Evidence Games: Lying Aversion and Commitment*

Elif B. Osun† Erkut Y. Ozbay‡

April 8, 2022

Abstract

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 his payoff, commitment has no value, as there is a theoretical equivalence of 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 has value. We also theoretically show that our experimental results are explained by accounting 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

In voluntary disclosure literature where there is an informed sender and an uninformed receiver, the case where the receiver moves first and commits to a reward policy corresponds to a mechanism setup and the optimal mechanism has been studied (see e.g. Green and Laffont, 1986; Bull and Watson, 2007; Deneckere and Severinov, 2008); alternatively, the case where the receiver decides on the reward after observing the sender's decision corresponds to a game setup and the equilibrium of the game has been studied (see e.g. Grossman and Hart, 1980; Grossman, 1981; Milgrom, 1981; Dye, 1985). The link between these two settings is an important question. Glazer and Rubinstein (2006) has studied a setting where commitment does not have any value, in other words, the outcome of the optimal mechanism could be obtained in the equilibrium of the game setup. This result has been extended and investigated in other settings (see e.g. Sher, 2011; Ben-Porath et al., 2019).[1] Particularly, Hart et al. (2017) extended this result to *evidence games*. A distinguishing feature of evidence games is that the sender's utility function is increasing in the reward independent of his type, the receiver's utility function depends on the sender's type and satisfies the single-peakedness condition. Furthermore, senders cannot lie about the pieces of evidence that they have, but they can choose not to disclose some pieces of their evidence. In this paper, we experimentally investigate whether commitment has any value in evidence games by testing the equivalence of the optimal mechanism and the game equilibrium outcomes.

Consider an agent (informed sender) who is asked to submit a self-evaluation for an ongoing project, and a principal (uninformed receiver) who decides on the agent's reward. If the agent conducted his part as planned, he does not have any evidence to report at this point. But if he made a mistake which can't be traced back to him unless he discloses it, he may choose to show his mistake or act as if he has no evidence.[2] The agent wants to have a reward as high as possible independent of his evidence but the principal wants to set the reward as close as possible to the agent's value. Our experimental setup mimics this motivating example and asks: Does it matter whether the principal commits to a reward policy and then the agent decides whether to reveal the evidence or not, or the principal moves second after observing the agent's decision?

First, to see the intuition of the theoretical result of why commitment does not have any value in evidence games, assume that with probability of 50%, the agent conducted his work without a mistake (High type) and his value is 100 but with probability of 50%, he

[1] In the closely related Bayesian persuasion literature initiated by Kamenica and Gentzkow (2011), the informed sender has commitment power that can be used to persuade the uninformed receiver. See Fréchette et al. (2019) for an experimental analysis of subject behavior when senders have commitment power.

[2] Alternatively, say a professor has submitted to a journal, and the dean, who decides on professor's salary increase, asks whether he got a desk rejection.

made a mistake (Low type) and his value is 0. In the mechanism setup (where there is commitment), the only way the principal can separate low and high types is to set a higher reward for low evidence (which can only be disclosed by low types) than for no-evidence, which is suboptimal. The optimal mechanism with commitment is that the principal sets a reward of 50 for no evidence and a reward lower than or equal to 50 for low evidence; which implies that the optimal mechanism cannot separate low and high type agents. So, the unique outcome of the optimal mechanism is that both agents still get 50 payoff. In the game setup (where there is no commitment), there is a unique sequential equilibrium where the low type hides his evidence and pretends as if he is a high type. Since neither type discloses any evidence, the principal sets the reward at $50 (= 50\% \times 100 + 50\% \times 0)$. Hence, commitment does not have any value.

In the simple example above, which is based on our experimental setup, there is a unique equilibrium. However, in a general evidence-game setup, the types and the evidences can be quite rich, and there may be multiple equilibria. Hart et al. (2017) identify a refinement (*truth-leaning refinement*) such that in evidence games, the outcomes coincide for the truth-leaning equilibria without commitment and the optimal mechanism with commitment. Since our aim is to test whether commitment has any value in evidence games, the environment in the experiment needs to be simple enough so that it is not affected from subjects' ability to do Bayesian updating or equilibrium selection. Indeed, the equilibrium with and without commitment in the aforementioned example do not require the subjects' ability to do complex Bayesian updating. Furthermore, the equilibrium outcome is unique; hence, the equilibrium selection is not a concern. Therefore, this simple environment is ideal to test the value of commitment in evidence games, and we used it in our experiment.

Despite this simple setup, our experimental results yield that commitment actually makes a difference. Particularly, the principals who choose the reward after observing the agent's action behave in line with equilibrium predictions, while the principals who commit to a reward scheme in advance choose a reward strictly higher than the optimal reward for no evidence. We then theoretically show that such a divergence between outcomes in the presence and absence of commitment is explained by accounting for lying aversion. Finally, in line with a lying aversion model, we show that percent of agents who withhold their evidence varies across these two setups and that when the agents move second, their decision to withhold evidence is affected by reward amounts even when there is no payoff gain from being truthful.

Lying aversion in games with strategic interactions has been well-documented in the literature.[3] Gneezy (2005) is the first one to experimentally measure people's aversion to

[3]Lying aversion has been widely investigated in setups that do not involve strategic interactions (see e.g.

lying in a sender-receiver game. His findings suggest that people sometimes act truthfully even if they have to forgo monetary payoffs to do so. Moreover, he shows that one's own earnings and the harm that lying causes to others are both important factors when deciding to lie. Sánchez-Pagés and Vorsatz (2007), Serra-Garcia et al. (2013), and Ederer and Fehr (2017) are other experimental studies documenting behavior consistent with truth-telling preferences in strategic environments. Along with the experimental papers presenting evidence for lying averse agents, there are also various theoretical papers incorporating aversion to lying in their models. Lacker and Weinberg (1989), Goldman and Slezak (2006), Guttman et al. (2006), Deneckere and Severinov (2017) are some examples studying optimal mechanism design with costly state misrepresentation. Kartik et al. (2007) and Kartik (2009) incorporate costly state misrepresentation to cheap talk setup of Crawford and Sobel (1982) and examine strategic communication with lying costs. Even though there is no study which explicitly accounts for lying averse agents in evidence games, according to the truth-leaning refinement of Hart et al. (2017), a sender prefers disclosing truthfully when the payoffs between disclosing the whole truth and withholding some evidence are equal. This refinement is justified by an infinitesimal increase in agent's utility for telling the whole-truth or equivalently by an infinitesimal decrease in agent's utility for withholding an evidence. We show that if this utility decrease for not revealing the whole truth is different than zero, even if it is small, the outcome equivalence result in evidence games no longer holds.

With this in mind, consider an agent who bears a small but strictly positive cost of lying such that his utility is reduced by this cost if he doesn't reveal the whole truth. When the principal commits to a reward scheme in advance, say in the aforementioned example, the principal sets the reward equal to \$50 if the agent presents no evidence and equal to \$49 if the agent presents evidence for his type. If a low type agent's cost of lying is higher than the additional \$1 he would earn by lying, then the agent would actually disclose his evidence and get the lower payoff. We show that in the presence of agents with a strictly positive cost of lying, the optimal reward set for no evidence when the principal commits to a reward scheme is higher than the reward when the agents do not have a cost of lying. On the other hand, when there is no commitment, the equilibrium outcome remains unchanged even in the presence of lying averse agents, just as we observe in the data.

The remainder of this paper is organized as follows. Section 2 presents the model. Section 3 describes the experimental design and protocol and states the hypotheses under the standard model. Section 4 presents the experimental results. Section 5 introduces a model with lying averse agents and discusses how predictions of the model compare to the experimental findings. Section 6 concludes.

Fischbacher and Föllmi-Heusi, 2013; Gneezy et al., 2018; Abeler et al., 2019).

2 Model

Following the model of Hart et al. (2017), there is an agent denoted by A and a principal denoted by P. The agent has a type $t \in T$ where T is a finite set. The agent's type is chosen according to a probability distribution $q = (q_t)_{t \in T} \in \Delta(T)$ where $q_t > 0$ for all $t \in T$. The agent knows his type, while the principal only knows the probability distribution of types. The agent has a value $v(t)$ associated with his type t.

Each type has access to a set of pieces of evidence $E_t \subseteq E$ where E is the set of all pieces of evidence. A type t agent can choose to reveal the whole truth (i.e. send E_t) or can send the evidence associated with a type who has less evidence (i.e. send $E_m \subseteq E_t$). In other words, agent can choose to withhold information. Denote the set of available messages to type t agent as $L(t) := \{m \in T : E_m \subseteq E_t\}$. In the general setting, the agent chooses a message $m \in L(t) \subseteq T$ to send to the principal, while the principal chooses a reward $x \in \mathbb{R}$ to send to the agent.

The agent's utility for reward x, $U^A(m, x; t) = u(x)$, does not depend on either the type t nor the message m. It is assumed to be a continuously differentiable and strictly increasing function. The principal's utility, $U^P(m, x; t) = w(x, v(t))$, depends on the reward and the value of the agent with type t but not on message m. We assume that the principal's utility is a continuously differentiable, strictly concave and single peaked function maximized at $x = v(t)$. These 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.

3 Experimental Procedures and Hypotheses

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 with 16 subjects in each. There were equal number of subjects in each treatment. Average session lasted about an hour and 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 consists 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 need 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 B.

The first part of the experiment consisted of 20 independent periods. Each subject was assigned the role "agent" or "principal" in the first period, which remained fixed throughout the experiment.[4] 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 the 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. In order 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 instead of letting the agents send $m \in$\{high, low\}. 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"\}.[5] The payoff of the agent was equal to the reward chosen by the principal. The payoff of the principal was $100 - |x - v(t)|$, where x is the reward that the agent receives 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.

There were two treatments which 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 among the messages that were available to his type. Once the agent chose which message to send, the principal observed the message and then chose a reward for the agent.[6] In the Commitment treatment, the principal chose a reward for each possible message that she could receive before she observed 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, identical in both of the treatments, we elicited the subjects' risk preferences and ability to do Bayesian updating using two incentivized activities. In the first activity, we asked the subjects to make choices in a menu of ordered

[4] 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.

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

[6] For studies in which the 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 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 on the strategy method).

lotteries following Holt and Laury (2002) to elicit their risk preference. In the second activity, following Charness and Levin (2005), we asked the subjects a Bayesian updating question which paid 10 ECUs if their answer was correct.

Hypotheses

Based on the model in Section 2, given the material payoffs, the principal's utility is $w(100 - |v(t) - x|)$ and the agent's utility is $u(x)$. Let x_0 and x_- denote the reward set for no evidence and low evidence, respectively.

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 problem of the principal 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 out of 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 equal to 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 $x_- > x_0$ (since incentive compatibility constraint is $u(x_-) \geq u(x_0)$), which is not optimal since expected utility of the principal is decreasing in x_-. So, 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 his evidence, the principal learns his type and gives $x_-^{NC} = 0$. However, if the agent withholds his evidence, the principal cannot learn his type and gives $x_0^{NC} = 50$. So, in the No-Commitment treatment, the low type agent always withholds his 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 his 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 in the No-Commitment and Commitment treatments are equal.*

Additionally, in the Commitment treatment, the agent is the second mover, so a low type agent decides whether to reveal his 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 an agent's decision to withhold evidence. For example, say 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. His decision solely depends on the highest reward rather than the amounts (Hypothesis 5).

Hypothesis 5 *In the Commitment treatment, provided that the reward for low evidence is not higher than the reward for no evidence, increasing or decreasing the reward amounts does not change the percentage of low type agents who withhold their evidence.*

4 Results

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

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
	(593)	(47)
Commitment	60.42	27.05
	(640)	(640)

Note: Number of observations are reported in parentheses.

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.[7] 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 to 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 ($p = 0.026$), falsifying Hypothesis 3. Treatment variable is significant after controlling for period, gender, risk attitudes, and ability to Bayesian update ($p = 0.029$).

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

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 that this difference is significantly higher than 0 ($p < 0.001$). Additionally, the average reward set by principals for low evidence is 27.05 and it is significantly less than 50 but significantly more than 0 ($p < 0.001$ for both). 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.

[7] 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.

Table 2: 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.

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 than one another ($p < 0.001$), 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 his 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 withholding information of 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 reward for no evidence is higher ($p < 0.001$); yet, they are less likely to withhold evidence when the reward for low evidence is higher ($p < 0.001$), falsifying Hypothesis 5. Reward for no evidence and reward for low evidence both have a significant effect on propensity to withhold evidence after controlling for period, gender, risk attitudes, and ability to Bayesian update ($p < 0.001$ for both rewards).[8]

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	(0.000)	(0.000)
Reward for	-0.026***	-0.027***
Low Evidence	(0.000)	(0.000)
Period		0.019**
		(0.035)
Gender		-0.289
		(0.190)
Risk aversion		-0.871***
		(0.000)
Ability to		0.067
Bayesian update		(0.773)
Constant	0.382**	1.155***
	(0.032)	(0.000)
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.

[8] Additionally, we report the results of a probit regression relating withholding information of low type agents to the difference between rewards in Table A.2. The difference between the reward for no evidence and 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.*

5 Discussion

Even though our experimental setup satisfies all the conditions in Hart et al. (2017), we fail to find equivalence of outcomes between the No-Commitment and the Commitment treatments. When the outcomes of an evidence game with and without commitment do not coincide, there may be multiple possible reasons which can explain such a divergence in a more general setting. For example, in presence of multiplicity of equilibria, the difference between the outcomes could be caused by the equilibrium selection. Another explanation could be that subjects have trouble doing Bayesian updating (Friedman, 1998, Charness and Levin, 2005) or that senders strategically act truthfully to exploit receivers' naivete (Jin et al., 2021). However, our experimental design is intentionally simple enough to rule out these alternative explanations. Nevertheless, our experimental results yield that commitment has value: even though the principals set rewards in line with equilibrium predictions when there is no-commitment, they consistently choose higher rewards when they commit on the reward scheme. Furthermore, despite the high reward for no-evidence, agents are still less likely to withhold their information when there is commitment. So, what accounts for our findings?

Our results highlight that when the agent is the second player, low type agents are less likely withhold their evidence even when it is profitable to do so. Hence, there should be an additional motive to payoff maximization. By allowing for lying costs in the framework of Hart et al. (2017), we show that our experimental findings are in line with the predictions of this costly lying model.[9]

Model with Lying Averse Agents

Lying costs have been widely studied in a cheap talk setup since the seminal work of Kartik (2009). Our paper is the first to investigate it in an evidence game framework. In evidence games, the agents bear a cost of lying when they withhold evidence. Although

[9] Alternatively, a low-type agent may feel guilty for disappointing the principal by withholding his evidence (see e.g. Charness and Dufwenberg, 2006; Battigalli and Dufwenberg, 2007). In our simple setup, it is straightforward to show that such a guilt aversion model does not predict our experimental findings accurately.

agents need to lie to withhold their evidence in our experiment,[10] it is possible that the agents can hide the whole truth without lying in other setups. The experimental literature shows that subjects have preferences for truth-telling as well as lying-aversion, even though the strength of preferences might differ in magnitude (see e.g. Sánchez-Pagés and Vorsatz, 2009; Serra-Garcia et al., 2011; Friesen and Gangadharan, 2013; Ertac et al., 2016).

In evidence games, the relevance of lying costs has already been hinted in Hart et al. (2017). As a motivation for truth-leaning refinement, Hart et al. (2017) make a limit argument that the agent's utility increases infinitesimally if and only if when he does not withhold any evidence. Equivalently, the agent's utility decreases infinitesimally if and only if when he withholds an evidence.[11] We call this decrease in utility as cost of lying, and we allow it to be small but positive. Formally, 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:

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

where $u(.)$ is continuously differentiable and strictly increasing.

On the other hand, since the principal does not send a message, her utility is assumed to be as in Section 2, $U^P(m, x; t) = w(x; v(t))$ where $w(.)$ is a continuously differentiable, strictly concave and single peaked function maximized when the reward is equal to the value of the agent, $x = v(t)$.

The agent can be one of two types: High or Low types with values $v(High) = H$ and $v(Low) = L$ such that $H > L \geq 0$. High type occurs with probability q, and Low type occurs with probability $1 - q$ where $q \in (0, 1)$. Let $I > 0$ be the additional compensation to the principal. Denote $\rho = \frac{1-q}{q}$. Recall that for the parameters used in the experiment (I=100, H=100, L=0 and q=0.5), in the absence of lying costs (i.e. $k = 0$), $x_0^{NC} = 50$ and $x_-^{NC} = 0$ in the No-Commitment treatment, and $x_0^C = 50$ and $x_-^C \leq 50$ in the Commitment treatment.

Before characterizing the optimal mechanism of the Commitment setup and the equilibrium of the No-Commitment setup under lying averse agents, let's illustrate that lying averse agents might behave differently than what is predicted in the model without lying

[10] In order to withhold evidence, agents need to lie about not having low evidence (i.e. send "I don't have evidence for my type" even though they do have evidence).

[11] It is not uncommon to use preference for truth-telling and cost of lying interchangeably (see e.g. Abeler et al., 2019).

costs. Consider, for example, two policies that a principal can commit. Say, in both of these policies, the payoff of an agent who withholds his information is 50, and the payoff of an agent who reveals his information is 0 in Policy 1 but it is 49 in Policy 2. An agent, who does not have a lying cost ($k = 0$), withholds his information in both of the policies since $u(50) > u(0)$ 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 his information in Policy 1 since $u(50) - k > u(0)$ but reveals his information in Policy 2 since $u(50) - k < u(49)$. Indeed, under 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.[12] Then, the principal's problem when she sees 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

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

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. That is in line with Result 1(a).

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 his low evidence since $u(x_{-}) < u(x_0)$. But this cannot be optimal 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

[12] 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 complications are not necessary, and our data can be explained by small cost of lying.

type agent may still reveal his low evidence since $u(x_-) > u(x_0) - k$. So, for $k > 0$, the principal's problem is

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

Then in the optimal mechanism

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

For the parameters of the experiment, Equation (2) becomes $w^{'}(x_0^C) = w^{'}(100 - x_-^C)$. It implies that for a concave $w(.)$, $x_-^C + x_0^C = 100$. So, $0 < x_-^C < 50 < x_0^C$ which is in line with Result 1b.

Additionally, in Proposition 1 we 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 a commitment than when there is no commitment (which is in line with Result 2).

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:

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

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), the inequality follows from strict concavity of $w(.)$. But $w^{'}(I - H + x_0^C) < w^{'}(I - H + x_0^{NC})$ 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 his 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 his evidence since $u(0) < u(50) - k$. However, the agent with the same cost of lying may reveal

his 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 $k = 20$. Say, a principal commits to the reward 60 if the agent provides no evidence and the reward 40 if the agent reveals low evidence. A low type agent with $u(x) = x$ and $k = 20$ will not withhold his evidence since $40 = 60 - 20$, but such an agent will withhold his 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 (in line with Result 3).

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. Still, any low type agent with $k > 0$ reveals his 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 (in line with Result 4). To see this, suppose $u(x_-^C) \geq u(x_0^C) - k$, i.e. agent reveals his evidence. If the reward for no evidence decreases to $\hat{x}_-^C$ such that $u(\hat{x}_-^C) < u(x_0^C) - k$, he withholds his evidence; or if the reward for low evidence increases to $\tilde{x}_0^C$ such that $u(x_-^C) < u(\tilde{x}_0^C) - k$, he withholds his evidence. Similarly, suppose $u(x_-^C) < u(x_0^C) - k$, i.e. he withholds his evidence. If the reward for no evidence increases to $\tilde{x}_-^C$ such that $u(\tilde{x}_-^C) \geq u(x_0^C) - k$, he reveals his evidence; or if the reward for low evidence decreases to $\hat{x}_0^C$ such that $u(x_-^C) \geq u(\hat{x}_0^C) - k$, he reveals his evidence.

Welfare Implications

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

Principal's expected utility when there is commitment is

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

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

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

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

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

For example, for the parameters of the experiment, with lying averse agents, in the unique equilibrium without commitment, $x_0^{NC} = 50$ and $x_-^{NC} = 0$. So, when the principal does not commit on a policy, her expected utility is $0.5 * w(100 - (100 - 50) + 0.5 * w(100 - (50 - 0)) = w(50)$. When there is commitment, the optimal mechanism separates the types with the rewards such that $x_0^C > 50 > x_-^C$ and $x_0^C + x_-^C = 100$. So, her 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.

6 Conclusion

The role of communication has been widely investigated as a form of cheap-talk and Bayesian persuasion games (see e.g. Fréchette et al., 2019 for an experimental analysis of subject behavior when senders have commitment power). Our experiment is the first to test whether commitment has any value in evidence games, in which an uninformed receiver who chooses a reward for an informed sender who can reveal pieces of evidence about his type. In a setup without commitment, the receiver moves second and chooses a reward after observing the sender's message; while in a setup with commitment, the receiver commits to a reward scheme in advance and the sender chooses which message to send after observing the rewards.

We design our experiment simple enough (a simpler version of Example 1 of Hart et al., 2017) to leave minimum room for subject mistakes. We experimentally falsify the equivalence of outcomes between these two setups, contrary to the theoretical predictions based on Hart et al. (2017). Our experimental results yield that although subjects behave in line with the equilibrium predictions when there is no commitment, they consistently choose higher rewards when they commit on the reward scheme. Hence, commitment has value. We show that the predictions of a model that includes cost of lying to the standard model are in line with our experimental findings. Additionally, when facing with a lying averse

agent, we theoretically demonstrate that the principal is better off from a committed policy. It may be interesting to experimentally investigate this theoretical prediction. Particularly, if a principal is given an option to decide whether to commit on a policy or not, is she willing to pay to commit on a policy? We leave this question for future work.

References

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

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.

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.

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.

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

Ederer, F. and E. Fehr (2017): "Deception and incentives: How dishonesty undermines effort provision," Tech. rep., Working paper.

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.

Fischbacher, U. and F. Föllmi-Heusi (2013): "Lies in disguise—An experimental study on cheating," *Journal of the European Economic Association*, 11, 525–547.

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.

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.

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.

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.

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

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

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.

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	(0.000)	(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 B 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

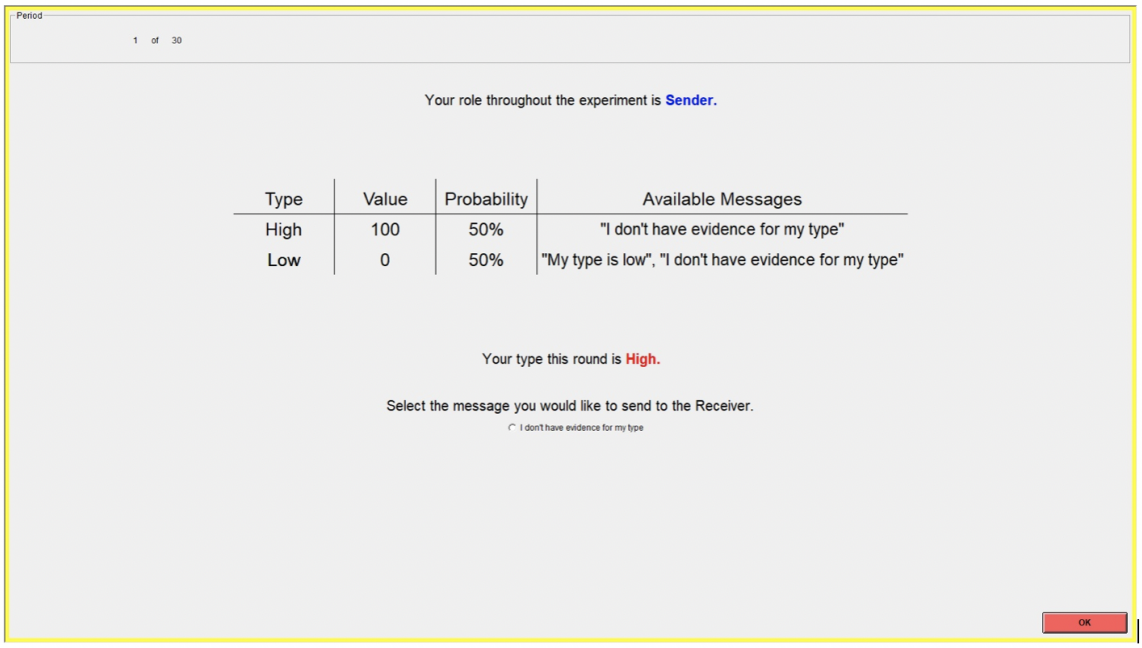


Fig 1: Screen of a High type Sender

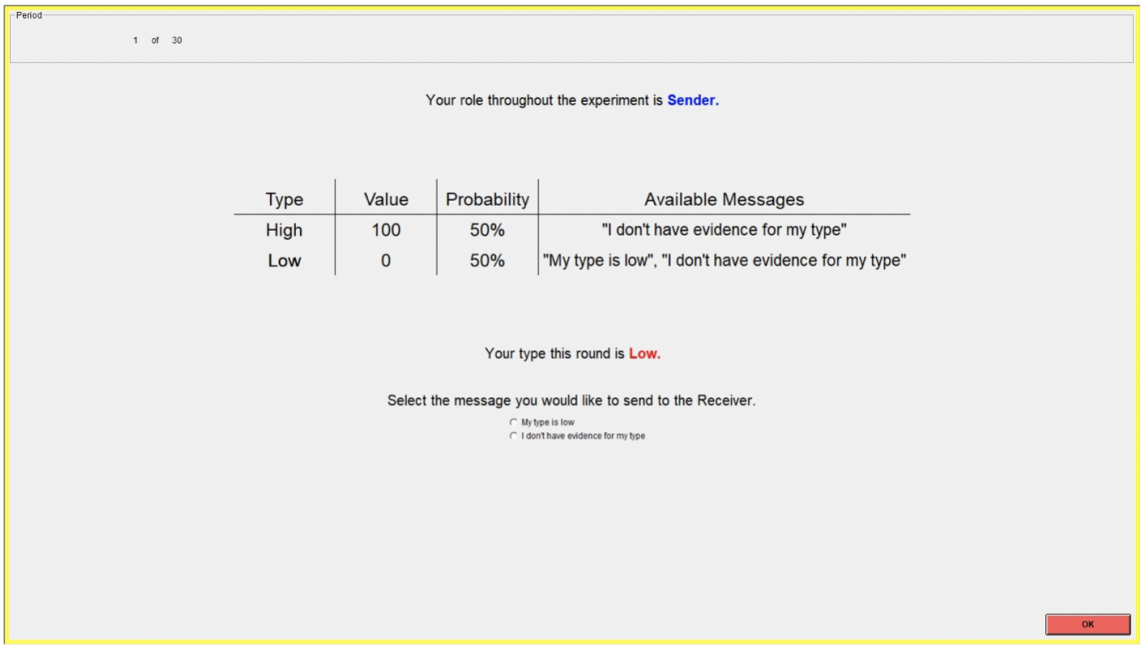


Fig 2: Screen of a Low type Sender

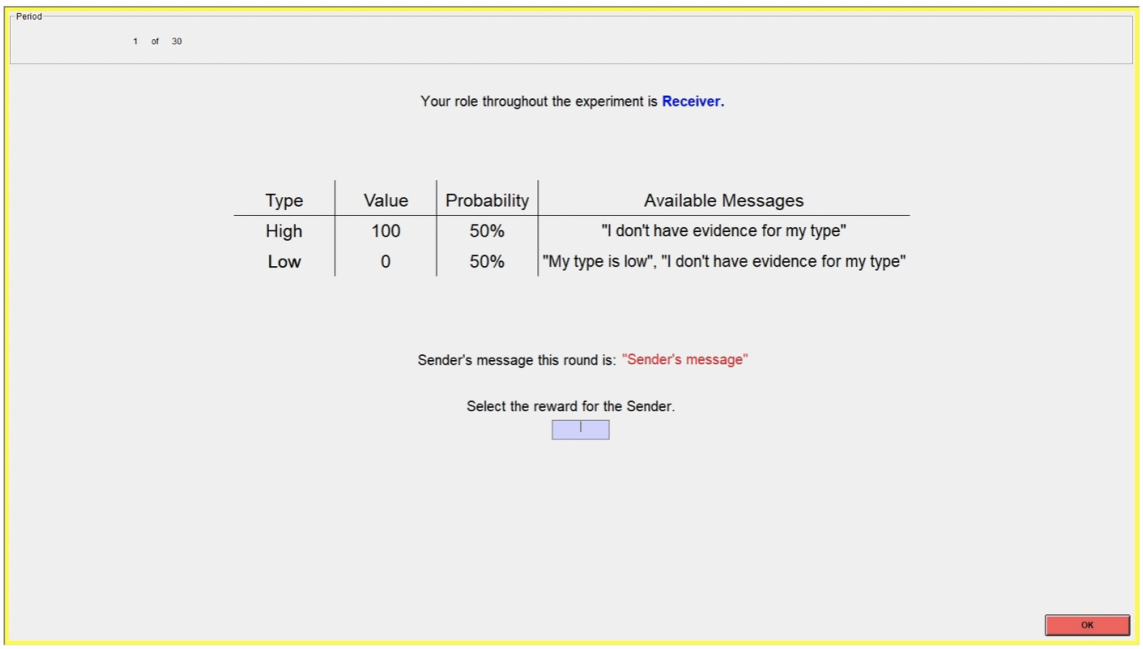
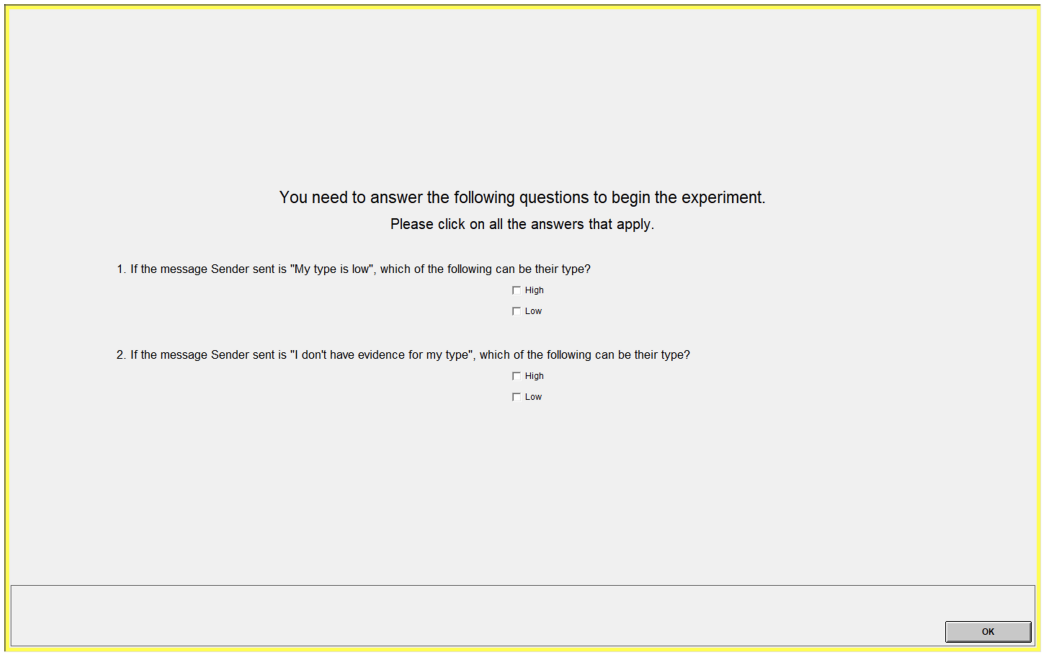


Fig 3: Screen of a Receiver (the message in the real experiment will be either "My type is low" or "I don't have evidence for my type" based on the Sender's choice)

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

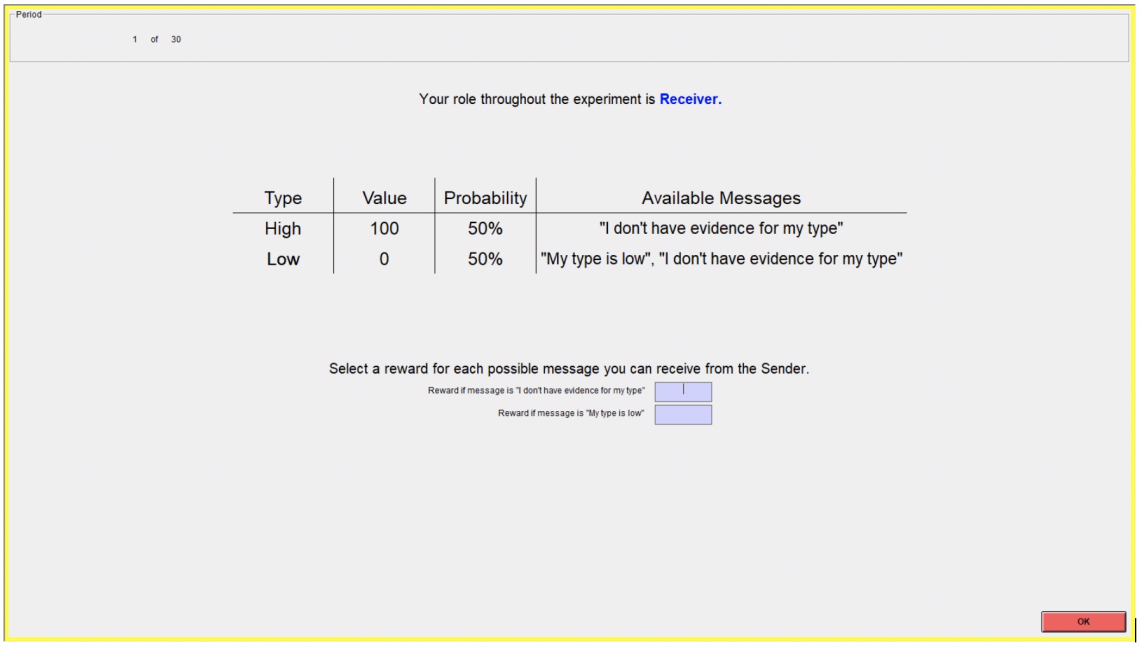


Fig 1: Screen of a Receiver

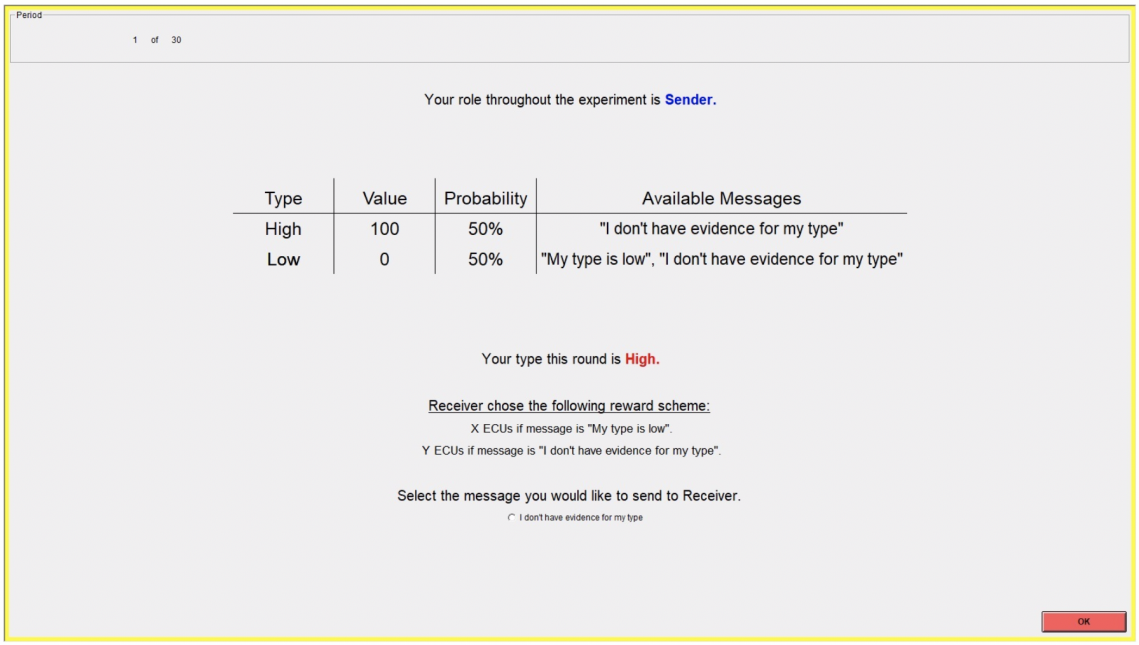


Fig 2: Screen of a High type Sender (the rewards in the real experiment will be numbers between 0 and 100 that the Receiver chose)

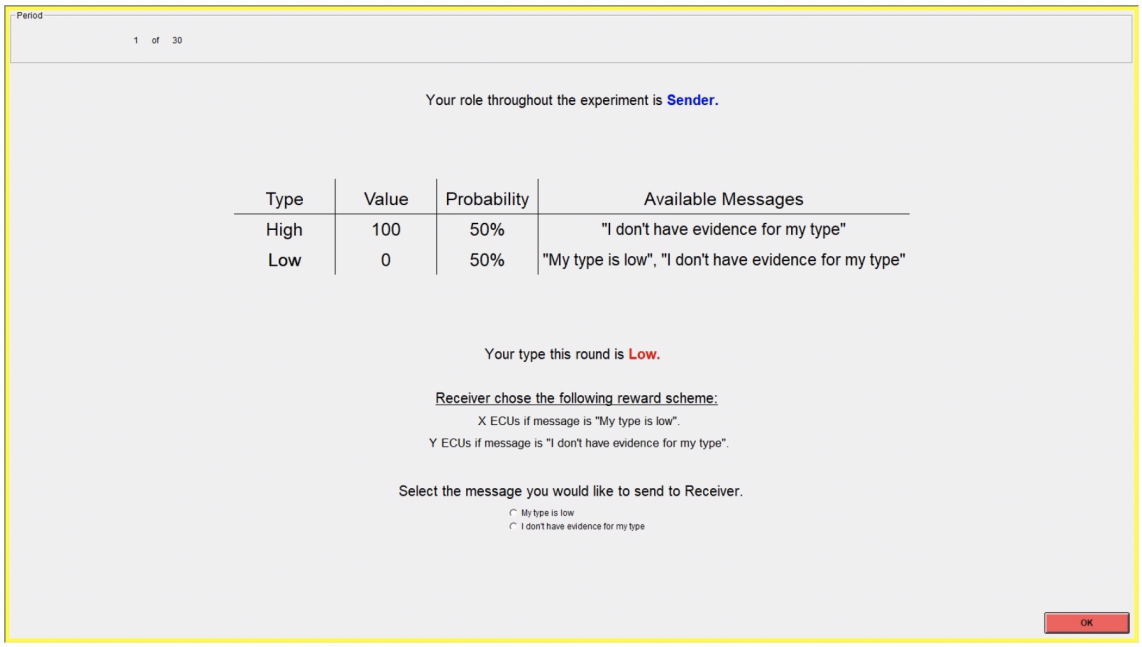
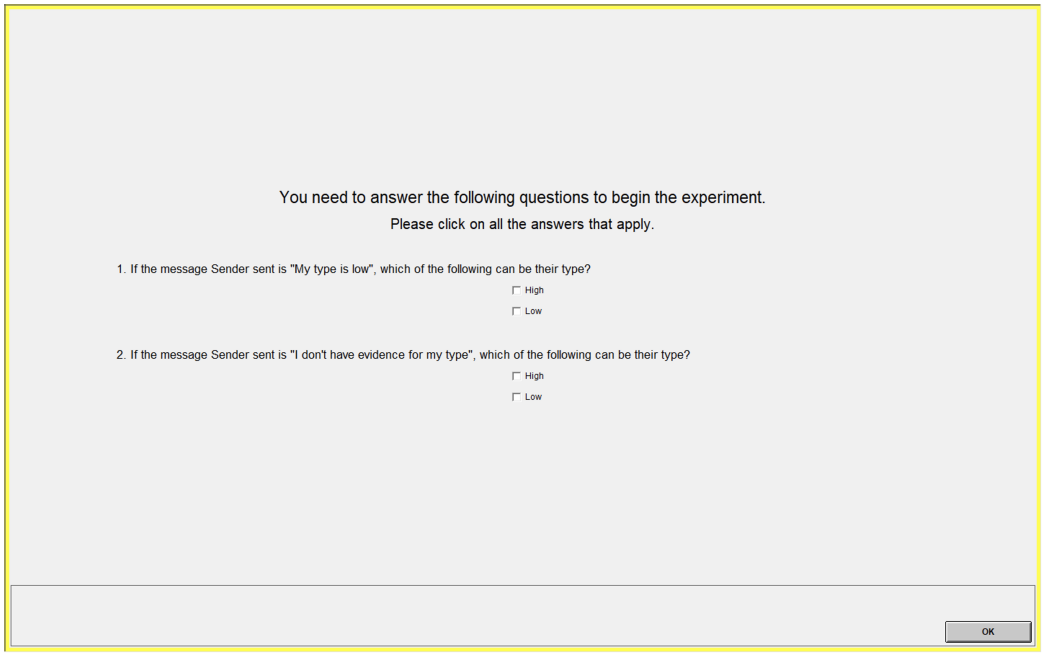


Fig 3: Screen of a Low type Sender (the rewards in the real experiment will be numbers between 0 and 100 that the Receiver chose)

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.

Urn X Urn Y

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