Can Communication Solve the Holdout Problem While Preserving Property Rights: Theory and Experimental Evidence [*]

Paul Schäfer
University of Mannheim

June 29, 2017

Abstract

Holdout problems may prevent political reforms, land-assembly and the assembly of complements. Mechanisms that respect participation constraints and are budget balanced cannot alleviate these problems. This paper explores exploiting lying aversion as a possible solution. Experimental evidence obtained on the related bilateral trade problem suggests that face-to-face communication can be exploited to find mechanisms that in practice extract almost all of the available first-best surplus. A close look at the available data suggests aversion to lying to an "identifiable victim" as a possible reason. This theory is tested experimentally on a simplified version of the holdout problem. The data from the experiment suggests that inducing lying aversion can be part of a solution to the holdout problem.

JEL Classification: C91, D82, D47
Keywords: private information games; lying aversion; communication; political reforms; public goods

[*] I am grateful to seminar participants at the University of Mannheim for helpful comments. This research was supported by the German Research Foundation (DFG) via SFB 884: ”Political Economy of Reforms" and SFB 504: "Concepts of Rationality, Behaviour in Decision-Making Processes and Economic Modelling". Contact: pauschae@mail.uni-mannheim.de

1 Introduction

Failing to compensate losers of a political reform may result in them preventing the reform. However, private information may prevent compensating the losers since each loser has an incentive to exaggerate his costs from the agreement. This problem is called the holdout problem. The holdout problem also occurs with land-assembly, spectrum-assembly, debt restructuring and a monopolist buying complements from multiple suppliers.

In the holdout problem it is generally impossible to find a mechanism that implements the ex-post efficient social choice function. Contrary to a regular market exchange setting the holdout problem does not converge to efficiency if more participants are added (Satterthwaite & Williams (1989) and Kominers & Weyl (2012a), Theorem 1). Therefore, other ways than increasing competition are needed in order to solve the holdout problem.

Under standard assumptions the holdout problem entails a trade-off between efficiency and preserving property rights. However, experimental results obtained for the bilateral trade problem suggest that face-to-face communication may increase efficiency dramatically. Drawing on these results I propose aversion to lying to an *identifiable victim* as a possible channel for this effect. This channel is then tested in an experiment.

I conduct this experiment on a simplified version of the holdout problem. A higher fraction of people chooses an action that reflects their true private information if the reverse would force them to lie to another participant. The effect is observed in a setting with reduced social distance. Here and in the remainder of this paper efficiency is measured by the surplus a mechanism is expected to generate as a fraction of the theoretically available first best surplus[1]. The results indicate a clear way in which framing a revelation mechanism differently can increase efficiency. The presence of lying aversion also implies that in some environments a "broken" revelation mechanism that implements an ex-post efficient outcome, conditional on telling the truth, can be more efficient than the second-best mechanism derived under standard assumptions. The results are potentially applicable to different incarnations of the holdout problem. However, in the experiment the setting of a political reform is chosen.

This paper is organized as follows. Section 2 defines the holdout problem formally and gives an overview of the theoretical and experimental literature concerned with finding solutions to it. Section 3 draws on evidence obtained on the bilateral trade

[1] This measure is conventional in the theoretical and empirical literature. For examples see: Valley et al. (1998); Myerson & Satterthwaite (1983).

problem to motivate why communication and lying aversion could solve the holdout problem. In the process of doing that the specific hypothesis which is tested in this paper is introduced. The experiment to test this hypothesis is presented in Section 4. Section 4 summarizes the theoretical predictions for the experiment. The results of this experiment are presented in Section 6. Section 7 assesses how the results could help in solving the holdout problem. Section 8 outlines the empirical and theoretical work that needs to be done in order to move closer towards practical applications.

2 The Holdout Problem

A modern formulation of the holdout problem as a mechanism design problem can be found in Kominers & Weyl (2012b). They trace the holdout problem back to Cournot (1838), who treats it in the context of a monopolist buying perfect complements as inputs.

The desired area of application for this paper is a government or politician buying consent to a reform from several people or interest groups. Since most of the literature is concerned with the assembly of complements and land-assembly the terminology from these settings is used in order to remain consistent. The buyer buys consent to a political reform from several sellers.

Consider a buyer that attaches a private valuation v to implementing a political reform. The valuation is distributed according to a continuous density on the interval $[\underline{v}, \bar{v}]$, where $\underline{v} > 0$. Sellers are indexed by $i \in I$. The sellers incur costs from the reform that are distributed according to a continuous density on the interval $[\underline{c_i}, \bar{c}_i]$. Values and costs are always positive. The property rights of the sellers have to be respected. That is, their participation constraints must hold. If the setting is a political reform this has the interpretation that losers must be fully compensated. The budget of the mechanism is required to be balanced. In order for the problem to be interesting there have to be some states of the world in which it is not efficient to implement the reform and some where it is:

$$\underline{v} - \Sigma_{i \in I} \bar{c}_i < 0 < \bar{v} - \Sigma_{i \in I} \underline{c_i}$$

Additionally there has to be more than one agent whose costs are actually private, i.e. for whom $\underline{v} < \bar{v}$ or $\underline{c_i} < \bar{c}_i$, respectively. Reformulating the seller's costs as negative values translates the holdout problem to the general public good problem discussed in Schweizer (1998), part 3. In Schweizer's public good problem the distribution of the valuations can vary by individual. The support of the distribution is not restricted. Thus, the holdout problem can be translated to a public good problem, where implementing the reform is a public bad for the sellers and a public good for the buyers. Therefore, the robust impossibility theorem derived by Schweizer applies.

Theorem 1 (Schweizer (1998)). *There is no mechanism for the holdout problem that implements an ex-post efficient social choice function.*

Most of the experimental evidence on the severity and solutions to the holdout problem is conducted on applied problems that are more complex and less abstract than the version formulated here. The experimental results indicate that the holdout problem occurs empirically and that direct negotiation may be a possible way out. Hoffman & Spitzer (1982) conduct multilateral face-to-face bargaining sessions with between 2 and 20 participants. Participants achieve close to full efficiency in perfect information environments. This is no longer the case if private information is introduced.

The name holdout problem comes from sellers rejecting an offer and holding out hoping for a more generous one. Cadigan et al. (2009) finds experimental evidence for the relevance of this behavior. Holding out even occurs if it is not a Nash equilibrium prediction.

Tanaka (2007) studies a land-assembly problem that is a more complicated version of the holdout problem described here. Participants trade parcels of land that are complements, but not perfect complements. In most cases direct negotiation outperforms other formal mechanisms.

The theoretical mechanism design literature concerning the holdout problem encompasses two approaches in order to achieve more efficient mechanisms. The first approach is making the problem easier by considering modifications of the environment. The second approach is weakening either the participation constraints or the requirement of ex-post efficiency. Kominers & Weyl (2012b) consider an example were a buyer wants to buy perfect complements some of which are sold by multiple sellers. In this case the probability of trade approaches one if competition increases. Kominers & Weyl (2012a) equip the buyer with information about the sellers' share of the total costs from conducting a land-assembly. They do not consider the trade-off between respecting participation constraints and efficiency explicitly, but propose a class of mechanisms that satisfy a weaker set of desired properties. Their mechanism converges towards full efficiency as the number of sellers increases. It is at least as efficient as the most efficient bilateral trade mechanism if sellers were able to collude perfectly. Each agent receives compensation according to his costs estimated by the costs of everyone else. Therefore, participation constraints only hold in an approximate sense. Posner & Weyl (2017) argue that the prevalence of the holdout problem necessitates partial common ownership of idiosyncratic goods. Grossman et al. (2010) go the traditional route and maximize efficiency under the constraint that agents participate voluntarily.

In conclusion theoretical and empirical results indicate that the holdout problem occurs. Theorem 1 implies that the only way to find more efficient mechanisms for the holdout problem with standard assumptions about human behavior is to weaken voluntary participation or budget balance. However, there is experimental evidence that direct negotiation may help. The channel through which this happens is unknown. Possible channels through which these happens are discussed in the following section.

3 Motivating Evidence

If there is only one seller the holdout problem becomes the bilateral trade problem. Despite the Myerson & Satterthwaite (1983) impossibility result there is an experimental literature that finds that pre-play communication can lead to close to optimal outcomes (Radner & Schotter (1989), Valley et al. (2002) and McGinn et al. (2003)). Simillar results can be found for the Lemons market (Valley et al. (1998)). In public good environments with free-form written negotiation the efficiency bounds implied by mechanism design theory hold empirically (Palfrey et al. (2015)).

In the following section these results are used in order to motivate an experiment on lying aversion in the holdout problem. Firstly, the set-up of the experiments undertaken by Valley et al. (2002) and McGinn et al. (2003) is explained. Secondly, several possible mechanisms for the effect of pre-play communication on efficiency are discussed. Finally an experiment is proposed that tests if one specific mechanism involving lying aversion is relevant in the holdout problem.

The experiments under discussion use the $1/2$-double auction from Chatterjee & Samuelson (1983). The $1/2$-double auction is a special case of the k-double auction ($k = 1/2$). In the k-double auction the buyer submits a bid (b) and the seller submits an ask (a). If $b \geq a$ the good is traded at price $p = k \cdot b + (1 - k) \cdot a$, where $k \in [0, 1]$. The $1/2$-double auction has a linear equilibrium which implements the second-best allocation (Myerson & Satterthwaite (1983); Chatterjee & Samuelson (1983)). In the experiments reported in Valley et al. (2002) the $1/2$-double auction was preceded by no communication, written communication and face-to-face communication. In all treatments except the face-to-face communication treatment the communication phase was conducted anonymously. In McGinn et al. (2003) the same communication treatments were used. However, anonymity was lifted in all three treatments. In both experiments in the treatment without communication approximately 80% of the available first-best surplus was extracted. In the written communication treatment of Valley et al. (2002) 77% of the available first-best surplus was extracted. However, the face-to-face treatment achieved a rate of surplus extraction of 94%. McGinn et al. (2003) only report a pooled measure of efficiency for the two communication treatments since they conclude based on a logistic regression that the probability of individual trades is not significantly different for those two communication treatments. They report a rate of surplus extraction of 98%.

Figure 1 gives an overview of different channels through which communication could affect efficiency in the reported experiments. Introducing pre-play communication expands the strategy space by adding an element of cheap talk to the game. Converting the game into a cheap talk game can increase the set of non-material outcomes, which in turn can make non-standard preferences more relevant. Communication in itself can also influence preferences. Face-to-face communication also suspends the anonymity of the participants. I follow Bohnet & Frey (1999b) in calling the absence of anonymity *identification.* Identification may affect preferences and adds the additional possibility of sanctions.

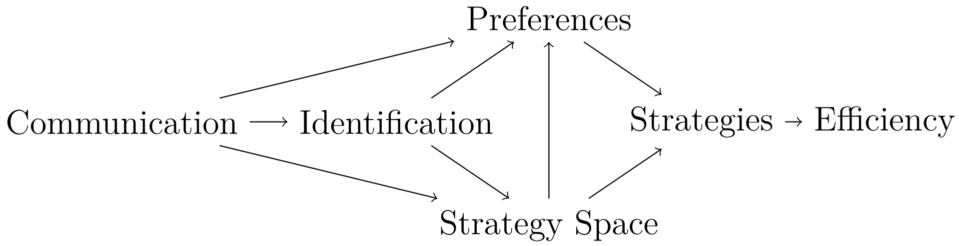


Figure 1: Possible channels for the effect of communication on efficiency.

Mathews & Postlewaite (1988) model cheap talk by an instantaneous exchange of messages before the $1/2$-double auction. This leads to a new class of equilibria. In these equilibria agents use the cheap talk round to announce their bid/ask in a specific k-double auction. People then use their actual bids to implement the outcome of this k-double auction. If the bid in the cheap talk phase is smaller than the ask the bid in the actual auction is also chosen to be smaller than the ask. If the bid in the cheap talk phase is weakly larger than the ask agents submit the resulting price of the k-double auction as a bid. If people actually played these equilibria one would observe them coordinating on one price or failing to trade. This is actually part of what is observed in the experiments of Valley et al. (2002) and McGinn et al. (2003). In 25 out of 50 observations of the communication treatment in McGinn et al. (2003) participants actually coordinate on one price, however it is rarer in Valley et al. (2002). Simply adding cheap talk is also unable to explain the increase in efficiency since the assumptions of Myerson & Satterthwaite (1983) still hold.

There is a large interaction effect between identification and communication. This can be seen by comparing the treatment effects from the McGill et al. and Valley et al. studies shown in figure 2. The lines connect the observations belonging to one experiment. The x-axis denotes the communication treatment. The y-axis shows the estimated probability that a trade is conducted conditional on it being a a Pareto-improvement. This measure is chosen since surplus extraction rates were not reported on a sufficiently disaggregated level in McGinn et al. (2003). The pyramids denote the observations from anonymous treatments, while the dots denote the treatments that were not anonymous. The no-communication treatments have roughly the same probability of trade despite one being anonymous and the other not being anonymous. The face-to-face treatments from both experiments are not anonymous and have probabilities of trade that are very close as well. However, the probabilities of trade in the written treatments differ by approximately 20 percentage points. A way to explain this would be that the treatment effects of written and face-to-face communication do not differ by much and that the difference is caused by an interaction between communication treatments and identification. A similar mechanism seems to be at work in social dilemma games with public information where face-to-face communication leads to a large increase in efficiency, while written communication is less effective in doing that (e.g. Frohlich & Oppenheimer (1998); Ostrom et al. (1992)).

One explanation for an effect of identification on bargaining behavior would be that if participants know each other, they can punish each other outside the laboratory. However, this does not explain the interaction effect. Further, because of the presence of private information it is not verifiable whether a participant behaved badly. There is a series of experiments on dictator games that explores whether anonymity versus identification affects a latent concept called *social distance*. These experiments are reported in Hoffman et al. (1996, 1999); Bohnet & Frey (1999a); Charness & Gneezy (2008). In the following I use the definition of Bohnet & Frey (1999a): “When social distance decreases, the ‘other’ is no longer some unknown individual from some anonymous crowd but becomes an ‘identifiable victim’ [Schelling (1968)]”.[2] A reduction of social distance is theorized to activate

[2] The definition of social distance in Hoffman et al. (1996, 1999); Bohnet & Frey (1999a) and Charness & Gneezy (2008) papers differs from that used in social psychology. Although there seems to be controversy around the definition within those papers, I side with Hoffman et al. (1999) that the points of contention are not important.

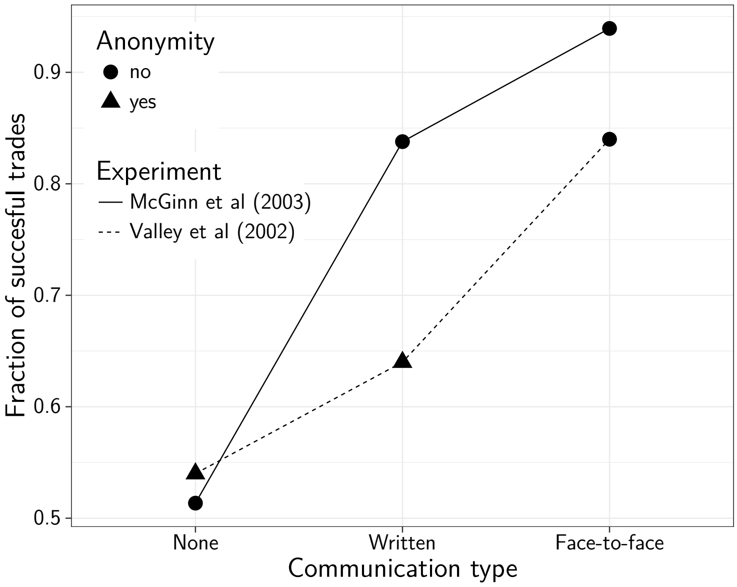


Figure 2: Effects of different communication treatments with and without anonymity (Data from Valley et al. (2002); McGinn et al. (2003)).

social norms (Hoffman et al. (1996)). The series of experiments gives evidence that the proposed mechanism works for fairness norms in the dictator game. However, how these norms actually work is not further specified (Hoffman et al. (1996, 1999); Bohnet & Frey (1999a); Charness & Gneezy (2008)). A further drawback is that the results could also be explained by sanctioning outside the laboratory (e.g. yelling or physical violence). Frohlich & Oppenheimer (1998); Bohnet & Frey (1999b) find evidence for these mechanisms in prisoner's dilemmas. Since the prisoner's dilemma can be interpreted as a public good game, this suggests that the mechanism may be relevant for the holdout problem as well.

Applied to the experimental evidence on communication in the bilateral trade problem social distance theory would say that identification causes a reduction in social distance. Your negotiation partner becomes more of an *identifiable victim*, so you are more willing to adhere to social norms in the contact with him. McGinn et al. (2003) attribute the increase in efficiency to a "process of disclosure and reciprocity". A large fraction of bidders coordinate on telling each other the truth. McGinn et al judge this to be reciprocal behavior, since participants reciprocate being told the truth by telling the truth. However, pay-off based theories of reciprocity like the fairness equilibrium (Rabin (1993); Bierbrauer & Netzer (2016)) are unlikely to be able to explain this concentration on truth-telling since the truth has no special role in these theories. Here the focus lies on disclosure since it seems to be the more promising mechanism. This includes norms-based reciprocity where telling the truth is reciprocated with telling the truth.

Not bidding your value in the 1/2-double auction could already be considered to be a lie. However, the efficiency gains only occur when social distance is reduced and participants have the opportunity to tell each other the truth or lie to each other. This suggests that it is important to lie to the *identifiable victim*. The experimental design presented in the next section tests if in the presence of a reduced social distance having to lie to the 'identifiable victim' can lead to a higher rate of truth-telling in the holdout problem.

The emphasis on disclosure connects this paper to the lying aversion literature. Mazar et al. (2008) theorize that people are averse to lying because they want to maintain an honest self-concept. Meub et al. (2016) find that it is harder for participants to maintain an honest self concept while lying if they have to lie to a participant instead of the experimenter. In this case participants also tell the truth more frequently. This evidence supports the *identifiable victim* mechanism. Gneezy (2005) theorizes that lying aversion is due to guilt aversion. That is, people do not lie because they feel guilty if they cause other people to have wrong beliefs about their pay-offs. Guilt aversion does not predict aversion to lying in the 1/2-double auction since beliefs about pay-offs are correct as long as players communicate the same information to the other player and to the mechanism. In this view lying is independent of consequences and people are also averse to white lies. White lies are lies that are good for the person that is being lied to. Mazar et al. (2008) and Fischbacher & Föllmi-Heusi (2013) suggest that the magnitude of a lie matters. People are less averse to lying by a little than to lying by much.

The closer seller's asks and buyer's bids are to their true costs or values the more efficient the outcome of the 1/2-double auction becomes. In the linear equilibrium of the 1/2-double auction seller's ask and the buyer's bid are linear functions of the seller's cost and the buyer's value. If the participants bid their true value or cost, the auction implements an ex-post efficient social choice function that splits the gains from trade equally. The 1/2-double auction is a broken revelation mechanism. In the sense that people can tell the truth and it implements a social choice function, but it is not an equilibrium. But it is also a constraint optimal mechanism. In conclusion being truthful is attractive for several reasons and it increases efficiency. The auction is the optimal mechanism if none of these reasons hold and agents behave according to classical theory. Since it is hard to find mechanisms with those properties it is important to find out through which specific channels communication increases efficiency.

Face-to-face communication leads to close to full extraction of surplus. Aversion to lying to an *identifiable victim* is a theory that explains this result and is consistent with the experimental data. Due to the complexity of the $1/2$-double auction with cheap talk it is not possible to confirm this mechanism using existing data. The results reported in this paper provide evidence that this mechanism is relevant in the holdout problem.

4 Experimental Design

In order to test the hypothesis that having to lie to an 'identifiable victim' can lead to a higher rate of truth-telling in the holdout problem, it is necessary to find a simple version of the holdout problem and a specific mechanism. The experimental design and the underlying game were chosen according to three criteria. Firstly, the game used in the environment should contain all essential features of the holdout problem. Secondly, the experiment should give theory its best shot in the sense of Plott (1982). That is I want to maximize the chance of finding the effect if it is there. Since the existing evidence on better than predicted mechanisms through communication was conducted in complicated environments there is a lot of uncertainty about the underlying mechanism. In this context it is helpful to be able to move on if the proposed theory failed its "best shot". Finally, if a higher rate of truth-telling actually occurs it should be clearly attributable to the proposed theory.

4.1 A Simplified Version of the Holdout Problem

The experiment uses the following simplified version of the holdout problem. There are two sellers that can block a reform. These sellers have private costs from the reform. It is publicly known that the reform generates a benefit of v. The seller's costs c_i are drawn from the set $\{0, \bar{c}\}$, where $\bar{c} \in (0.5v, v]$. Seller i has positive cost $\bar{c}$ with probability $p \in (0, 1)$.

Since the problem is binary participants can either lie or tell the truth. Concentrating only on the sellers simplifies the problem in two ways. Firstly, all observations can be pooled. This assures maximal power given a fixed number of participants. Secondly, there is only one person that can be the *identifiable victim.* The holdout problem is a public good problem among the sellers nested into a bilateral trade problem. The experiment focuses on the public good problem.

Besides the general structure a defining feature of the holdout problem is than an impossibility theorem holds. Since costs in the experimental game follow a discrete distribution the assumptions of Theorem 1 no longer hold. Therefore, it is necessary to choose the parameters of the simplified holdout problem so that an impossibility theorem holds. It is ex-post efficient to implement the reform if at most one seller has the high costs. The goal is to implement a social choice function that fulfills the participation constraints of the two sellers, is budget balanced and ex-post efficient. The social choice function f consists of a transfer rule for each of the two agents (t_i) and an outcome function q. The transfer that player i receives is denoted by $t_i(c_i, c_{-i})$. The outcome function maps the vector of costs into the decision if the reform should be implemented or not:

$$q(c_i, c_{-i}) = \begin{cases} 1 \text{ if: } c_i + c_{-i} < v \\ 0 \text{ if: } c_i + c_{-i} \geq v \end{cases}$$

Consider a revelation mechanism that implements the ex-post efficient outcome rule. If the ex-post efficient outcome is implementable under the desired conditions it satisfies the following incentive compatibility constraints for all $i \in \{1, 2\}$:

$$p \cdot t_i(0, \bar{c}) + (1 - p) \cdot t_i(0, 0) \geq p \cdot t_i(\bar{c}, \bar{c}) + (1 - p) \cdot t_i(\bar{c}, 0), \tag{IC1}$$

$$p \cdot t_i(\bar{c}, \bar{c}) + (1-p) \cdot (-\bar{c} + t_i(\bar{c}, 0)) \geq p \cdot (-\bar{c} + t_i(0, \bar{c})) + (1-p) \cdot (-\bar{c} + t_i(0, 0)). \tag{IC2}$$

The Participation constraints are given by

$$p \cdot t_i(0, \bar{c}) + (1 - p) \cdot t_i(0, 0) \geq 0, \tag{PC1}$$

$$p \cdot t_i(\bar{c}, \bar{c}) + (1 - p) \cdot (-\bar{c} + t_i(\bar{c}, 0)) \geq 0 \tag{PC2}$$

and the budget balance constraint is

$$t_1(c_1, c_2) + t_2(c_2, c_1) \leq q(c_1, c_2) \cdot v \qquad \forall (c_i, c_{-i}) \in \{0, \bar{c}\} \times \{0, \bar{c}\}. \tag{BC}$$

Checking these constraints results in the following theorem, the proof of which is given in the appendix.

Theorem 2 (Impossibility)**.** *If* $\frac{1+p}{2} v < \bar{c}$ *any social choice function that satisfies BC, PC1 and PC2 also satisfies IC2 and violates IC1.*

The parameters for the experiments are chosen in order to satisfy this constraint.

4.2 The Experimental Game

To keep the mechanism simple it should be deterministic. To give theory its best shot requires the possibility of a large increase in truth-telling due to the treatment. In order to leave room for a large increase in truth-telling the Nash equilibrium of the mechanism used in the experiment should predict a small probability of truth-telling. An increase in truth-telling should lead to an increase in efficiency. A second desired property is that switching to telling the truth should also not be overly expensive.

The experimental game is a "broken" revelation mechanism of the following social choice function:

$$f(c_1, c_2) = (q(c_1, c_2), t(c_1, c_2), t(c_2, c_1)), \text{ where:}$$
$$q(c_1, c_2) = \begin{cases} 1 & \text{if: } c_1 + c_2 < v \\ 0 & \text{if: } c_1 + c_2 > v \end{cases}$$

and $t(0, \bar{c}) = v - \bar{c}$, $t(\bar{c}, \bar{c}) = 0$, $t(\bar{c}, 0) = \bar{c}$, $t(0, 0) = v/2$. Let the strategy of player i in the revelation mechanism which is generated by the above social choice function be denoted by $s_i : \{0, \bar{c}\} \mapsto \{0, \bar{c}\}$. Then Theorem 3 describes the pure strategy equilibria of that mechanism. The proof is given in the appendix.

Theorem 3. *The strategy profiles (s_t, s_l) and (s_l, s_t) are the unique pure strategy equilibria of the revelation mechanism induced by social choice function f if: $s_t(c_i) = c_i$ and $s_l(c_i) = \bar{c}$. Further, the participation constraints hold.*

The mechanism is not incentive compatible on purpose since a low predicted probability of truth-telling is needed. If truth-telling were an equilibrium the mechanism would be ex-post efficient. This ensures that an increase in truth-telling is also a move towards efficiency. The transfers of the mechanism are structured in a way that if the good is provided and a participant claims high costs she is compensated by exactly her costs. The remaining surplus is distributed to the other participant. So the mechanism is budget balance by construction. By Theorem 2 incentive compatibility for high cost types follows at no additional costs. Hence all the remaining surplus can be used to make telling the truth less costly for the low-cost types.

Since this equilibrium is not symmetric and in the experiment there is no plausible way for the two players to coordinate the symmetric mixed strategy Bayes Nash equilibrium probably yields a better prediction of observed behavior.

Theorem 4. *Let $s_t(c_i) = c_i$ and $s_l(c_i) = \bar{c}$. There is a unique symmetric mixed strategy Bayes Nash equilibrium of the revelation mechanism induced by social choice function f in which each player plays s_t with probability $\rho = \frac{2}{1-p}\left(1 - \frac{\bar{c}}{v}\right)$ and s_l with the converse probability. If participants incur a cost c_l from playing strategy s_l they play s_l with probability $\rho(c_l) = \frac{2}{1-p}\left(1 - \frac{\bar{c}-c_l}{v}\right)$*

Using the results from above the parameters for the experiment can be determined. In the following discussion v is assumed to be fixed. This can be done since all equations only depend on the relative sizes of v and $\bar{c}$. The absolute size of monetary incentives can be adjusted, using the exchange rate between real and experimental currency.

In order to allow for a potentially large treatment effect it is necessary to have a high fraction of liars in the mixed strategy Nash equilibrium. On the other hand if ρ falls by enough coordinating on the asymmetric pure strategy Bayes Nash equilibrium becomes more desirable. Therefore, it must be ensured that the communication channels provided by the experiment cannot be used in order to coordinate on an equilibrium. There are two ways to reduce ρ: reducing p or increasing $\bar{c}$. The parameters have to satisfy the conditions given in Theorem 2.

It would also be desirable that the mechanism has a good shot to beat the second best mechanism. The impossibility result critically depends on conducting the reform being optimal if only one of the players has high costs. If the surplus in this case $(v - \bar{c})$ is low enough a revelation mechanism that only implements the reform if both players have low costs becomes close to optimal. This also happens if p falls, making this case very improbable.

In the experiment participants in the lying treatment are able to observe the other player's announcement of their costs. This causes a further problem with a low probability of having high costs. Since in this case it is very likely that your partner is lying if he announces high costs a given player can judge relatively well if he is being lied to. This is something that is not always given when deciding about reforms in the real world.

In conclusion there is a trade-off between keeping the experiment relevant to applications and having enough power to determine the treatment effect. Since this paper is the first try for determining if the proposed approach is practical at all this paper errs on the side of high statistical power.

With respect to the impossibility theorem the surplus in the case where only one player has high cost, $v - \bar{c}$, and the probability with which a player has low costs, p, are complements. Both determine the importance of this case. However, having a low p also makes it more probable that someone who says that she has low costs is lying. I sacrifice some surplus in order to get a higher p since I do not want it to be obvious that claiming high costs is a lie.

Keeping all these considerations in mind the parameters are chosen to be $v = 100$, $\bar{c} = 80$ and $p = 0.2$.

4.3 Experimental Procedures

The experiment was programed in oTree (Chen et al. (2016)). In order to test the hypothesis two versions of the experimental game are needed. A control game where participants simply choose a strategy without any connection to lying and a treatment game where the participants have to lie to each other. The treatment is forcing the participants to lie to each other about their private information if they want to misrepresent it towards the mechanism. For the purpose of this experiment lying is communicating someone something that is not true. The screens containing the main parts of the instructions are shwon in Appendix E. The source-code for the experiment is available on-line. [3].

In order to avoid the results being confounded by converting the game into a cheap talk game the treatment game should be strategically equivalent to the control game. Table 1 shows the material pay-offs of the mechanism used in the experiment depending on the actions taken by the two players. The rows show the actions player 1 takes. The column labels show the actions of player 2. For now the actions are labeled 80 and 0. In the experiment they work slightly different depending on the treatment. However, the underlying material pay-offs and the information structure remain unchanged.

Table 1: Pay-off matrix of the game used in the experiment.

		Player 2	
		0	80
Player 1	0	$50 - c_1$, $50 - c_2$	$20 - c_1$, $80 - c_2$
	80	$80 - c_1$, $20 - c_2$	0, 0

In the beginning of the experiment the general structure of the experiment is explained to all participants. Participants are not told the rules of the experimental game yet. In the next task the participants are able to communicate with each other. In order to avoid them using this communication phase to coordinate or exchange information they have to remain oblivious of their tasks. In order to decrease social distance participants are given the opportunity to introduce themselves by writing a message (introduction message) to the other players that they will be grouped with. They are instructed not to send identifying information (whether people could be identified is checked later by asking the recipients of these messages). Anonymity should be preserved in order to prevent sanctioning outside the lab.

There is no control group for the introduction message. This means that the treatment is tested only in the presence of a decreased social distance to the other

[3] The source-code can be found on GitHub (https://github.com/pauschae/experiment_holdouts).

player. Thus, the theorized interaction between decreased social distance and having to lie is not tested. This is not done here for the reason that I want to test first whether lying aversion is relevant in a setting with a large predicted effect (low social distance) instead of sacrificing statistical power or resources to test details about an effect that may not be there. This strategy corresponds to McGinn et al. (2003) testing pre-play communication only in the presence of identification.

The setting including the distribution of costs is explained to all participants. The problem is framed explicitly as deciding on a political reform. The explicit framing was chosen to make the problem easier to explain.

Following recommendations by Normand (2016) a within-subject design is chosen in order to increase statistical power. The treatment requires sending information to the other player. This is ideally done directly after the decisions is made in order to preserve the connection the decision and its consequences. This makes it necessary to always run the control game first in order to avoid the result being confounded by learning effects. A disadvantage of this procedure is that treatments cannot be counterbalanced in order to avoid confoundment, by order effects. In order to avoid order effects despite this limitation participants are not given any feedback between the treatment and control game.

Both games start with an explanation of the rules. After that understanding is checked by three questions. Since the games differ control questions were asked twice. Asking the questions in the control game decreases the risk that participants misunderstand the rules and think that the games are not strategically equivalent. The questions were equivalent for both games. If those questions are answered wrongly the participants get additional explanations. After each game probabilistic beliefs about the other player's action when having low costs are elicited using a quadratic scoring rule. The quadratic scoring rule is presented in a format analogous to Vanberg (2015). Participants are asked to judge whether the other player is likely to announce high costs given that he has low costs. They can enter their judgments using a slider. They can choose probabilities in steps of 0.1. While doing that they see the pay-offs for both cases calculated by the quadratic scoring rule. The pay-offs change dynamically with the probabilities they announce. In order to avoid hedging final pay-offs are chosen randomly from either treatment or control game. Rewards for the probability judgments are taken from the other game.

Before the control game participants get randomly matched to a partner. Then the introduction message of the other player is shown. After that participants get told their costs. Costs stay constant during both games in order to facilitate within-subject comparisons. People are asked if they want to demand a compensation payment. Demanding compensation corresponds to the action of announcing 80 in the experimental game summarized in Table 1. Not demanding a compensation payment corresponds to announcing 0.

For the treatment game participants get matched to another partner. This is done in order to avoid interactions between treatment and control game. Then the introduction message of the other player is shown and participants are reminded of their costs. People have the choice between one of two messages (corresponding actions in brackets):

- "I have no costs from the reform" (0)
- "The reform costs me 80 Taler [the experimental currency]." (80)

People are informed that their chosen message will be send to their partner after both games. Additionally the chosen message triggers the corresponding action in the game summarized in table 1. Participants know that.

After the two rounds participants have to answer demographic questions and get paid a fixed fee of 60 plus their pay-off from one round chosen at random plus their pay-off from the belief elicitation in the round which was not chosen.

5 Predictions

In the control game the labels of the actions do not involve lying and no information is sent, therefore lying costs should play no role in the control game. Since lying costs should play no role in the control group the results should be predictable by a Nash equilibrium. The absence of opportunities for coordination makes the mixed strategy equilibrium the best prediction. Introducing an aversion to lying should lead to less lying. The simplest way to model that is a mixed Nash equilibrium with homogeneous costs for lying. A model with private heterogeneous lying costs would also be possible, but it is hard to compare to the prediction for the control.

5.1 Control

Recall that the parameters for the experimental game are chosen to be $v = 100$, $\bar{c} = 80$ and $p = 0.2$. Since the control game does not involve lying and participants have no way to coordinate a good prediction for the control group seems to be the mixed strategy Nash equilibrium. This leads to a probability of playing s_t (always telling the truth) of $\frac{2}{1-p}(1 - \frac{\bar{c}}{v}) = \frac{2}{1-0.2}(1 - \frac{90}{100}) = 0.25$.

If players play the mixed equilibrium the reform is implemented if at least one player announces costs of 0. An individual player announces costs of 0 with probability $0.8 \cdot 0.25 = 0.2$. Both players announce costs of 0 with probability $0.8^2 \cdot 0.25^2 = 0.04$. At least one player announces cost 0 with probability $0.4 - 0.04 = 0.36$. In the pure strategy equilibrium the reforms is implemented if the player that plays s_t has the low valuation. That happens with probability 0.8.

If only one player has the low valuation the first best surplus is 20. This situation occurs with $p = 2 \cdot 0.8 \cdot 0.2 = 0.32$. In the pure strategy equilibrium in half of these cases the reform is implemented. In the mixed strategy equilibrium the reform is implemented in this case if the player with the low costs plays s_t, this occurs with probability 0.25.

If both players have costs of 0 the first-best surplus is 100. This occurs with probability $0.8^2 = 0.64$. In the pure strategy Bayes Nash equilibrium the reform is always implemented in this case. In the mixed strategy equilibrium the reform is implemented if at least one player plays s_t. This happens with probability $1 - 0.75^2 = 0.4375$.

Taking this together the expected available surplus is given by: $W_{fb} = 0.32 \cdot 20 + 0.64 \cdot 100 = 70.4$. The expected realized surplus in the pure strategy equilibrium is: $W_{ps} = 0.32 * 0.5 * 20 + 0.64 * 100 = 67.2$. The expected surplus in the mixed strategy equilibrium is $W_{ms} = 0.32 * 0.25 * 20 + 0.64 * 0.4375 * 100 = 29.6$. The pure strategy equilibrium extracts $\approx 95\%$ of available surplus, whereas the mixed equilibrium only extracts $\approx 29\%$.

So for the pure strategy equilibrium and the chosen parameters the impossibility result is not actually that strong. This is slightly problematic since it encourages coordination. But there seems to be no good way around it.

The binary setting necessarily leads to the asymmetric equilibrium, which makes the mixed equilibrium more credible. In turn the mixed equilibrium needs the case with the mixed valuations to be relatively inconsequential in order to produce a large fraction of liars. The final consequence is a large fraction of extracted surplus in the pure strategy equilibrium.

If players are able to coordinate on one of the two pure strategy equilibria one should expect half of the players with costs of 0 to report their costs truthfully. If the mixed strategy equilibrium is played one should expect approximately 25% of players to do so.

5.2 Treatment

The treatment involves lying to an 'identifiable victim'. This should make the participants more averse to misrepresent their private information. The simplest way to model this is by a homogeneous cost of lying. Announcing costs of 0 when having costs of 80 remains dominated since lying costs make announcing costs of 0 even worse.

If lying costs remain small enough the best responses for the pure strategy equilibrium remain unchanged. The best response to s_l is s_t, even without lying costs. Lying costs do not change that. If lying costs become large the incentive to tell the truth dominates the material incentives and s_t becomes the new best response to s_t. This yields the strategy profile (s_t, s_t) as an ex-post efficient pure strategy equilibrium.

Forcing participants to lie to an identifiable victim instead of simply misrepresenting their private information should lead to a cost, c_l, associated with lying, i.e. playing strategy s_l. Using the equation from Theorem 4 yields $\rho(c_l) = \frac{2}{1-p}(1 - \frac{\bar{c}-c_l}{v})$ as the probability of telling the truth. Since this rises in c_l the mixed Nash equilibrium predicts an increase in the fraction of truth-telling.

In conclusion the treatment probably leads to a higher fraction of truth-telling. On the off-chance that players are able to coordinate on an asymmetric pure strategy equilibrium the fraction of truth-telling should remain the same or change to one. An increase in truth-telling also increases efficiency.

6 Results

The experiment was conducted in 5 sessions with 88 participants in total. Sessions were conducted between mid December 2016 and mid January 2017. The experiment was conducted in Mannheim Laboratory for Experimental Economics (mLab). Participants were recruited using the ORSEE system (Greiner (2004)). The experiment lasted 30 minutes and the participants on average received a payoff of 4.66 Euro. There were 4 participants that had to be removed from the data since they were able to identify their partner (as indicated by their answer in the questionnaire). The average age of the remaining 84 participants was 22. Slightly less than half of the students studied economics for at least one semester. Slightly more than half of the participants were men. Three fourths of the participants were bachelor's students.

Figure 3 shows the fraction of participants that choose a strategy that accurately reflects their cost. The left panel shows the participants with costs of 0 and the right panel shows the participants with costs of 80. The x-axis indicates whether the observations are from the treatment or the control group. The top left corner of each panel shows the number of participants with the corresponding costs. Mixed Nash equilibrium predictions are indicated by the black vertical lines with the "N.e"

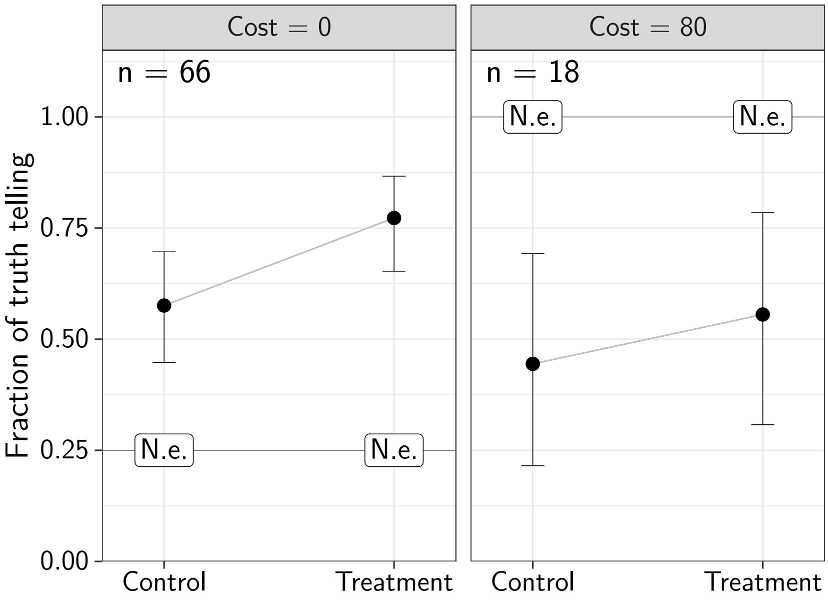


Figure 3: Fraction of participants that choose a strategy that accurately reflects their cost in the treatment and control group. The plot is split by actual cost.

labels. The confidence intervals are exact binomial. Note, that since treatment and control group observations are from the same participants comparing two confidence intervals is meaningless. In order to test if there is a significant difference between the two treatments a within-subject test has to be conducted.

In both treatments and for both types the fraction of truth-telling deviates significantly from the Nash equilibrium predictions. This is especially remarkable for the participants with costs of 80 because claiming costs of 0 is a dominated action for them, which does not depend on lying aversion. Taking the data from the participants with cost 80 into account the higher than predicted fraction of truth-telling for the participants with costs 0 does not necessarily mean that there is lying aversion in the control group. For both groups the amount of truth-telling rises with the treatment. This again is remarkable for the participants with costs 80 since theory with or without incorporating lying aversion predicts a constant rate of truth-telling of 1.

Insofar as it can be captured by wrong answers to the control questions one can control for misunderstanding the rules. For the following analysis one observation is the behavior of one person in one round. Recall that the control questions were asked in the treatment and in the control game. All control questions were answered correctly in approximately 68% of observations. Restricting the sample to those observations and redrawing figure 3 does not change any of the findings. The figure for the restricted sample can be seen in Appendix D.

Table 2: Contingency table for strategies of participants with cost of zero.

		Treatment Game	
		0	80
Control Game	0	34	4
	80	17	11

Table 2 shows the relationship between actions in the control and the treatment game for participants with costs of 0. The row labels of the table indicate whether the participant chose the action corresponding to costs of 80 or the action corresponding to costs of 0 in the control game. The column labels indicate the same for the treatment game. As you can see from the table, 34 participants (approximately half) chose to play a truthful action in the control as well as in the treatment game. The other 28 participants chose to claim costs of 80 in the control game. More than half of those changed to claiming costs of 0 in the treatment game ($\approx 60\%$). Almost no participants (4) switched from claiming costs of 0 to claiming costs of 80.

Since the outcome is binary and I use a within-subject design I follow the recommendations of Moffatt (2015) and conduct a McNemar change test with continuity correction. The null hypothesis that an equal number of people changed from truthful to untruthful actions than did the reverse can be rejected ($p \approx 0.01$). A permutation test allows for checking the hypothesis that the fraction of truthful actions increased more directly. The null hypothesis that the fraction of truthful actions stayed the same when switching from control to the treatment group could be rejected in favor of the alternative that the fraction of truthful actions increased ($p \approx 0.00$). To account for within-participant correlation treatment status was permuted on participant level. Participants did not randomize independently between the two games.

To test whether participants are able to coordinate I run a chi-squared test with the null hypothesis that the actions of one player are independent from the actions of the other player. The sample is pooled by costs and the players are arbitrarily assigned to one side. The test is unable to reject the null ($p \approx 0.65$).

In the control games 77% of available surplus were extracted. In the treatment games 93% of available surplus were extracted. Surplus extraction rates are computed on the whole sample and thus include the cases were trade was conducted despite causing a welfare loss.

In their introduction messages almost all of the participants told each other how old they were and what hobbies they have. Some described their character traits. Two of the participants made a promise too cooperate, even though they did not know the rules yet.

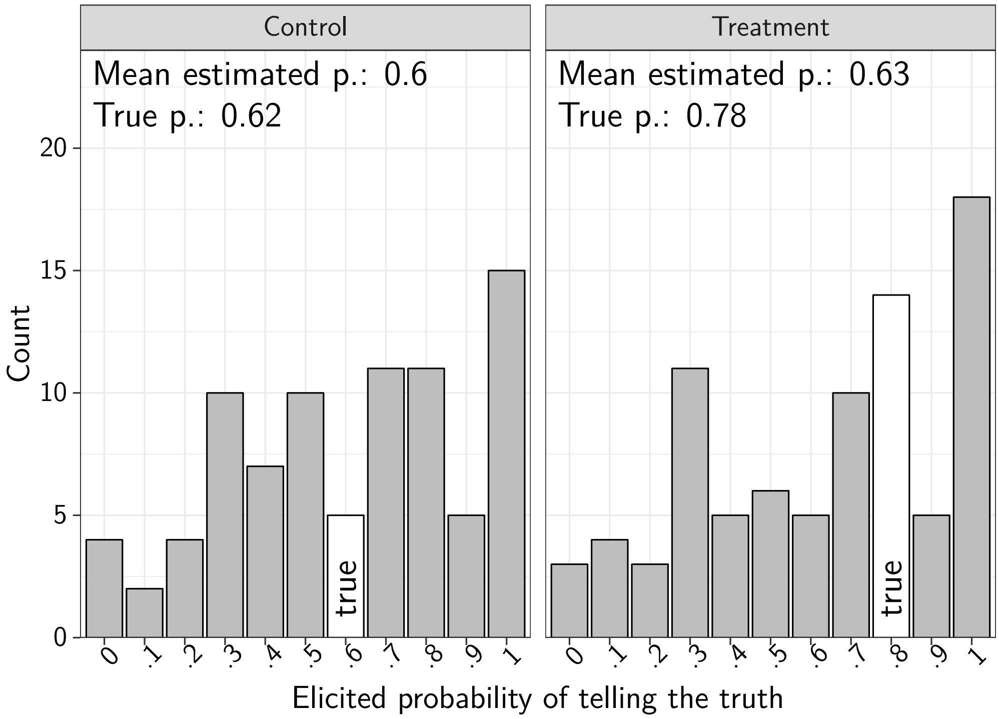


Figure 4: Beliefs about truth-telling for treatment and control group. The sample is restricted to participants with cost 0.

Figure 4 shows the distribution of elicited beliefs about strategies grouped by treatment status. The x-axis shows the elicited probability that a participant choose the action corresponding to 0, given that he has actually has costs of 0. The x-axis shows how many participants had that belief about their partner. The population frequency is given by the white bars labeled "true". The belief distributions are quite dispersed. There is some concentration around the true value. The mode of both distributions is the belief that the partner always tells the truth. Players that have costs of 0 and choose action 0 belief that their partner is telling the truth conditional on having costs of 0 with an average probability of approximately 0.7 independently of treatment status. Players that have costs of 0 and choose action 80 when being in the control believe that their partner is telling the truth conditional on having costs of 0 with an average probability of approximately 0.5. When they are in the treatment the average probability is 0.55. In conclusion elicited beliefs are not very accurate and there is a positive dependency between telling the truth and believing that others tell the truth. The positive dependency does not depend on treatment status.

7 Discussion

The experimental results indicate that having to lie to an "identifiable victim" leads to more private information being revealed in the broken revelation mechanism used in the experiment. This suggests that mechanisms, where people have to lie to an "identifiable victim'" lead to more efficient outcomes. However, the mechanism in the experiment under-performs the theoretically most efficient mechanism, but not by much. The constraint optimal mechanism must extract at least as much surplus as the asymmetric pure strategy Nash equilibrium of the broken revelation mechanism which extracts 95% of available surplus. In the treatment group 93% of available surplus were extracted.

The sub-optimal performance of the broken revelation mechanism is not surprising because the environment was optimized in order to test lying to an *identifiable victim* as the specific channel through which efficiency improves. This was done at the cost of a high predicted efficiency of the second best mechanism. Still, the efficiency gain was very high and the results can be used in order to extrapolate in which settings the broken revelation mechanism exploiting lying aversion outperforms the second-best mechanism.

Assuming that participants are rational, i. e. the dominated actions are due to some kind of non-material utility from having the reform implemented. Something about the optimal mechanism can be learned by looking at the incentive compatibility constraint of the type with costs 0. Theorem 3 indicates that the constraint for the high cost type always holds. Adding a lying cost c_l to the constraint for the low cost type yields the following constraint:

$$p \cdot (v - \bar{c}) + (1 - p) \cdot 0.5v \geq (1 - p) \cdot \bar{c} - c_l, \tag{$IC1^*$}$$

This can be re-arranged to:

$$c_l \geq \bar{c} - 0.5v(p + 1), \tag{$IC1^*$}$$

Re-arranging the condition from Theorem 2 tells us that the impossibility theorem does not hold if:

$$0 \geq \bar{c} - 0.5v(p + 1).$$

This shows that lying costs act as a subsidy that reduces the regret a seller incurs from telling the truth. If lying costs are large enough for all participants a truth-telling equilibrium can be sustained. This is easier if the gains from lying become smaller. The gains from lying become smaller if there is less money available when claiming to have high costs, which happens when $\bar{c}$ is low. The gains from lying also become smaller when it becomes more likely that lying prevents the reform. This happens when p is high. Given fixed lying costs $IC1^*$ holds when the benefit to implementing the reform in the case when one seller has high costs and the other seller has low costs is large.

A challenge to extrapolating from the experimental results is that they do not seem to be easily captured by a rationality-based model. First order beliefs are not very accurate and vary greatly across participants. Additionally, participants play dominated strategies. However, playing dominated strategies is something that also occurs in other mechanisms like the second price auction (Kagel & Roth (1995), chapter 7) and Nash equilibrium is still useful there. Some participants may value being perceived as honest higher than actually telling the truth. Since high cost types are rare and empirically mostly truthful the identifiable victim should belief that a player told the truth with a higher probability if he announced low costs. As a consequence if participants want to be perceived as honest they should announce low cost even if they have high costs. A similar behavior is observed in Fischbacher & Föllmi-Heusi (2013). Alternative reasons for the observed behavior are decision errors or framing. The game was framed as a decision on a reform with a public benefit. It could be the case that participants believe that the reform being implemented is a good thing in itself. Another possibility is that participants make random errors. There is no evidence for the hypothesis that participants did not understand the rules sufficiently well.

Lying costs probably depend on the parameters of the game. Fischbacher & Föllmi-Heusi (2013) suggest that participants care about others believing that they tell the truth. In the context of the game discussed here this means that lying costs should become smaller if the probability of having high cost rises. If it is more likely to have high costs it becomes less likely that someone that is saying that she has high costs is lying. Because it is dominated for high cost participants to lie low cost participants can hide among the increasing population of high cost participants. Lying costs may also depend on the consequences of lying for the other participant. As a consequence, even though lying aversion seems to be one channel through which communication increases efficiency other-regarding utility functions may still be relevant.

In the treatment condition telling the truth is equivalent to telling the identifiable victim that one announced low costs. In a revelation mechanism participants are asked for their private information. Answering this question truthfully may be seen as fair play akin to abstaining from fouling in a soccer game. Recall that if a player wants the identifiable victim to belief that he played fairly he should announce low costs. This is less costly if he actually has low costs. Thus wanting to be seen as someone who plays fairly could motivate participants to be truthful. Since this theory is very close to lying aversion it could also explain among others the findings of Fischbacher & Föllmi-Heusi (2013) and Abeler et al. (2014).

Participants who tell the truth believe with a higher probability that others will do the same. If participants would play a mixed strategy equilibrium the first order beliefs should be consistent and there should be no correlation between beliefs and actions. If participants best respond to their beliefs and individual lying costs are independent of beliefs there should be a negative relationship between telling the truth and believing that others tell the truth. This happens because if the probability that the other player tells the truth rises it becomes less likely that claiming high costs while having low costs prevents trade. Participants believing that others act similarly as they do explains the dependency between beliefs and actions. A rival explanation would be that participants want to reciprocate if others tell the truth. If this where true participants would react stronger to changes that increase lying costs since there is a second-order effect through reciprocity.

Half of all participants tell the truth in the control game. This is twice as much as predicted by the mixed strategy equilibrium. Since it is half of all participants it may be the case that participants simply use a rule of thumb. Another explanation could be that participants have some non-material utility from telling the truth, even in the control group. One reason would be that participants also have lying costs in the control group. Other reasons are social preferences or a preference for efficiency. Risk aversion could also be a reason. If low cost types tell the truth they either get $v - \bar{c}$ or $0.5v$. If they lie they get 0 or $\bar{c}$. The first gamble has a lower expectation, but also a lower variance. So risk-aversion could lead to more truth-telling.

8 Conclusion

Having to lie to an *identifiable victim* decreases misrepresentation of preferences in a broken revelation mechanism for the holdout problem. The given mechanism does not breach the upper bound for efficiency. However, there is some indication that it may do that for different parameters. The results suggest that exploiting lying aversion by a mechanism in which participants have to lie to each other in order to misrepresent their private information may be a good way to find more efficient mechanisms.

There are several building blocks that are still missing to move these kinds of mechanisms towards a real world application in the holdout problem. Checking whether the introduction message is necessary would help in judging how familiar participants in the mechanism have to be with each other. It should also be checked whether it is more effective to force participants to lie to fellow seller instead of the buyer. In order to make better predictions about the outcomes of mechanisms involving lying aversion a better understanding of how lying costs are influenced by reciprocity is needed. Further, it should be investigated how the ability to hide lies behind a high fraction of high cost buyers influences the fraction of truth-telling. The broken revelation mechanism should be compared to the second-best mechanism in a setting where the second-best mechanism is predicted to be less efficient than here. In order to be practically applicable the results of this paper would also have to translate to continuous environments, where lying is a matter of degrees.

Appendices

A Proof of Theorem 2

Proof. Without loss of generality we can assume that the transfer function is the same for both players. This holds since the incentive compatibility constraints and the participation constraints do not depend on the other player's transfer function. Therefore, if these constraints are fulfilled for a pair of transfer functions that differ on a pair of announcements we can use this to construct a mechanism with symmetric transfer functions by using the cheaper transfer at all those points. This automatically leads to the budget balance constraint being satisfied, since it was satisfied for the original more expensive transfer scheme.

First some conditions on transfers are established using the participation and budget balance constraints. Then it is shown that the incentive compatibility constraint for the high cost types always holds. This result leads to concrete values of transfer for which the incentive compatibility constraint of the low cost types has to hold if it holds at all. From that observation the desired result follows. Recall the budget balance constraint:

$$t_1(c_1, c_2) + t_2(c_2, c_1) \leq q(c_1, c_2) \cdot v \qquad \forall (c_i, c_{-i}) \in \{0, \bar{c}\} \times \{0, \bar{c}\}. \tag{BC}$$

If both players have high costs BC implies that:

$$t_1(\bar{c}, \bar{c}) + t_2(\bar{c}, \bar{c}) \leq 0 \Leftrightarrow t_i(\bar{c}, \bar{c}) \leq 0 \qquad \forall i \in \{1, 2\}.$$

The implication follows by the symmetry of the transfer functions. Recall the participation constraint of the type with high costs:

$$p \cdot t_i(\bar{c}, \bar{c}) + (1 - p) \cdot (-\bar{c} + t_i(\bar{c}, 0)) \geq 0 \tag{PC2}$$

Combining the two equations above yields:

$$(1 - p) \cdot (-\bar{c} + t_i(\bar{c}, 0)) \geq 0 \Leftrightarrow t_i(\bar{c}, 0) \geq \bar{c}.$$

Since $\bar{c} > 0.5v \Leftrightarrow v < 2\bar{c}$. If one player has high cost and the other has low cost, the budget constraint is:

$$t_i(\bar{c}, 0) + t_{-i}(0, \bar{c}) \leq v < 2\bar{c}$$
$$\Leftrightarrow$$
$$t_{-i}(0, \bar{c}) \leq v - t_i(\bar{c}, 0) \leq v - \bar{c} < 2\bar{c} - \bar{c} = \bar{c}$$

Since the transfers are symmetric this also holds for t_i. If both players have low costs BC becomes:

$$t_1(0, 0) + t_2(0, 0) \leq v \Leftrightarrow t_i(0, 0) \leq 0.5v < \bar{c}$$

Using these ingredients it can be shown that IC2 always holds. Recall IC2:

$$p \cdot t_i(\bar{c}, \bar{c}) + (1-p) \cdot (-\bar{c} + t_i(\bar{c}, 0)) \geq p \cdot (-\bar{c} + t_i(0, \bar{c})) + (1-p) \cdot (-\bar{c} + t_i(0, 0)). \quad \text{(IC2)}$$

The right side of IC2 is always smaller than 0, since $t_i(0, \bar{c}) \leq \bar{c}$ and $t_i(0, 0) < \bar{c}$. The left side of IC2 is weakly bigger than 0 if PC2 holds. Therefore IC2 always holds. Recall IC1:

$$p \cdot t_i(0, \bar{c}) + (1-p) \cdot t_i(0, 0) \geq p \cdot t_i(\bar{c}, \bar{c}) + (1-p) \cdot t_i(\bar{c}, 0). \quad \text{(IC1)}$$

Every additional compensation that the high cost sellers receive, $t_i(\bar{c}, \bar{c})$ and $t_i(\bar{c}, 0)$ only increases the right side of IC1. This makes it harder for IC1 to hold. Therefore if IC1 holds for PC2 holding with inequality, IC1 also holds for PC2 holding with equality. PC2 holding with equality implies: $t_i(\bar{c}, \bar{c}) = \frac{1-p}{p}(\bar{c} - t_i(\bar{c}, 0))$. Inserting into IC1:

$$p \cdot t_i(0, \bar{c}) + (1-p) \cdot t_i(0, 0) \geq p \cdot \frac{1-p}{p}(\bar{c} - t_i(\bar{c}, 0)) + (1-p) \cdot t_i(\bar{c}, 0)$$
$$\Leftrightarrow$$
$$p \cdot t_i(0, \bar{c}) + (1-p) \cdot t_i(0, 0) \geq (1-p)\bar{c}$$

Using $t_i(0, \bar{c}) \leq v - \bar{c}$ and $t_i(0, 0) \leq 0.5v$ produces the following implication of the above statement:

$$p \cdot (v - \bar{c}) + (1-p) \cdot 0.5v \geq p \cdot t_i(0, \bar{c}) + (1-p) \cdot t_i(0, 0) \geq (1-p)\bar{c}$$

Using the contrapositive statement IC1 is violated if the following inequality is violated:

$$p \cdot (v - \bar{c}) + (1 - p) \cdot 0.5v \geq (1 - p)\bar{c} \Leftrightarrow \frac{1 + p}{2} v \geq \bar{c}$$

□

B Proof of Theorem 3

Proof. First it is shown that $s_i(\bar{c}) = \bar{c}$ for every i in every equilibrium. Then the asymmetric equilibrium is established. Finally uniqueness is proven by checking if the remaining possible strategy profiles form an equilibrium.

For a player with $c_i = \bar{c}$ misrepresenting her value is dominated.

If the other player announces costs of 0, player i's utility from truthfully reporting costs of $\bar{c}$ is given by $t_i(\bar{c}, 0) - \bar{c} = 0$. Her utility from lying and reporting low costs is given by $t(0, 0) - \bar{c} = v/2 - \bar{c} < 0$.

If the other player reports costs of $\bar{c}$, truthful reporting prevents the good from being provided and results in a utility of 0. Reporting 0 costs results in a utility of $t(0, \bar{c}) - \bar{c} = v - \bar{c} - \bar{c} = v - 2\bar{c} < 0$. Therefore, reporting the high valuation is also dominated in this case for the low cost buyer. Therefore, it holds that in all equilibria $s_i(\bar{c}) = \bar{c} \quad \forall i \in \{1, 2\}$.

Since it is a dominant strategy for the high cost buyer to report truthfully and she gets utility 0 in all cases in which she does that the participation constraint holds ex-post for the high cost buyer. Since all transfers are positive the participation constraint for a player with $c_i = 0$ holds ex-post. Using this observation it directly follows from Theorem 2 that there is no Bayes Nash equilibrium in which the low cost Buyer always tells the truth. However, there is an asymmetric equilibrium in pure strategies were one buyer always tells the truth and one buyer does so only if he has costs of $\bar{c}$. If one player always reports her actual cost it follows by the violation of IC1 that the best response of the other player is to report costs of $\bar{c}$ when having costs of 0. It follows from the observation that reporting costs of 0 when having costs of $\bar{c}$ is dominated that the best response to one player fully revealing his costs is to always report high costs.

It remains to show that it is a best response to a player that always reports high costs to always report the true costs. Reporting the true costs when having high costs follows again by the fact that reporting low costs is dominated in this case. It remains to check that conditional on the other player always reporting $\bar{c}$ it is optimal to report low costs when having low costs, i.e. that:

$$q(0,\bar{c})t(0,\bar{c}) \geq q(0,0)t(0,0) \Leftrightarrow v - \bar{c} \geq 0,$$

which is true. It turns out that these are the only pure strategy Bayes Nash equilibria of this mechanism. Since reporting low costs when actually having high costs is a dominated action there is no equilibrium in which this occurs. It remains to check for all equilibria left after eliminating this possibility. Always reporting the actual valuations is not an equilibrium by direct implication of Theorem 2. All players always reporting high is not an equilibrium since always reporting high costs is not a best response to always reporting high costs. Therefore, the only pure strategy equilibrium that remains is the asymmetric one calculated above, which is the desired result. □

C Proof of Theorem 4

Proof. Since all strategies where $s_i(\bar{c}) = 0$ are dominated the players mix between s_t and s_l. Looking for symmetric mixed strategy equilibria the probability that a player plays strategy s_t is denoted by ρ. In a mixed strategy equilibrium ρ is chosen so that each player is indifferent between his pure strategies. Since both strategies prescribe the same action for the high cost types the high cost types are automatically indifferent between the two strategies. Therefore it remains to check indifference for the low cost types.

Assuming that player 1 mixes and $c_2 = 0$ the value for ρ that guarantees a symmetric mixed strategy equilibrium can be found by making player 2 indifferent between s_t and s_l. Player 1 announces costs of 0 if he plays strategy s_t and $c_1 = 0$. This happens with probability $\rho \cdot (1-p)$. If player 2 plays strategy s_t the reform is always implemented and player 2 gets an expected transfer of $\rho(1-p)0.5v + (1 - \rho(1-p))(v - \bar{c})$. If player 2 plays strategy s_l the reform is only implemented if player 1 announces costs of 0. In this case the expected transfer to the low cost player 2 is given by: $\rho(1-p)\bar{c}$. Player 2 is indifferent if:

$$\rho(1-p)0.5v + (1-\rho(1-p))(v-\bar{c}) = \rho(1-p)\bar{c} \Leftrightarrow \rho = \frac{2}{1-p}(1 - \frac{\bar{c}}{v})$$

If both players incur a cost from lying, i.e. playing s_l the equation becomes:

$$\rho(1-p)0.5v + (1-\rho(1-p))(v-\bar{c}) = \rho(1-p)\bar{c} - c_l$$

$$\Leftrightarrow$$

$$\rho = \frac{2}{1-p}(1 - \frac{\bar{c}-c_l}{v}) \qquad \square$$

D Subsample Analysis

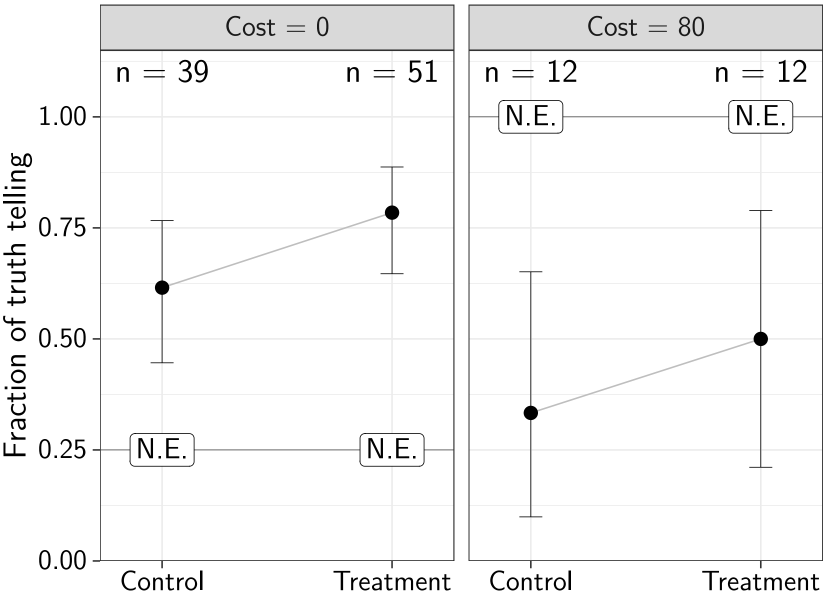
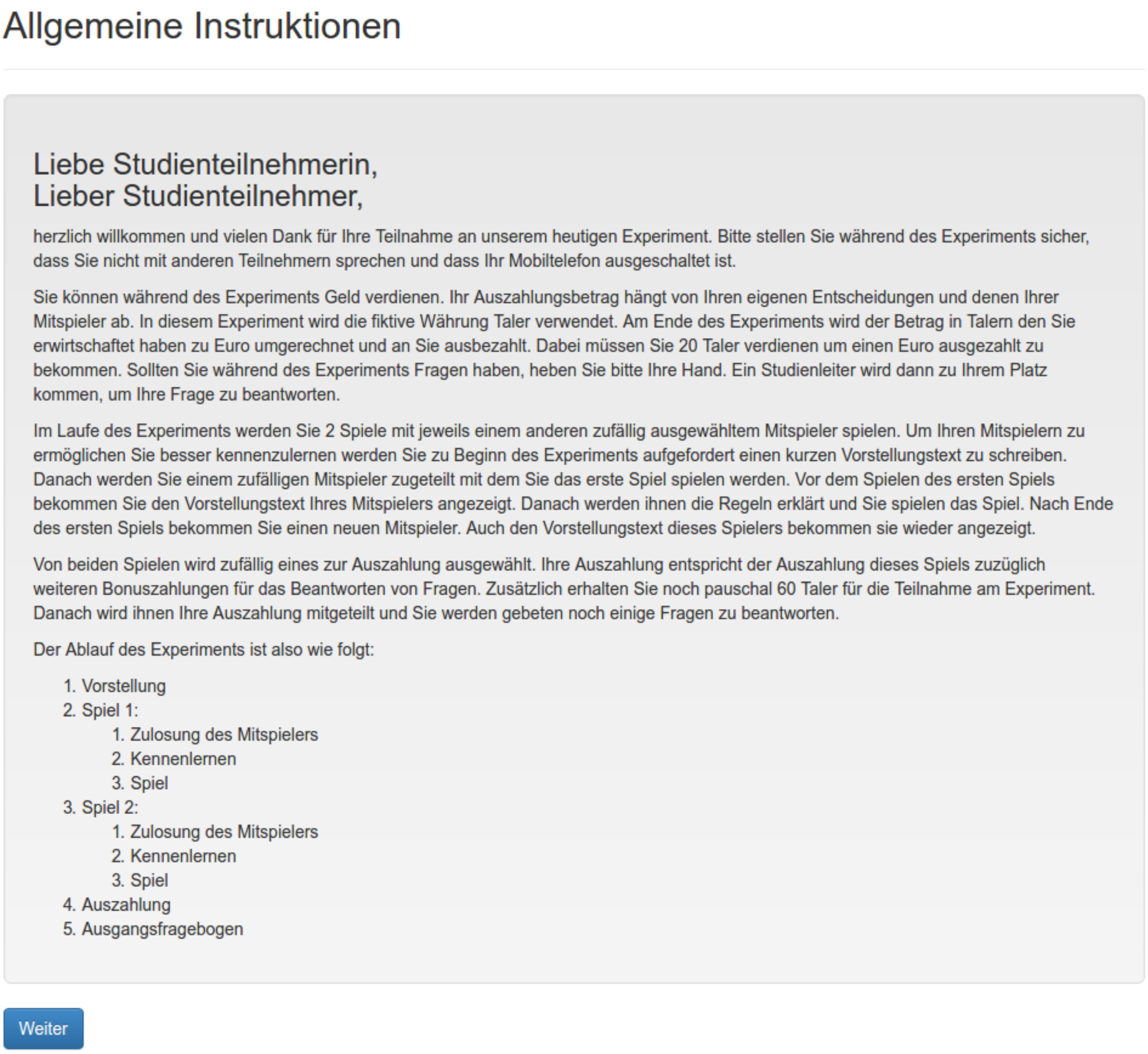


Figure 5: Fraction of participants that choose a strategy that accurately reflects their cost in the treatment and control group. The plot is split by actual cost. The sample is restricted to observations where all control questions were answered correctly.

E Instructions



Allgemeine Instruktionen

Liebe Studienteilnehmerin,
Lieber Studienteilnehmer,

herzlich willkommen und vielen Dank für Ihre Teilnahme an unserem heutigen Experiment. Bitte stellen Sie während des Experiments sicher, dass Sie nicht mit anderen Teilnehmern sprechen und dass Ihr Mobiltelefon ausgeschaltet ist.

Sie können während des Experiments Geld verdienen. Ihr Auszahlungsbetrag hängt von Ihren eigenen Entscheidungen und denen Ihrer Mitspieler ab. In diesem Experiment wird die fiktive Währung Taler verwendet. Am Ende des Experiments wird der Betrag in Talern den Sie erwirtschaftet haben zu Euro umgerechnet und an Sie ausbezahlt. Dabei müssen Sie 20 Taler verdienen um einen Euro ausgezahlt zu bekommen. Sollten Sie während des Experiments Fragen haben, heben Sie bitte Ihre Hand. Ein Studienleiter wird dann zu Ihrem Platz kommen, um Ihre Frage zu beantworten.

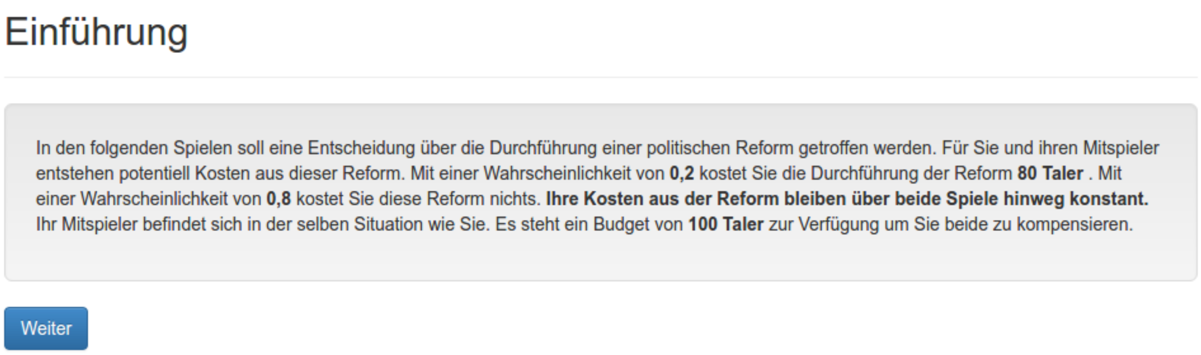
Im Laufe des Experiments werden Sie 2 Spiele mit jeweils einem anderen zufällig ausgewähltem Mitspieler spielen. Um Ihren Mitspielern zu ermöglichen Sie besser kennenzulernen werden Sie zu Beginn des Experiments aufgefordert einen kurzen Vorstellungstext zu schreiben. Danach werden Sie einem zufälligen Mitspieler zugeteilt mit dem Sie das erste Spiel spielen werden. Vor dem Spielen des ersten Spiels bekommen Sie den Vorstellungstext Ihres Mitspielers angezeigt. Danach werden ihnen die Regeln erklärt und Sie spielen das Spiel. Nach Ende des ersten Spiels bekommen Sie einen neuen Mitspieler. Auch den Vorstellungstext dieses Spielers bekommen sie wieder angezeigt.

Von beiden Spielen wird zufällig eines zur Auszahlung ausgewählt. Ihre Auszahlung entspricht der Auszahlung dieses Spiels zuzüglich weiteren Bonuszahlungen für das Beantworten von Fragen. Zusätzlich erhalten Sie noch pauschal 60 Taler für die Teilnahme am Experiment. Danach wird ihnen Ihre Auszahlung mitgeteilt und Sie werden gebeten noch einige Fragen zu beantworten.

Der Ablauf des Experiments ist also wie folgt:

1. Vorstellung
2. Spiel 1:
    1. Zulosung des Mitspielers
    2. Kennenlernen
    3. Spiel
3. Spiel 2:
    1. Zulosung des Mitspielers
    2. Kennenlernen
    3. Spiel
4. Auszahlung
5. Ausgangsfragebogen

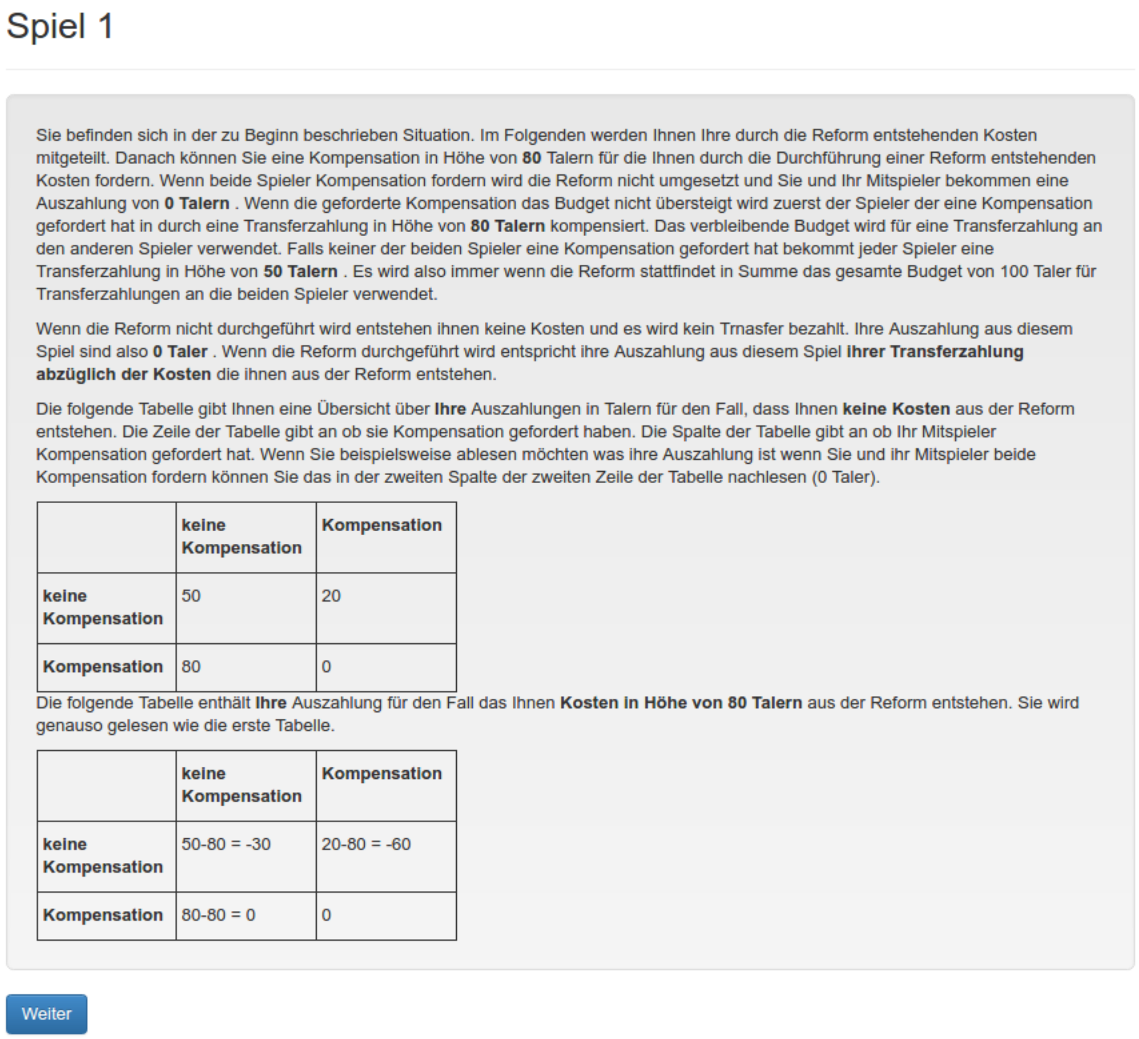
Instructions that were shown before the experiment.



Einführung

In den folgenden Spielen soll eine Entscheidung über die Durchführung einer politischen Reform getroffen werden. Für Sie und ihren Mitspieler entstehen potentiell Kosten aus dieser Reform. Mit einer Wahrscheinlichkeit von **0,2** kostet Sie die Durchführung der Reform **80 Taler** . Mit einer Wahrscheinlichkeit von **0,8** kostet Sie diese Reform nichts. **Ihre Kosten aus der Reform bleiben über beide Spiele hinweg konstant.** Ihr Mitspieler befindet sich in der selben Situation wie Sie. Es steht ein Budget von **100 Taler** zur Verfügung um Sie beide zu kompensieren.

Explanation of the setting.



Spiel 1

Sie befinden sich in der zu Beginn beschrieben Situation. Im Folgenden werden Ihnen Ihre durch die Reform entstehenden Kosten mitgeteilt. Danach können Sie eine Kompensation in Höhe von **80** Talern für die Ihnen durch die Durchführung einer Reform entstehenden Kosten fordern. Wenn beide Spieler Kompensation fordern wird die Reform nicht umgesetzt und Sie und Ihr Mitspieler bekommen eine Auszahlung von **0 Talern** . Wenn die geforderte Kompensation das Budget nicht übersteigt wird zuerst der Spieler der eine Kompensation gefordert hat in durch eine Transferzahlung in Höhe von **80 Talern** kompensiert. Das verbleibende Budget wird für eine Transferzahlung an den anderen Spieler verwendet. Falls keiner der beiden Spieler eine Kompensation gefordert hat bekommt jeder Spieler eine Transferzahlung in Höhe von **50 Talern** . Es wird also immer wenn die Reform stattfindet in Summe das gesamte Budget von 100 Taler für Transferzahlungen an die beiden Spieler verwendet.

Wenn die Reform nicht durchgeführt wird entstehen ihnen keine Kosten und es wird kein Trnasfer bezahlt. Ihre Auszahlung aus diesem Spiel sind also **0 Taler** . Wenn die Reform durchgeführt wird entspricht ihre Auszahlung aus diesem Spiel **ihrer Transferzahlung abzüglich der Kosten** die ihnen aus der Reform entstehen.

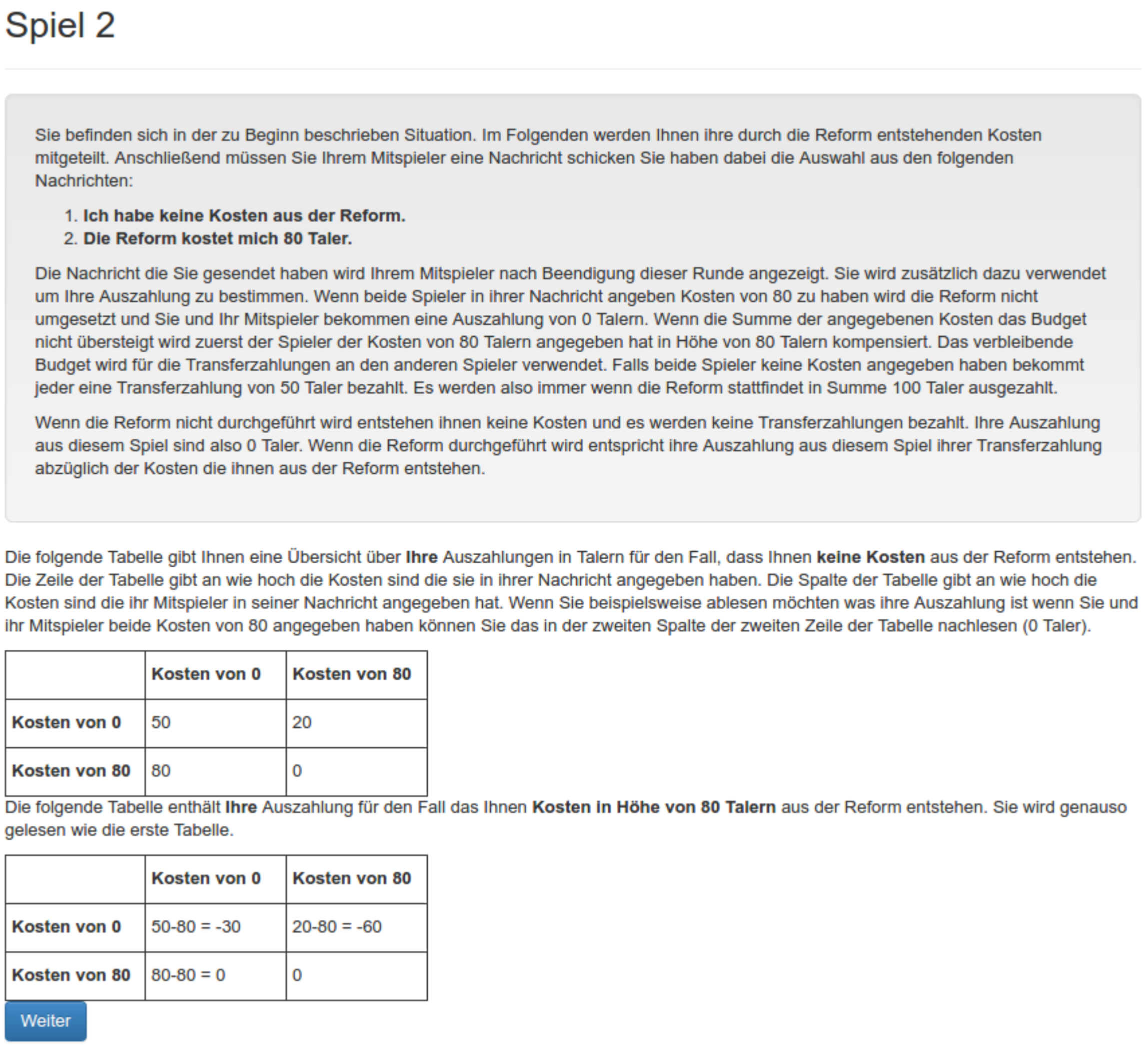
Die folgende Tabelle gibt Ihnen eine Übersicht über **Ihre** Auszahlungen in Talern für den Fall, dass Ihnen **keine Kosten** aus der Reform entstehen. Die Zeile der Tabelle gibt an ob sie Kompensation gefordert haben. Die Spalte der Tabelle gibt an ob Ihr Mitspieler Kompensation gefordert hat. Wenn Sie beispielsweise ablesen möchten was ihre Auszahlung ist wenn Sie und ihr Mitspieler beide Kompensation fordern können Sie das in der zweiten Spalte der zweiten Zeile der Tabelle nachlesen (0 Taler).

	keine Kompensation	**Kompensation**
keine Kompensation	50	20
Kompensation	80	0

Die folgende Tabelle enthält **Ihre** Auszahlung für den Fall das Ihnen **Kosten in Höhe von 80 Talern** aus der Reform entstehen. Sie wird genauso gelesen wie die erste Tabelle.

	keine Kompensation	**Kompensation**
keine Kompensation	50-80 = -30	20-80 = -60
Kompensation	80-80 = 0	0

Explanation of the control game.



Spiel 2

Sie befinden sich in der zu Beginn beschrieben Situation. Im Folgenden werden Ihnen ihre durch die Reform entstehenden Kosten mitgeteilt. Anschließend müssen Sie Ihrem Mitspieler eine Nachricht schicken Sie haben dabei die Auswahl aus den folgenden Nachrichten:

1. **Ich habe keine Kosten aus der Reform.**
2. **Die Reform kostet mich 80 Taler.**

Die Nachricht die Sie gesendet haben wird Ihrem Mitspieler nach Beendigung dieser Runde angezeigt. Sie wird zusätzlich dazu verwendet um Ihre Auszahlung zu bestimmen. Wenn beide Spieler in ihrer Nachricht angeben Kosten von 80 zu haben wird die Reform nicht umgesetzt und Sie und Ihr Mitspieler bekommen eine Auszahlung von 0 Talern. Wenn die Summe der angegebenen Kosten das Budget nicht übersteigt wird zuerst der Spieler der Kosten von 80 Talern angegeben hat in Höhe von 80 Talern kompensiert. Das verbleibende Budget wird für die Transferzahlungen an den anderen Spieler verwendet. Falls beide Spieler keine Kosten angegeben haben bekommt jeder eine Transferzahlung von 50 Taler bezahlt. Es werden also immer wenn die Reform stattfindet in Summe 100 Taler ausgezahlt.

Wenn die Reform nicht durchgeführt wird entstehen ihnen keine Kosten und es werden keine Transferzahlungen bezahlt. Ihre Auszahlung aus diesem Spiel sind also 0 Taler. Wenn die Reform durchgeführt wird entspricht ihre Auszahlung aus diesem Spiel ihrer Transferzahlung abzüglich der Kosten die ihnen aus der Reform entstehen.

Die folgende Tabelle gibt Ihnen eine Übersicht über **Ihre** Auszahlungen in Talern für den Fall, dass Ihnen **keine Kosten** aus der Reform entstehen. Die Zeile der Tabelle gibt an wie hoch die Kosten sind die sie in ihrer Nachricht angegeben haben. Die Spalte der Tabelle gibt an wie hoch die Kosten sind die ihr Mitspieler in seiner Nachricht angegeben hat. Wenn Sie beispielsweise ablesen möchten was ihre Auszahlung ist wenn Sie und ihr Mitspieler beide Kosten von 80 angegeben haben können Sie das in der zweiten Spalte der zweiten Zeile der Tabelle nachlesen (0 Taler).

	Kosten von 0	**Kosten von 80**
Kosten von 0	50	20
Kosten von 80	80	0

Die folgende Tabelle enthält **Ihre** Auszahlung für den Fall das Ihnen **Kosten in Höhe von 80 Talern** aus der Reform entstehen. Sie wird genauso gelesen wie die erste Tabelle.

	Kosten von 0	**Kosten von 80**
Kosten von 0	50-80 = -30	20-80 = -60
Kosten von 80	80-80 = 0	0

Explanation of the treatment game.

References

Abeler, J., Becker, A. & Falk, A. (2014). Representative evidence on lying costs. *Journal of Public Economics* 113: 96–104.

Bierbrauer, F.J. & Netzer, N. (2016). Mechanism design and intentions. *J. Economic Theory* 163: 557–603.

Bohnet, I. & Frey, B.S. (1999a). Social Distance and Other-Regarding Behavior in Dictator Games: Comment. *American Economic Review* 89(1): 335–339.

——— (1999b). The sound of silence in prisoner's dilemma and dictator games. *Journal of Economic Behavior & Organization* 38(1): 43 – 57.

Cadigan, J., Schmitt, P., Shupp, R. & Swope, K. (2009). An Experimental Study of the Holdout Problem in a Multilateral Bargaining Game. *Southern Economic Journal* 76(2): 444–457.

Charness, G. & Gneezy, U. (2008). What's in a name? Anonymity and social distance in dictator and ultimatum games. *Journal of Economic Behavior & Organization* 68(1): 29 – 35.

Chatterjee, K. & Samuelson, W. (1983). Bargaining under incomplete information. *Operations Research* 31(5): 835–851.

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

Cournot, A.A. (1838). *Recherches sur les principes mathématiques de la théorie des richesses par Augustin Cournot.* chez L. Hachette.

Fischbacher, U. & Föllmi-Heusi, F. (2013). Lies in Disguise - An Experimental Study on Cheating. *Journal of the European Economic Association* 11(3): 525–547.

Frohlich, N. & Oppenheimer, J. (1998). Some consequences of e-mail vs. face-to-face communication in experiment. *Journal of Economic Behavior & Organization* 35: 389–403.

Gneezy, U. (2005). Deception: The Role of Consequences. *The American Economic Review* 95(1): 384–394.

Greiner, B. (2004). The Online Recruitment System ORSEE 2.0 - A Guide for the Organization of Experiments in Economics. Working Paper Series in Economics 10, University of Cologne, Department of Economics.

Grossman, Z., Pincus, J. & Shapiro, P. (2010). A Second-Best Mechanism for Land Assembly. Working Paper qt1dn8g6vk, Department of Economics, UC Santa Barbara.

Hoffman, E. & Spitzer, M.L. (1982). The Coase theorem: Some experimental tests. *The Journal of Law & Economics* 25(1): 73–98.

Hoffman, E., McCabe, K. & Smith, V.L. (1996). Social Distance and Other-Regarding Behavior in Dictator Games. *The American Economic Review* 86(3): 653–660.

——— (1999). Reply: Social Distance and Other-Regarding Behavior in Dictator Games: Reply. *American Economic Review* 89(1): 340–341.

Kagel, J.H. & Roth, A.E. (1995). *The Handbook of Experimental Economics*. Princeton: Princeton University Press.

Kominers, S.D. & Weyl, E.G. (2012a). Concordance Among Holdouts. Working paper, Harvard Institute of Economic Research.

——— (2012b). Holdout in the Assembly of Complements: A Problem for Market Design. *The American Economic Review* 102(3): 360–365.

Mathews, S.A. & Postlewaite, A. (1988). Pre-Play Communication in Two-Person Sealed-Bid Double Auctions. *Journal of Economic Theory* (48): 238–263.

Mazar, N., Amir, O. & Ariely, D. (2008). The Dishonesty of Honest People: A Theory of Self-Concept Maintenance. *Journal of Marketing Research* 45(6): 633–644.

McGinn, K.L., Thompson, L. & Bazerman, M.H. (2003). Dyadic processes of disclosure and reciprocity in bargaining with communication. *Journal of Behavioral Decision Making* 16(1): 17–34.

Meub, L., Proeger, T., Schneider, T. & Bizer, K. (2016). The victim matters – experimental evidence on lying, moral costs and moral cleansing. *Applied Economics Letters* 23(16): 1162–1167.

Moffatt, P.G. (2015). *Experimetrics: Econometrics for Experimental Economics.* Palgrave Macmillan.

Myerson, R.B. & Satterthwaite, M.A. (1983). Efficient mechanisms for bilateral trading. *Journal of economic theory* 29(2): 265–281.

Normand, M.P. (2016). Less Is More: Psychologists Can Learn More by Studying Fewer People. *Frontiers in Psychology* 7: 934.

Ostrom, E., Walker, J. & Gardner, R. (1992). Covenants With and Without a Sword: Self-Governance is Possible. *The American Political Science Review* 86(2): 404–417.

Palfrey, T., Rosenthal, H. & Nilanjan, R. (2015). How Cheap Talk Enhances Efficiency in Threshold Public Goods Games .

Plott, C.R. (1982). Industrial Organization Theory and Experimental Economics. *Journal of Economic Literature* 20(4): 1485–1527.

Posner, E.A. & Weyl, E.G. (2017). Property Is Only Another Name for Monopoly. *Journal of Legal Analysis* 9(1): 51–123.

Rabin, M. (1993). Incorporating Fairness into Game Theory and Economics. *The American Economic Review* 83(5): 1281–1302.

Radner, R. & Schotter, A. (1989). The sealed-bid mechanism: An experimental study. *Journal of Economic Theory* 48(1): 179–220.

Satterthwaite, M.A. & Williams, S.R.. (1989). The Rate of Convergence to Efficiency in the Buyer's Bid Double Auction as the Market Becomes Large. *The Review of Economic Studies* 56(4): 477–498.

Schelling, T.C. (1968). *The Life You Save May Be Your Own.* Washington, DC: Brookings Institution, 127–162.

Schweizer, U. (1998). Robust Possibility and Impossibility Results. Working paper, Universität Bonn.

Tanaka, T. (2007). Resource allocation with spatial externalities: Experiments on land consolidation. *The BE Journal of Economic Analysis & Policy* 7(1).

Valley, K., Thompson, L., Gibbons, R. & Bazerman, M.H. (2002). How communication improves efficiency in bargaining games. *Games and Economic Behavior* 38(1): 127–155.

Valley, K.L., Moag, J. & Bazerman, M.H. (1998). A matter of trust: Effects of communication on the efficiency and distribution of outcomes. *Journal of Economic Behavior and Organization* 34(2): 211–238.

Vanberg, C. (2015). Who never tells a lie? Working Paper 581, Universität Heidelberg, Department of Economics.