. The two-by-two design is implemented *between subjects.*

### 3.2.1 Variation 1: *Neutral* vs. *Loaded* type

We vary exogenously whether the type $t$ is *Loaded* or *Neutral.*

In the *Loaded* treatments, the type $t$ corresponds to a measure of the relative performance of the Receiver in the IQ test completed in Part 1. Let us call this loaded type the *IQ-rank* of the Receiver, and explain how it is generated. In *Step 1* of each Sender-Receiver game, the computer randomly selects four subjects from the benchmark group of 99 subjects who did the IQ test previously. For each Receiver, the IQ-rank $t \in \{1, 2, 3, 4, 5\}$ is then computed by comparing his performance to the performance of these four randomly-selected subjects (with ties broken at random):

- ◇ $t = 1$ when the Receiver has the highest performance in the group of five subjects or, said differently, when he ranks first,
- ◇ $t = 2$ when the Receiver has the second highest performance in the group of five subjects or, said differently, when he ranks second,
- ◇ etc.

Importantly, a new IQ-rank is computed in every game: four new subjects are randomly selected from the benchmark group in *Step 1* of each game.

[17] This payoff function is not exactly the same as in the theory section. In the experimental disclosure literature, experimenters usually consider relatively complex Receivers' payoff functions and show them payoff matrices in which they see an integer payoff for every pair of type and action. Instead of doing that, which usually requires restricting the set of actions available to the Receiver, we presented the function with the distance because of its simplicity. In Section 1 of the Online Appendix, we show that, with this payoff function, the Receiver's optimal action is equal to the median of his beliefs (Proposition 4) and that the results of the theory section are still valid.

In the *Neutral* treatments, the type $t$ also corresponds to a rank but it has no particular meaning. Let us call this type a *neutral rank*, and explain how it is generated. At the very beginning of Part 2, that is, before the 10 Sender-Receiver games are played, an integer between 0 and 15 is randomly attributed to each Receiver.[18] In *Step 1* of each Sender-Receiver game, the computer randomly selects four other integers between 0 and 15. For each Receiver, the neutral rank $t \in \{1, 2, 3, 4, 5\}$ is then computed by comparing his integer to the four randomly-selected other integers (with ties broken at random):

- ⋄ $t = 1$ when the Receiver's integer is the highest in the group of five integers,
- ⋄ $t = 2$ when the Receiver's integer is the second highest in the group of five integers,
- ⋄ etc.

Importantly, a new neutral rank is computed in every game: four new integers are randomly selected in *Step 1* of each game.

We now discuss three important aspects of the first experimental variation.

**Receivers' preferences over beliefs.** In both the *Loaded* and the *Neutral* treatments, the Receiver precisely knows how the type is generated. In particular, he knows the type is a measure of relative intelligence in *Loaded* and a rank with no particular meaning in *Neutral*. The main hypothesis behind the *Loaded* / *Neutral* manipulation is that it affects whether or not the Receiver has intrinsic preferences over what he believes about the type. In the *Loaded* treatments, as in Subsection 2.3 of the theory, we assume that the Receiver prefers to have a higher expectation of his IQ-rank (closer to 1). In the *Neutral* treatment, there is no such preference over the rank.

In the *Loaded* treatments, the Receiver forms beliefs about his IQ-rank knowing that it is correlated to the payment for the IQ test. It may therefore be that Receivers prefer to believe that their IQ-rank is high because they prefer to anticipate a high payment and not because this is ego-relevant. In our experiment, subjects know that they will later learn their payment but that the uncertainty about their relative intelligence will never be resolved.[19] According to Drobner (2022), this should ensure that agents form motivated beliefs about the loaded type because it is ego-relevant. For an additional check, we have run a treatment in which the payment of the IQ test was fixed and show that our results are unaffected by this variation (see Online Appendix 3 for details about this additional treatment).

[18]This procedure is meant to parallel the fact that, in *Loaded* treatments, Receivers start Part 2 with a fixed IQ performance between 0 and 15.

[19]In Part 1, we write *"You will never learn how many correct answers you gave in the test."* In Part 2, we write *"You will never receive more information about your IQ-ranks than the information given by the Senders."*

**What do Senders know about the type generation?** In *Step 1* of every game, the Sender is fully informed about the type $t$. However, in both the *Loaded* and *Neutral* treatments, the type is simply presented to the Sender as a "secret number" from the set {1,2,3,4,5}. In other words, the Sender does not know how $t$ is generated in the different treatments. The Receiver knows that the Sender does not know how the type is generated. Because Senders do not know that the type has a different meaning for Receivers in the *Loaded* and *Neutral* treatments, their communication strategy are comparable across treatments (provided that the Sender's payoff is the same). This allows us to focus on the Receivers' reactions facing comparable messages in *Neutral* and *Loaded* treatments. If Senders knew they were transmitting information about Receivers' relative IQ, it could affect their communication strategies in potentially complex ways (they could derive pleasure or discomfort disclosing good or bad news about Receivers relative IQ, etc.).

**Comparability between the *Loaded* and *Neutral* treatments.** We designed the *Neutral* treatments to be as comparable as possible to the *Loaded* treatments. In the *Loaded* treatments, as the 10 games are played, the Receiver may learn something about his performance in the IQ test even if the IQ-rank is newly computed in each round. In the *Neutral* treatments, a similar process can occur because the Receiver is initially attributed an integer that is fixed for the 10 games. In both treatments, we do not give Receivers feedback between the games to limit this learning. One difference however remains. In the *Loaded* treatment, the Receiver may have some information about his IQ-rank because he has experienced the IQ test. In the *Neutral* treatment, he only knows his integer has been selected according to a uniform distribution. We examine the impact of Receivers' prior beliefs in Subsection 5.3.

### 3.2.2 Variation 2: *High* vs. *Low* guess

We vary exogenously whether the Sender's objective is to induce a *High* or a *Low* guess from the Receiver. Payoffs are common knowledge, so the Receiver knows the Sender's objective.

In the *High* treatments, the Sender's payoff is equal to the Receiver's guess $a$. According to theory, when $m$ is vague, the Receiver should be skeptical and, as stated in Definition 2, believe the lowest type / highest rank in $m$. For instance, if $m = \{3, 4, 5\}$, the skeptical belief assigns probability one to $t = 3$. In the *Low* treatments, the Sender's payoff is equal to $6 - a$. According to theory, when $m$ is vague, the Receiver should be skeptical and, as stated in Definition 3, believe the highest type / lowest rank in $m$. For instance, if $m = \{1, 2, 3\}$, the skeptical belief assigns probability one to $t = 3$.

### 3.2.3 Two-by-two design

By crossing variations 1 and 2, we affect whether, for any given message, the Receiver's skeptical belief is neutral, self-serving or self-threatening.

Table 1: Summary of treatments

| Treatment | Skeptical belief facing $m$ | Intrinsic preferences over beliefs? | Skeptical belief is |
|---|---|---|---|
| *High_ Neutral* | Highest rank in $m$ | No | Neutral |
| *High_ Loaded* | Highest rank in $m$ | Yes | Self-serving |
| *Low_ Neutral* | Lowest rank in $m$ | No | Neutral |
| *Low_ Loaded* | Lowest rank in $m$ | Yes | Self-threatening |

## 3.3 Main hypotheses

Our main hypotheses relate to the Receivers' levels of skepticism in the different treatments. We construct a measure of skepticism that captures the idea that a Receiver, while being consistent, is less skeptical when the distance between his guess and the skeptical guess (Definition 4) is larger. To make the measure comparable across games, we normalize this distance by the maximal distance to the skeptical guess that any consistent guess could have, that is, by the size of the message. The measure is denoted $Sk(a,m)$ and constructed for each game in which the Receiver made a consistent guess $a$ when facing a vague message $m$.[20] The measure is in $[0,1]$ and equals 1 when the Receiver makes the skeptical guess.

**Definition 5** In *High* treatments, for every vague $m$ and consistent $a$, the measure of skepticism is given by:

$$Sk(a,m) = 1 - \frac{a - t_{inf}(m)}{t_{sup}(m) - t_{inf}(m)}$$

In *Low* treatments, for every vague $m$ and consistent $a$, the measure of skepticism is given by:

$$Sk(a,m) = 1 - \frac{t_{sup}(m) - a}{t_{sup}(m) - t_{inf}(m)}$$

In the data, skepticism will be evaluated using averages of $Sk(a,m)$ over games and frequencies of skeptical guesses. We use *Neutral* treatments as benchmarks and formulate our main hypotheses in terms of differences in skepticism between the *Loaded* and *Neutral* treatments. We also formulate an hypothesis on *Neutral* treatments. All the hypotheses presented below have been pre-registered (AEARCTR-0007541).

[20]When considering precise messages, consistent guesses are necessarily skeptical.

According to Proposition 1 in Section 2, Receivers should be skeptical in *High_ Neutral* and in *Low_ Neutral*. Since we see no reason a priori to think that skepticism is harder to exercise when it consists in assigning probability one to the lowest or to the highest type disclosed, we formulate the first hypothesis:

**Hypothesis 1:** *Facing vague messages, Receivers are as skeptical in High_ Neutral as in Low_ Neutral.*

Proposition 1 and Proposition 2 of Section 2 predict that Receivers should be skeptical in *High_ Neutral* and in *High_ Loaded*. This however relies on the theoretical result that individuals are fully skeptical in *High_ Neutral*, which leaves no room to increase skepticism when it is self-serving. We know from previous experiments that subjects are not always skeptical, so the next hypothesis incorporates the possibility that skepticism could be higher in *High_ Loaded*:

**Hypothesis 2:** *Facing vague messages, Receivers are at least as skeptical in High_ Loaded as in High_ Neutral.*

Proposition 1 and Proposition 2 of Section 2 predict that Receivers should be skeptical in *Low_ Neutral* but not in *Low_ Loaded*, which leads to the last hypothesis:

**Hypothesis 3:** *Facing vague messages, Receivers are less skeptical in Low_ Loaded than in Low_ Neutral.*

## 3.4 Implementation

A total of 464 subjects participated in the experiment, 232 Receiver subjects and 232 Sender subjects.[21] Subjects belonged to the pool of the WZB-TU Lab in Berlin, which is mostly made up of students from the Technical University of Berlin. They were invited to virtual experimental sessions, of about 20 subjects each, on Zoom. Once checked in, each subject received a link to run the experiment on his own computer (while staying on Zoom in the presence of the experimenter). The experiment was programmed with z-Tree (Fischbacher, 2007) and run using z-Tree unleashed (Duch et al., 2020). Sessions took around one hour and subjects earned on average 19.51 euros ($s.d. = 2.13$).

## 3.5 Data analysis

**Sample.** In the data, we drop 16 Receivers for whom at least one game out of 10 could not be played because of computer bugs (either the Sender they were matched with could not send a message, or the Receiver himself could not make a guess). We also drop 16 Receivers who, in

[21] 150 additional subjects participated in two complementary treatments detailed in Online Appendix 3.

more than half of the games they played, made a guess which was not consistent with the evidence contained in the message received. We believe that these subjects did not understand that Senders were disclosing hard evidence about the type.[22] Overall, this leaves 200 Receiver subjects. Each of them played 10 games, so our dataset is made up of 2000 games. Each of these 2000 observations consists of a true type $t$, a message $m$ sent by the Sender and a Receiver's guess $a$. Among these games, 540 correspond to the *High_ Neutral* treatment, 480 to the *High_ Loaded* treatment, 500 to the *Low_ Neutral* treatment and 480 to the *Low_ Loaded* treatment.

**Statistical tests.** Unless noted otherwise, for all statistical tests we report p-values obtained from random-effects linear regressions on panel data with the Senders' or Receivers' identifiers as the group variable and the rounds as the time variable.[23] Standard errors are clustered at the session level using bootstrapping. In Appendix B, we provide robustness checks of the tests reported in the main text by exploring alternative specifications. They include (i) accounting for the bounded (sometimes binary) nature of the dependent variable by using Probit or Tobit models when appropriate, (ii) using linear regressions without considering panel data structure, and (iii) clustering at the individual level rather than at the session level.

# 4 Experimental results: first steps

## 4.1 Senders' communication strategies

We first describe Senders' strategies and, in particular, check that they account for the Senders' reversed objective in *High* and *Low* treatments.

**Senders' strategy in *High* treatments.** Table 2 reports the frequency with which each message is sent conditionally on the Sender observing each type $t$. In the *High* treatments, in which Senders want Receivers to make a high guess (closer to 5), Senders of type $t$ most often disclose that the type is *at least* $t$: they do so 67.84% of the time over all types. This disclosure strategy corresponds to selective disclosure described in Subsection 2.4.[24]

[22] As shown in Online Appendix 4, our main results are unaffected by the inclusion of these Receivers. Following Jin et al. (2021), we additionally checked whether Receivers could be inconsistent because they care about the Senders' payoffs and find no support for this hypothesis.

[23] When studying Senders' communication strategies, the group identifier variable is the Senders' identity. When studying Receivers' behavior, the group identifier variable is the Receivers' identity.

[24] The experiment of Deversi et al. (2021) in which Senders can use a rich/flexible language corresponds to our *High_ Neutral* treatment with 6 instead of 5 possible types. The authors report that the messages sent most often by Senders almost perfectly coincide with the prediction of selective disclosure.

Table 2: Senders' communication strategy in *High* treatments

| | Message | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Type | {1} | {2} | {3} | {4} | {5} | {1,2} | {2,3} | {3,4} | {4,5} | {1,2,3} | {2,3,4} | {3,4,5} | {1,2,3,4} | {2,3,4,5} | {1,2,3,4,5} | Total |
| 1 | 5.24 | - | - | - | - | 2.62 | - | - | - | 9.17 | - | - | 15.28 | - | **67.69** | 100 |
| 2 | - | 6.11 | - | - | - | - | 3.93 | - | - | 0.87 | 17.90 | - | 2.18 | **54.59** | 14.41 | 100 |
| 3 | - | - | 6.97 | - | - | - | 1.49 | 11.94 | - | 0.50 | 4.98 | **65.17** | 1.49 | 2.49 | 4.98 | 100 |
| 4 | - | - | - | 16.76 | - | - | - | 1.16 | **72.83** | - | 1.16 | 5.78 | 1.16 | - | 1.16 | 100 |
| 5 | - | - | - | - | **82.45** | - | - | - | 9.04 | - | - | 4.26 | - | 2.13 | 2.13 | 100 |
| Total | 1.18 | 1.37 | 1.37 | 2.84 | 15.20 | 0.59 | 1.18 | 2.55 | 14.02 | 2.35 | 5.20 | 14.61 | 4.41 | 13.14 | 1.99 | 100 |

*Note:* The Table reports the frequency with which each message is sent conditionally on the Sender observing each type $t$, in the *High* treatments. Numbers in red highlight the most frequently sent message for each type. For instance, Senders of type $t = 5$ send the precise message $m = \{5\}$ 82.45% of the time.

Regarding the comparison between *High_Neutral* and *High_Loaded*, we remind that Senders see the same instructions. For each type, the message sent most often is the same in these two treatments, and the frequency with which this message is sent is never significantly different (except marginally when $t = 4$, $p = 0.076$).[25]

**Senders' strategy in *Low* treatments.** Table 3 gives the frequency with which each message is sent conditionally on the Sender observing each type $t$. In the *Low* treatments, in which Senders want Receivers to make a low guess (closer to 1), Senders of type $t$ most often disclose that the type is *at most* $t$: they do so 70.82% of the time over all types. This disclosure strategy also corresponds to selective disclosure.

Table 3: Senders' communication strategy in *Low* treatments

| | Message | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Type | {1} | {2} | {3} | {4} | {5} | {1,2} | {2,3} | {3,4} | {4,5} | {1,2,3} | {2,3,4} | {3,4,5} | {1,2,3,4} | {2,3,4,5} | {1,2,3,4,5} | Total |
| 1 | **72.82** | - | - | - | - | 10.77 | - | - | - | 6.15 | - | - | 3.59 | - | 6.67 | 100 |
| 2 | - | 7.10 | - | - | - | **71.01** | 1.78 | - | - | 7.69 | 0.59 | - | 2.37 | 5.33 | 4.14 | 100 |
| 3 | - | - | 2.76 | - | - | - | 5.52 | 1.10 | - | **75.69** | 3.31 | 2.21 | 2.76 | 3.87 | 2.76 | 100 |
| 4 | - | - | - | 1.22 | - | - | - | 3.27 | 0.41 | - | 15.51 | 1.63 | **62.45** | 4.90 | 10.61 | 100 |
| 5 | - | - | - | - | 0.53 | - | - | - | 3.68 | - | - | 6.32 | - | 14.74 | **74.74** | 100 |
| Total | 14.49 | 1.22 | 0.51 | 0.31 | 0.10 | 14.39 | 1.33 | 1.02 | 0.82 | 16.53 | 4.59 | 2.04 | 17.24 | 5.71 | 19.69 | 100 |

*Note:* The Table reports the frequency with which each message is sent conditionally on the Sender observing each type $t$, in the *Low* treatments. Numbers in red highlight the most frequently sent message for each type. For instance, Senders of type $t = 1$ send the precise message $m = \{1\}$ 72.82% of the time.

Regarding the comparison between *Low_Neutral* and *Low_Loaded*, we remind again that Senders see the same instructions. For each type, the message sent most often is the same in these two treatments, and the frequency with which this message is sent is never significantly different (except marginally when the type is 4, $p = 0.089$).[26]

[25] Online Appendix 2 gives the frequencies of each $m$ conditional on $t$ in *High_Neutral* and *High_Loaded* separately.
[26] Online Appendix 2 gives the frequencies of each $m$ conditional on $t$ in *Low_Neutral* and *Low_Loaded* separately.

**Result 1.** Senders' strategies are well-structured: in *High* treatments, Senders of type $t$ most often disclose that the type is *at least* $t$; in *Low* treatments, Senders of type $t$ most often disclose that the type is *at most* $t$.

## 4.2 Messages seen by Receivers

The messages seen by Receiver in *High* and *Low* treatments are different because of Senders' reversed objectives. We however can check whether Receivers face comparable messages in the four treatments by considering two objects. The first is the size of the messages seen by Receivers in the various treatments, since it may be easier to make skeptical inferences when a smaller set of types is disclosed. The average size of the messages seen is not different between the *High_Neutral* and *High_Loaded* treatments (2.90 and 3.01, $p = 0.463$). Messages are slightly vaguer in the *Low_Neutral* treatment than in the *Low_Loaded* (average sizes are 3.25 and 2.98; $p = 0.014$). When studying Receivers' skepticism, we will take into account the size of the messages they saw to control for these variations.[27] The second object is the skeptical guesses which correspond to the vague messages that Receivers see. Considering all vague messages, the average skeptical guess is neither different between *High_Neutral* and *High_Loaded* (2.21 and 2.24, $p = 0.686$), nor different between *Low_Neutral* and *Low_Loaded* (3.85 and 3.71, $p = 0.362$).

**Observation 1.**[28] When making their guesses, Receivers are in comparable situations in all treatments. In the two *High* treatments, the messages seen by Receivers are similar in terms of size and of skeptical guesses that these messages should induce. In the two *Low* treatments, the messages seen by Receivers are similar in terms of size and skeptical guesses that these messages should induce, with slightly vaguer messages in *Low_Neutral*.

## 4.3 Preliminary checks about Receivers' guesses

Before studying whether Receivers' guesses are skeptical, we make two preliminary checks.

**Consistency of Receivers' guesses.** Applying Definition 1 of the theory section, a guess is consistent with message $m$ if it is in $m$. Over all treatments, the percentage of consistent guesses is 98.10%, which is very high.[29] It ranges from 95.74% in *High_Neutral* to 99.37% in *Low_Loaded*, with

[27] The measure of skepticism $Sk(a, m)$ is normalized by the size of the message. When considering the frequency of skeptical guesses, we will use message size as a control variable.

[28] We use the term "result" for findings linked to our experimental treatments, and the term "observation" otherwise.

[29] This percentage is in line with the levels of consistency reported in Hagenbach and Perez-Richet (2018) or Schipper and Li (2020) who also allow for vague messages.

no significant differences between the treatments. The rate of consistency is lower when Receivers received a precise message (92.25%) than a vague message (99.50%, $p = 0.004$) as is mechanically expected. When looking at the very few inconsistent guesses made by Receivers, we find no evidence that inconsistent guesses correspond to higher ranks.

**Observation 2.** Receivers' guesses are consistent with the evidence provided by Senders.

**Receivers' guesses in *Neutral* and *Loaded* treatments.** The Receivers' monetary payoff function is the same in all treatments. In *Loaded* treatments, we assume that Receivers additionally prefer to hold higher beliefs about their IQ-rank, so they should guess higher ranks in these treatments. Pooling all messages, Figure D.1 in Appendix D reports the distribution of Receivers' guesses in the *Neutral* and *Loaded* treatments. It shows that the frequency with which Receivers guess they ranked 2 or 3 is indeed higher in *Loaded* than in *Neutral* ($p = 0.024$ and $p = 0.022$, resp.). On the contrary, the frequency with which Receivers guess they ranked 5 is lower in *Loaded* than in *Neutral* ($p = 0.093$). We take this as an indication that our *Neutral/Loaded* experimental variation was effective.

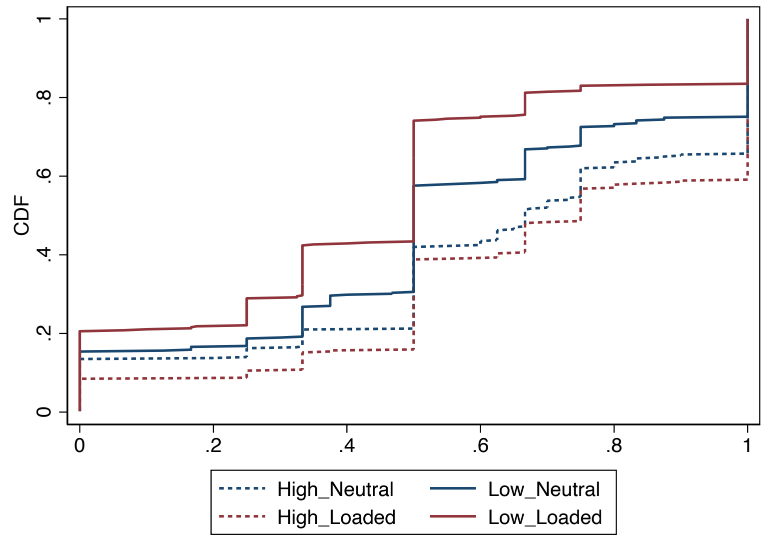
**Observation 3.** Receivers guess higher ranks in *Loaded* than in *Neutral* treatments.

# 5 Receivers' skepticism

In this section, we study Receivers' skepticism by testing the three hypotheses presented in Subsection 3.3. For every vague message $m$ and consistent guess $a$, skepticism is measured using the formula $Sk(a, m)$ from Definition 5. Overall, we look at the 1605 games in which the message is vague and the guess consistent, which corresponds to 80.25% of our data set.

## 5.1 Motivated skepticism

Figure 1 gives an overview of our main findings. It displays the cumulative distribution function of the measure of skepticism for each treatment. The closer the line is to the bottom-right of the box, the more skeptical the Receivers' guesses are. On that figure, we see that the level of skepticism is rather high in the *High* treatments, without a big difference between the *High_Neutral* and the *High_Loaded* treatments. The level of skepticism seems lower in the *Low* treatments, with a bigger difference between the *Low_Neutral* and the *Low_Loaded* treatments. Figure 1 suggests that the levels of skepticism in the four treatments is indeed ordered as hypothesized in Subsection 3.3.




*Note:* The Figure displays the cumulative distribution function of skepticism, by treatment. The closer the line is to the bottom-right of the box, the more skeptical the Receivers' guesses are. On the contrary, the closer the line is to the top-left of the box, the less skeptical the Receivers' guesses are.

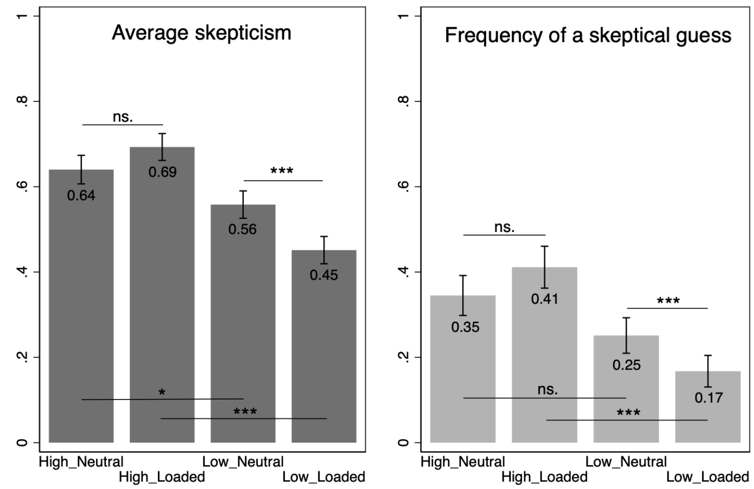
Figure 1: Cumulative distribution of skepticism, by treatment

The left part of Figure 2 displays the average value of $Sk(a, m)$ by treatment and gives another representation of our findings. First, the average level of skepticism is higher in *High_ Loaded* than in *High_ Neutral* but the difference is not significant ($p = 0.341$). This indicates that Receivers' skepticism is not significantly enhanced when the skeptical belief is self-serving. Second, the average level of skepticism in *Low_ Loaded* is strongly and significantly lower than in *Low_ Neutral* ($p = 0.005$). On average, Receivers are less skeptical when skepticism involves believing a low IQ-rank than when it involves believing a low neutral rank. Third, the average levels of skepticism in *Low_ Neutral* and *High_ Neutral* are marginally different from each other ($p = 0.076$). This small difference could be explained by the fact that it is easier for Receivers to understand the objective of a Sender whose payoff equals their guess than the one of a Sender whose payoff equals 6 minus their guess.[30]

The above findings are confirmed when looking at the frequencies of skeptical guesses, reported on the right part of Figure 2.[31] They are also confirmed when looking at the 10 guesses of each individual: individual behaviors are similar in all but the *Low_ Loaded* treatment in which Receivers make significantly fewer guesses that are close to the skeptical ones (see Online Appendix 6 for details about skepticism at the individual level).

[30]On the Sender side, an observation from previous Tables 2 and 3 points to the same direction: there is a higher rate of full disclosure of type 5 in *High* treatments than of type 1 in *Low* treatments ($p = 0.056$).

[31]It is interesting to note that around 30% of guesses are skeptical in *Neutral* treatments, a rate which corresponds almost exactly to the one found in Deversi et al. (2021) who present an experiment similar to *High_ Neutral*.




*Note:* The left part of the Figure displays the average level of skepticism, by treatment. The right part of the Figure displays the average frequency of a skeptical guess, by treatment. Black segments are 95% confidence intervals. P-values are from random-effects linear regressions on panel data with the Receivers' id as the group variable and the rounds as the time variable. Standard errors are clustered at the session level using bootstrapping. See columns (1), (4) and (7) in Table 4 (left part of the Figure) and columns (1), (4) and (7) in Table C.1 (right part of the Figure). $^*p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Figure 2: Receivers' average skepticism, by treatment

Table 4 shows the determinants of the Receivers' level of skepticism. Columns (1) to (3) focus on the *High* condition. The coefficient of the *Loaded* dummy is small and insignificant in all three specifications. Columns (4) to (6) focus on the *Low* condition. The estimated negative coefficient of the *Loaded* dummy confirms that the level of skepticism is substantially and significantly lower in the *Low_ Loaded* than in the *Low_ Neutral* treatment. Columns (7) to (9) report coefficients of the full difference-in-differences specification. The coefficients of the interaction term is negative and significant, confirming the findings from columns (1)-(6). The estimated negative coefficient of the *Low* dummy shows the marginal difference between the *Neutral* treatments.

Columns (2), (5) and (8) include two additional control variables. First, they include the Receivers' performance in the IQ test, which we measure for all Receivers in Part 1 of the experiment. We see that better cognitive abilities are significantly and positively correlated to Receivers' skepticism, a correlation already established in Schipper and Li (2020).[32] Second, they include dummy variables for each round of play, which shows no overall learning trend. In fact, learning may be hindered by the fact that there are relatively few rounds, that Receivers do not face the same message in every round and do not get any feedback between the rounds (see Online Appendix 5 for more details). Columns (3), (6) and (9) additionally control for various demographic variables which include age, gender, educational attainment, socio-professional category, social class and experience in participating in experiments.

[32]We discuss the role of IQ on skepticism in more details in Online Appendix 7.

Table 4: Determinants of skepticism

| *Dep. Var.* | Skepticism | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | *High* treatments | | | *Low* treatments | | | *Difference-in-difference* | | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| 1 if *Loaded* | 0.048 | 0.017 | 0.016 | -0.109*** | -0.124*** | -0.156*** | 0.048 | 0.022 | 0.019 |
| | (0.050) | (0.046) | (0.050) | (0.036) | (0.037) | (0.059) | (0.052) | (0.049) | (0.050) |
| 1 if *Low* | | | | | | | -0.062* | -0.065* | -0.051 |
| | | | | | | | (0.035) | (0.034) | (0.039) |
| 1 if *Low_ Loaded* | | | | | | | -0.157** | -0.149** | -0.157** |
| | | | | | | | (0.063) | (0.061) | (0.065) |
| IQ performance | | 0.026*** | 0.026*** | | 0.019*** | 0.017* | | 0.023*** | 0.021*** |
| | | (0.006) | (0.009) | | (0.007) | (0.009) | | (0.005) | (0.006) |
| Rounds dummies | | ✓ | ✓ | | ✓ | ✓ | | ✓ | ✓ |
| *Demo.* | | | ✓ | | | ✓ | | | ✓ |
| Cons. | 0.639*** | 0.348*** | 0.458** | 0.577*** | 0.391*** | 0.164 | 0.639*** | 0.399*** | 0.464*** |
| | (0.020) | (0.064) | (0.187) | (0.028) | (0.061) | (0.261) | (0.020) | (0.056) | (0.154) |
| *N* | 789 | 789 | 789 | 816 | 816 | 816 | 1605 | 1605 | 1605 |

*Note:* The Table reports random-effects linear regressions on panel data with the Receivers' id as the group variable and the rounds as the time variable. Standard errors (in parentheses) are clustered at the session level using bootstrapping. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table C.1 in Appendix C replicates Table 4 considering whether or not a guess is skeptical. When looking at the Receivers' likelihood to make a skeptical guess rather than at the distance between the Receiver's guess and the skeptical guess, the results of Table 4 are confirmed.

Overall, our results validate Hypothesis 2 (no difference between *High_ Neutral* and *High_ Loaded*) and Hypothesis 3 (lower skepticism in *Low_ Loaded* than in *Low_ Neutral*). Regarding Hypothesis 1 about the two *Neutral* treatments, we observe a small difference in skepticism. In Appendix B, we report alternative specifications to test the robustness of the effects presented in Table 4 and Table C.1. The validation of Hypotheses 2 and 3 is robust to all specifications considered. Regarding Hypothesis 1, the level of skepticism between the two *Neutral* treatments appears either not significantly or marginally significantly different depending on the specification. Our results can be summarized as follows:

**Result 2.** (a) Receivers' skepticism is not significantly higher when it is self-serving than when it is not. (b) Receivers' skepticism is significantly lower when it is self-threatening than when it is not. (c) The average level of skepticism is marginally lower in *Low_ Neutral* than in *High_ Neutral*.

Result 2 points to the following asymmetry: compared to the *Neutral* setting, the *Loaded* setting hinders skepticism in the *Low* treatments more than it enhances it in the *High* treatments. One reason for this asymmetry could be that the level of skepticism in *High_ Neutral* and *Low_ Neutral* is so high that it leaves little room for increase in *High_ Loaded* (as predicted by Propositions 1 and 2). Since only 30% of guesses are skeptical in *Neutral* treatments, this explanation does not seem entirely convincing. Another possible reason for this asymmetry is that it is more unpleasant to make the skeptical guess in *Low_ Loaded* than it is pleasant to make it in *High_ Loaded*. More precisely, the effect on well-being of believing one's relative intelligence is low may be inherently stronger in magnitude than the effect of believing one's relative intelligence is high.

Several experiments on the formation of motivated beliefs also report asymmetries in the way individuals process good and bad news. The literature on selective recall shows that individuals remember with a higher accuracy positive feedback than negative feedback. The literature on asymmetric updating shows that subjects' inferences conform more closely to Bayes' rule in response to good than to bad signals. However, these findings are established in situations which clearly differ from the one we consider. First, in these works, subjects usually can directly assess whether a noisy signal is good or bad news. In our work, Receivers cannot assess whether a message is good or bad from its face value; it is the strategic setting – *High* or *Low* treatment – and skeptical inferences in this setting, that allow them to assess how good or bad a message is for them. Second, our subjects update their prior beliefs using a message which is strategically transmitted to them by another player. It follows that, in all treatments alike, Receivers need to reason about the Sender's strategy before using it to update their beliefs.

## 5.2 More or less self-serving or self-threatening skepticism

We now dig further into our findings by evaluating the extent to which a skeptical guess is self-threatening or self-serving in *Loaded* treatments. Table 5 presents the skeptical guess for every vague message in the *High* and *Low* treatments.

Table 5: Skeptical guess for every possible vague message in the *High* and *Low* treatments

| Message | $\{1,2\}$ | $\{2,3\}$ | $\{3,4\}$ | $\{4,5\}$ | $\{1,2,3\}$ | $\{2,3,4\}$ | $\{3,4,5\}$ | $\{1,2,3,4\}$ | $\{2,3,4,5\}$ | $\{1,2,3,4,5\}$ |
|---|---|---|---|---|---|---|---|---|---|---|
| *High* | 1 | 2 | **3** | **4** | 1 | 2 | **3** | 1 | 2 | 1 |
| *Low* | 2 | 3 | **4** | **5** | 3 | **4** | **5** | **4** | **5** | **5** |

*Note:* The Table reports the skeptical guess for every possible vague message in the *High* and *Low* treatments. For instance, in the *High* treatments, the skeptical guess corresponding to message $m = \{1, 2\}$ is $a = 1$. It is $a = 2$ in the *Low* treatments. Strongly self-serving and strongly self-threatening skeptical guesses appear in bold in the *High* and *Low* lines respectively.

In the *High_ Loaded* treatment, we say that skeptical guesses that correspond to ranks 3 or 4 are *strongly self-serving*, while skeptical guesses that correspond to 1 or 2 are *mildly self-serving*. This may seem a bit counter-intuitive but the extent to which a skeptical guess is self-serving is evaluated in comparison to other consistent guesses. In fact, when messages are $\{3, 4\}$, $\{4, 5\}$ or $\{3, 4, 5\}$, making a consistent guess which is not skeptical means guessing a particularly low rank (4 or 5). In that sense, the skeptical guess is strongly self-serving after these messages. For other messages, making a guess that is not skeptical means guessing a relatively high rank such as 2 or

3. In that sense, for messages such as $\{1,2\}$ or $\{1,2,3\}$, the skeptical guess is mildly self-serving. Similarly, in the *Low_ Loaded* treatment, we say that skeptical guesses that correspond to ranks 2 or 3 are *mildly self-threatening*, while skeptical guesses that correspond to 4 and 5 are *strongly self-threatening*. Strongly self-serving and strongly self-threatening skeptical guesses appear in bold in Table 5.

Figure D.2 in Appendix D gives the average skepticism and the frequency of skeptical guesses, for mildly/strongly self-serving/self-threatening skeptical guesses. In *High_ Loaded*, Receivers' make the skeptical guess significantly more frequently when it is strongly self-serving than when it is mildly self-serving ($p = 0.014$). In *Low_ Loaded*, the average level of skepticism and the frequency of a skeptical guess are significantly lower when the skeptical guess is strongly self-threatening than when it is mildly self-threatening ($p = 0.031$ and $p < 0.001$). These findings demonstrate that Receivers' exercise of skepticism depends finely on the desirability of the conclusion that skepticism leads to.

**Result 3.** (a) In the *High_ Loaded* treatment, Receivers make skeptical guesses significantly more often when skepticism is strongly self-serving than when it is mildly so. (b) In the *Low_ Loaded* treatment, the average level of skepticism and the frequency of skeptical guesses is lower when skepticism is strongly self-threatening than when it is mildly so.

## 5.3 The role of prior beliefs

We designed the *Neutral* and *Loaded* treatments with the objective to make them as comparable as possible except for the meaning of types to Receivers. One difference, directly related to the meaning of types, remains between these treatments: the Receivers' prior beliefs about their types, which we do not control directly. In *Loaded* treatments, Receivers form priors about their IQ-ranks after having experienced the IQ test in Part 1 of the experiment. They may therefore have some sense of how they performed in the test, which potentially affects how they read messages. In particular, if individuals are confident about their relative performance and tend to stick to this prior view, they may mechanically appear not very skeptical in the *Low_ Loaded* treatment and rather skeptical in the *High_ Loaded* treatment.

We investigate the role of prior beliefs by running two analyses. First, we test whether our main results are robust to the exclusion of the most confident subjects. Second, we run a complementary *Neutral* treatment in which we exogenously push Receivers' optimism about their neutral ranks to mimic Receivers' confidence about their relative performance in *Loaded* treatments.

**Exclusion of the most confident subjects.** In all treatments, we elicited Receivers' beliefs about their relative performance right after the IQ test. Precisely, we inserted each subject into a group of 99 subjects who had done the same IQ test previously. We then divided the group of 100 individuals into five quintiles – from the 20% of individuals with the best performance to the 20% of individuals with the worst performance – and asked subjects to give an estimate of the likelihood that they belong to each quintile. This beliefs elicitation procedure was incentivized using a quadratic scoring rule.[33] Even if these beliefs about relative performance do not correspond exactly to prior beliefs about the IQ-ranks generated in the Sender-Receiver games, they give an indication of how well the Receiver thinks he performed compared to other individuals. On average, Receivers consider that they are more likely to belong to the second and third quintiles than to the fourth and fifth quintiles (see Figure D.3 in Appendix D). This shows that individuals are rather confident regarding their performance in the IQ test, consistently with the large literature on overconfidence (e.g. Moore and Healy, 2008; Merkle and Weber, 2011).

To investigate whether the lower skepticism in *Low_Loaded* could be driven by the average tendency to confidence, we check that our results are robust to the exclusion of the most confident subjects.[34] There are several ways to identify these subjects. A first, simple way to evaluate confidence is to consider the quintile that each Receiver estimated most likely. We consider that the most confident Receivers are either the ones who estimate that they are most likely to belong to the first quintile, or the ones who estimate that they are most likely to belong to one of the first two quintiles. Excluding these subjects, we see that both the average level of skepticism and the frequency of a skeptical guess stay lower in *Low_Loaded* than in *Low_Neutral* ($p = 0.035$ and $p < 0.001$, resp.). A second way to evaluate Receivers' confidence is to examine whether their beliefs about performance, i.e., the distribution of the estimated likelihood to belong to each quintile, is skewed to the left (they estimate the first and second quintiles more likely than the fourth and fifth). We consider that the most confident subjects are the ones whose distribution of estimates differs significantly from the uniform distribution at the 1% level (Chi2 test) and is skewed to the left. Excluding these subjects, we see that both the average level of skepticism and the frequency of skeptical guesses stays lower in *Low_Loaded* than in *Low_Neutral* ($p = 0.039$ and $p = 0.012$, resp.).[35]

[33] Precisely, we explain subjects that their payoff is highest if they estimate most accurately their chance to belong to each quintile, and they can see the exact method used to compute their payoff. At the end of the experiment, one quintile is selected at random. The subject's payoff is based on the following formula: Payoff$= 2 - 2 * [(I - p/100)]^2$, where $I$ is an indicator variable that takes value 1 if the quintile the subject actually belongs to is equal to the selected quintile and 0 otherwise, and $p$ is the subject's estimate in percent.

[34] We apply our exclusion criterion to all treatments alike. Even if confidence does not play a role in *Neutral* treatments (as Receivers guess neutral ranks), it will ensure that we compare Receivers with similar IQ levels, types and messages across treatments.

[35] Figures D.4 and D.5 in Appendix D replicate Figure 2, without confident subjects.

**Observation 4.** Skepticism is significantly lower when it is self-threatening than when it is not, even when excluding confident Receivers who may lack skepticism simply because they stick to their optimistic prior beliefs.

**Complementary *Neutral* treatment with modified priors.** As reported above, Receivers are confident about their relative performance, which may play a role in *Loaded* settings. To mimic this confidence in *Neutral* settings, we run some complementary experimental sessions in which we do the following twist: the integer number between 0 and 15 which is associated to every Receiver (at the beginning of Part 2, that is, before neutral ranks are generated) is not uniformly distributed anymore but has a 20% chance to be low (between 0 and 7) and a 80% chance to be high (between 8 and 15). To determine neutral ranks, the modified program now compares the Receiver's integer to four other integers which are still uniformly drawn from $\{0, 1, \ldots, 15\}$. As in the main treatments, Receivers are informed about how types are generated, so the twist should make them more optimistic about their neutral ranks. We implement this change in the *Low_ Neutral* treatment and call the new treatment *Low_ Neutral_ ModifiedPriors*.[36] If the lack of skepticism in *Low_ Loaded* is driven by the fact that subjects stick to their optimistic priors, we should also see this lack of skepticism in *Low_ Neutral_ ModifiedPriors*.

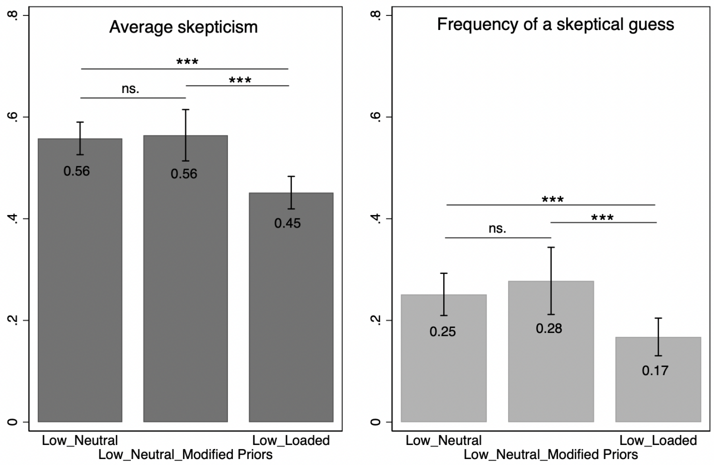
Figure 3 (and Table 3.2 in Online Appendix 3) compares the average level of skepticism and the frequency of skeptical guesses in the *Low_ Neutral*, *Low_ Neutral_ ModifiedPriors* and *Low_ Loaded* treatments. Both measures of skepticism stay lower in *Low_ Loaded* than in *Low_ Neutral_ Modified Priors* ($p = 0.005$ and $p < 0.001$), and skepticism is not significantly different between *Low_ Neutral* and *Low_ Neutral_ Modified Priors* ($p = 0.894$ and $p = 0.852$). We take this as evidence that confident priors are not the driver of the lower skepticism we observe in *Low_ Loaded* relative to *Low_ Neutral*.

**Observation 5.** Skepticism in *Low_ Neutral* is not affected by an exogenous variation that makes Receivers' prior beliefs about their neutral ranks more optimistic.

# 6 Receivers' depth of reasoning

In this section, we take a different look at Receivers' behavior by considering the steps of reasoning that could lead to the guesses they made.

[36]We collected data for 340 games. Details of the implementation are given in Online Appendix 3. We did not implement the change in *High_ Neutral* because, if prior beliefs impact the level of skepticism, confidence should push skepticism in *High_ Loaded* compared to *High_ Neutral*, and this is not what we observe.




*Note*: The left part of the Figure displays the average level of skepticism. The right part of the Figure displays the average frequency of a skeptical guess. Black segments are 95% confidence intervals. P-values are from random-effects linear regressions on panel data with the Receivers' id as the group variable and the rounds as the time variable. Standard errors are clustered at the session level using bootstrapping. *p $< 0.10$, **p $< 0.05$, ***p $< 0.01$.

Figure 3: Receivers' average skepticism with modified priors

## 6.1 Steps of reasoning

Disclosure games can be solved in a relatively intuitive way by using iterative elimination of weakly dominated strategy, an approach taken in Schipper and Li (2020).[37] We explain the elimination procedure by considering, as an example, message $\{3, 4, 5\}$ seen in a *High* treatment. The procedure alternates elimination on the Sender and on the Receiver sides and, without loss of generality, we start on the Sender side. At step zero of the procedure, the Sender sends messages that are constrained by the truth (the true type must belong to the set of types disclosed). At the first step, the Receiver understands that the Sender's messages are constrained by the truth, so consistent guesses are rational. At the second step, the Sender draws conclusions from the Receiver's consistency. In particular, the Sender of type 5 understands that the message $\{5\}$ induces guess 5 while the message $\{3, 4, 5\}$ can induce a guess in $[3, 5]$. Message $\{5\}$ then weakly dominates $\{3, 4, 5\}$ for a Sender of type 5. At the third step, the Receiver updates his possible interpretation of the message $\{3, 4, 5\}$ and eliminates type 5 as possible, so he makes a guess in $[3, 4]$. At the fourth step, the Sender of type 4 is then weakly better-off fully disclosing than sending $\{3, 4, 5\}$. At the fifth step, the Receiver eliminates 4 as a possible type and makes the skeptical guess 3. For every message, this procedure terminates in a finite number of steps.

In this procedure, odd steps correspond to belief updating and elimination of strategies on the

[37]The authors characterize iterated elimination of weakly dominated strategies by prudent rationalizability and its extensive-form analogue. For a formal description of the procedure, we refer the reader to Section 2 in Schipper and Li (2020). For an alternative procedure which consists of iterated elimination of obviously dominated strategies, see Hagenbach and Perez-Richet (2018).

Receiver's side. For tractability, we relabel the odd steps 1, 3, 5, etc. as the Receiver's steps of reasoning 1, 2, 3, etc. For every message $m$, the Receiver's maximal number of steps of reasoning is given by $|m|$ and the only Receiver's guess that survives theses $|m|$ steps is the skeptical guess. Table 6 below gives, for every message seen by the Receiver, the guesses that correspond to each of his reasoning steps in *High* and in *Low*.[38] The steps of reasoning for the Sender are given in Table E.1 in Appendix E.

Table 6: Reasoning steps for the Receiver

| Message | *High* treatments step 1 | step 2 | step 3 | step 4 | step 5 |
|---|---|---|---|---|---|
| $\{1\}$ | 1 | - | - | - | - |
| $\{2\}$ | 2 | - | - | - | - |
| $\{3\}$ | 3 | - | - | - | - |
| $\{4\}$ | 4 | - | - | - | - |
| $\{5\}$ | 5 | - | - | - | - |
| $\{1, 2\}$ | [1,2] | 1 | - | - | - |
| $\{2, 3\}$ | [2,3] | 2 | - | - | - |
| $\{3, 4\}$ | [3,4] | 3 | - | - | - |
| $\{4, 5\}$ | [4,5] | 4 | - | - | - |
| $\{1, 2, 3\}$ | [1,3] | [1,2] | 1 | - | - |
| $\{2, 3, 4\}$ | [2,4] | [2,3] | 2 | - | - |
| $\{3, 4, 5\}$ | [3,5] | [3,4] | 3 | - | - |
| $\{1, 2, 3, 4\}$ | [1,4] | [1,3] | [1,2] | 1 | - |
| $\{2, 3, 4, 5\}$ | [2,5] | [2,4] | [2,3] | 2 | - |
| $\{1, 2, 3, 4, 5\}$ | [1,5] | [1,4] | [1,3] | [1,2] | 1 |

| Message | *Low* treatments step 1 | step 2 | step 3 | step 4 | step 5 |
|---|---|---|---|---|---|
| $\{1\}$ | 1 | - | - | - | - |
| $\{2\}$ | 2 | - | - | - | - |
| $\{3\}$ | 3 | - | - | - | - |
| $\{4\}$ | 4 | - | - | - | - |
| $\{5\}$ | 5 | - | - | - | - |
| $\{1, 2\}$ | [1,2] | 2 | - | - | - |
| $\{2, 3\}$ | [2,3] | 3 | - | - | - |
| $\{3, 4\}$ | [3,4] | 4 | - | - | - |
| $\{4, 5\}$ | [4,5] | 5 | - | - | - |
| $\{1, 2, 3\}$ | [1,3] | [2,3] | 3 | - | - |
| $\{2, 3, 4\}$ | [2,4] | [3,4] | 4 | - | - |
| $\{3, 4, 5\}$ | [3,5] | [4,5] | 5 | - | - |
| $\{1, 2, 3, 4\}$ | [1,4] | [2,4] | [3,4] | 4 | - |
| $\{2, 3, 4, 5\}$ | [2,5] | [3,5] | [4,5] | 5 | - |
| $\{1, 2, 3, 4, 5\}$ | [1,5] | [2,5] | [3,5] | [4,5] | 5 |

*Note:* The Table for the *High* treatments is on the left and for the *Low* treatments on the right. In each cell of a table, we report the guesses that correspond, for a given message (in row), to a given step of reasoning (in column). For example, in the *High* treatments, a guess equal to 4 conditional on message $\{4, 5\}$ corresponds to the step 2 of reasoning for the Receiver.

Table 6 also shows that, for the Receiver and $k \geq 2$, every $k$-th step of reasoning in the *High* treatments consists in updating beliefs by eliminating the lowest rank still believed with positive probability at step $k-1$ (provided that there are two such ranks at least). On the contrary, in the *Low* treatments, every $k$-th step of reasoning in *Low* consists in eliminating the highest rank still believed with positive probability at step $k-1$ (provided that there are two such ranks at least). In short, in our games, as the Receiver reasons stepwise, the way he updates his beliefs has a very clear structure.

## 6.2 Motivated depth of reasoning

Alaoui and Penta (2015) propose a framework in which an individual's depth of reasoning is determined by considering that the individual compares, at each step of reasoning, the benefit from doing this additional step to the cognitive cost of doing it. We follow this approach and detail the assumptions we make about reasoning costs and benefits in our games.

In Alaoui and Penta (2015), reasoning costs are exogenous in that they capture how cognitively complex it is to reason at a given step for a given individual. We consider that, for a Receiver facing

[38] We assume that reasoning steps are the same in the two *High* treatments, as well as in the two *Low* treatments. In Subsection 6.2, we develop the idea that the *Loaded vs.Neutral* variation however affects the number of these steps that Receivers effectively do.

a message $m$, it is cognitively equivalent to reason in *High_ Neutral* and in *High_ Loaded*: possible types, available guesses, and steps of reasoning are the same in these games after any given $m$. Similarly, for a Receiver facing a message $m$, it is cognitively equivalent to reason in *Low_ Neutral* and in *Low_ Loaded*.[39] We denote $c_m^g(k)$ the Receiver's cognitive cost of doing the $k$-th step of reasoning in game $g$ when facing message $m$, and make the following assumption:

**Assumption 1** For every message $m$ and step $k \in [1, |m|]$, $c_m^{High\_Neutral}(k) = c_m^{High\_Loaded}(k)$ and $c_m^{Low\_Neutral}(k) = c_m^{Low\_Loaded}(k)$.

While reasoning costs are exogenous, Alaoui and Penta (2015) connect the value of reasoning to the payoffs of the game. In our games, doing a step $k$ of reasoning when facing message $m$ has the same effect on material payoffs in *High_ Neutral* and in *High_ Loaded* since material payoffs are the same in both games. As explained above, in these games, doing the $k$-th reasoning step consists in updating beliefs by eliminating the lowest rank still believed. We assume that, in *High_ Loaded*, this update has a direct, separable, positive effect on the Receivers' utility that adds to the effect on the material utility. Intuitively, we say that it is psychologically pleasant for the Receiver to eliminate the most unpleasant disclosed rank. We make a similar argument for *Low_ Neutral* and *Low_ Loaded*: the effect of doing the $k$-th step of reasoning has the same effect on material utilities in both treatments; in these games, doing the $k$-th step of reasoning consists in eliminating the highest rank still believed. We assume that, in *Low_ Loaded*, this has a direct, separable, negative effect on the Receiver's utility. Intuitively, we say that it is psychologically unpleasant to eliminate the most pleasant rank disclosed. We denote $v_m^g(k)$ the material and psychological value of doing the $k$-th step of reasoning in game $g$ when facing message $m$, and make the following assumption:

**Assumption 2** For every message $m$ and step $k \in [1, |m|]$, $v_m^{High\_Loaded}(k) > v_m^{High\_Neutral}(k)$ and $v_m^{Low\_Loaded}(k) < v_m^{Low\_Neutral}(k)$.

Having defined the reasoning costs and benefits in each game, we can identify the step at which it becomes too costly to reason relative to the benefit, i.e. the reasoning depth of a Receiver facing $m$ in game $g$ (following Definition 2 in Alaoui and Penta, 2015).

**Definition 6** Given $(c_m^g, v_m^g)$, the depth of reasoning of a Receiver facing message $m$ in game $g$ is given by $\mathcal{K}(c_m^g, v_m^g) = \min\{k \in [1, |m|] : c_m^g(k) \leq v_m^g(k) \text{ and } c_m^g(k+1) > v_m^g(k+1)\}$.

We can directly derive the following predictions about a Receiver's depth of reasoning in our treatments. It captures the idea, key to the theory of endogenous depth of reasoning, that *Loaded* and *Neutral* treatments give Receivers different incentives to reason:

[39] Cognitive equivalence between problems is defined similarly in Alaoui and Penta (2022), which also provides an axiomatic foundation for the cost-benefit approach.

**Proposition 3** *The reasoning of a Receiver facing m is deeper in High_ Loaded than in High_ Neutral. The reasoning of a Receiver facing m is less deep in Low_ Loaded than in Low_ Neutral.*

## 6.3 Depth of reasoning in the data

We make a few remarks before bringing Proposition 3 to the data. First, we have not run our treatments within subjects and compare depth of reasoning in *Loaded* and *Neutral* between subjects. This comparison is valid to the extent that subject populations are similar in the four treatments, which is the case.[40] Second, the depth of reasoning that Receivers can reach is constrained by the size of the messages they see. As argued in Subsection 4.2, Receivers receive similar messages in *High_ Loaded* and in *High_ Neutral*, as well as in *Low_ Loaded* and *Low_ Neutral*. In addition, even if Receivers do not all see messages of every possible size, 84% of them see messages of size 4 or 5.

In Table 7, we pool the messages seen by each Receiver and consider the following: a Receiver makes $k$ steps of reasoning if at least one of his 10 guesses required $k$ steps of reasoning. In every column of Table 7, we give the fraction of Receivers who make $k$ steps.

Table 7: Fraction of Receivers (in %) reaching each reasoning step, by treatment

| | 1 step | 2 steps | 3 steps | 4 steps | 5 steps |
|---|---|---|---|---|---|
| *High_ Neutral* | 100 | 92.59 | 72.22 | 44.44 | 22.22 |
| *High_ Loaded* | 100 | 93.75 | 85.42 | 50.00 | 14.58 |
| *p-value* | - | *0.817* | *0.106* | *0.575* | *0.323* |
| *Low_ Neutral* | 100 | 96.00 | 86.00 | 54.00 | 24.00 |
| *Low_ Loaded* | 100 | 93.75 | 68.75 | 18.75 | 4.17 |
| *p-value* | - | *0.613* | *0.041* | *<0.001* | *0.005* |
| *p-value (Neutral)* | - | *0.456* | *0.086* | *0.330* | *0.830* |

*Note:* The Table reports the fraction of Receivers who make $k$ steps of reasoning. P-values come from proportion tests. One observation by individual.

Table 7 shows that all Receivers make the first reasoning step, and more than 90% of them perform two steps. Receivers' make similar numbers of reasoning steps in the two *High* treatments. If we look at *Low* treatments however, a significantly larger fraction of Receivers perform three, four or five steps in *Low_ Neutral* than in *Low_ Loaded*. This observation suggests that Receivers make fewer steps of reasoning towards the skeptical guess when this guess is self-threatening than when it is not. In Figure E.1 of Appendix E, we report the Receivers' depth of reasoning for each message size and observe the same tendency.

[40]Using Kolmogorov-Smirnov tests, we checked that IQ and every demographic variable that we include in our analysis - age, gender, educational attainment, socio-professional category, social class and experience in participating in experiments - is distributed in the same way in the two *High* treatments and in the two *Low*.

**Result 4.** (a) Receivers make a similar number of reasoning steps towards the skeptical guess when skepticism is self-serving and when it is not. (b) Receivers make fewer steps of reasoning towards the skeptical guess when skepticism is self-threatening than when it is not. (c) Receivers make a similar number of reasoning steps in *High_ Neutral* and *Low_ Neutral*.

Note that Result 4 points to the same asymmetry as the one present in Result 2. Proposition 3 does not predict this asymmetry because it relies on the assumption that it is as psychologically pleasant to eliminate the lowest rank believed as it is unpleasant to eliminate the highest rank believed.

## 6.4 Discussion on the behavioral approaches

The behavioral approaches taken in Section 2 and in Section 6 are different. In Section 2, the Receiver chooses his beliefs freely by trading-off the material benefit of holding skeptical beliefs, accurate in equilibrium, and the psychological impact of these beliefs. In Section 6, the Receiver evaluates reasoning at each step by incorporating the psychological impact of stepwise belief updating. Both approaches predict less skepticism in *Low_ Loaded* than in *Low_ Neutral*, which we clearly find. Proposition 3 (Section 6) predicts more skepticism in *High_ Loaded* than in *High_ Neutral* whereas Proposition 2 (Section 2) does not. The asymmetry observed in our data may then suggest a higher validity of the approach taken in Section 2. However, as explained, Proposition 2 relies on the strong assumption that the Receiver reaches the skeptical belief before manipulating it further. If it were not the case, the approach taken in Section 2 could also predict more skepticism in *High_ Loaded* than in *High_ Neutral* because of the self-serving nature of skeptical beliefs in the former treatment. In short, both approaches similarly predict motivated skepticism but differ in how it can emerge. Since our treatments were designed to study *why* subjects are sometimes skeptical and sometimes not, it is hard to use them to understand exactly *how* subjects adapt their level of skepticism to the context.

We now discuss whether the approaches make different predictions regarding the effect of the size of the messages on skepticism.[41] For simplicity, consider a Receiver facing a vague message $m$ in *Low* treatments. According to Section 2, a psychological Receiver never makes a guess equal to the skeptical guess. Said differently, there should be a difference in skepticism between *Low_ Loaded* and *Low_ Neutral* for vague messages of any size. In Section 6, if we assume that the costs to reason are smaller when facing messages of smaller size, it could be that there is no difference in skepticism between *Low_ Loaded* and *Low_ Neutral* when messages are small but a difference

[41] In a related exercise, Schipper and Li (2020) compare the fully-revealing equilibrium approach and the procedure of iterated elimination of strategies in disclosure games. The authors experimentally vary the number of possible types with the idea that skepticism requires more steps of reasoning when the message gets larger while the size of messages has no effect with the equilibrium approach.

when they get large. Comparing skepticism in *Low_ Loaded* and *Low_ Neutral*, we indeed observe significant differences for messages of size 5 and no differences for messages of size 2.[42] Even if the evidence is more mixed for messages of size 3 and 4, this suggests that motivated depth of reasoning can be a mechanism at play. We leave for further research the development of an experiment specifically designed to evaluate the relative importance of the two approaches proposed.

# 7 Conclusion

We designed an experiment to study how individuals read strategically-transmitted information when they have preferences over what they will learn. In our disclosure games, the reading of information precisely consists in making skeptical inferences when faced with vague statements. We vary whether Receivers make inferences about loaded or neutral types and, importantly, whether inferences about the loaded types lead to attractive or unattractive conclusions. When skeptical inferences are self-threatening, skepticism is low. This is true while Receivers are able to exercise skepticism in neutral settings or when skeptical inferences are self-serving.

We consider disclosure of ego-relevant types but our work speaks more generally to the disclosure literature. Since its early development, this literature aims at analyzing communication about product quality from sellers to buyers. As explained in Dranove and Jin (2010), quality disclosure is pervasive and impacts individuals in almost all their purchasing decisions. Our work pushes forward the idea that, if consumers care *per se* about what they will learn, it may affect how they read product information, and the relevance of the unraveling result. In this vein, Mathios (2000) empirically reports a lack of unraveling of nutritional information on the market of salad dressings, which could occur because consumers prefer not to see they consume fat dressings. Bederson et al. (2018) study disclosure of hygiene grades of restaurants, and implicitly assume there is a value in believing that one eats in a clean place. The next question, which we do not answer in this work, is whether information providers expect motivated skepticism from consumers, and whether they can exploit it, for instance to keep producing low quality.[43]

Partially connected to works on disclosure, a relatively large literature studies environmental

[42]The levels of skepticism in *Low_ Neutral* and *Low_ Loaded* are respectively: 0.61 vs 0.48 when message size is 2 ($p = 0.361$), 0.56 vs 0.48 when the message size is 3 ($p = 0.081$), 0.51 vs 0.41 when the message size is 4 ($p = 0.004$) and 0.57 vs 0.42 when the message size is 5 ($p < 0.001$). The frequencies of skeptical guesses in *Low_ Neutral* and *Low_ Loaded* are respectively: 0.51 vs 0.39 when the message size is 2 ($p = 0.436$), 0.26 vs 0.18 when the message size is 3 ($p = 0.144$), 0.12 vs 0.06 when the message size is 4 ($p = 0.268$) and 0.21 vs 0.03 when the message size is 5 ($p = 0.001$).

[43]In theory, the equilibrium effect of the Receivers' lack of skepticism has already been studied in Milgrom and Roberts (1986), Hagenbach and Koessler (2017) or Deversi et al. (2021). These papers however adopt a reduced-form approach in that they do not root the lack of skepticism in a particular behavioral theory. An exception is Manili (2023) who develops a model in which firms exploit consumers who are wishful thinkers.

or ethical labeling, widely and increasingly used. On the one hand, this literature points at the efficiency of mandatory labels in affecting consumers behaviors. For instance, Teisl et al. (2002) report that dolphin-safe labels significantly affect the purchase of tuna towards more animal-friendly products; Hainmueller et al. (2015) provide similar conclusions about the effect of fair-trade labels on the sales of coffee; Camilleri et al. (2019) show that consumers underestimate the greenhouse gas emissions associated with food, but that labels help them shift choices towards lower-emission options. These findings, as our work, suggest that it may be particularly important and effective to mandate information disclosure in markets in which some truths about products are disturbing. On the other hand, the literature on green labeling discusses what information labels should exactly contain (see, for instance, Cohen and Vandenbergh, 2012; De Haan and Knutsen, 2023), and shows that the uncertainty regarding the standards needed to obtain some labels can impact their reading (Harbaugh et al., 2011). Our work suggests that individuals can use the room left by complex or imprecise labels to reach comfortable beliefs. If the reading of labels is motivated, their impact on helping individuals achieve sustainable consumption may be limited.

From a more abstract point of view, we contribute to the literature on the formation of motivated beliefs by bringing it to a setting of strategic communication. We show that Receivers read information in a motivated way, even if they are well aware of the Senders' objective and probably pushed by the experimental environment to think about what Senders want them to guess. The open question is how exactly Receivers reach the favored conclusions, sometimes by going against a game-theoretic logic they otherwise understand. In Section 2, Receivers freely manipulate their beliefs after having made equilibrium inferences. In Section 6, Receivers think in steps and their depth of reasoning is endogenous. Another interesting possibility, which we do not explore, is that Receivers perceive Senders' strategies in a way that helps them form pleasant beliefs. When skepticism leads to bad news, Receivers could for instance convince themselves that Senders did a mistake, disclosing more low ranks than what is optimal for them. We leave for further research the study of the extent to which strategic uncertainty leaves wiggle room to form motivated beliefs in games.

# References


Alaoui, L. and Penta, A. (2015). Endogenous Depth of Reasoning. *The Review of Economic Studies*, 83(4):1297–1333.

Alaoui, L. and Penta, A. (2022). Cost-benefit analysis in reasoning. *Journal of Political Economy*, 130(4):881–925.

Ali, S. N., Lewis, G., and Vasserman, S. (2021). Voluntary disclosure and personalized pricing. *Review of Economic Studies*, forthcoming.

Amelio, A. and Zimmermann, F. (2023). Motivated memory in economics—a review. *Games*, 14(1):15.

Bederson, B. B., Jin, G. Z., Leslie, P., Quinn, A. J., and Zou, B. (2018). Incomplete disclosure: Evidence of signaling and countersignaling. *American Economic Journal: Microeconomics*, 10:41–66.

Bénabou, R. (2015). The economics of motivated beliefs. *Revue d'économie politique*, 125(5):665–685.

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

Bénabou, R. and Tirole, J. (2011). Identity, morals and taboos: Beliefs as assets. *Quarterly Journal of Economics*, 126(2):805–855.

Benndorf, V., Kübler, D., and Normann, H.-T. (2015). Privacy concerns, voluntary disclosure of information, and unraveling: an experiment. *European Economic Review*, 75:43–59.

Brown, A. L., Camerer, C. F., and Lovallo, D. (2012). To review or not to review? limited strategic thinking at the movie box office. *American Economic Journal: Microeconomics*, 4(2):1–26.

Brunnermeier, M. and Parker, J. (2005). Optimal expectations. *American Economic Review*, 95(4):1092–1118.

Burro, G. and Castagnetti, A. (2022). Will i tell you that you are smart (dumb)? deceiving others about their iq or about a random draw. *Journal of Behavioral and Experimental Economics*, 100:101920.

Burro, G. and Castagnetti, A. (2023). Ego is no fool: Absence of motivated belief formation in strategic interactions. *Working paper.*

Camerer, C. F., Ho, T.-H., and Chong, J. K. (2004). A cognitive hierarchy model of games. *Quarterly Journal of Economics*, 119(3):861–98.

Camilleri, A. R., Larrick, R. P., Hossain, S., and Patino-Echeverri, D. (2019). Consumers underestimate the emissions associated with food but are aided by labels. *Nature Climate Change*, 9(1):53–58.

Caplin, A. and Eliaz, K. (2003). Aids policy and psychology: A mechanism-design approach. *The RAND Journal of Economics*, 34:631–646.

Caplin, A. and Leahy, J. (2019). Wishful thinking. *Working paper.*

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

Chen, S. and Heese, C. (2023). Fishing for good news: Motivated information acquisition. *Working paper.*

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

Cohen, M. A. and Vandenbergh, M. P. (2012). The potential role of carbon labeling in a green economy. *Energy Economics*, 34:S53–S63.

Costa-Gomes, M. A. and Crawford, V. P. (2006). Cognition and behavior in two-person guessing games: An experimental study. *American Economic Review*, 96(5):1737–68.

Crawford, V. P., Costa-Gomes, M. A., and Iriberri, N. (2013). Structural models of nonequilibrium strategic thinking: Theory, evidence, and applications. *Journal of Economic Literature*, 51(1):5–62.

Crawford, V. P. and Iriberri, N. (2007). Level-k auctions: Can a nonequilibrium model of strategic thinking explain the winner's curse and overbidding in private-value auctions? *Econometrica*, 75(6):1721–1770.

De Haan, T. and Knutsen, M. V. (2023). Product ratings and externalities. *Working paper.*

Deversi, M., Ispano, A., and Schwardmann, P. (2021). Spin doctors: An experiment on vague disclosure. *European Economic Review*, 193.

Dickhaut, J. W., Ledyard, M., Mukherji, A., and Sapra, H. (2003). Information management and valuation: An experimental investigation. *Games and Economic Behavior*, 44(1):25–53.

Dranove, D. and Jin, G. Z. (2010). Quality disclosure and certification: Theory and practice. *Journal of Economic Literature*, 48:935–963.

Drobner, C. (2022). Motivated beliefs and anticipation of uncertainty resolution. *American Economic Review: Insights*, 4(1):89–105.

Drobner, C. and Goerg, S. J. (2021). Motivated belief updating and rationalization of information. *Working paper.*

Duch, M. L., Grossmann, M. R., and Lauer, T. (2020). z-tree unleashed: A novel client-integrating architecture for conducting z-tree experiments over the internet. *Journal of Behavioral and Experimental Finance*, 28:100400.

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.

Epley, N. and Gilovich, T. (2016). The mechanics of motivated reasoning. *Journal of Economic perspectives*, 30(3):133–40.

Exley, C. L. and Kessler, J. B. (2021). Information avoidance and image concerns. *Economic Journal*, Forthcoming.

Fischbacher, U. (2007). z-tree: Zurich toolbox for ready-made economic experiments. *Experimental economics*, 10(2):171–178.

Giovannoni, F. and Seidmann, D. J. (2007). Secrecy, two-sided bias and the value of evidence'. *Games and Economic Behavior*, 59(2):296–315.

Golman, R., Hagmann, D., and Loewenstein, G. (2017). Information avoidance. *Journal of Economic Literature*, 55(1):96–135.

Grossman, S. J. (1981). The informational role of warranties and private disclosure about product quality. *Journal of Law and Economics*, 24:461–483.

Grossman, Z. (2014). Strategic ignorance and the robustness of social preferences. *Management Science*, 60(11):2659–2665.

Grossman, Z. and Van der Weele, J. J. (2017). Self-image and willful ignorance in social decisions. *Journal of the European Economic Association*, 15(1):173–217.

Hagenbach, J. and Koessler, F. (2017). Simple versus rich language in disclosure games. *Review of Economic Design*, 21(3):163–175.

Hagenbach, J. and Koessler, F. (2022). Selective memory of a psychological agent. *European Economic Review*, page 104012.

Hagenbach, J., Koessler, F., and Perez-Richet, E. (2014). Certifiable pre-play communication: Full disclosure. *Econometrica*, 82(3):1093–1131.

Hagenbach, J. and Perez-Richet, E. (2018). Communication with evidence in the lab. *Games and Economic Behavior*, 112:139–165.

Hainmueller, J., Hiscox, M. J., and Sequeira, S. (2015). Consumer demand for fair trade: Evidence from a multistore field experiment. *Review of Economics and Statistics*, 97(2):242–256.

Harbaugh, R., Maxwell, J. W., and Roussillon, B. (2011). Label confusion: The groucho effect of uncertain standards. *Management science*, 57(9):1512–1527.

Hestermann, N., Le Yaouanq, Y., and Treich, N. (2020). An economic model of the meat paradox. *European Economic Review*, 129:103569.

Jin, G. Z. (2005). Competition and disclosure incentives: an empirical study of hmos. *Rand journal of Economics*, pages 93–112.

Jin, G. Z., Luca, M., and Martin, D. (2021). Is no news (perceived as) bad news? an experimental investigation of information disclosure. *American Economic Journal: Microeconomics*, 12:141–173.

King, R. R. and Wallin, D. E. (1991). Market induced information disclosure: An experimental markets investigation. *Contemporary Accounting Research*, 8:170–197.

Kogan, S., Schneider, F., and Weber, R. (2021). Self-serving biases in beliefs about collective outcomes. *Working paper.*

Kőszegi, B. (2006). Emotional agency. *The Quarterly Journal of Economics*, 121(1):121–155.

Kunda, Z. (1990). The case for motivated reasoning. *Psychological bulletin*, 108(3):480.

Loewenstein, G., Sunstein, C. R., and Golman, R. (2014). Disclosure: Psychology changes everything. *Annual Review of Economics*, 6:391–419.

Luca, M. and Smith, J. (2015). Strategic disclosure: The case of business school rankings. *Journal of Economic Behavior & Organization*, 112:17–25.

Manili, J. (2023). Beyond pigou: Disclosure and the green transition of markets. *Working paper.*

Mathios, A. D. (2000). The impact of mandatory disclosure on product choices: An analysis of the salad dressing market. *Journal of Law and Economics*, 43(2):651–77.

Merkle, C. and Weber, M. (2011). True overconfidence: The inability of rational information processing to account for apparent overconfidence. *Organizational Behavior and Human Decision Processes*, 116(2):262–271.

Milgrom, P. (1981). Good news and bad news: Representation theorems and applications. *Bell Journal of Economics*, 12:380–391.

Milgrom, P. and Roberts, J. (1986). Relying on the information of interested parties. *Rand Journal of Economics*, 17(1):18–32.

Möbius, M. M., Niederle, M., Niehaus, P., and Rosenblat, T. S. (2022). Managing self-confidence: Theory and experimental evidence. *Management Science.*

Moore, D. A. and Healy, P. J. (2008). The trouble with overconfidence. *Psychological review*, 115(2):502.

Nagel, R. (1995). Unraveling in guessing games: An experminental study. *American Economic Review*, 85(5):1313–26.

Okuno-Fujiwara, A., Postlewaite, M., and Suzumura, K. (1990). Strategic information revelation. *Review of Economic Studies*, 57:25–47.

Oprea, R. and Yuksel, S. (2021). Social exchange of motivated beliefs. *Journal of the European Economic Association.*

Raven, J. C. (1936). Mental tests used in genetic studies: The performance of related individuals on tests mainly educative and mainly reproductive. *MSc Thesis, University of London.*

Schipper, B. C. and Li, Y. X. (2020). Strategic reasoning in persuasion games: an experiment. *Games and Economic Behavior*, 121:329–367.

Schwardmann, P. (2019). Motivated health risk denial and preventative health care investments. *Journal of Health Economics*, 65:78–92.

Schwardmann, P., Tripodi, E., and Van der Weele, J. J. (2022). Self-persuasion: Evidence from field experiments at two international debating competitions. *American Economic Review*, 112(4):1118–46.

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

Seidmann, D. J. and Winter, E. (1997). Strategic information transmission with verifiable messages. *Econometrica*, 65(1):163–169.

Serra-Garcia, M. and Szech, N. (2021). The (in)elasticity of moral ignorance. *Management Science*.

Sharot, T., Korn, C. W., and Dolan, R. J. (2011). How unrealistic optimism is maintained in the face of reality. *Nature neuroscience*, 14(11):1475–1479.

Solda, A., Ke, C., Page, L., and Von Hippel, W. (2020). Strategically delusional. *Experimental Economics*, 23(3):604–631.

Teisl, M. F., Roe, B., and Hicks, R. L. (2002). Can eco-labels tune a market? evidence from dolphin-safe labeling. *Journal of environmental Economics and Management*, 43(3):339–359.

Thaler, M. (2022). The fake news effect: Experimentally identifying motivated reasoning using trust in news. *American Economic Journal: Microeconomics*, Forthcoming.

Thompson, L. and Loewenstein, G. (1992). Egocentric interpretations of fairness and interpersonal conflict. *Organizational Behavior and Human Decision Processes*, 51(2):176–197.

Zimmermann, F. (2020). The dynamics of motivated beliefs. *American Economic Review*, 110(2):337–61.

# Appendices

## A The unraveling result

**Proof of Proposition 1.** First, consider by contradiction an equilibrium in which at least two types of the Sender, say $t$ and $t'$ with $t > t'$, send the same (necessarily vague) message $m$. Facing $m$, the Receiver's beliefs $\beta_m$ assign a positive probability to both types $t$ and $t'$. In *High* games, the higher type $t$ then has a strict interest in deviating and sending the precise message $m' = \{t\}$ because $\sigma_R(m') = t > \sigma_R(m)$. In *Low* games, the lower type $t'$ then has a strict interest in deviating and sending the precise message $m' = \{t'\}$ because $\sigma_R(m') = t' < \sigma_R(m)$. This demonstrates that the Sender uses a fully-revealing strategy in every equilibrium.

Second, assume that, when facing the vague message $m$, the Receiver's beliefs are not skeptical. In *High* games, it means that $\beta_m$ assign some positive probability to a type in $m$ which is not the lowest. Then, the lowest type in $m$, say $t$, has an interest in deviating from sending the message prescribed by any fully-revealing strategy and sending message $m$ (to get action $\sigma_R(m) > t$). In *Low* games, it means that $\beta_m$ assigns some positive probability to a type in $m$ which is not the highest. Then, the highest type in $m$, say $t$, has an interest in deviating from sending the message prescribed by any fully-revealing strategy and sending message $m$ (to get action $\sigma_R(m) < t$). This is in contradiction with the former point which establishes that the Sender's strategy is fully-revealing in every equilibrium.

## B Statistical tests

The p-values reported in the main text come from random-effects linear regressions on panel data with the Senders' or Receivers' id as the group variable and the rounds as the time variable. Standard errors are clustered at the session level using bootstrapping. While this specification has the advantage of being portable (we can use the same throughout the paper), it does not directly account for the sometimes limited nature of our dependent variables (for instance, *skepticism* is bounded between 0 and 1, *is_ skeptical* is a dummy, etc.).

Table B provides robustness checks of the tests reported in the main text by exploring alternative specifications. These include (i) directly accounting for the bounded nature of the dependent variable by using Probit or Tobit models when appropriate, (ii) using linear regressions without considering panel data structure, and (iii) clustering at the individual rather than at the session level.

Overall, our main results are fairly robust: Hypotheses 2 and 3 are rejected in none of the 24 specifications considered. Evidence are more mixed regarding Hypothesis 1: 11 specifications reject Hypothesis 1 while 13 fail to reject it. This is consistent with our main result 2(c) which concludes that the aggregate level of skepticism is lower in *Low_ Neutral* than in *High_ Neutral*, but only marginally.

Table B.1: P-values of statistical tests

| Model | Linear | Linear | Linear | Linear | Tobit | Tobit | Probit | Probit |
|---|---|---|---|---|---|---|---|---|
| Subject | RE | RE | RE | RE | RE | RE | RE | RE |
| Cluster | Session | Id | Session | Id | Session | Id | Session | Id |
| Panel | ✓ | ✓ | | | | | | |
| *Without controls* | | | | | | | | |
| $Skept_{High\_Neu} = Skept_{Low\_Neu}$ | 0.075 | 0.159 | 0.031 | 0.066 | 0.022 | 0.086 | | |
| $Skept_{High\_Loa} \geq Skept_{High\_Neu}$ | 0.335 | 0.278 | 0.334 | 0.227 | 0.285 | 0.203 | | |
| $Skept_{Low\_Loa} < Skept_{Low\_Neu}$ | 0.004 | 0.017 | 0.006 | 0.018 | 0.002 | 0.030 | | |
| $Skept_{Low\_Loa} < Skept_{High\_Loa}$ | <0.001 | <0.001 | <0.001 | <0.001 | 0.001 | <0.001 | | |
| $IsSkept_{High\_Neu} = IsSkept_{Low\_Neu}$ | 0.181 | 0.401 | 0.057 | 0.163 | | | 0.033 | 0.174 |
| $IsSkept_{High\_Loa} \geq IsSkept_{High\_Neu}$ | 0.271 | 0.299 | 0.370 | 0.347 | | | 0.347 | 0.341 |
| $IsSkept_{Low\_Loa} < IsSkept_{Low\_Neu}$ | 0.005 | 0.033 | 0.013 | 0.036 | | | 0.006 | 0.028 |
| $IsSkept_{Low\_Loa} < IsSkept_{High\_Loa}$ | <0.001 | <0.001 | 0.002 | <0.001 | | | <0.001 | <0.001 |
| *With controls* | | | | | | | | |
| $Skept_{High\_Neu} = Skept_{Low\_Neu}$ | 0.126 | 0.239 | 0.046 | 0.126 | 0.050 | 0.168 | | |
| $Skept_{High\_Loa} \geq Skept_{High\_Neu}$ | 0.736 | 0.729 | 0.918 | 0.909 | 0.808 | 0.793 | | |
| $Skept_{Low\_Loa} < Skept_{Low\_Neu}$ | 0.002 | 0.003 | 0.008 | 0.006 | 0.002 | 0.010 | | |
| $Skept_{Low\_Loa} < Skept_{High\_Loa}$ | <0.001 | <0.001 | 0.001 | <0.001 | <0.001 | <0.001 | | |
| $IsSkept_{High\_Neu} = IsSkept_{Low\_Neu}$ | 0.272 | 0.489 | 0.080 | 0.205 | | | 0.047 | 0.219 |
| $IsSkept_{High\_Loa} \geq IsSkept_{High\_Neu}$ | 0.587 | 0.642 | 0.982 | 0.984 | | | 0.957 | 0.961 |
| $IsSkept_{Low\_Loa} < IsSkept_{Low\_Neu}$ | 0.002 | 0.010 | 0.013 | 0.020 | | | 0.005 | 0.012 |
| $IsSkept_{Low\_Loa} < IsSkept_{High\_Loa}$ | <0.001 | 0.001 | 0.004 | 0.001 | | | <0.001 | <0.001 |

# C Alternative measure of skepticism

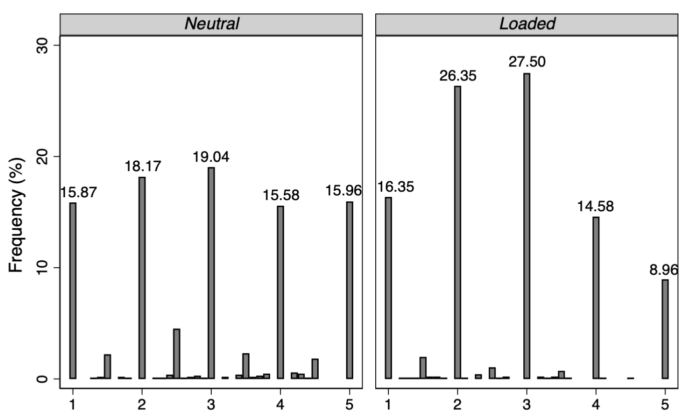
Table C.1 replicates Table 4 considering a dependant variable which equals 1 if the guess is skeptical and 0 if not. The coefficient of the treatment dummy in columns (1) to (3) is small and insignificant, which shows that the Receivers' likelihood to make the skeptical guess is not significantly different in *High_ Neutral* and *High_ Loaded*. We also see no significant difference in the likelihood to make a skeptical guess between the *Neutral* treatments. On the contrary, the estimated negative coefficient of the treatment dummy in columns (4) to (6) reveals that the Receivers' likelihood to make the skeptical guess is substantially and significantly lower in *Low_ Loaded* than in *Low_ Neutral*. All specifications control for the size of the message received by the Receiver. Whatever the size, the skeptical guess always corresponds to guessing one specific rank among the ones disclosed. When the number of disclosed ranks gets larger, it may become mechanically less likely or cognitively harder to make the skeptical guess. The coefficient of the message size is negative and significant indicating that the likelihood to make a skeptical guess indeed decreases as more ranks are disclosed.

Table C.1: Determinants of a skeptical guess

| *Dep. Var.* | = 1 if the guess is skeptical, 0 if not | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | *High* treatments | | | *Low* treatments | | | *Difference-in-difference* | | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| 1 if *Loaded* | 0.065 | 0.038 | 0.028 | -0.118*** | -0.135*** | -0.173*** | 0.065 | 0.042 | 0.032 |
| | (0.059) | (0.053) | (0.063) | (0.044) | (0.032) | (0.059) | (0.059) | (0.058) | (0.060) |
| 1 if *Low* | | | | | | | -0.043 | -0.045 | -0.033 |
| | | | | | | | (0.034) | (0.031) | (0.040) |
| 1 if *Low_ Loaded* | | | | | | | -0.191*** | -0.183*** | -0.187** |
| | | | | | | | (0.068) | (0.068) | (0.081) |
| Mess. size | -0.158*** | -0.159*** | -0.161*** | -0.089*** | -0.090*** | -0.090*** | -0.124*** | -0.125*** | -0.126*** |
| | (0.018) | (0.019) | (0.016) | (0.014) | (0.013) | (0.014) | (0.013) | (0.013) | (0.013) |
| IQ performance | | 0.022* | 0.023* | | 0.022*** | 0.019* | | 0.019*** | 0.018** |
| | | (0.0134) | (0.014) | | (0.007) | (0.011) | | (0.007) | (0.008) |
| Rounds dummies | | ✓ | ✓ | | ✓ | ✓ | | ✓ | ✓ |
| *Demo.* | | | ✓ | | | ✓ | | | ✓ |
| Cons. | 0.889*** | 0.578*** | 0.593*** | 0.604*** | 0.357*** | 0.0482 | 0.773*** | 0.523*** | 0.588*** |
| | (0.060) | (0.113) | (0.198) | (0.073) | (0.096) | (0.337) | (0.046) | (0.067) | (0.214) |
| *N* | 789 | 789 | 789 | 816 | 816 | 816 | 1605 | 1605 | 1605 |

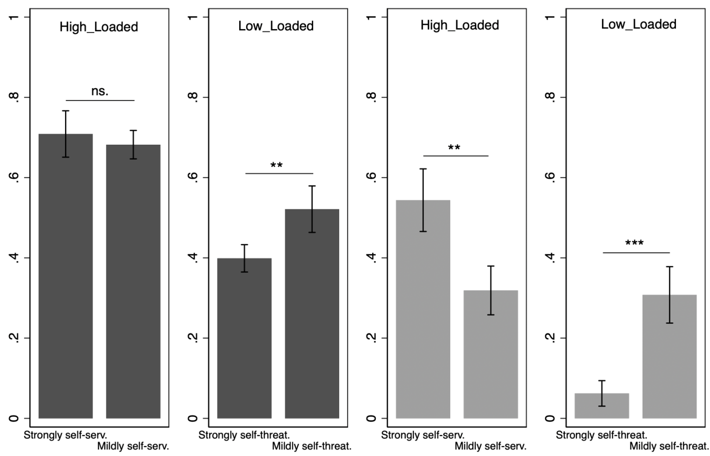
*Note:* The Table reports random-effects linear regressions on panel data with the Receivers' id as the group variable and the rounds as the time variable. Standard errors (in parentheses) are clustered at the session level using bootstrapping. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

# D Additional figures



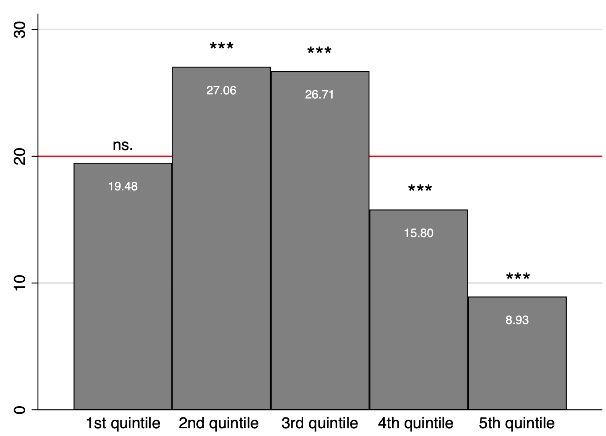

*Note:* The Figure displays the frequency of Receivers' guesses in the *Neutral* and *Loaded* treatments. For instance, 19.04% of the guesses in the *Neutral* treatments were $a = 3$. It is 27.50% in the *Loaded* treatments.

Figure D.1: Distribution of Receivers' Guesses



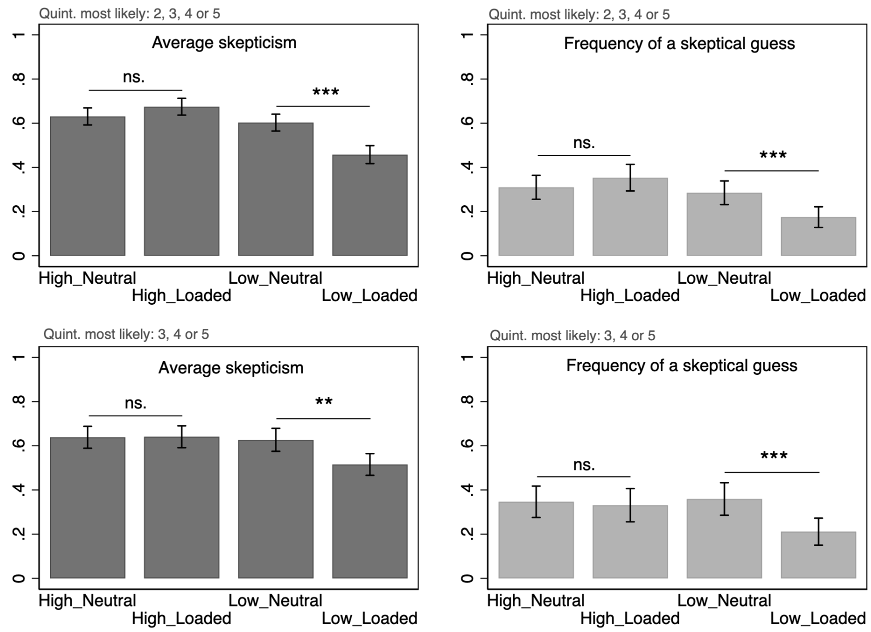

*Note:* The Figure displays the average level of skepticism (dark grey) and the frequency of skeptical guesses (light grey) in the *High_ Loaded* treatment (resp. *Low_ Loaded*), when the skeptical guess is mildly or strongly self-serving (resp. self-threatening). Black segments are 95% confidence intervals. P-values are from clustered t-tests with the Receivers' id as the group variable. $^{**}p < 0.05$, $^{***}p < 0.01$.

Figure D.2: Skepticism when it is mildly/strongly self-threatening/self-serving



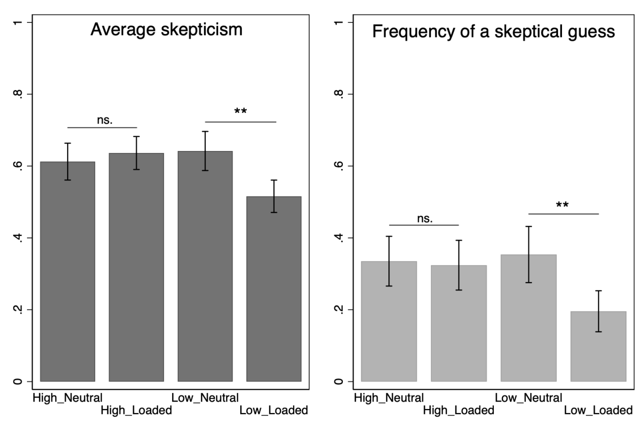

*Note:* The Figure displays the averages of the estimated likelihood to belong to each quintile. For instance, individuals estimate that they have on average 19.48% chances to belong to the first quintile. Significance levels are from two-sided t-tests that the estimated percentage is different from 20%. *** $p < 0.01$.

Figure D.3: Receivers' average likelihood to belong to each quintile




*Note*: The top-left part of the Figure displays the average level of skepticism, by treatment, excluding subjects who estimate most likely that they belong to the first quintile. The top-right part of the Figure displays the average frequency of a skeptical guess, by treatment, excluding subjects who estimate most likely that they belong to the first quintile. The bottom-left and bottom-right parts of the Figure exclude subjects who estimate most likely that they belong to the first or to the second quintile. Black segments are 95% confidence intervals. P-values are from random-effects linear regressions on panel data with the Receivers' id as the group variable and the rounds as the time variable. Standard errors are clustered at the session level using bootstrapping. *p $< 0.10$, **p $< 0.05$, ***p $< 0.01$.

Figure D.4: Receivers' average skepticism by treatment, without confident subjects (quintile estimated as most likely)

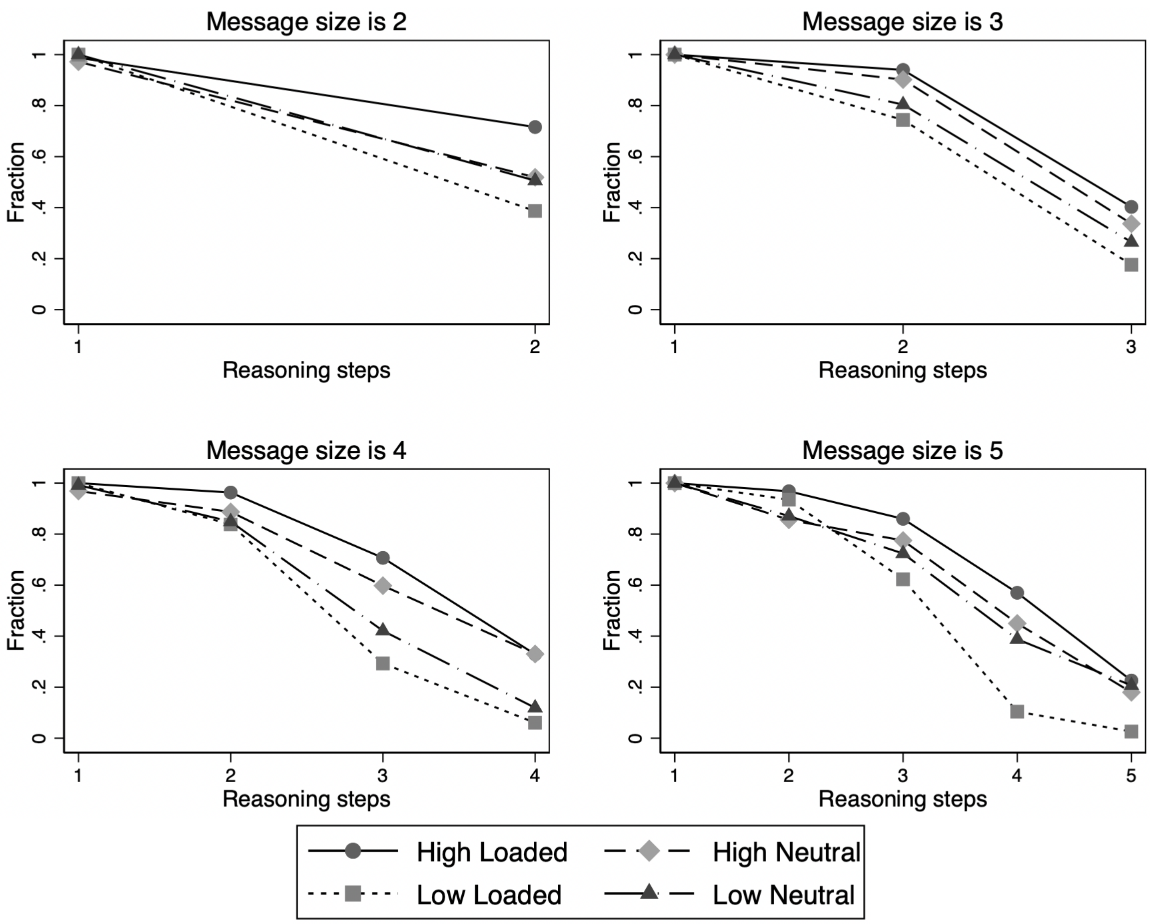



*Note*: The left part of the Figure displays the average level of skepticism, by treatment, excluding confident subjects. The right part of the Figure displays the average frequency of a skeptical guess, by treatment, excluding confident subjects. Black segments are 95% confidence intervals. P-values are from random-effects linear regressions on panel data with the Receivers' id as the group variable and the rounds as the time variable. Standard errors are clustered at the sessionlevel using bootstrapping. *p $< 0.10$, **p $< 0.05$, ***p $< 0.01$.

Figure D.5: Receivers' average skepticism by treatment, without confident subjects (distribution of priors)

# E Additional material regarding depth of reasoning

In the procedure described in Section 6.1, even steps correspond to elimination of strategies on the Sender's side. We relabel these even steps 0, 2, 4, etc. as the Sender's steps of reasoning 1, 2, 3, etc. Table E.1 on the next page gives, for every type, the messages that correspond to each step of reasoning of the Sender. Figure E.1 below displays, for every message size, the fraction of Receivers reaching each reasoning step, by treatment.




*Note*: The Figure displays, for every message size, the fraction of Receivers reaching each reasoning step, by treatment. For instance, when the message is of size 2, 72% of Receivers in *High_ Loaded* make guesses that correspond to step 2 of reasoning while it is only 39% in *Low_ Loaded*.

Figure E.1: Fraction of Receivers (in %) reaching each reasoning step, by message size and treatment

## Table E.1: Reasoning steps for the Sender

*High* treatments

**Step 1**

| Message | $t=1$ | $t=2$ | $t=3$ | $t=4$ | $t=5$ |
|---|---|---|---|---|---|
| {1} | ✓ | | | | |
| {2} | | ✓ | | | |
| {3} | | | ✓ | | |
| {4} | | | | ✓ | |
| {5} | | | | | ✓ |
| {1, 2} | ✓ | ✓ | | | |
| {2, 3} | | ✓ | ✓ | | |
| {3, 4} | | | ✓ | ✓ | |
| {4, 5} | | | | ✓ | ✓ |
| {1, 2, 3} | ✓ | ✓ | ✓ | | |
| {2, 3, 4} | | ✓ | ✓ | ✓ | |
| {3, 4, 5} | | | ✓ | ✓ | ✓ |
| {1, 2, 3, 4} | ✓ | ✓ | ✓ | ✓ | |
| {2, 3, 4, 5} | | ✓ | ✓ | ✓ | ✓ |
| {1, 2, 3, 4, 5} | ✓ | ✓ | ✓ | ✓ | ✓ |

**Step 2**

| Message | $t=1$ | $t=2$ | $t=3$ | $t=4$ | $t=5$ |
|---|---|---|---|---|---|
| {1} | | | | | |
| {2} | | | | | |
| {3} | | | | | |
| {4} | | | | | |
| {5} | | | | | ✓ |
| {1, 2} | ✓ | | | | |
| {2, 3} | | ✓ | | | |
| {3, 4} | | | ✓ | | |
| {4, 5} | | | | ✓ | |
| {1, 2, 3} | ✓ | ✓ | | | |
| {2, 3, 4} | | ✓ | ✓ | | |
| {3, 4, 5} | | | ✓ | ✓ | |
| {1, 2, 3, 4} | ✓ | ✓ | ✓ | | |
| {2, 3, 4, 5} | | ✓ | ✓ | ✓ | |
| {1, 2, 3, 4, 5} | ✓ | ✓ | ✓ | ✓ | |

**Step 3**

| Message | $t=1$ | $t=2$ | $t=3$ | $t=4$ | $t=5$ |
|---|---|---|---|---|---|
| {1} | | | | | |
| {2} | | | | | |
| {3} | | | | | |
| {4} | | | | | |
| {5} | | | | | ✓ |
| {1, 2} | | | | | |
| {2, 3} | | | | | |
| {3, 4} | | | | | |
| {4, 5} | | | | ✓ | |
| {1, 2, 3} | ✓ | | | | |
| {2, 3, 4} | | ✓ | | | |
| {3, 4, 5} | | | ✓ | | |
| {1, 2, 3, 4} | ✓ | ✓ | | | |
| {2, 3, 4, 5} | | ✓ | ✓ | | |
| {1, 2, 3, 4, 5} | ✓ | ✓ | ✓ | | |

**Step 4**

| Message | $t=1$ | $t=2$ | $t=3$ | $t=4$ | $t=5$ |
|---|---|---|---|---|---|
| {1} | | | | | |
| {2} | | | | | |
| {3} | | | | | |
| {4} | | | | | |
| {5} | | | | | ✓ |
| {1, 2} | | | | | |
| {2, 3} | | | | | |
| {3, 4} | | | | | |
| {4, 5} | | | | ✓ | |
| {1, 2, 3} | | | | | |
| {2, 3, 4} | | | | | |
| {3, 4, 5} | | | ✓ | | |
| {1, 2, 3, 4} | ✓ | | | | |
| {2, 3, 4, 5} | | ✓ | | | |
| {1, 2, 3, 4, 5} | ✓ | ✓ | | | |

**Step 5**

| Message | $t=1$ | $t=2$ | $t=3$ | $t=4$ | $t=5$ |
|---|---|---|---|---|---|
| {1} | | | | | |
| {2} | | | | | |
| {3} | | | | | |
| {4} | | | | | |
| {5} | | | | | ✓ |
| {1, 2} | | | | | |
| {2, 3} | | | | | |
| {3, 4} | | | | | |
| {4, 5} | | | | ✓ | |
| {1, 2, 3} | | | | | |
| {2, 3, 4} | | | | | |
| {3, 4, 5} | | | ✓ | | |
| {1, 2, 3, 4} | | | | | |
| {2, 3, 4, 5} | | ✓ | | | |
| {1, 2, 3, 4, 5} | ✓ | | | | |

*Low* treatments

**Step 1**

| Message | $t=1$ | $t=2$ | $t=3$ | $t=4$ | $t=5$ |
|---|---|---|---|---|---|
| {1} | ✓ | | | | |
| {2} | | ✓ | | | |
| {3} | | | ✓ | | |
| {4} | | | | ✓ | |
| {5} | | | | | ✓ |
| {1, 2} | ✓ | ✓ | | | |
| {2, 3} | | ✓ | ✓ | | |
| {3, 4} | | | ✓ | ✓ | |
| {4, 5} | | | | ✓ | ✓ |
| {1, 2, 3} | ✓ | ✓ | ✓ | | |
| {2, 3, 4} | | ✓ | ✓ | ✓ | |
| {3, 4, 5} | | | ✓ | ✓ | ✓ |
| {1, 2, 3, 4} | ✓ | ✓ | ✓ | ✓ | |
| {2, 3, 4, 5} | | ✓ | ✓ | ✓ | ✓ |
| {1, 2, 3, 4, 5} | ✓ | ✓ | ✓ | ✓ | ✓ |

**Step 2**

| Message | $t=1$ | $t=2$ | $t=3$ | $t=4$ | $t=5$ |
|---|---|---|---|---|---|
| {1} | ✓ | | | | |
| {2} | | | | | |
| {3} | | | | | |
| {4} | | | | | |
| {5} | | | | | |
| {1, 2} | | ✓ | | | |
| {2, 3} | | | ✓ | | |
| {3, 4} | | | | ✓ | |
| {4, 5} | | | | | ✓ |
| {1, 2, 3} | | ✓ | ✓ | | |
| {2, 3, 4} | | | ✓ | ✓ | |
| {3, 4, 5} | | | | ✓ | ✓ |
| {1, 2, 3, 4} | | ✓ | ✓ | ✓ | |
| {2, 3, 4, 5} | | | ✓ | ✓ | ✓ |
| {1, 2, 3, 4, 5} | | ✓ | ✓ | ✓ | ✓ |

**Step 3**

| Message | $t=1$ | $t=2$ | $t=3$ | $t=4$ | $t=5$ |
|---|---|---|---|---|---|
| {1} | ✓ | | | | |
| {2} | | | | | |
| {3} | | | | | |
| {4} | | | | | |
| {5} | | | | | |
| {1, 2} | | ✓ | | | |
| {2, 3} | | | | | |
| {3, 4} | | | | | |
| {4, 5} | | | | | |
| {1, 2, 3} | | | ✓ | | |
| {2, 3, 4} | | | | ✓ | |
| {3, 4, 5} | | | | | ✓ |
| {1, 2, 3, 4} | | | ✓ | ✓ | |
| {2, 3, 4, 5} | | | | ✓ | ✓ |
| {1, 2, 3, 4, 5} | | | ✓ | ✓ | ✓ |

**Step 4**

| Message | $t=1$ | $t=2$ | $t=3$ | $t=4$ | $t=5$ |
|---|---|---|---|---|---|
| {1} | ✓ | | | | |
| {2} | | | | | |
| {3} | | | | | |
| {4} | | | | | |
| {5} | | | | | |
| {1, 2} | | ✓ | | | |
| {2, 3} | | | | | |
| {3, 4} | | | | | |
| {4, 5} | | | | | |
| {1, 2, 3} | | | ✓ | | |
| {2, 3, 4} | | | | | |
| {3, 4, 5} | | | | | |
| {1, 2, 3, 4} | | | | ✓ | |
| {2, 3, 4, 5} | | | | | ✓ |
| {1, 2, 3, 4, 5} | | | | ✓ | ✓ |

**Step 5**

| Message | $t=1$ | $t=2$ | $t=3$ | $t=4$ | $t=5$ |
|---|---|---|---|---|---|
| {1} | ✓ | | | | |
| {2} | | | | | |
| {3} | | | | | |
| {4} | | | | | |
| {5} | | | | | |
| {1, 2} | | ✓ | | | |
| {2, 3} | | | | | |
| {3, 4} | | | | | |
| {4, 5} | | | | | |
| {1, 2, 3} | | | ✓ | | |
| {2, 3, 4} | | | | | |
| {3, 4, 5} | | | | | |
| {1, 2, 3, 4} | | | | ✓ | |
| {2, 3, 4, 5} | | | | | |
| {1, 2, 3, 4, 5} | | | | | ✓ |

*Note:* Tables for the *High* treatments are on the left and for the *Low* treatments on the right. Each table corresponds to a step $k$ of reasoning for the Sender. In a table corresponding to step $k$, there is a check mark in the cell corresponding to column $t$ and message $m$ if the Sender of type $t$ sends $m$ if he is at the $k$th step of reasoning.