Evidence Games: Lying Aversion and Commitment*

Elif B. Osun† Erkut Y. Ozbay‡

May 25, 2023

Abstract

The voluntary disclosure literature suggests that in evidence games, where the informed sender chooses which pieces of evidence to disclose to the uninformed receiver who determines their payoff, commitment does not matter, as there is a theoretical equivalence between the optimal mechanism and the game equilibrium outcomes. In this paper, we experimentally investigate whether the optimal mechanism and the game equilibrium outcomes coincide in a simple evidence game. Contrary to the theoretical equivalence, our results indicate that outcomes diverge and that commitment changes the outcomes. Our experimental results are in line with the predictions of a model that accounts for lying-averse agents. (JEL: C90, D82, D91)

*We thank Emel Filiz-Ozbay, Navin Kartik, Barton Lipman, and Marta Serra-Garcia for their helpful comments.

†Department of Economics, University of Maryland, Email: elif@umd.edu

‡Department of Economics, University of Maryland, Email: ozbay@umd.edu

1 Introduction

Decisions in various fields, such as medicine, finance, and law, often involve hard evidence. Suppose an informed sender (agent), who possesses hard evidence regarding their value, aims to maximize their benefits regardless of their evidence, while an uninformed receiver (principal), who sets the agent's reward, wants it to closely match the agent's value. When the agent is asked to disclose their evidence, they can do so fully or partially but cannot fabricate false evidence. These games are known as "evidence games" (Hart et al., 2017). For example, imagine an agent asked to submit a self-evaluation for an ongoing project. If the agent completed their work as planned, they have no evidence to report yet. However, if they made a mistake that cannot be traced back to them unless they disclose it, they may reveal it or act as if they have no evidence. In this example, the agent has an incentive to withhold their negative evidence, so the principal doesn't learn the agent's true value and there is no full disclosure in the equilibrium. As an alternative, suppose a principal commits to a reward policy before the agent reveals their evidence. This binding reward policy specifies the rewards for each piece of evidence and also for the case of no evidence. Does commitment to such a policy lead to different outcomes for the principal?

The role of commitment has been a central question in economics. Commitment takes various forms, such as binding agreements, rules, and laws, and it can serve multiple important purposes. Commitment can assist individuals and organizations in overcoming self-control issues, reducing risk, and signaling credibility or trustworthiness (e.g. Laibson, 1997; Maskin and Tirole, 1999; Bernheim and Whinston, 1990). Despite limiting one's future behavior, having commitment power can be beneficial in theory. For example, by committing, a leader can obtain the more advantageous Stackelberg outcome instead of the Cournot outcome.[1] Similarly, in the cheap talk setup of Crawford and Sobel (1982), commitment leads to a different outcome favoring the principal.[2] However, despite its broad applicability in many

[1] Bagwell (1995) shows that the commitment power is destroyed in any pure-strategy equilibrium if the leader's action is imperfectly observed. On the other hand, Van Damme and Hurkens (1997) show that when mixed-strategy equilibria are allowed, this conclusion is reversed. Huck and Müller (2000) experimentally test these results, finding no support for Bagwell (1995), but instead supporting Van Damme and Hurkens (1997).

[2] E.g. Example 3 in Appendix B in Hart et al. (2017).

economic settings, commitment does not matter in evidence games from a theoretical standpoint. Specifically, Hart et al. (2017) show that the unique equilibrium outcome of the game, where the principal is not committed to a reward policy, is the same as the optimal mechanism outcome with commitment. We experimentally test this outcome equivalence result to investigate the role of commitment in evidence games.

In an experiment on evidence games, there are several reasons why the outcome-equivalence result may not hold. First, the difference between outcomes could be due to equilibrium selection in the presence of multiple equilibria. Second, individuals may have difficulty with Bayesian updating (Friedman, 1998; Charness and Levin, 2005) or they may even make calculation mistakes. To test the role of commitment in evidence games, the experimental design needs to be simple enough to eliminate these factors. The following example, which demonstrates the intuition behind the outcome-equivalence result, is simple enough that neither subjects' ability to do complex Bayesian updating nor equilibrium selection is required. Hence, this simple environment is ideal for testing the role of commitment in evidence games, and we used it in our experiment.

Example: Assume that with a 50% probability, the agent completed their work without mistakes (called "high" type with a value of 100) and has no evidence. But with a 50% probability, the agent made a mistake (called "low" type with a value of 0) and has evidence to prove it. Without commitment, if the agent reveals their evidence, this evidence discloses their low type and the principal gives a reward of 0. Therefore, in the unique sequential equilibrium, the low-type agent hides their evidence and pretends to be a high type. Since neither type discloses any evidence, the principal sets the reward for no evidence at 50 ($50\% \times 100 + 50\% \times 0$). On the other hand, when there is commitment, the principal sets a reward of 50 for no evidence and a reward lower than or equal to 50 for evidence. This means that the principal cannot separate low and high type agents. The only way the principal can separate them is to set a strictly higher reward for evidence (which can only be disclosed by low types) than for no evidence, which is suboptimal. Therefore, whether there is a commitment or not, the outcomes coincide: the low-type agent hides their evidence, and both types of agents receive a payoff of 50.[3]

[3]This example is a simpler variant (with two types) of Example 1 in Hart et al. (2017). In

Despite the intuitiveness of this outcome-equivalence result, it relies on the assumption that the agent is solely motivated by maximizing their payoff. While it is almost always possible to provide no evidence in any evidence game, there may be cases where lying is unavoidable in order to do so.[4] Experimental literature has consistently shown a preference for truth-telling (e.g. Abeler et al., 2019), which aligns with models that factor in lying costs (e.g. Gneezy et al., 2018). Additionally, a considerable number of subjects in sender-receiver setups have exhibited a preference for truth-telling (e.g. Gneezy, 2005; Sánchez-Pagés and Vorsatz, 2007; Serra-Garcia et al., 2013). Since the existing literature has shown that one's preference for truth-telling can influence their behavior, it is reasonable to expect that an agent may disclose their evidence even when it results in a lower reward. In evidence games, if an agent incurs a cost for hiding evidence, they may choose to disclose it when the cost of hiding it outweighs its benefits. In the above example, if the principal commits to a reward of 50 for no evidence and a reward slightly smaller than 50 for disclosed evidence, a low-type agent might disclose their evidence to avoid the cost of hiding it. However, when there is no commitment, the equilibrium outcome remains unchanged since hiding evidence offers a benefit of 50, which may be higher than the cost of hiding it. Exploring whether an individual's behavior is influenced by another person's aversion to lying is a novel question. In this study, we investigate whether a principal considers the lying cost of an agent when committing to a reward policy.

Theoretically, lying costs have been widely studied in a cheap talk setup of Crawford and Sobel (1982) since the seminal work of Kartik (2009) (see also Kartik et al., 2007). Even in evidence games, truth-telling is a crucial aspect of the model. When there is a multiplicity of equilibria in an evidence game with no commitment, all *truth-leaning* equilibria without commitment yield the commitment outcome (Hart et al., 2017). According to the truth-leaning refinement, when the rewards for revealing the whole truth and partial truth are the same, the agent prefers to reveal the whole truth. That is, telling the whole truth leads to an infinitesimal

this scenario, consider a professor who has submitted a paper to a journal, and the dean, who decides on the professor's salary increase, asks whether the paper has been desk-rejected. If it has, the professor may either reveal the rejection letter (negative evidence) or hide this evidence by pretending not to have received a desk rejection (no evidence).

[4] We discuss the difference between deception with and without lying in detail in the Discussion Section.

increase in the agent's utility, or equivalently, hiding some evidence leads to an infinitesimal decrease in the agent's utility. Our theoretical analysis demonstrates that if this decrease in utility is strictly positive, even if it is small as in Serra-Garcia et al. (2013), commitment may matter. Specifically, we show that: (i) commitment increases the reward for no evidence, (ii) the expected utility of a principal is higher when there is commitment, and (iii) when there is commitment, the smaller the difference between rewards, the less likely a low-type agent withholds their evidence. Indeed, our experimental results are in line with these predictions of a model with lying averse agents,[5] falsifying the outcome-equivalence result in evidence games.

Related Literature

In the voluntary disclosure literature, the commitment case where the principal moves first and commits to a reward policy corresponds to a mechanism setup (e.g. Green and Laffont, 1986; Bull and Watson, 2007; Deneckere and Severinov, 2008); the no-commitment case where the receiver decides on the reward after observing the sender's decision corresponds to a game setup (e.g. Grossman and Hart, 1980; Grossman, 1981; Milgrom, 1981; Dye, 1985). Glazer and Rubinstein (2006) show that the outcome of the optimal mechanism could be obtained in the equilibrium of the game setup. This equivalence result has been extended and investigated in other settings (e.g. Sher, 2011; Ben-Porath et al., 2019). Hart et al. (2017) extend this result to evidence games. However, this outcome-equivalence result rests on the assumption of solely payoff-maximizing agents despite the optimal mechanism design with costly state misrepresentation has been already investigated in the literature (e.g. Lacker and Weinberg, 1989; Goldman and Slezak, 2006; Guttman et al., 2006; Deneckere and Severinov, 2017). We provide an example where in the presence of lying-averse agents this outcome-equivalence result fails in evidence games. The distinguishing factor of our setup is that the principal needs to take into account the agent's lying aversion.

In a closely related paper, Fréchette et al. (2019) experimentally investigate the role of commitment in communication. Their framework includes the setups of cheap

[5] Alternatively, a low-type agent may experience guilt for disappointing the principal by withholding their evidence (e.g. Charness and Dufwenberg, 2006; Battigalli and Dufwenberg, 2007). However, in Appendix C, we show that such a guilt aversion model does not accurately predict our experimental findings.

talk (Crawford and Sobel, 1982) and Bayesian persuasion game (Kamenica and Gentzkow, 2011). Fréchette et al. (2019) find that some senders over-communicate when information is verifiable and under-communicate when it is not. Our setup complements theirs by investigating the role of commitment in evidence games that lie outside of their framework. We find that receivers (principals) benefit from their commitment power if informed senders (agents) have lying costs.

Our paper also relates to the experimental literature on disclosure games. An important question in this literature is whether full information is disclosed when it is predicted by the theory that a separating equilibrium exists (e.g. Forsythe et al., 1989; King and Wallin, 1991; Deversi et al., 2018; Hagenbach and Perez-Richet, 2018; Li and Schipper, 2020; Jin et al., 2022). Jin et al., 2021 show that the receivers are not skeptical enough to perceive no news as bad news, which contributes to the lack of full information disclosure by the agents contrary to the separating equilibrium outcome predicted by theory. In our setup, we are interested in whether commitment changes the outcome when only a pooling equilibrium exists.

2 Experiment

We conducted the experiment at the Experimental Economics Laboratory at the University of Maryland (EEL-UMD). We recruited 128 subjects from the University of Maryland's undergraduate student pool via ORSEE (Greiner, 2015). None of the subjects participated in more than one session. We used the experimental software zTree (Fischbacher, 2007) to design the experiment. We conducted 8 sessions, each with 16 subjects. There was an equal number of subjects in each treatment. The average session lasted about an hour, and the average payment was \$15.4, including the \$7 show-up fee. Payoffs in the experiment were in Experimental Currency Units (ECUs) with a conversion rate of 10 ECUs for \$1. Each session of our experiment consisted of two parts. Paper instructions were distributed and read aloud prior to the start of each part. Before the experiment began, each subject was required to answer two questions that checked their understanding. If a subject failed to answer either of these questions correctly, they received a pop-up message informing them that they needed to correct their relevant answer. The experiment started only after every subject answered these questions correctly. The instructions, sample screenshots, and the two understanding questions are in Appendix D.

The first part of the experiment consisted of 20 independent periods. In the first period, each subject was assigned the role of "agent" or "principal", which remained fixed throughout the experiment.[6] In each period, subjects were randomly matched to another subject who was of the other role and played a single-shot game where the agent sends a message regarding their type and the principal chooses a reward between 0 and 100 for the agent.

The agent could be one of two types: *high* or *low* with values 100 and 0, respectively. Each type occurred with probability 50%. Low-type agents had evidence for their type, whereas high-type agents did not have evidence. To ensure that the subjects understood the difference between "evidence" and "type", we used sentences associated with each type of agent to be sent as messages: Low-type agents had access to the messages $m \in${"My type is low", "I don't have evidence for my type"} whereas high-type agents only had access to the message $m \in${"I don't have evidence for my type"}.[7] The agent's payoff was equal to the reward chosen by the principal. The principal's payoff was $100 - |x - v(t)|$, where x is the reward that the agent received, and $v(t)$ is the true value of the agent of type t. Note that the principal's payoff is maximized when the reward is equal to the true value of the agent. The probability distribution of the agent's type, available messages for each type, and payoff functions for both roles were common knowledge to both agents and principals. All subjects knew that this information was common knowledge for both roles.

There were two treatments that differed in whether the agent [No-Commitment Treatment] or the principal [Commitment Treatment] was the first mover. In the No-Commitment Treatment, the agent chose which message to send to the principal from the messages that were available to their type. Once the agent chose which message to send, the principal observed the message and then chose a reward for the agent.[8] In the Commitment Treatment, the principal chose a reward for each

[6] In the experiment, we stated the role of the agent as "sender" and the role of the principal as "receiver". We continue referring to the roles as "agent" and "principal" for the remainder of this paper for ease of reading.

[7] Information about agent types is summarized in Table A.1.

[8] For studies in which off-equilibrium behavior is important, one may use a strategy method for the principal's decision. Since our aim is to investigate the outcome equivalence between

possible message that they could receive before observing the message. The agent chose which message to send after observing the reward policy set by the principal. The type of the agent was randomly determined in each period.

In the second part of the experiment, which was identical in both treatments, we elicited subjects' risk preferences and ability to do Bayesian updating using two incentivized activities. In the first activity, we asked subjects to make choices from a menu of ordered lotteries following Holt and Laury (2002) to elicit their risk preferences. In the second activity, following Charness and Levin (2005), we asked subjects a Bayesian updating question which paid 10 ECUs if their answer was correct.

3 Theoretical Framework

Following the model of Hart et al. (2017), there is an agent denoted by A and a principal denoted by P. The agent can be one of two types, denoted by t, High or Low types with values $v(High) = H$ and $v(Low) = L$ such that $H > L \geq 0$. The probability of an agent being of High type is q, and the probability of an agent being of Low type is $1 - q$, where $q \in (0, 1)$. Low type agents have evidence for their type, while High type agents do not have evidence. As agents can choose to withhold their evidence, Low type agents have access to the messages $m \in${"Low evidence", "No evidence"}, whereas high type agents only have access to the message $m \in${"No evidence"}. Let $I > 0$ be the additional compensation to the principal. The principal chooses a reward $x \in [0, I]$. The probability distribution of the agent's type, available messages for each type, and payoff functions for both roles were common knowledge to both the agent and the principal.

The utility functions capture the idea that the agent wants as much reward as possible, while the principal wants the reward to match the value of the agent. The principal's utility depends on the reward and the value of the agent with type t but not on the message m: $U^P(m, x; t) = w(I - |v(t) - x|)$, where $w(.)$ is a continuously differentiable, strictly concave, and single-peaked function that is maximized at the

treatments, it is sufficient to observe the on-equilibrium behavior, and hence we use the direct-response method as in many sequential game experiments (see Brandts and Charness, 2011 for a detailed survey of the strategy method).

point where the reward equals the value of the agent, $x = v(t)$.

The agent's utility does not depend on either the type t or the message m; it depends on the reward x. Additionally, the agent may face a cost of lying if they withhold their evidence. Simplifying Kartik (2009), we follow Serra-Garcia et al. (2013) such that the utility of an agent with type t and cost of lying $k \geq 0$, sending a message m, and receiving a reward $x \geq 0$, $\hat{u}^A(m, x; t, k)$ takes the form:

$$\hat{u}^A(m, x; k, t) = \begin{cases} u(x) & \text{if whole-truth} \\ u(x) - k & \text{if withhold evidence} \end{cases}$$

where $u(.)$ is continuously differentiable and strictly increasing. Obviously, when $k = 0$, $u(.)$ becomes a standard payoff-maximizing utility function.

There are two setups that differ in whether the agent is the first-mover [No-Commitment] or the principal is the first-mover [Commitment]:

No-Commitment (NC): The agent chooses which message to send among the messages that are available to their type. After observing the agent's message, the principal chooses a reward for the agent.

Commitment (C): The principal, before the agent sends their message, sets a reward policy specifying the reward for each possible message. After observing the reward policy set by the principal, the agent chooses which message to send among the messages that are available to their type.

Let x_0^{NC}, x_-^{NC} denote the reward set for no evidence and low evidence in NC, respectively; and let x_0^C, x_-^C denote the reward set for no evidence and low evidence in C, respectively.

3.1 Payoff-Maximizing Agents

Before analyzing the general case, we will formulate our hypothesis based on the standard payoff-maximizing agents for the parameters used in the experiment. Recall that the parameters used in the experiment are I=100, H=100, L=0, and q=0.5.

First, let's consider the No-Commitment (NC) setup. In the unique sequential equilibrium of the game, both low and high type agents send no evidence. If the principal were ever to observe low evidence, the best response would be to set the reward equal to 0 since the principal's problem in this case is to choose $x_- \in [0, 100]$ to maximize $w(100 - x_-)$. So, $x_-^{NC} = 0$ in the No-Commitment setup. If the principal observes no evidence, the principal does not gain any new information from this message since all low type agents will pretend to be high type as long as $(x_0 > 0)$. The principal's problem upon observing no evidence is then to choose $x_0 \in [0, 100]$ to maximize $0.5 \cdot w(100 - x_0) + 0.5 \cdot w(x_0)$, which results in $x_0^{NC} = 50$ (Hypothesis 1).

Hypothesis 1 *The reward set for no evidence in the No-Commitment Treatment is 50.*

In the Commitment (C) setup, commitment does not help the principal. The only way to separate low type agents from high types is to set $x_- > x_0$ (as incentive compatibility constraint is $u(x_-) \geq u(x_0)$), which is not optimal because the expected utility of the principal is decreasing in x_-. Therefore, the problem of the principal is still to choose $x_0 \in [0, 100]$ to maximize $0.5 \cdot w(100 - x_0) + 0.5 \cdot w(x_0)$, which results in $x_0^C = 50$ and $x_-^C \leq 50$ (Hypothesis 2) in the optimal mechanism. Hence, commitment should not matter (Hypothesis 3).

Hypothesis 2 *The reward set for no evidence is equal to 50 in the Commitment Treatment.*

Hypothesis 3 *The reward set for no evidence in the No-Commitment Treatment is equal to the reward set for no evidence in the Commitment Treatment.*

Next, we turn to the agents. Since high type agents do not have any evidence to disclose or withhold, we will look at the behavior of low type agents. In the unique sequential equilibrium of the No-Commitment Treatment, if the agent reveals their evidence, the principal learns their type and gives $x_-^{NC} = 0$. However, if the agent withholds their evidence, the principal cannot learn their type and gives $x_0^{NC} = 50$. So, in the No-Commitment Treatment, the low type agent always withholds their evidence to get a higher reward. Similarly, in the optimal mechanism of the Commitment Treatment, the principal offers a higher reward for no evidence, $x_0^C = 50$, than for low evidence, $x_-^C \leq 50$, and the low type agent chooses not to reveal their type in any sequential equilibrium of the Commitment Treatment. Hence, withholding evidence behavior should be identical in both treatments (Hypothesis 4).

Hypothesis 4 *The percentage of low-type agents who withhold their low evidence is equal in both the No-Commitment Treatment and the Commitment Treatment.*

Additionally, in the Commitment Treatment, the agent is the second mover, so a low type agent decides whether to reveal their evidence or not after seeing the rewards committed by the principal. Unless the reward for low evidence is higher than the reward for no evidence, the reward amounts should not affect the agent's decision to withhold evidence. For instance, suppose the reward for no evidence is 50. Then, whether the reward for low evidence is 49 or 0 should not affect the agent's decision. Their decision solely depends on the highest reward rather than the amount of each reward (Hypothesis 5).

Hypothesis 5 *In the Commitment Treatment, as long as the reward for low evidence is not higher than the reward for no evidence, increasing or decreasing the reward amounts will not affect the percentage of low type agents who withhold their evidence.*

3.2 Lying Averse Agents

Before we characterize the optimal mechanism of the Commitment setup and the equilibrium of the No-Commitment setup under lying averse agents, let's illustrate how lying averse agents might behave differently than what is predicted in the model without lying costs. For example, consider two policies that a principal can commit to. In both of these policies, the payoff for an agent who withholds their information is 50, but in Policy 1, the payoff for an agent who reveals their information is 0, while in Policy 2 it is 49. An agent who does not have a lying cost ($k = 0$) withholds their information in both policies because $u(50) > u(0)$ and $u(50) > u(49)$. However, an agent with a small but positive cost of lying such that $u(49) > u(50) - k > u(0)$ withholds their information in Policy 1 because $u(50) - k > u(0)$, but reveals their information in Policy 2 because $u(50) - k < u(49)$. Thus, under the lying aversion model, the outcomes of the equilibrium when there is no commitment and the optimal mechanism when there is commitment may not coincide.

Rewards for No Evidence and Low Evidence

When there is no commitment, in the unique sequential equilibrium, the principal sets $x_-^{NC} = L$ for any $k \geq 0$ since the low evidence could be provided only by the agent with low value. The agent with no evidence does not have any evidence to send, and the agent with low evidence sends no evidence if k is small enough.[9] Then, the principal's problem, when they see no evidence, is:

$$\max_{x_0} \quad q \cdot w(I - |H - x_0|) + (1 - q) \cdot w(I - |L - x_0|)$$

The solution to this problem results in:

$$w'(I - H + x_0^{NC}) = \rho \cdot w'(I + L - x_0^{NC}) \tag{1}$$

where $\rho = (1 - q)/q$.

For the parameters of the experiment, Equation (1) becomes $w'(x_0^{NC}) = w'(100 - x_0^{NC})$, which implies that $x_0^{NC} = 50$. In other words, in the unique sequential equilibrium, the principal sets the reward for no evidence equal to 50. Hence, we expect Hypothesis 1 to be satisfied even if we account for lying-averse agents.

When there is commitment, for any strictly positive cost of lying, $k > 0$, the optimal mechanism *can* separate the types. Recall that in the absence of cost of lying, in order to separate the types, the principal needs to give distinct rewards, i.e. $x_- \neq x_0$. Also, the reward for low evidence needs to be higher than the reward for no evidence, i.e. $x_- > x_0$. Otherwise, i.e. $x_- < x_0$, the low type agent will withhold their low evidence since $u(x_-) < u(x_0)$. However, setting $x_- > x_0$ cannot be optimal for the principal since the value of the low type agent is smaller than the value of the high type agent. On the other hand, in the presence of cost of lying, the reward for low evidence can be lower than the reward for no evidence, i.e. $x_- < x_0$, and the low type agent may still reveal their low evidence since it is possible that

[9] Note that there should be an upper bound on the cost of lying, since if k were very large, the rewards would have become irrelevant for the subjects, and they would always reveal their evidence no matter what the rewards are. Such behavior is not observed in our data. We will show that this additional complication is not necessary, and our data can be explained by a small cost of lying.

$u(x_-) > u(x_0) - k$. Therefore, for $k > 0$, the principal's problem is:

$$\max_{x_0, x_-} \quad q \cdot w(I - |H - x_0|) + (1 - q) \cdot w(I - |L - x_-|)$$
$$\text{s.t.}$$
$$u(x_-) \geq u(x_0) - k$$
$$u(x_0) \geq u(x_-) \geq 0$$

Then, in the optimal mechanism:

$$u(x_-^C) = u(x_0^C) - k, \text{ and}$$
$$w^{'}(I - H + x_0^C) = \rho \cdot w^{'}(I + L - x_-^C) \tag{2}$$

For the parameters of the experiment, Equation (2) becomes $w^{'}(x_0^C) = w^{'}(100 - x_-^C)$. For a concave $w(.)$, $x_-^C + x_0^C = 100$, which implies $0 < x_-^C < 50 < x_0^C$. Therefore, we do not expect Hypothesis 2 to be satisfied when we account for lying-averse agents.

In Proposition 1, we additionally show that when the cost of lying is strictly positive, the commitment matters such that the principal sets a higher reward for no evidence when there is commitment than when there is no commitment, which is in contrast with Hypothesis 3.

Proposition 1 $x_0^C > x_0^{NC}$.

Proof: The only important assumption regarding $u(.)$ that it is a strictly increasing function. So, w.l.o.g., let $u(x) = x$. Equation (1) remains the same, and Equation (2) becomes:

$$x_-^C = x_0^C - k, \text{ and}$$
$$w^{'}(I - H + x_0^C) = \rho \cdot w^{'}(I + L - x_0^C + k) \tag{3}$$

For contradiction, assume $x_0^C \leq x_0^{NC}$. Then, for any $k > 0$, $-x_0^C + k > -x_0^{NC}$. So,

$$w^{'}(I - H + x_0^C) = \rho \cdot w^{'}(I + L - x_0^C + k) < \rho \cdot w^{'}(I + L - x_0^{NC}) = w^{'}(I - H + x_0^{NC})$$

where the first and the last equalities follow from Equations (1) and (3), and the inequality follows from the strict concavity of $w(.)$. However, $w^{'}(I - H + x_0^C) < w^{'}(I - H + x_0^{NC})$ implies that $x_0^C > x_0^{NC}$, which is a contradiction. ■

Withholding Evidence

Next, we look at the effect of rewards on subjects' decision to withhold their evidence. A low-type agent withholds their evidence if $u(x_-) < u(x_0) - k$. As argued above, in the No-Commitment Treatment, every low-type agent with a small cost of lying withholds their evidence since $u(0) < u(50) - k$. However, the agent with the same cost of lying may reveal their evidence in the Commitment Treatment since the optimal mechanism can incentivize not withholding the evidence by setting rewards such that $u(x_-^C) = u(x_0^C) - k$. To see this, for example, consider a low-type agent with $u(x) = x$ and $k = 20$. Suppose a principal commits to a reward of 60 if the agent provides no evidence and a reward of 40 if the agent reveals low evidence. A low-type agent with $u(x) = x$ and $k = 20$ will not withhold their evidence since $40 = 60 - 20$. However, such an agent will withhold their evidence in the No-Commitment Treatment since $0 < 50 - 20$. Hence, there will be fewer low-type agents withholding their evidence in the Commitment Treatment (contrary to the prediction in Hypothesis 4).

Additionally, in the Commitment Treatment, consider the cases where the reward for no evidence, x_0^C, is higher than the reward for low evidence, x_-^C. In these cases, a low-type agent with $k > 0$ reveals their evidence if and only if $u(x_-^C) \geq u(x_0^C) - k$. Since by changing the rewards it is possible to change the direction of the inequality, the decision of the agent may be altered. In particular, for any low-type agent with $k > 0$, there is a positive relation between the reward for no evidence and the likelihood of withholding the evidence, and a negative relation between the reward for low evidence and the likelihood of withholding the evidence (contrary to the prediction in Hypothesis 5). To see this, suppose $u(x_-^C) \geq u(x_0^C) - k$, i.e. agent reveals their evidence. If the reward for no evidence decreases to $\hat{x}_-^C$ such that $u(\hat{x}_-^C) < u(x_0^C) - k$, they withhold their evidence, or if the reward for low evidence increases to $\tilde{x}_0^C$ such that $u(x_-^C) < u(\tilde{x}_0^C) - k$, they withhold their evidence. Similarly, suppose $u(x_-^C) < u(x_0^C) - k$, i.e. they withhold their evidence. If the reward for no evidence increases to $\tilde{x}_-^C$ such that $u(\tilde{x}_-^C) \geq u(x_0^C) - k$, they reveal their evidence, or if the reward for low evidence decreases to $\hat{x}_0^C$ such that $u(x_-^C) \geq u(\hat{x}_0^C) - k$, they reveal their evidence.

Welfare Implications

Finally, if the outcome equivalence between two setups does not hold, does the principal prefer to commit to a policy when they face a lying-averse agent? We have shown that when the agent is lying-averse, the principal sets higher rewards for both low evidence and no evidence in a committed policy than in no-commitment. On the other hand, the principal can only separate the types with commitment. In this trade-off, it turns out that the principal is better off with a committed policy.

Proposition 2 *For $k > 0$, principal's expected utility when there is commitment is higher than that when there is no commitment.*

Proof: Since $x_0^C > x_0^{NC}$, $w'(I - H + x_0^C) < w'(I - H + x_0^{NC})$ due to strict concavity of $w(.)$. Plugging in the corresponding expressions from Equations (1) and (3), we get $\rho \cdot w'(I + L - x_0^C + k) < \rho \cdot w'(I + L - x_0^{NC})$, which in turn results in $x_0^C - k < x_0^{NC}$ by strict concavity of $w(.)$.

The principal's expected utility when there is commitment is:

$$q \cdot w(I - H + x_0^C) + (1 - q) \cdot w(I + L - x_-^C)$$

$$= q \cdot w(I - H + x_0^C) + (1 - q) \cdot w(I + L - x_0^C + k) \text{ (since } x_-^C = x_0^C - k)$$

$$> q \cdot w(I - H + x_0^{NC}) + (1 - q) \cdot w(I + L - x_0^C + k) \text{ (since } x_0^C > x_0^{NC} \text{ by Proposition 1)}$$

$$> q \cdot w(I - H + x_0^{NC}) + (1 - q) \cdot w(I + L - x_0^{NC}) \text{ (since } x_0^{NC} > x_0^C - k)$$

which is the principal's expected utility when there is no commitment. Hence, principal is better off in a setup with commitment. ■

For example, with lying averse agents and the experiment parameters, in the unique equilibrium without commitment, $x_0^{NC} = 50$ and $x_-^{NC} = 0$. Therefore, when the principal does not commit to a policy, their expected utility is $0.5 * w(100 - (100 - 50) + 0.5 * w(100 - (50 - 0)) = w(50)$. However, with commitment, the optimal mechanism separates the types with the rewards such that $x_0^C > 50 > x_-^C$ and $x_0^C + x_-^C = 100$. As a result, the principal's expected utility with commitment is $0.5 * w(100 - (100 - x_0^C)) + 0.5 * w(100 - x_-^C) = w(x_0^C) > w(50)$. Hence, when the agent is lying averse, the principal prefers to have a committed policy.

4 Results

In this section, we report the experimental results on the reward for no evidence and low evidence set by the principals, the truthful behavior of agents, and the payoffs of the subjects. We compare the results with the hypotheses discussed in the previous section.

Our data is independent at the session level, but there are 8 independent session clusters. Therefore, for more reliable inferences, throughout the analysis, we use non-parametric tests and the wild cluster bootstrap method for regression analysis (see Cameron et al., 2008). In particular, we follow the wild cluster bootstrap procedure of Cameron and Miller (2015) for OLS regressions, and the score wild bootstrap procedure of Kline and Santos (2012) for tobit and probit regressions. We compute 95% confidence intervals and p-values by using the wild bootstrap algorithms developed by Roodman et al. (2019) with 9,999 bootstrap replications and clustering at the session level.[10]

We begin our analysis with the reward decision of the subjects in the role of a principal. First, we compute the average reward set for no evidence and the average reward set for low evidence in each treatment by all principals (see Table 1).

Table 1: Average Rewards by Treatment

Treatment	Reward for No Evidence	Reward for Low Evidence
No-Commitment	50.58	19.36
	[85.4%]	[14.6%]
	(593)	(47)
Commitment	60.42	27.05
	[72.2%]	[27.8%]
	(640)	(640)

Note: Percent of low type subjects who chose the corresponding message in each cell are reported in brackets, number of observations are reported in parentheses.

[10] All results are robust to conducting the regression analyses without the wild cluster bootstrap method. Tables obtained without the wild cluster bootstrap method are reported in Appendix B.

Reward for No Evidence:

Theoretically, the reward for no evidence should be equal to 50 in both No-Commitment and Commitment treatments. The experimental data shows that the average reward set by principals for no evidence is 50.58 in the No-Commitment Treatment and 60.42 in the Commitment Treatment (see Table 1). By using a Wilcoxon signed-rank test, we compare the estimated constant to the theoretical prediction.[11] We find that the reward for no evidence in the No-Commitment Treatment is not significantly different than 50 ($p = 0.133$), which is in line with Hypothesis 1; yet it is significantly more than 50 in the Commitment Treatment ($p < 0.001$), which falsifies Hypothesis 2. These results are robust when we condition on the reward set by subjects who are classified as risk averse ($p = 0.162$ in the No-Commitment Treatment and $p < 0.001$ in the Commitment Treatment).

Result 1 *(a) In the No-Commitment Treatment, the reward set for no evidence is not significantly different from the equilibrium reward. (b) In the Commitment Treatment, the reward set for no evidence is significantly higher than the optimal reward.*

To measure treatment effects, we use a Tobit regression relating reward for no evidence on the treatment dummy (depicted in Table 2). The coefficient of the commitment variable is positive and significant ($p = 0.005$), falsifying Hypothesis 3. The treatment variable remains significant after controlling for period, gender, risk attitudes, and ability to Bayesian update ($p = 0.002$).

Result 2 *The reward set for no evidence in the Commitment Treatment is significantly higher than that in the No-Commitment Treatment.*

[11] Unless otherwise stated, all p-values to compare distributions are obtained using the Mann Whitney U-test and all p-values to compare measures to benchmarks are obtained using the Wilcoxon signed-rank test in non-parametric analysis.

Table 2: Tobit Regressions Relating Reward for No Evidence to Treatment

	(1)	(2)
Commitment	15.32***	15.06***
	(0.005)	(0.002)
Period		-0.47
		(0.327)
Gender		-1.6
		(0.899)
Risk aversion		-0.89
		(0.885)
Ability to		-7.0
Bayesian update		(0.602)
Constant	50.3***	59.9***
	(0.003)	(0.003)
Observations	1,233	1,233

Notes: Dependent variable is *reward for no evidence*, bounded between 0 and 100. *Commitment* is a dummy variable that takes the value 1 if subject is in the Commitment Treatment and 0 if subject is in the No-Commitment Treatment. *Period* takes values from 1 to 20 and represents the period. *Gender* is a dummy variable that takes the value 1 if subject is female and 0 otherwise. *Risk Aversion* takes the value 1 if the subject is classified as risk averse based on the number of safe options they chose in Activity 1 and 0 otherwise. *Ability to Bayesian update* is a dummy variable that takes the value 1 if subject answered the Activity 2 question of Part II correctly and 0 otherwise. p-values computed by the score wild bootstrap procedure are in parentheses (clusters are at the session level); * p<0.1, ** p<0.05, *** p<0.01.

Reward for Low Evidence:

In the Commitment Treatment, as expected, the reward for low evidence is rarely higher than the reward for no evidence (only 3.9%). For each policy, we take the difference between the reward for no evidence and the reward for low evidence. We find that this difference is significantly higher than 0 ($p < 0.001$). Additionally, the average reward set by principals for low evidence is 27.05, significantly less than 50 ($p < 0.001$), but significantly more than 0 ($p < 0.001$). On the other hand, in the No-Commitment Treatment, observing low evidence is an off-equilibrium behavior. As expected, when the principals observe low evidence, 59.57% of the reward for low evidence is equal to 0 in the No-Commitment Treatment.

Withholding Information

Next, we examine the percentage of subjects withholding their information (i.e. sending no evidence when they are low type) across treatments. In the No-Commitment Treatment, 85.4% of low-type subjects withhold their low evidence, while this ratio is 72.2% in the Commitment Treatment. These percentages are significantly different from each other ($p < 0.001$, both with a test of proportions and with a Mann–Whitney test), falsifying Hypothesis 4. The difference in withholding information across treatments may be due to the principal's reward choice or due to the agent's behavior. In the Commitment Treatment, if a principal sets the reward for low evidence strictly higher than the reward for no evidence, it becomes optimal even for a payoff-maximizing low-type agent to reveal their evidence. Even when we exclude those rare cases, the percentage of low-type agents withholding their evidence in the Commitment Treatment (74.4%) is still significantly lower ($p < 0.001$).

Result 3 *The subjects with low evidence are significantly less likely to withhold their evidence in the Commitment Treatment than those in the No-Commitment Treatment.*

To test Hypothesis 5, we use a probit regression relating the withholding of information by low-type agents to the rewards for no evidence and low evidence in the Commitment Treatment, conditioning on the cases in which the reward for low evidence is not higher than the reward for no evidence. Table 3 shows that agents are more likely to withhold evidence when the reward for no evidence is higher ($p = 0.012$), yet they are less likely to withhold evidence when the reward for low evidence is higher ($p = 0.013$), falsifying Hypothesis 5. The reward for no evidence and the reward for low evidence both continue to have a significant effect on the propensity to withhold evidence after controlling for period, gender, risk attitudes, and the ability to Bayesian update ($p = 0.011$ and $p = 0.017$, respectively).[12]

[12] Additionally, we report the results of a probit regression relating the withholding of information by low-type agents to the difference between rewards in Table A.2. The difference between the reward for no evidence and the reward for low evidence has a significant effect on low-type agents' propensity to withhold evidence in the Commitment Treatment.

Result 4 *In the Commitment Treatment, subjects with low evidence are significantly more likely to withhold evidence as the reward for no evidence increases, and are significantly less likely to withhold evidence as the reward for low evidence increases, even when the reward for low evidence is not higher than the reward for no evidence.*

Table 3: Probit Regressions Relating Withholding Information to the Rewards in the Commitment Treatment Conditioning on the Difference being Positive

	(1)	(2)
Reward for	0.018**	0.019**
No Evidence	(0.012)	(0.011)
Reward for	-0.026**	-0.027**
Low Evidence	(0.013)	(0.017)
Period		0.019*
		(0.081)
Gender		-0.289
		(0.194)
Risk aversion		-0.871
		(0.173)
Ability to		0.067
Bayesian update		(0.865)
Constant	0.382**	1.155
	(0.028)	(0.129)
Observations	320	320

Notes: Dependent variable *withhold evidence* is equal to 1 if the low-type agent sent no evidence in the Commitment Treatment and 0 if they sent low evidence. *Period* takes values from 1 to 20 and represents the period. *Gender* is a dummy variable that takes the value 1 if subject is female and 0 otherwise. *Risk Aversion* takes the value 1 if the subject is classified as risk averse based on the number of safe options they chose in Activity 1 and 0 otherwise. *Ability to Bayesian update* is a dummy variable that takes the value 1 if subject answered the Activity 2 question of Part II correctly and 0 otherwise. p-values computed by the score wild bootstrap procedure are in parentheses (clustered at the session level); * p<0.1, ** p<0.05, *** p<0.01.

Next, we investigate the source of the withholding of evidence by low-type agents. The difference in agents' behavior in terms of withholding evidence across treatments may be due to the rewards being different across treatments or due to a difference in agents' behavior in the presence of commitment. Recall that the principals set different rewards in different treatments, and hence the low-type agents received different rewards in different treatments (the average rewards for a low-type agent are 46.1 and 58.1 in NC and C, respectively).[13].

In order to disentangle these effects, we examine the withholding behavior while controlling for treatment and rewards. Note that in the Commitment Treatment, the agents observe both rewards before deciding whether to withhold their evidence, whereas they do not observe the rewards in the No-Commitment Treatment at the time of making a decision. Therefore, we use the reward for no evidence and the reward for low evidence as controls for rewards in the Commitment Treatment, but we use the theoretical predictions (50 for no evidence, 0 for low evidence) as controls for rewards in the No-Commitment Treatment.[14]

We report the results of a probit regression that relates the withholding of information by low-type agents to the treatment dummy and the rewards, as explained above, in Table 4. We find that the decision to withhold evidence is significantly related to rewards, while commitment by itself has no significant effect on the decision to withhold evidence. Hence, we conjecture that the difference in agents' behavior in terms of withholding evidence across treatments is driven by the difference in rewards across treatments, and not by a psychological effect of commitment that leads agents to behave differently.

[13]Also, the average rewards for a high-type agent are 50.5 and 61.7 in NC and C, respectively. Theoretically, in the absence of lying aversion, these rewards should have been equal to 50. However, in the presence of lying aversion, a high-type agent should expect to receive 50 in NC but higher than 50 in C, and a low-type agent should expect to receive lower than 50 in NC but higher than 50 in C.

[14]We will denote the rewards described in this paragraph as Reward* for No Evidence and Reward* for Low Evidence.

Table 4: Probit Regressions Relating Withholding Information to Commitment and Rewards*

	(1)	(2)
Commitment	0.204	0.105
	(0.435)	(0.687)
Reward* for	0.018**	0.020***
No Evidence	(0.011)	(0.003)
Reward* for	-0.025***	-0.025***
Low Evidence	(0.006)	(0.010)
Controls	No	Yes
Observations	655	655

Notes: Dependent variable *withhold evidence* is equal to 1 if the low-type agent sent no evidence and 0 if they sent low evidence. *Commitment* is a dummy variable that takes the value 1 if subject is in the Commitment Treatment and 0 if subject is in the No-Commitment Treatment. *Reward* for No Evidence* is equal to 50 in the No-Commitment Treatment and equal to the reward for no evidence set by the principal in the Commitment Treatment. *Reward* for Low Evidence* is equal to 0 in the No-Commitment Treatment and equal to the reward for low evidence set by the principal in the Commitment Treatment. Column (1) does not include any additional controls, Column (2) additionally controls for *Period*, *Gender*, *Risk Aversion*, and *Ability to Bayesian update*. p-values computed by the score wild bootstrap procedure are in parentheses (clustered at the session level); * p<0.1, ** p<0.05, *** p<0.01.

Payoff of Subjects

Last, we turn our attention to payoffs of agents and principals. We begin by calculating the expected payoff of principals. In the Commitment Treatment, we calculate a principal's expected payoff by:

$$E[\pi^C] = 0.5 \times x_0^C + 0.5 \times p_{lie}^C \times (100 - x_0^C) + 0.5 \times (1 - p_{lie}^C) \times (100 - x_-^C)$$

where x_0^C is the reward for no evidence, x_-^C is the reward for low evidence, and p_{lie}^C is the probability of lying, which is a function of x_0^C and x_-^C, since the agent observes both rewards when deciding whether to withhold their evidence in the Commitment Treatment.

We estimate p_{lie}^C by regressing the low-type agents' decision to withhold their evidence in the Commitment Treatment on both rewards. Using a logit regression in which the intercept is β_0, the coefficient of the reward for no evidence is β_{NE}, and the coefficient of the reward for low evidence is β_{LE}, we estimate the probability to withhold evidence as $1/(1 + e^{-(\beta_0+\beta_{NE}\times x_0^C+\beta_{LE}\times x_-^C)})$. Using this estimated probability of lying, we calculate the expected reward for the principals as 50.4 in the Commitment Treatment.

In the No-Commitment Treatment, we calculate a principal's expected payoff by:

$$E[\pi^{NC}] = 0.5 \times x_0^{NC} + 0.5 \times p_{lie}^{NC} \times (100 - x_0^{NC}) + 0.5 \times (1 - p_{lie}^{NC}) \times (100 - x_-^{NC})$$

where x_0^C is the reward for no evidence, x_-^C is the reward for low evidence, and p_{lie}^C is the probability of lying, which is not a function of the rewards, since the agent doesn't observe the rewards when deciding whether to withhold their evidence in the No-Commitment Treatment.

We estimate p_{lie}^{NC} using the frequency of low-type agents' decision to withhold their evidence in the No-Commitment Treatment. Using this estimated probability of lying, we calculate the expected reward for the principals as 48.9 in the No-Commitment Treatment. Using a Mann-Whitney test, we find that the difference in principals' expected reward across treatments is significant ($p < 0.001$).

Result 5 *Principals' expected earning is higher in the Commitment Treatment compared to the No-Commitment Treatment.*

Finally, looking at agents' realized earnings, experimental results show that the average payoff of agents is equal to 48.3 in the No-Commitment Treatment versus 59.8 in the Commitment Treatment. The difference is statistically significant ($p < 0.001$). A breakdown of agents by type shows that both types earn significantly less in the No-Commitment Treatment. The payoff of low-type agents is 46.1 in the No-Commitment Treatment versus 58.1 in the Commitment Treatment. The average payoff of high-type agents is 50.5 in the No-Commitment Treatment versus 61.7 in the Commitment Treatment. Both differences are statistically significant ($p < 0.001$).

Result 6 *Both low and high types of agents have a higher payoff in the Commitment Treatment compared to the No-Commitment Treatment.*

5 Discussion and Further Directions

The role of commitment in information disclosure has been a central question. While it has been studied in various setups such as cheap-talk and Bayesian persuasion games (Fréchette et al., 2019), our experiment is the first to the role of commitment in evidence games, in which an uninformed principal chooses a reward for an informed agent who can reveal pieces of evidence about their type. We design our experiment to be simple enough to leave minimum room for subject mistakes. Nevertheless, we falsify the outcome-equivalence between settings with and without commitment contrary to the predictions based on payoff-maximizing agents (Hart et al., 2017). Our experimental results yield commitment increases the rewards set by the principal. On the other hand, agents are more likely to reveal their evidence when there is a commitment. Moreover, even when the reward for disclosing the evidence is lower than providing no-evidence, as these rewards get closer the agents are more likely to disclose their evidence. These results are in line with the predictions of a model with lying averse agents.

Truthtelling is a crucial aspect of evidence games. In these games, evidence is verifiable, so agents can either disclose all of their evidence (the whole truth) or withhold some pieces of evidence (partial truth), but they cannot fabricate false evidence (Hart et al., 2017). In other words, messages are restricted to a subset of their available evidence. In an evidence game setup, while an agent with evidence can pretend to have less evidence, they cannot pretend to have more. However, this truthtelling requirement does not impose any restrictions on the language of how the messages are communicated.

As defined in Sobel (2020), deception involves attempting to induce incorrect beliefs, while lying involves deliberately providing false information. In any evidence game, agents who withhold evidence are purposefully trying to make the principal believe that they have less evidence with a positive probability. Therefore, deception occurs in all evidence games with the exception of trivial evidence games where telling the whole truth for all types is the equilibrium. For example, in the equilibrium of our experimental setting, when an agent with evidence provides no evidence, they deceive the principal into believing they do not have evidence with a probability of 50 percent. In evidence games, although agents cannot lie by providing false evidence, they can still lie by sending a false message. Our experimental setting is in the framework of evidence games since agents without evidence are limited to providing no evidence, whereas agents with evidence can convey either their evidence or provide no-evidence although they need to lie by sending the message "I don't have evidence" in order to provide no-evidence. Indeed, in many real-life settings, perhaps with the exception of criminal cases, an agent needs to lie in order to provide no-evidence. For example, one may provide no-evidence by using their right to remain silent in a trial, but when a dean is deciding on a professor's salary and asks the professor the outcome of their most recent submission, the professor cannot realistically say that they want to remain silent. Or, when a used car seller is asked whether the car had an accident, they cannot hide the accident information without lying.[15] Moreover, by design, evidence may only be withheld by lying. For example, a car-selling website forces to disclose whether the car had an accident with a yes-no answer.

In a laboratory experiment, it is possible to have a setup where agents withhold their evidence without lying. This can be achieved for instance, if the agent makes an uninformative but true statement, such as "I may or may not have evidence" or "I want to remain silent", to provide no evidence. By making such a statement, an agent with evidence is involved in deception without lying. The experimental literature has already shown that individuals are still averse to deceiving in situations where they do not have to lie to deceive, but merely need to withhold their information to do so (e.g. Sánchez-Pagés and Vorsatz, 2009; Serra-Garcia et al., 2011; Friesen and Gangadharan, 2013; Ertac et al., 2016). However, aversion to deceiving without lying is not as strong as lying aversion (e.g. Sánchez-Pagés and Vorsatz, 2009; Jin et al., 2021). A novel aspect of our setup is that the principal needs to anticipate the agents' aversion by committing more generous rewards. Hence, we conjecture, that in a setup where agents withhold their evidence without lying, the results will be qualitatively similar to our results, but they will be less pronounced.[16]

Even if an agent can provide no-evidence without lying, such as remaining silent, in reality, sending the message, "I do not have any evidence" is also available.

[15] Since the agent knows their evidence, saying "I do not know" is also a lie.

[16] Additionally, the norms against lying differ across cultures and the strength of a norm has an effect on the lying behavior (Aycinena et al., 2022). Investigating how the behavior of the principals changes if the principals learn the lying norm within their session may be interesting.

Hence, the principal may infer remaining silent as hiding negative evidence. A further interesting question may be to investigate a setup where both providing no-evidence with and without lying are possible. For instance, an agent, with low evidence, can send one of the three messages: "I have low evidence", "I do not have any evidence" or "I want to remain silent" and an agent, with no-evidence, can send either 'I do not have any evidence" or "I want to remain silent". In such a situation, it may be interesting to see how the principal commits to a reward policy for these messages of an agent.[17]

Finally, our experimental result yields the power of commitment in evidence games: the expected payoff of the principal is higher when there is commitment. We also find that agents are better off when the principal commits to a policy. Therefore, it would be interesting to investigate if a principal is given the option to decide whether to commit to a policy or not, would they be willing to pay to commit to a policy? Additionally, would the different types of agents be willing to pay for the principal to commit to a reward policy? We leave these questions for future work.

[17]In the game setting where there is no commitment, to study the rewards for 'I do not have any evidence" and "I want to remain silent" messages, one may use a strategy method.

References

Abeler, J., D. Nosenzo, and C. Raymond (2019): "Preferences for truth-telling," *Econometrica*, 87, 1115–1153.

Aycinena, D., L. Rentschler, B. Beranek, and J. F. Schulz (2022): "Social norms and dishonesty across societies," *Proceedings of the National Academy of Sciences*, 119, e2120138119.

Bagwell, K. (1995): "Commitment and observability in games," *Games and Economic Behavior*, 8, 271–280.

Battigalli, P. and M. Dufwenberg (2007): "Guilt in games," *American Economic Review*, 97, 170–176.

Ben-Porath, E., E. Dekel, and B. L. Lipman (2019): "Mechanisms with evidence: Commitment and robustness," *Econometrica*, 87, 529–566.

Bernheim, B. D. and M. D. Whinston (1990): "Multimarket contact and collusive behavior," *Rand Journal of Economics*, 21, 1–26.

Brandts, J. and G. Charness (2011): "The strategy versus the direct-response method: a first survey of experimental comparisons," *Experimental Economics*, 14, 375–398.

Bull, J. and J. Watson (2007): "Hard evidence and mechanism design," *Games and Economic Behavior*, 58, 75 – 93.

Cameron, A. C., J. B. Gelbach, and D. L. Miller (2008): "Bootstrap-based improvements for inference with clustered errors," *Review of Economics and Statistics*, 90, 414–427.

Cameron, A. C. and D. L. Miller (2015): "A practitioner's guide to cluster-robust inference," *Journal of Human Resources*, 50, 317–372.

Charness, G. and M. Dufwenberg (2006): "Promises and partnership," *Econometrica*, 74, 1579–1601.

Charness, G. and D. Levin (2005): "When optimal choices feel wrong: A laboratory study of Bayesian updating, complexity, and affect," *American Economic Review*, 95, 1300–1309.

Crawford, V. P. and J. Sobel (1982): "Strategic information transmission," *Econometrica*, 1431–1451.

Deneckere, R. and S. Severinov (2008): "Mechanism design with partial state verifiability," *Games and Economic Behavior*, 64, 487–513.

——— (2017): "Screening, signalling and costly misrepresentation," Tech. rep., Working paper.

Deversi, M., A. Ispano, and P. Schwardmann (2018): "Spin doctors: A model and an experimental investigation of vague disclosure," .

Dye, R. A. (1985): "Disclosure of nonproprietary information," *Journal of Accounting Research*, 123–145.

Ertac, S., L. Koçkesen, and D. Ozdemir (2016): "The role of verifiability and privacy in the strategic provision of performance feedback: Theory and experimental evidence," *Games and Economic Behavior*, 100, 24–45.

Fischbacher, U. (2007): "z-Tree: Zurich toolbox for ready-made economic experiments," *Experimental Economics*, 10, 171–178.

Forsythe, R., R. M. Isaac, and T. R. Palfrey (1989): "Theories and tests of" blind bidding" in sealed-bid auctions," *The Rand Journal of Economics*, 214–238.

Fréchette, G. R., A. Lizzeri, and J. Perego (2019): "Rules and commitment in communication," *CEPR Discussion Paper No. DP14085.*

Friedman, D. (1998): "Monty Hall's three doors: Construction and deconstruction of a choice anomaly," *American Economic Review*, 88, 933–946.

Friesen, L. and L. Gangadharan (2013): "Designing self-reporting regimes to encourage truth telling: An experimental study," *Journal of Economic Behavior & Organization*, 94, 90–102.

Glazer, J. and A. Rubinstein (2006): "A study in the pragmatics of persuasion: a game theoretical approach," *Theoretical Economics*, 1, 395–410.

Gneezy, U. (2005): "Deception: The role of consequences," *American Economic Review*, 95, 384–394.

Gneezy, U., A. Kajackaite, and J. Sobel (2018): "Lying aversion and the size of the lie," *American Economic Review*, 108, 419–53.

Goldman, E. and S. L. Slezak (2006): "An equilibrium model of incentive contracts in the presence of information manipulation," *Journal of Financial Economics*, 80, 603–626.

Green, J. R. and J.-J. Laffont (1986): "Partially verifiable information and mechanism design," *Review of Economic Studies*, 53, 447–456.

Greiner, B. (2015): "Subject pool recruitment procedures: organizing experiments with ORSEE," *Journal of the Economic Science Association*, 1, 114–125.

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

Grossman, S. J. and O. D. Hart (1980): "Disclosure laws and takeover bids," *Journal of Finance*, 35, 323–334.

Guttman, I., O. Kadan, and E. Kandel (2006): "A rational expectations theory of kinks in financial reporting," *Accounting Review*, 81, 811–848.

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

Hart, S., I. Kremer, and M. Perry (2017): "Evidence games: Truth and commitment," *American Economic Review*, 107, 690–713.

Holt, C. A. and S. K. Laury (2002): "Risk aversion and incentive effects," *American Economic Review*, 92, 1644–1655.

Huck, S. and W. Müller (2000): "Perfect versus imperfect observability—an experimental test of Bagwell's result," *Games and Economic Behavior*, 31, 174–190.

Jin, G. Z., M. Luca, and D. Martin (2021): "Is no news (perceived as) bad news? An experimental investigation of information disclosure," *American Economic Journal: Microeconomics*, 13, 141–73.

——— (2022): "Complex disclosure," *Management Science*, 68, 3236–3261.

Kamenica, E. and M. Gentzkow (2011): "Bayesian persuasion," *American Economic Review*, 101, 2590–2615.

Kartik, N. (2009): "Strategic communication with lying costs," *Review of Economic Studies*, 76, 1359–1395.

Kartik, N., M. Ottaviani, and F. Squintani (2007): "Credulity, lies, and costly talk," *Journal of Economic theory*, 134, 93–116.

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

Kline, P. and A. Santos (2012): "A score based approach to wild bootstrap inference," *Journal of Econometric Methods*, 1, 23–41.

Lacker, J. M. and J. A. Weinberg (1989): "Optimal contracts under costly state falsification," *Journal of Political Economy*, 97, 1345–1363.

Laibson, D. (1997): "Golden eggs and hyperbolic discounting," *The Quarterly Journal of Economics*, 112, 443–478.

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

Maskin, E. and J. Tirole (1999): "Unforeseen contingencies and incomplete contracts," *The Review of Economic Studies*, 66, 83–114.

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

Roodman, D., M. Ø. Nielsen, J. G. MacKinnon, and M. D. Webb (2019): "Fast and wild: Bootstrap inference in Stata using boottest," *The Stata Journal*, 19, 4–60.

Sánchez-Pagés, S. and M. Vorsatz (2007): "An experimental study of truth-telling in a sender–receiver game," *Games and Economic Behavior*, 61, 86–112.

——— (2009): "Enjoy the silence: an experiment on truth-telling," *Experimental Economics*, 12, 220–241.

Serra-Garcia, M., E. Van Damme, and J. Potters (2011): "Hiding an inconvenient truth: Lies and vagueness," *Games and Economic Behavior*, 73, 244–261.

——— (2013): "Lying about what you know or about what you do?" *Journal of the European Economic Association*, 11, 1204–1229.

Sher, I. (2011): "Credibility and determinism in a game of persuasion," *Games and Economic Behavior*, 71, 409.

Sobel, J. (2020): "Lying and deception in games," *Journal of Political Economy*, 128, 907–947.

Van Damme, E. and S. Hurkens (1997): "Games with imperfectly observable commitment," *Games and Economic Behavior*, 21, 282–308.

Appendix A Additional Tables

Table A.1: Types of an Agent

Type t	**Value** $v(t)$	**Probability** q_t	**Available Messages**
High	100	50%	{"I don't have evidence for my type"}
Low	0	50%	{"My type is low", "I don't have evidence for my type"}

Table A.2: Probit Regressions Relating Withholding Information to the Difference Between Rewards in the Commitment Treatment Conditioning on the Difference being Positive

	(1)	(2)	(3)
Difference Between	0.022***	0.023***	0.019**
Rewards	(0.009)	(0.009)	(0.011)
Reward for			-0.007*
Low Evidence			(0.083)
Period		0.026**	0.019*
		(0.014)	(0.081)
Gender		-0.263	-0.289
		(0.263)	(0.194)
Risk Aversion		-0.848	-0.871
		(0.159)	(0.173)
Ability to		0.114	0.067
Bayesian update		(0.784)	(0.865)
Constant	0.046	0.698	1.155
	(0.814)	(0.110)	(0.129)
Observations	320	320	320

Notes: Dependent variable *withhold evidence* is equal to 1 if the low-type agent sent no evidence in the Commitment Treatment and 0 if they sent low evidence. *Difference Between Rewards* is the difference between Reward for No Evidence and Reward for Low Evidence. *Period* takes values from 1 to 20 and represents the period. *Gender* is a dummy variable that takes the value 1 if subject is female and 0 otherwise. *Risk Aversion* takes the value 1 if the subject is classified as risk averse based on the number of safe options they chose in Activity 1 and 0 otherwise. *Ability to Bayesian update* is a dummy variable that takes the value 1 if subject answered the Activity 2 question of Part II correctly and 0 otherwise. p-values computed by score wild bootstrap procedure are in parentheses (clustered at the session level); * p<0.1, ** p<0.05, *** p<0.01.

Appendix B Regressions Without Bootstrapping Procedure

Table B.1: Tobit Regressions Relating Reward for No Evidence to Treatment

	(1)	(2)
Commitment	15.32**	15.06**
	(0.026)	(0.029)
Period		-0.47
		(0.213)
Gender		-1.6
		(0.853)
Risk aversion		-0.89
		(0.897)
Ability to		-7.0
Bayesian update		(0.427)
Constant	50.3***	59.9***
	(0.000)	(0.000)
Observations	1,233	1,233

Notes: Dependent variable is *reward for no evidence*, bounded between 0 and 100. *Commitment* is a dummy variable that takes the value 1 if subject is in Commitment Treatment and 0 if subject is in No-Commitment Treatment. *Period* takes values from 1 to 20 and represents the period. *Gender* is a dummy variable that takes the value 1 if subject is female and 0 otherwise. *Risk Aversion* takes the value 1 if the subject is classified as risk averse based on the number of safe options they chose in Activity 1 and 0 otherwise. *Ability to Bayesian update* is a dummy variable that takes the value 1 if subject answered the Activity 2 question of Part II correctly and 0 otherwise. Standard errors are clustered at the individual level. p-values are in parentheses; * p<0.1, ** p<0.05, *** p<0.01.

Table B.2: Probit Regressions Relating Withholding Information to the Rewards in the Commitment Treatment Conditioning on the Difference being Positive

	(1)	(2)
Reward for	0.018***	0.019***
No Evidence	(0.000)	(0.000)
Reward for	-0.026***	-0.027***
Low Evidence	(0.000)	(0.000)
Period		0.019
		(0.204)
Gender		-0.289
		(0.275)
Risk aversion		-0.871***
		(0.000)
Ability to		0.067
Bayesian update		(0.799)
Constant	0.382**	1.155***
	(0.031)	(0.008)
Observations	320	320

Notes: Dependent variable *withhold evidence* is equal to 1 if the low-type agent sent no evidence in the Commitment Treatment and 0 if they sent low evidence. *Period* takes values from 1 to 20 and represents the period. *Gender* is a dummy variable that takes the value 1 if subject is female and 0 otherwise. *Risk Aversion* takes the value 1 if the subject is classified as risk averse based on the number of safe options they chose in Activity 1 and 0 otherwise. *Ability to Bayesian update* is a dummy variable that takes the value 1 if subject answered the Activity 2 question of Part II correctly and 0 otherwise. Standard errors are clustered at the individual level. p-values are in parentheses; * p<0.1, ** p<0.05, *** p<0.01.

Table B.3: Probit Regressions Relating Withholding Information to the Difference Between Rewards in the Commitment Treatment Conditioning on the Difference being Positive

	(1)	(2)	(3)
Difference Between	0.022***	0.023***	0.019***
Rewards	(0.000)	(0.000)	(0.000)
Reward for			-0.007**
Low Evidence			(0.017)
Period		0.026*	0.019
		(0.071)	(0.204)
Gender		-0.263	-0.289
		(0.316)	(0.275)
Risk Aversion		-0.848***	-0.871***
		(0.000)	(0.000)
Ability to		0.114	0.067
Bayesian update		(0.669)	(0.799)
Constant	0.046	0.698*	1.155***
	(0.751)	(0.092)	(0.008)
Observations	320	320	320

Notes: Dependent variable *withhold evidence* is equal to 1 if the low-type agent sent no evidence in the Commitment Treatment and 0 if they sent low evidence. *Difference Between Rewards* is the difference between Reward for No Evidence and Reward for Low Evidence. *Period* takes values from 1 to 20 and represents the period. *Gender* is a dummy variable that takes the value 1 if subject is female and 0 otherwise. *Risk Aversion* takes the value 1 if the subject is classified as risk averse based on the number of safe options they chose in Activity 1 and 0 otherwise. *Ability to Bayesian update* is a dummy variable that takes the value 1 if subject answered the Activity 2 question of Part II correctly and 0 otherwise. Standard errors are clustered at the individual level. p-values are in parentheses; * p<0.1, ** p<0.05, *** p<0.01.

Appendix C Model With Guilt

Alternative to the lying aversion model in which the agent was lying averse, we consider a model in which the agent may be guilt averse. Using the simple guilt model of Battigalli and Dufwenberg (2007), a principal who accounts for the agent's guilt aversion solves the following problem:

$$\begin{aligned} \max_{x_0, x_-} \quad & q \cdot (I - (H - x_0)) + (1 - q) \cdot (I - (x_- - L)) \\ \text{s.t.} \quad & x_- \geq x_0 - G \cdot \beta \cdot max\{\xi - (I - (x_0 - L)), 0\} \end{aligned}$$

where $G > 0$ is the agent's guilt parameter, $\beta \in [0, 1]$ is the agent's second order belief on the principal's belief that the agent is high type when he sees no evidence, $\xi \in [I - H, I]$ is the principal's expected payoff when he sees no evidence.

Using the parameters of the experiment, the problem is:

$$\begin{aligned} \max_{x_0, x_-} \quad & 0.5 \cdot x_0 + 0.5 \cdot (100 - x_-) \\ \text{s.t.} \quad & x_- \geq x_0 - G \cdot \beta \cdot max\{\xi - (100 - x_0), 0\} \end{aligned}$$

In the optimal mechanism: $x_-^C = x_0^C - G \cdot \beta \cdot max\{\xi - (100 - x_0^C), 0\}$, and the principal's maximization problem becomes:

$$\max_{x_0} \quad x_0 + (100 - (x_0 - G \cdot \beta \cdot max\{\xi - (100 - x_0), 0\}))$$

Case I: If $\xi \leq 100 - x_0$
Then, the principal's maximization problem reduces to the model without guilt.

Case II: If $\xi > 100 - x_0$, the principal's maximization problem:

$$\max_{x_0} \quad x_0 + (100 - x_0 + G \cdot \beta \cdot (\xi - 100 + x_0))$$

Fist order condition, $G \cdot \beta$, is strictly increasing in x_0, since $G > 0$ and $\beta \geq 0$. Additionally, $\xi = x_0^C$ in equilibrium. So, optimal rewards are:

$$\begin{aligned} x_0^C = 100 \ , \ x_-^C = 100 \cdot (1 - G \cdot \beta) \quad & \text{if } G \cdot \beta \leq 1 \\ x_0^C = 100 \ , \ x_-^C = 0 \quad & \text{if } G \cdot \beta > 1 \end{aligned}$$

Since we find that the reward for no evidence in the Commitment Treatment, 60.42, is significantly lower than 100 ($p < 0.001$), the simple guilt model does not explain our experimental findings.

Appendix D Instructions

[*Part I Instructions for No-Commitment Treatment*]

Welcome, and thank you for coming today to participate in this experiment. This is an experiment in decision making. You will receive a $7 participation fee. In addition to that, if you follow the instructions and are careful with your decisions, you can earn a significant amount of money, which will be paid to you privately at the end of the session.

The experiment is expected to finish in 120 minutes. The experiment consists of two independent paying parts and a questionnaire. This is the instructions for Part 1.

In this part of the experiment, you will participate in 20 independent decision periods. At the end of the experiment, the computer will randomly select one decision period for payment. The period selected depends solely upon chance and each period is equally likely. Your final earnings in the experiment will be your earnings in the selected period plus your earnings in Part II and the $7 show-up fee.

Your earnings in this experiment will be calculated in Experimental Currency Units (ECUs). At the end of today's session, all your earnings will be converted to US dollars at a rate of 10 ECUs=$1

During the experiment, it is important that you do not talk to any other subjects. Please turn off your cell phones. If you have a question, please raise your hand, and the experimenter will come by to answer your question. Food or drink is not allowed in the lab; if you have food or drink with you, please keep it stored away in your bags. Failure to comply with these instructions means that you will be asked to leave the experiment and all your earnings will be forfeited.

Instructions

You will be informed of your role as the Sender or the Receiver in the first round of the experiment. Your role will be fixed throughout this part of the experiment. In each period, you will be randomly matched with another subject in this room who will be assigned the other role. There will be a new random matching at the beginning of each period, so you will potentially be matched with different people in different rounds. In each round, the Sender will be randomly assigned a type: High or Low. Each type is equally likely. The value of High type to the Receiver is 100, while the value of the Low type is 0.

The Low type Sender has evidence about their type, while the High type sender doesn't. At the beginning of each round, each Sender will choose a message to send to the Receiver they are matched with in that round. The Low type Sender has a choice between telling the truth or pretending that they don't have evidence. The messages available to the Low type Sender are: "My type is low" and "I don't have evidence for my type". The High type Sender, on the other hand, can only send the message "I don't have evidence for my type". The information is summarized in Table 1.

Type (t)	Value (v)	Probability (p)	Available Messages
High	100	50%	"I don't have evidence for my type"
Low	0	50%	"My type is low", "I don't have evidence for my type"

Table 1

After observing the message that the Sender sent, the Receiver will choose a reward between 0 and 100 to send to the Sender.

Payoffs in Each Round

The Sender's payoff in each round will be equal to the reward chosen by the Receiver for the message the Sender sent.

$$\pi_{Sender} = reward$$

The payoff of the Receiver is:

$$\pi_{Receiver} = 100 - |value - reward|$$

where "value" is the value associated with the Sender's type and "reward" is the reward the Receiver chose for the message the Sender sent. The payoff to the Receiver will be 100 minus the distance between the chosen reward and the value of the Sender. So, the Receiver's ideal point for the reward is equal to the value associated with the Sender's type. Notice that the Receiver can choose any number between 0 and 100 as the reward.

At the end of each round, the Sender's type, the message the Sender chose, and the payoffs of the matched Sender and Receiver will be shown to both players. Then, there will be a new random matching and a new round will begin.

Earnings

Once the experiment is finished, the computer will randomly pick 1 round out of the 20 rounds that you completed. The earnings you made on that round will be your earnings in this part of the experiment. Hence, you should make careful decisions in each round because it might be the paying round.

Questions for Checking Understanding

The first screen in the experiment consists of 2 questions that you need to answer correctly to begin the actual experiment. If you answer any of the questions incorrectly, you will receive a pop-up indicating which question you need to correct. Once you answer both questions correctly, you will be directed to the first period of the experiment.

Are there any questions?

Sample Screenshots

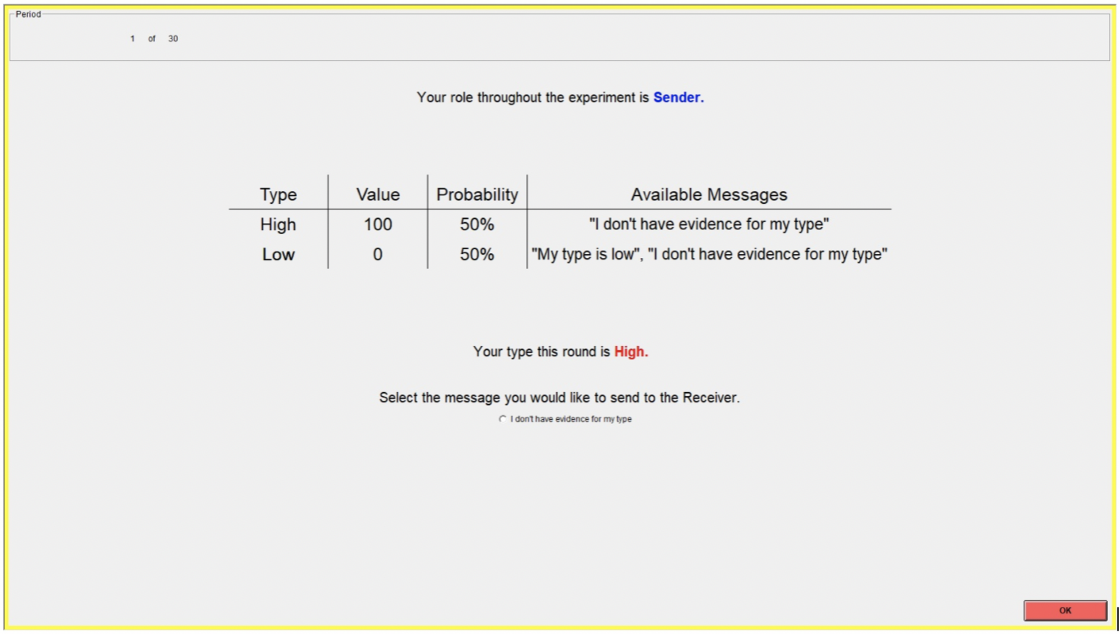


Fig 1: Screen of a High type Sender

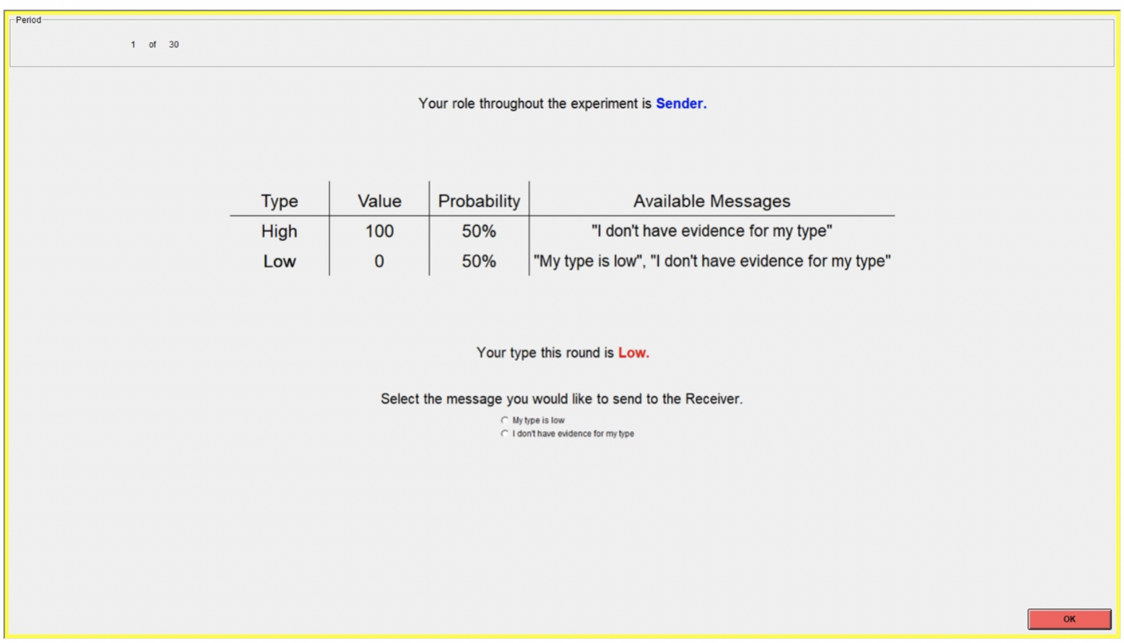


Fig 2: Screen of a Low type Sender

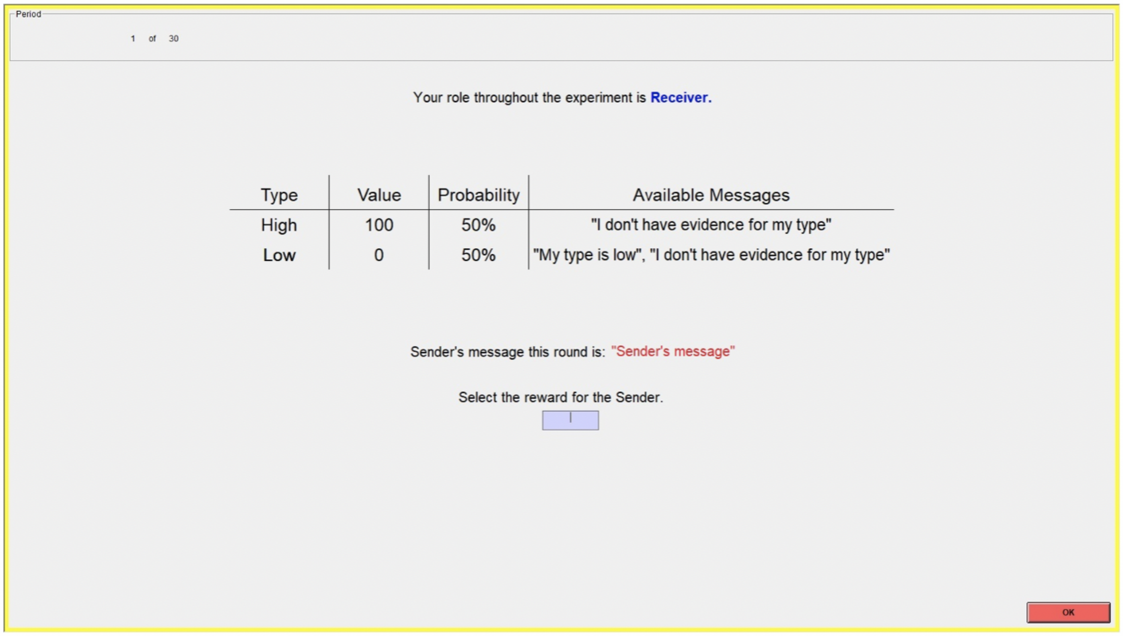
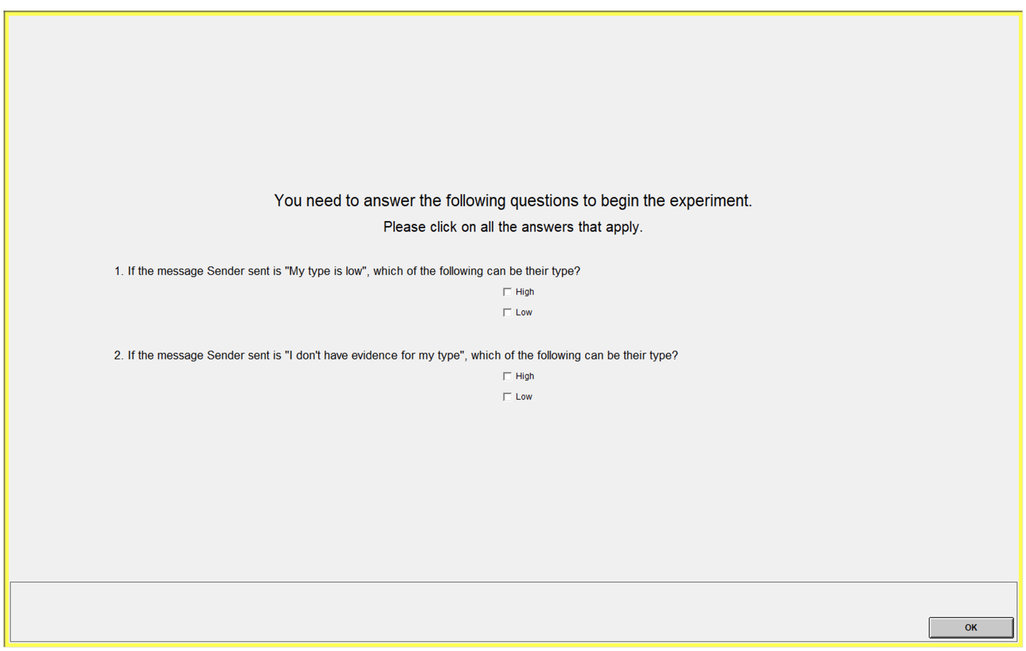


Fig 3: Screen of a Receiver (the message in the real experiment will be either "My type is low" or "I don't have evidence for my type" based on the Sender's choice)

Questions for Checking Understanding



[Part I Instructions for Commitment Treatment]

Welcome, and thank you for coming today to participate in this experiment. This is an experiment in decision making. You will receive a $7 participation fee. In addition to that, if you follow the instructions and are careful with your decisions, you can earn a significant amount of money, which will be paid to you privately at the end of the session.

The experiment is expected to finish in 120 minutes. The experiment consists of two independent paying parts and a questionnaire. This is the instructions for Part 1.

In this part of the experiment, you will participate in 20 independent decision periods. At the end of the experiment, the computer will randomly select one decision period for payment. The period selected depends solely upon chance and each period is equally likely. Your final earnings in the experiment will be your earnings in the selected period plus your earnings in Part II and the $7 show-up fee.

Your earnings in this experiment will be calculated in Experimental Currency Units (ECUs). At the end of today's session, all your earnings will be converted to US dollars at a rate of 10 ECUs=$1

During the experiment, it is important that you do not talk to any other subjects. Please turn off your cell phones. If you have a question, please raise your hand, and the experimenter will come by to answer your question. Food or drink is not allowed in the lab; if you have food or drink with you, please keep it stored away in your bags. Failure to comply with these instructions means that you will be asked to leave the experiment and all your earnings will be forfeited.

Instructions

You will be informed of your role as the Sender or the Receiver in the first round of the experiment. Your role will be fixed throughout this part of the experiment. In each period, you will be randomly matched with another subject in this room who will be assigned the other role. There will be a new random matching at the beginning of each period, so you will potentially be matched with different people in different rounds.

In each round, the Sender will be randomly assigned a type: High or Low. Each type is equally likely. The value of High type to the Receiver is 100, while the value of the Low type is 0.

At the beginning of each round, the Receiver will choose a reward between 0 and 100 for each message that they can possibly receive. After observing the reward scheme, the Sender will choose which message to send.

The Low type Sender has evidence about their type, while the High type sender doesn't. After observing the reward scheme, each Sender will choose a message to send to the Receiver they are matched with in that round. The Low type Sender has a choice between telling the truth or pretending that they don't have evidence. The messages available to the Low type Sender are: "My type is low" and "I don't have evidence for my type". The High type Sender, on the other hand, can only send the message "I don't have evidence for my type". The information is summarized in Table 1.

Type (t)	Value (v)	Probability (p)	Available Messages
High	100	50%	"I don't have evidence for my type"
Low	0	50%	"My type is low", "I don't have evidence for my type"

Table 1

Payoffs in Each Round

The Sender's payoff in each round will be equal to the reward chosen by the Receiver for the message the Sender sent.

$$\pi_{Sender} = reward$$

The payoff of the Receiver is:

$$\pi_{Receiver} = 100 - |value - reward|$$

where "value" is the value associated with the Sender's type and "reward" is the reward the Receiver chose for the message the Sender sent. The payoff to the Receiver will be 100 minus the distance between the chosen reward and the value of the Sender. So, the Receiver's ideal point for the reward is equal to the value associated with the Sender's type. Notice that the Receiver can choose any number between 0 and 100 as the reward. At the end of each round, the Sender's type, the message the Sender chose, and the payoffs of the matched Sender and Receiver will be shown to both players. Then, there will be a new random matching and a new round will begin.

Earnings

Once the experiment is finished, the computer will randomly pick 1 round out of the 20 rounds that you completed. The earnings you made on that round will be your earnings in this part of the experiment. Hence, you should make careful decisions in each round because it might be the paying round.

Questions for Checking Understanding

The first screen in the experiment consists of 2 questions that you need to answer correctly to begin the actual experiment. If you answer any of the questions incorrectly, you will receive a pop-up indicating which question you need to correct. Once you answer both questions correctly, you will be directed to the first period of the experiment.

Are there any questions?

Sample Screenshots

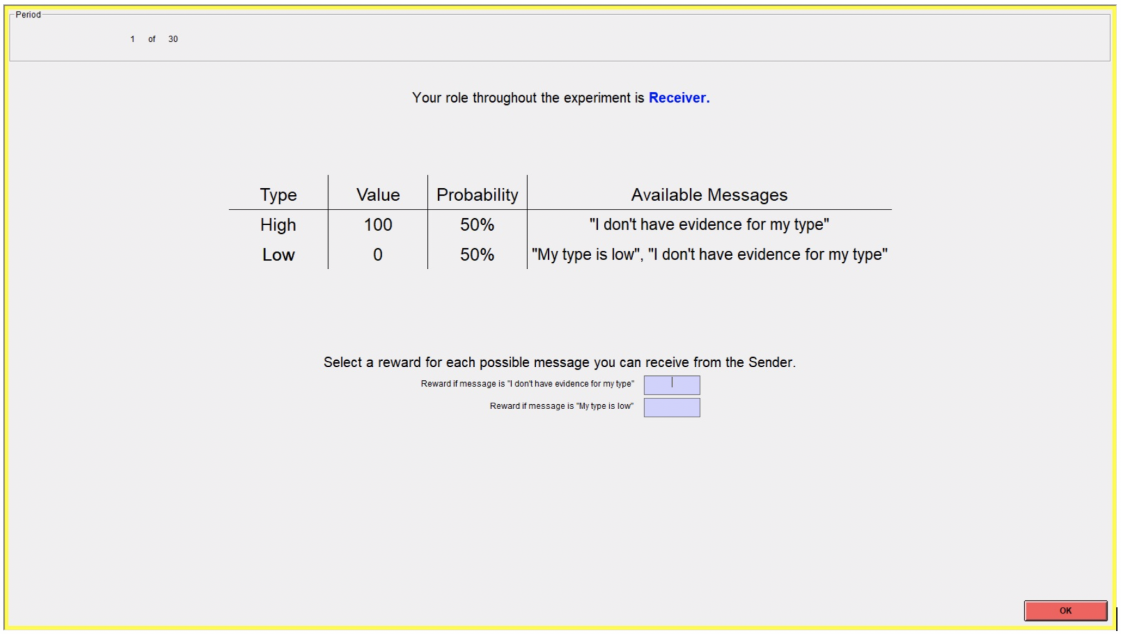


Fig 1: Screen of a Receiver

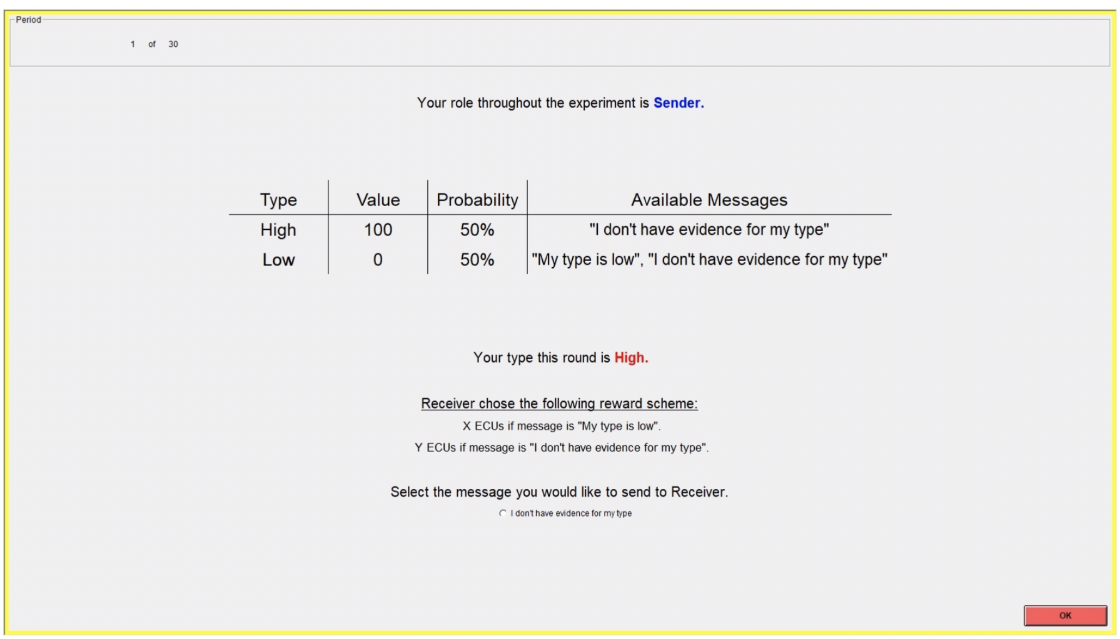


Fig 2: Screen of a High type Sender (the rewards in the real experiment will be numbers between 0 and 100 that the Receiver chose)

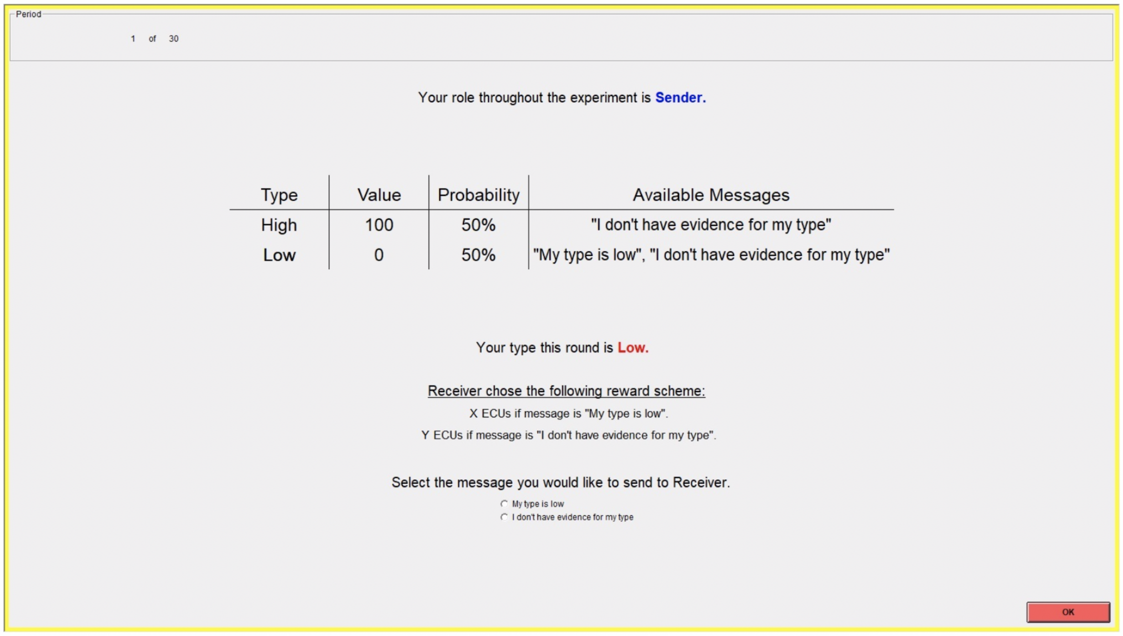
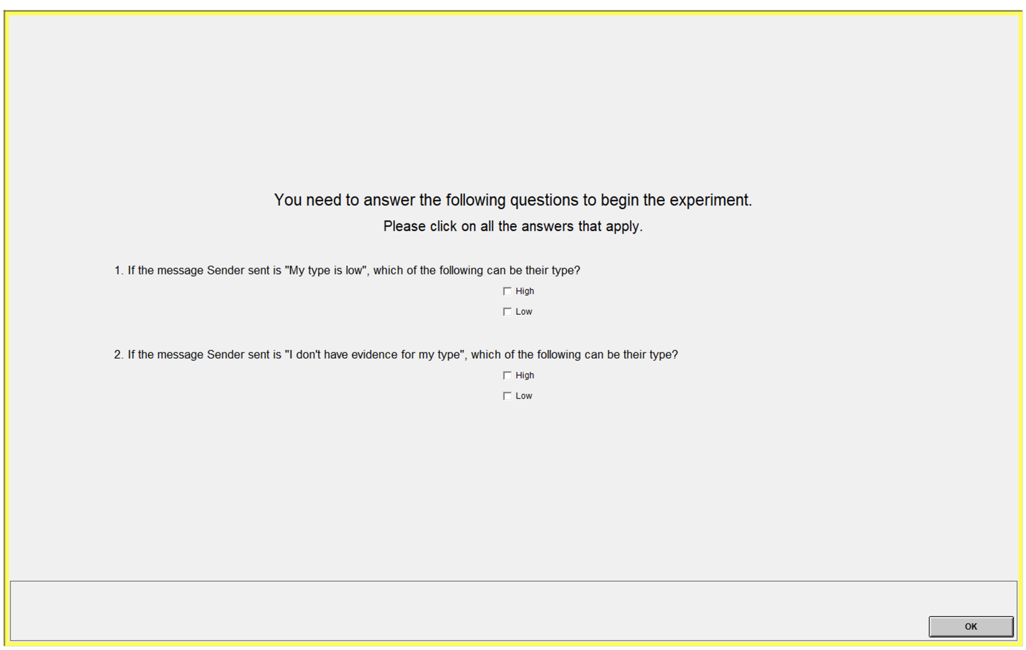


Fig 3: Screen of a Low type Sender (the rewards in the real experiment will be numbers between 0 and 100 that the Receiver chose)

Questions for Checking Understanding



Part II Instructions

This part of the experiment consists of two activities. Your income in Part 2 is the sum of your earnings in both activities. Once you finish an activity you will not be able to go back.

Activity 1

Your earnings in Activity 1 depend on your decisions and also on chance. In this activity, you are asked to choose between Option A and Option B for the following 10 gambles. You will make 10 choices, but only one of these questions will be implemented. After you submit all your choices, the computer will generate two random numbers. The first number will determine which question is implemented, and the second number will determine which outcome is realized. Notice that in each of the questions, you're choosing between two gambles: Option A, which pays 20 ECUs as the high outcome and 16 ECUs as the low outcome and Option B, which pays 38.5 ECUs as the high outcome and 1 ECU as the low outcome. The probability of getting the high outcome is the same for options A and B in each one of the questions and this probability increases as you move down the table. For example, the probability of getting the high outcome is 10% for both options in question 1, it's 20% in question 2, and so on. Please make each one of your choices carefully, as each question is equally likely to be selected for implementation.

Activity 2

In Activity 2, you will be asked a math question that has one correct answer. If your answer is correct, you will earn 10 ECUs in this activity. Otherwise, you will not make any profits from this activity.

Final Earnings

At the end of the experiment, in addition to $7 participation fee, you will receive your earnings based on a randomly selected round in Part 1, a randomly selected question in Activity 1, and your answer from Activity 2.

Are there any questions?

Screen of Activity 1

For each question, please choose between Option A and Option B.

Please choose by clicking on the box of your choice. The box associated with your current choice will turn black.

Question	Option A	Option B
1	☐ 10% chance of 20 ECUs, 90% chance of 16 ECUs	☐ 10% chance of 38.5 ECUs, 90% chance of 1 ECU
2	☐ 20% chance of 20 ECUs, 80% chance of 16 ECUs	☐ 20% chance of 38.5 ECUs, 80% chance of 1 ECU
3	☐ 30% chance of 20 ECUs, 70% chance of 16 ECUs	☐ 30% chance of 38.5 ECUs, 70% chance of 1 ECU
4	☐ 40% chance of 20 ECUs, 60% chance of 16 ECUs	☐ 40% chance of 38.5 ECUs, 60% chance of 1 ECU
5	☐ 50% chance of 20 ECUs, 50% chance of 16 ECUs	☐ 50% chance of 38.5 ECUs, 50% chance of 1 ECU
6	☐ 60% chance of 20 ECUs, 40% chance of 16 ECUs	☐ 60% chance of 38.5 ECUs, 40% chance of 1 ECU
7	☐ 70% chance of 20 ECUs, 30% chance of 16 ECUs	☐ 70% chance of 38.5 ECUs, 30% chance of 1 ECU
8	☐ 80% chance of 20 ECUs, 20% chance of 16 ECUs	☐ 80% chance of 38.5 ECUs, 20% chance of 1 ECU
9	☐ 90% chance of 20 ECUs, 10% chance of 16 ECUs	☐ 90% chance of 38.5 ECUs, 10% chance of 1 ECU
10	☐ 100% chance of 20 ECUs, 0% chance of 16 ECUs	☐ 100% chance of 38.5 ECUs, 0% chance of 1 ECU

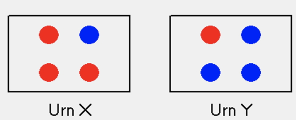
Continue

Screen of Activity 2

Please answer the following question. **If your answer is correct, you will receive 10 ECUs.**

There are two urns containing colored balls. Urn X contains 3 red balls and 1 blue ball. Urn Y contains 1 red ball and 3 blue balls.

One of the two urns is randomly chosen (both urns have probability 50% of being chosen) and then a ball is drawn at random from the chosen urn.



If you learn that the drawn ball is red, what is the probability that it comes from Urn X?:

Please write your answer as a percentage between 0 and 100.

If needed, round your answer to the closest percentage

Continue