WORKING PAPERS

Economics Department

UNIVERSITAT JAUME·I

An experimental test of reporting systems for deception

Sascha Behnk
Iván Barreda-Tarrazona
Aurora García-Gallego

2017 / 11

Castellón (Spain)

An experimental test of reporting systems for deception

Sascha Behnk
University of Zurich
Department of Banking and Finance
sascha.behnk@bf.uzh.ch.

Iván Barreda-Tarrazona
Universitat Jaume I & LEE
Department of Economics
ivan.barreda@eco.uji.es

Aurora García-Gallego
Universitat Jaume I & LEE
Department of Economics
mgarcia@eco.uji.es

2017 / 11

Abstract

We use a repeated sender-receiver game in which sender behavior is revealed to future counterparts by (i) standardized computer reports or by (ii) individual reports composed by the receivers, representing a common form of consumer feedback. Compared to our baseline without reporting, computer reports reduce deception in all payoff scenarios while the effect of individually written reports is lower and in some scenarios only marginal. This comparably weaker impact can be explained by the senders' anticipation of a high number of missing or deficient receiver reports that we find. We conclude that the precision of a reporting system has a higher importance for reducing deception than its personal character via individual feedback. Surprisingly, the reliability of computer reports is not correctly anticipated by receivers, who trust individually written reports more in the beginning and hence seem to back the wrong horse initially.

Keywords: deception; trust; reporting systems; reputation; experiment

JEL classification: D03; D63; K42

An experimental test of reporting systems for deception

Sascha Behnk[a]*
Iván Barreda-Tarrazona[b]
Aurora García-Gallego[b]

[a] *University of Zurich, Switzerland*

[b] *Universitat Jaume I, Castellón, Spain*

Abstract: We use a repeated sender-receiver game in which sender behavior is revealed to future counterparts by (i) standardized computer reports or by (ii) individual reports composed by the receivers, representing a common form of consumer feedback. Compared to our baseline without reporting, computer reports reduce deception in all payoff scenarios while the effect of individually written reports is lower and in some scenarios only marginal. This comparably weaker impact can be explained by the senders' anticipation of a high number of missing or deficient receiver reports that we find. We conclude that the precision of a reporting system has a higher importance for reducing deception than its personal character via individual feedback. Surprisingly, the reliability of computer reports is not correctly anticipated by receivers, who trust individually written reports more in the beginning and hence seem to back the wrong horse initially.

Key words: deception; trust; reporting systems; reputation; experiment
JEL Classification: D03; D63; K42

* Corresponding author. Contact information: Department of Banking and Finance, University of Zurich, Plattenstrasse 14, 8032 Zurich, Switzerland. Phone: +41 44 634 19 03. Email: sascha.behnk@bf.uzh.ch.

1. INTRODUCTION

Moral hazard is a prevalent problem in economic relationships with asymmetric information as shown by Akerlof (1970). A large body of experimental literature has therefore investigated how deception can be reduced in one-shot interactions using a variety of pecuniary incentives from positive reinforcements such as voluntary payments (Angelova and Regner, 2013) to punishment (e.g. Church and Kuang, 2009). Other studies found that intrinsic motivations can also keep agents from deception, for instance though lying aversion (e.g., Gneezy et al., 2013) or image concerns (Behnk et al., 2014).[1]

In environments with recurring interactions, an additional factor has proved to foster pro-social behavior among the involved parties: reputation (Rockenbach and Milinski, 2006). Danilov and Sliwka (2013) find that individuals are willing to increase their effort substantially in principal-agent relationships when principals receive information about the agents' past behavior. In public good settings, ex post communication about the appropriateness of the players' behavior can lead to more cooperation (Zylbersztejn, 2014). With regard to deception, Kimbrough and Rubin (2013) have shown that reputation can indeed reduce dishonest behavior over time and that it works as a complement to precuniary punishment.[2] Moreover, Koch and Schmidt (2010) find that reputation is an effective measure against the deception-increasing effect of ex ante disclosure of conflicts of interest (Cain et al., 2011).

In this paper we address a crucial condition for an effective reputation building to reduce deception in principal-agent relationships: an agent's reputation can only be established if the underlying information is both credible and accessible for her future principals (Sobel, 1985). For instance, Gino et al. (2013) show that in a system with voluntary monitoring, dishonest behavior even increases compared to a setting without a monitoring regulation. We propose that a promising way to provide sufficient access to reliable information about past behavior is through the implementation of reporting systems. In general, the systematic provision of reports can be designed in two ways. One of them is through a central authority. Examples for such *exogenous* reports are the published results of product tests, typically conducted by consumer protection organizations, as well as the work of credit rating agencies evaluating the solvency of countries and corporations. These report types have in common that they are based on pre-defined criteria and are often provided in a standardized form, for instance by a point system (White, 2010). Furthermore, such exogenous reports are usually provided by a neutral player, who can make a more objective assessment of the agents' behavior.

The opposite case is the provision of *endogenous* reports, which are directly filed by the agents' counterparts. An example for this report type is the provision of consumer recommendations on commercial online platforms (Dellarocas, 2003). Since these reports reflect the principals' personal experience with specific agents and since they are provided in an individual form, they can

[1] Other factors that can lead to less deception are guilt aversion (Charness and Dufwenberg, 2006), the temporal distance between decisions and pay-outs (Ruffle and Tobol, 2014) and pre-play communication (Bicchieri and Sontuoso, 2015). Furthermore, there exists also the possibility to avoid deception from the beginning through the use of random devices (Kimbrough et al., 2013).

[2] Repeated sender-receiver games have also been used in other studies, such as Blume et al. (1998, 2002), Sánchez-Pagés and Vorsatz (2007 & 2009) and Peeters et al. (2013).

potentially enhance the perceived social proximity among the agent's previous and future counterparts, especially regarding victims of the agent's potential deception.

We are particularly interested in examining which of these two approaches is most effective in reducing deception in repeated principal-agent relationships. The comparison of both report types allows us to shed more light on the trade-off between the standardized and objective nature of exogenous reporting and the personal and subjective character of individually written, text-based reports.[3]

In this paper, we use a sender-receiver game with three different payoff scenarios, similar to the ones used by Gneezy (2005), to investigate how sender behavior changes over time with each report type. In our baseline without reporting, a receiver finds out about the honesty or dishonesty of the sender with 50% probability at the end of each round, representing the fact that moral hazard is not always observable in the direct aftermath of an economic interaction. Receivers do not get information about the senders' past behavior in the baseline. In a second treatment, we test the effect of exogenous reporting on deception, that is, the computer-generated reports regarding the sender's honesty or dishonesty. All reports regarding the respective sender are stored in his personal report list, which was available to his counterparts in future rounds. The third treatment was identical to the second one with one exception: the reports were now endogenously generated by the receivers themselves, in form of an individual text.

Our results show that reputation building through reporting significantly reduces deception compared to the baseline. With regard to the question whether the two report types have different effects on senders, we find that computer reports lead to a significantly lower level of deception than receiver reports in all payoff scenarios. Regarding the dynamics of sender behavior, we find that deception even increases over time with endogenous reports, whereas the fractions of deceptive messages are stable when reports are provided exogenously. Our econometric models including individual beliefs show that deceptive senders are rather driven by first-order beliefs (whether the receiver accepts their proposal), which play an important role for strategic considerations, while honest senders are mainly influenced by their second-order beliefs (whether the receiver expects a relatively higher payoff), which are related to guilt aversion. Furthermore, we use the three payoff scenarios to analyze how senders react to different payoff temptations and losses for their counterpart from successful deception. In line with the findings of Gneezy (2005), more senders transmit deceptive messages when their earnings are comparatively high, while the lowest level of deception occurs when successful deception leads to the most unfair outcome for the receiver.

A qualitative analysis of the receiver reports provides a potential explanation for the higher impact of exogenous reports. It turns out that a substantial fraction of the receivers did not use the opportunity to report dishonest senders in a proper way. Altogether, the inadequate reporting seems to be symptomatically due to the nature of the endogenous reports and is a possible

[3] Abraham et al. (2016) investigate the effect of players' subjective reports without a personal character, that is, in terms of a number-based rating of their satisfaction with counterpart behavior. They find that, in contrast to objective information, the subjective ratings increase investments and reciprocity in trust games only when information transmission is public.

reason for the substantially lower effect of receiver reports on sender behavior, who might have anticipated this lack of reliability from the beginning.

Receivers do not seem to be aware of the endogenous reports' lower reliability and their lower effect on deception. Compared to the baseline, they show relatively more trust when reports are provided by them and their peers. It turns out that this difference is mainly driven by changes in their second-order beliefs about the senders' relative payoff expectations. On the other hand, there is no significant difference in the trust levels between the baseline and computer reports. This finding reveals an interesting discrepancy between the agents' and their principals' perception of each report type, which leads to a suboptimal situation for all involved parties. We conclude that the reliability of a reporting system has a higher impact on pro-social behavior in our setting than its personal character.

We organized the paper as follows: In section 2 we present the experimental design, treatments and procedures. In section 3 we state our hypotheses. In section 4 the results are presented and discussed. Section 5 concludes. The experimental instructions can be found in Appendix A.

2. EXPERIMENTAL DESIGN

2.1 Repeated sender-receiver game

Our study is based on a repeated version of the sender-receiver game used in Behnk et al. (2014). In this experiment, we let subjects play the two-player game over ten rounds in all treatments. Therefore, eight subjects were randomly matched to form a group at the beginning of each session. Within each group, four subjects were randomly assigned the role of a sender while the remaining four subjects played the game as receivers. The roles were neutrally framed "Player 1" and "Player 2". Both senders and receivers were numbered from I to IV so that they could be recognized by their counterparts in later stages. Subjects remained in the same group and kept their role as well as their number during the whole experiment.

In the beginning of each round, each sender was randomly matched with one of the receivers in his group and both players learned the assigned number of their counterparts. Except for this number, the game was played anonymously, i.e. none of the subjects learned the real identity of the others during the experiment. After that, three options A, B and C, which contained payoffs for both players, were presented to the sender. His task was to recommend an option to the receiver he was matched with by choosing one of the following three messages.

> Message 1: Option A will earn you more money than the other two options.
>
> Message 2: Option B will earn you more money than the other two options.
>
> Message 3: Option C will earn you more money than the other two options.

We used the strategy method by letting senders choose a message in each of the three payoff scenarios illustrated in Table 1. This procedure enabled us to investigate their behavior regarding different monetary incentives and various consequences for their counterparts. Independently of the scenario, option A paid the receiver more money than the other two options. Therefore, only message 1 is an *honest* message, while the other two messages represent lies. We differentiate

two types of lies by the payoff allocation they refer to. With the *deceptive* message 2, the sender recommended option B which provided him with the highest payoff at the expense of a comparatively lower payoff for his counterpart, whereas the *payoff-equalizing* message 3 recommended option C that led to an equal but Pareto-dominated outcome for the players.

Scenario	Option	Payoff sender	Payoff receiver
1 (low+;low-)	A	5	6
	B	6	5
	C	3	3
2 (low+;high-)	A	5	15
	B	6	5
	C	3	3
3 (high+;high-)	A	5	15
	B	15	5
	C	3	3

Table 1
Sender and receiver payoffs by scenario and option (in euros).

In scenario 1, a successful deception led to a comparatively low additional gain of one euro for the sender compared to an equally low loss for the receiver. We label this scenario (low+,low-), indicating how much senders earn (+) and receivers lose (-) from successful deception. In scenario 2, the sender obtained the same profit from successful deception as in scenario 1 but now at the expense of a higher comparative loss for his counterpart (low+,high-). In scenario 3, the sender was able to gain a profit from successful deception that was higher than in the other two scenarios, at the cost of her counterpart's loss that was equally high as in scenario 2 (high+,high-). We presented the scenarios on different screens and randomized the appearance of the scenarios and options to control for order effects.

After the sender made a choice in each scenario, one scenario was randomly selected by the computer and the respectively chosen message was sent to the receiver. The receiver knew that there were three options available but she was not informed about the payoffs in each option and did not know that the sender faced a conflict of interest regarding his recommendation task. After receiving the sender's message, the receiver decided whether to accept or reject it. In the case of acceptance, both players earned the payoffs from the option mentioned in the message. In case the receiver rejected the message, one of the two remaining options was randomly selected by the computer and determined the payoffs for each player. With this procedure, in addition to the presence of the third, Pareto-dominated option, we limited the sender's possibility to be strategically honest, in the sense of sending a truthful message while expecting the receiver to reject it, as in Sutter (2009).

In the end of each round, we presented the final outcome to the players. Senders always received information about the acceptance or rejection of the sent message, the implemented option and the earnings of both players. The final information shown to receivers was randomized in each round in the sense that, in principle, they only received information about their own payoffs.

Furthermore, a possibility existed that all potential payoffs for both players in each option in the implemented scenario was shown to the receiver, which happened with 50% probability. In the latter case, the receiver could find out about the honest or dishonest behavior of her counterpart. Both player types were informed about this possible ex post disclosure and its probability in the experimental instructions as shown in Appendix A. We introduced this procedure to simulate the fact that negative consequences of moral hazard are not always observable in the direct aftermath of an economic interaction.[4] At the end of the experiment, we presented the subjects' own payoffs from each of the 10 rounds, their respective payoffs from the belief elicitation described in the following subsection as well as their final earnings, which consisted of their total payoffs from two randomly selected rounds.

2.2 Belief elicitation

After the subjects made their decisions and before showing them the final outcome of the round, both player types were asked to answer a series of belief-related questions. That the elicitation of individual beliefs can shed light on the different motivations behind the subjects' decisions has been shown by previous studies such as Charness and Dufwenberg (2006), Peeters et al. (2012) and López-Pérez and Spiegelman (2013). We elicited subjects' first-order beliefs about their counterparts' behavior by asking senders to estimate the fraction of receivers in their session who accepted the message and receivers were correspondingly asked about the fraction of truthful senders. In addition, we elicited an more direct type of these beliefs by asking senders if they expected the respective receiver with whom they were matched to accept the message they had sent and receivers were asked if they believed that the sender they were matched with had sent a truthful message. We also elicited how much the subjects expected to earn from the sent message as well as their second-order beliefs by asking how much they think their counterpart in the specific round expected to earn in comparison to their own payoffs. For both questions we used a five-point Likert scale with the categories "much less", "less", "equal", "more" and "much more".

We elicited all before-mentioned beliefs in each round in order to account for possible changes in the subjects' expectations over time. One exception was the peer beliefs which we elicited only in rounds one and five as well as in the last two rounds, since these moments gave us sufficient information about how the players expected their peers to behave. This procedure is based on our experience from a pilot and enabled us to reduce the duration of the sessions substantially. In the mentioned rounds, senders faced again the three payoff scenarios on different screens and were asked to estimate the percentage of other senders who had chosen a message that favored them in the specific scenario. On the other hand, receivers were asked about how likely they believed it was that other receivers had accepted the message sent to them in that specific round.

Since the belief elicitation was a substantial part of the subjects' tasks during the experiment, we paid the players an additional amount of money according to the accuracy of their answers. In each round, subjects earned one additional euro for each answer that coincided with the other players' behavior or beliefs, depending on the question. For the first-order and peer beliefs

[4] Receivers could eventually learn the payoff structure of all scenarios in case all scenarios had been implemented and also disclosed to them over the 10 rounds. However this effect was limited since there were three scenarios and only a 50% probability of disclosure. This learning process was possible in all of our treatments.

regarding which we asked for a specific percentage, we used a simple quadratic scoring rule following Brier (1950) to calculate the additional earnings for each player.[5] According to the scoring rule we applied, subject *i* gained one additional euro in case her estimation $f_{i,j}$ of the fraction in question *j* was correct. The more $f_{i,j}$ deviated from the actual fraction f_t, the more we discounted from the subject's maximum payoff α = *1€,* arriving at her final payoff π_i as shown in equation (1).

$$\pi_i(f_{i,j}) = \alpha - (f_{i,j} - f_t)^2 \quad (1)$$

2.3 Treatments

Our experiment consisted of three treatments which we used to test the effect of different reporting systems on deceptive behavior over time. Our baseline treatment, henceforth called *T_base*, follows the unmodified game structure described in the previous subsection. In this setting it was possible for senders to gain bilateral reputation since they interacted more than once with some or even all of the four receivers in their group, due to the random stranger matching, and learned the number of their counterparts at the beginning of each round. This design provided receivers with the possibility to remember the senders' behavior from previous rounds in which the payoff structure was disclosed to them. However, this information was neither stored for future rounds nor available to other players in the group since there were no reports available in this treatment.

This bilateral sender reputation was also present in the two other treatments, but in these settings we enhanced the reputation building through different reporting systems. In our second treatment, we tested the effect of report provided by a central authority. In the end of each round in which the payoff structure had been revealed to the receiver, the computer generated a report regarding the sender's honesty or dishonesty. Since these reports were exogenously produced, we henceforth call this treatment *T_exo*. The computer did not distinguish between deceptive and payoff-equalizing messages but reported them as dishonest messages. The reason for this was to avoid any further disclosure of the payoff structure to future counterparts and to circumvent an unnecessary labeling of the two kinds of lies. Hence, there were two reports with which the computer classified the sender behavior:

"Player 1 has sent an honest message"

"Player 1 has sent a dishonest message"

The computer generated a personal report list for each of the senders. All reports regarding the sender's own behavior were stored in his report list, which contained ten rows, one for each round. In rounds in which the sender's behavior was revealed to his counterpart, the text "no information" was displayed in the respective rows. From the second round on and before making decisions, the sender's current report list was shown to his counterpart. In this way, the receiver had the possibility to observe the sender's past behavior before deciding whether to accept or reject his message. In order to ensure that the sender was fully aware of the information provided to the receiver, his personal report list was also shown to him in the beginning of each round. We present

[5] The advantages and limitations of using quadratic scoring rules to elicit beliefs in economic experiments are discussed, for instance, in Nyarko & Schotter (2002) and Palfrey & Wang (2009).

an example of how the information was presented to the receiver including all possible information types for the first six rounds in Figure 1.

Round	Player 1 ...
1	- No information -
2	- ... has sent a dishonest message -
3	- ... has sent an honest message -
4	- No information -
5	- ... has sent a dishonest message -
6	- ... has sent an honest message -

Figure 1
Example report list for the first six rounds in T_exo.

Our third treatment *T_endo* was identical to the second one with one exception. The reports were now endogenously generated by the receivers with whom the senders were matched in the respective rounds, in form of an individually composed text with up to 140 characters.[6] Both player types knew from the instructions that the purpose of the reports was to inform the following receivers about whether the sender had sent an honest or a dishonest message in the current round. But in contrast to the standardized computer reports in our second treatment, receivers were not obliged to provide a report and were free to give any information, except, they knew from the instructions that they would receive a final earning of zero in case they used swearwords or insulted other players in their report. In line with the other treatments, receivers could not write a report in rounds in which the sender behavior had not been revealed to them. Accordingly, the text "no information" was shown in the respective rows of the sender's report list. As in T_exo, we showed each sender his personal list in the beginning of each round.

We summarize our treatments in Table 2. In each of the treatments, there were ten independent groups of eight players per treatment. Altogether, 40 subjects played in the role of a sender and another 40 subjects took on the role of a receiver. The samples were almost balanced between men and women.

Treatments	Subjects	Percentages of females
T_base Repeated sender-receiver game	80	49%
T_exo Repeated game with computer-generated reports	80	43%
T_endo Repeated game with receiver free-text reports	80	54%

Table 2
Treatments, number of subjects and percentages of female subjects.

[6] We limited the report text to 140 characters in order to urge the receiver to provide efficient information regarding the sender's behavior, referring to the established text format on social media platforms.

2.4 Procedures

We conducted the experiment at the LEE-Laboratory of Experimental Economics at the Universitat Jaume I in Castellón, Spain, and recruited 240 undergraduate students from different faculties through the Online Recruitment System for Economic Experiments ORSEE (Greiner, 2004). We ran six sessions, two sessions per treatment with 40 players each. The sessions lasted around two hours. Upon arrival, subjects entered the laboratory one by one, sat down in front of the computers and read the instructions. After that, the instructions were read aloud by the experimenter and the subjects answered a quiz question in order to find out if they understood the rules of the game. In case a subject did not answer the question correctly, the experimenters explained the instructions again to him or her in order to ensure a full understanding of the game. The experiment was programmed in z-Tree (Fischbacher, 2007). After the experiment, we paid the subjects anonymously in cash. Subjects' final earnings were around €21 on average.

3. Hypotheses

With our repeated sender-receiver game, we seek to answer the question if reputation through reporting systems is able to reduce deceptive behavior over time compared to a setting in which only bilateral reputation building is possible among the players. We know from studies like Kimbrough and Rubin (2013) that reputation can actually lead to less deception. Furthermore, Behnk et al. (2014) have shown that the provision of ex post information about honest or dishonest behavior can reduce the rates of deceptive messages even in one-shot interactions, due to the agents' image concerns. Therefore, we hypothesize that this image effect is even stronger in a setting in which an agent interacts several times with multiple principles who are given access to information about his past behavior. In addition to the image concerns, strategic considerations play an important role in our design, since receivers, although being blind regarding the payoffs in the beginning, had a say in the outcome of the game by accepting or rejecting the transmitted message. We suppose that reported dishonesty could therefore lead to less trust in the information transmitted by the sender in future rounds and, hence, to a lower probability of the option mentioned in his message being implemented. A payoff-maximizing sender should take this into account, since the possibility to lie strategically was limited in our game due to the fact that in case of a rejection the finally implemented option depended on a random process. Therefore, we hypothesize that fewer senders try to deceive their counterparts when their behavior can be reported to his future counterparts.

H1*: The rate of deceptive messages is lower in T_exo and T_endo compared to the baseline without reporting.*

The second aim of our study is to investigate if exogenous and endogenous reports have different effects on deceptive behavior. The reports in T_exo were standardized and automatically generated by the computer. This procedure guaranteed an objective classification of the sender's behavior in two categories, honesty and dishonesty, and assured at the same time that a report was filed in each and every round in which the payoff information was revealed to the receiver. The reporting system in T_endo differed from this system in two aspects. On the one hand, the

reports were composed by the receivers as involved parties and, in case of successful deception, as actual victims of the respective senders. We assume that this fact turned the nature of the receiver reports into a more personal one and could have created a more pronounced social proximity between the receiver and the sender's future counterparts than the computer-generated reports. This assumption is supported by a comparison of the freely written pre-play messages in the trust game of Charness and Dufwenberg (2006) and the pre-fabricated text in their otherwise identical design in Charness and Dufwenberg (2010). Altogether, the standardized text led to a substantially lower level of trust compared to the individual messages.

On the other hand, receivers were not obliged to provide a report and were free to give any information in their individual text, regardless of the actual sender behavior. In this sense, the receiver reports may suffer from a lack of objectivity in their classification of sender behavior and could therefore lead to less meaningful information for other players. Furthermore, manipulations were possible in the sense that receivers report a dishonest (an honest) message while the sender was actually honest (dishonest). The phenomenon of biased word-to-mouth information can be observed, for instance, in the popular field of online reputation through consumer feedback (Dellarocas and Wood, 2008). Interestingly, the effect of negative (positive) feedback does not necessarily lead to lower (higher) prices and sale rates on commercial online platforms, as shown in the review of Dellarocas (2003). Furthermore, in case the receiver did not provide a report at all, the personal character of the written text could simply not unfold its potential, which is supported by the theoretical analysis of Muehlheusser and Roider (2008). The authors find in their model that honest individuals refrain from reporting norm defections in equilibrium since they fear that damaging the reputation of others might backfire and lead to lower benefits from interactions with them in the future. Based on these considerations, we assume that receiver reports are on average less reliable than computer-generated reports, which senders anticipate, and hypothesize therefore that exogenous reports lead to a comparatively higher reduction in deception.

***H2**: The reporting in T_exo leads to a higher reduction in deception than the reporting in T_endo.*

Previous studies have shown that some people do not only exhibit lying aversion per se but also that their willingness to deceive is heterogeneous and that it depends on the severity of the lie (see for instance Fischbacher and Föllmi-Heusi, 2013, or Gibson et al., 2013). In order to test the effect of reporting sender behavior regarding different monetary incentives for deception and different consequences for the receivers, we use a within-subject comparison of three payoff scenarios similar to Gneezy's (2005) scenarios. In this setup, the second scenario (low+;high-) is the most extreme one since the sender can obtain a relatively low payoff from successful deception at the cost of a comparatively high loss for the receiver. This payoff distribution can therefore be characterized as a "mean" scenario. Since previous studies have shown in different settings that deception rates are substantially lower with this payoff distribution than in the other scenarios (see for instance Behnk et al., 2014), we expect to find the same tendency in the presence of reporting systems over time.

***H3**: In T_exo and T_endo, fewer deceptive messages are sent in scenario 2 than in the other payoff scenarios.*

Although the main focus of this study is to observe the behavior of senders, we seek to answer the question if the use of different repot types increases trust in the information transmitted by the senders. This is important since the receivers' trust is a crucial condition for the establishment and functioning of principal-agent relationships. Consider, for instance, the case in which an uninformed private investor is reluctant to accept the service of a sophisticated market participant, who potentially faces conflicts of interest, which leads to a situation with a Pareto-dominated outcome for both parties. Since we assume that a sender's future counterparts' access to the information about his past behavior lowers the rate of deceptive messages, this access should also lead to a higher trust level over time. We therefore hypothesize that a higher rate of accepted messages appears in our treatments with reporting in comparison to the baseline with only bilateral reputation.

***H4**: The message acceptance rate is higher in T_exo and T_endo compared to the baseline.*

We are further interested in the potentially different effects that exogenous and endogenous reports have on receiver trust. In line with our considerations regarding the lower reliability of endogenous reports, we hypothesize that receivers expect a higher probability of receiving a dishonest message and therefore accept, on average, a lower number of messages with endogenous reporting compared to computer-generated reports.

***H5**: The message acceptance rate is higher in T_exo than in T_endo.*

4. Results and discussion

We start this section by presenting the aggregated sender behavior over time and the test results regarding the respective differences across treatments and payoff scenarios. Afterwards, we have a detailed look at which factors drive sender decisions by analyzing the role of individual beliefs in our setting. Consistently, we use the same procedure for receiver behavior and beliefs before we present our qualitative analysis of the individual receiver reports in the last subsection.

4.1 Sender behavior

We illustrate the senders' behavior in Figure 2 in form of the fractions of messages sent in each of the ten rounds. In the first row, the rates of deceptive messages in each treatment are shown per payoff scenario. Observe that in almost every round of the three scenarios, the highest level of deception is reached in the baseline without reporting. On average, around 50% of the senders' messages were deceptive in scenarios 1 and 2 compared to a deception level of almost 70% in the third scenario. With computer-generated reports, the fractions of deceptive messages are substantially lower in all rounds of each scenario with around 30% deceptive message in scenarios 1 and 2 and again a higher fraction of over 50% deception in scenario 3. However, receiver reports reduce deception only to an intermediate level of 40% deceptive messages in scenarios 1 and 2 as well as 60% in scenario 3.

Figure 2
Fractions of messages chosen by senders per scenario and treatment (x-axis: rounds; y-axis: message fractions).

In order to test for general differences in deception among our three treatments, we compared the average fractions of deceptive messages sent between the first round and round nine per scenario and treatment.[7] These fractions and the respective results of the McNemar tests, which we used in order to control for the repeated measures, are presented in Table 3. In each of the scenarios, we find a signficantly lower deception rate in both treatments with reporting compared to the baseline, except for the difference between T_base and T_endo in scenario 1, which is marginally significant. The decline in deception with reporting is substantial and amounts up to over 18 percentage points. Therefore, we can confirm our hypothesis H1.

Result 1: *Compared to the baseline, reputation building through reporting systems significantly reduces the average fraction of deceptive messages.*

In order to answer the question if the two report types have different effects on deception, we compared the average message fractions between T_exo and T_endo in the last column of Table

[7] Subjects knew from the instructions that there were no interactions after round 10 and, hence, that the final report would not have any consequences for their further reputation. It is possible that this fact led to an exaggerated behavior in the last round of our experiment. Observe in Figure 1 that the highest fraction of deceptive messages in our baseline is reached in the final round in scenarios 1 and 2 in addition to a comparatively high fraction in scenario 3. On the other hand, the rate of deception decreases substantially in the last round in the reporting treatments, except for scenario 2 in T_endo. In order to account for this possible endgame effect, we excluded the last round from all our analyses of sender and receiver behavior. When we include the final round into the analysis, we obtain results that are overall similar to the ones presented in this section.

3. We find that computer reports lead to a significantly lower level of deception than receiver reports in all payoff scenarios, which confirms our hypothesis H2.

Result 2: *The fractions of deceptive messages are significantly lower with exogenous reports than with endogenous reports.*

Message type	Scenario	*Average fractions (%)*			*Treatment differences*		
		T_base	**T_exo**	**T_endo**	**T_base vs. T_exo**	**T_base vs. T_endo**	**T_exo vs. T_endo**
Deception	1	49.7	32.8	42.5	21.02***	3.56*	7.61***
	2	46.1	27.8	37.5	23.42***	5.03**	7.52***
	3	67.8	51.9	60.0	17.01***	4.40**	4.75**
Honesty	1	30.6	41.7	37.5	10.00***	4.37**	1.32
	2	36.9	48.6	46.4	9.28***	7.22***	0.33
	3	18.6	26.1	23.3	5.93**	2.60	0.79
Payoff-equalization	1	19.7	25.6	20.0	3.32*	0.01	2.99*
	2	16.9	23.6	16.1	4.88**	0.09	6.13***
	3	13.6	21.9	16.7	7.76***	1.27	3.25*

Note: *** p-value < 0.01; ** p-value < 0.05; * p-value < 0.1

Table 3
Average message fractions and McNemar test results for differences in sender behavior across treatments in each payoff scenario (rounds 1-9).

Honesty is not the only alternative to deception in our design since senders could also recommend the option with equal but Pareto-dominated payoffs. Therefore, we also run McNemar tests to analyze the differences among the rates of honest messages and find that exogenous reports lead to a signficant increase in honest behavior in all scenarios. The fractions of honest messages are significantly higher with endogenous reports in scenarios 1 and 2 but not in scenario 3. Hence, the provision of receiver reports is not a sufficient incentive to increase honesty when senders can obtain a comparatively high payoff from successful deception. These findings are in line with our previous assumptions that reports have a positive effect on sender behavior and that this effect is comparatively stronger with computer-generated reports. However, we do not find significant differences in honest messages between the two report types.

Interestingly, a substantial part of up to 26% of the senders decided to transmit payoff-equalizing messages, which is similar to the findings in the one-shot version of the game in Behnk et al. (2014). Our tests show a significantly higher fraction of payoff-equalizing messages for exogenous reports in scenarios 2 and 3 compared to the baseline but no significant effects for endogenous reports.

We ran additional tests regarding the within-subject differences across payoff scenarios. Observe in Table 3 that the highest rates of deceptive messages are reached in scenario 3. In all treatments, the differences between scenario 3 and the other scenarios are significant at the 0.01 level according to McNemar tests. While we do not find a significant difference between scenarios 1 and 2 in the baseline, significantly fewer deceptive messages are send in the mean scenario 2 compared to scenario 1 in T_exo (p=0.042). The same appears in T_endo, yet the difference is only marginally significant (p=0.089). We can summarize that senders are tempted by the

comparatively higher earnings in scenario 3 regardless of the presence of a reporting system, which is in line with the pattern that has been found in Gneezy (2005). However, the receiver's relative loss from successful deception seems to play a role for the sender's decision in our repeated game only when his behavior is reported to his future counterparts. Receiver reports show again a comparatively weaker effect. We can therefore only partly confirm our hypothesis H3.

Result 3: *When reporting systems are present, fewer deceptive message are sent in scenario 2 compared to the other scenarios. However, this difference is only marginally significant for receiver reports.*

With regard to honest messages we find a pattern that mirrors the effects of different payoff distributions on deception. The level of honesty is significantly lower in scenario 3 than in the other two scenarios in all treatments. Significances are at the 0.01 level, except for the difference between honest messages in scenarios 1 and 2 in the baseline ($p=0.020$). Furthermore, we find a significantly lower rate of payoff-equalizing messages in scenario 3 compared to in scenario 1 in the baseline ($p<0.01$). Other differences are not significant at conventional levels.

In a last step we have a look at the dynamics of sender behavior in each combination of treatments and scenarios. In order to test for possible trends, we calculate the fraction of senders who transmitted the respective message type at the group level in each round, obtaining the ordered values {0, 0.25, 0.5, 0.75, 1}, and use the Cuzick trend test (Cuzick, 1985). In T_base, the tests show a moderate upward trend for the fraction of deceptive messages in scenario 1 ($p<0.01$). Since reputation is only bilateral in the baseline, senders might have learned over time that receivers are not fully aware of their previous behavior. On the other hand, we also find a significant upward trend in scenario 3 ($p<0.01$) and marginally significant upward trends in scenario 1 ($p=0.071$) and scenario 2 ($p=0.079$) in T_endo. Furthermore, we do not find such trends in T_exo, reflecting the relatively stronger effect of computer reports on deceptive behavior.

The fractions of honest messages show a comparatively higher dynamic over time. Although both report types lead to a significantly higher level of honesty, senders gradually become less honest over the rounds in almost all treatments and scenarios with significances at conventional levels. The only exceptions are the fractions of honest messages in scenario 3 of T_base and T_exo where we do not find significant trends, possibly due to the strong monetary temptations from successful deception in this scenario which lead to an already substantially higher level of deception from the beginning. Since the fraction of honest messages decreases over time regardless of the treatment, this effect cannot be attributed to the reporting but seems to be due to the general reputation building in our design.

Result 4: *While the fractions of deceptive messages are stable over time with exogenous reports, deceptive message rates follow an upward trend with endogenous reports. On the other hand, the rates of honest messages decrease over time in almost all cases.*

A potential explanation for this effect is that some senders take the accumulated earnings of both players into account and decide to claim a bigger slice of the pie after recommending options that

favor their counterparts instead of themselves in several rounds. A second explanation would be that some senders, who are already stigmatized by reported acts of dishonesty, might expect their future counterparts to trust their messages less anyway and hence tend to refrain from being honest more often in future rounds. However, further investigations are needed to find out why fewer senders behave less honestly over time even when reports about their behavior are available. Finally, we obtain mixed trend results regarding the payoff-equalizing messages. While the fraction of senders who transmit these messages follows a significant upward trend in all scenarios of T_exo, these trends turned out to be significant only in scenario 1 of T_endo and in scenario 2 of T_base.

4.2 Sender beliefs

We turn now to an analysis of the senders' beliefs as potential determinants of their behavior in our experiment. In Table 4, we present the averages of both the first-order beliefs on message acceptance and the peer beliefs about other senders transmitting deceptive messages as well as the respective percentages of the senders' second-order beliefs about their counterparts' relative payoff expectations.

Sender beliefs	**Treatments**		
	T_base	**T_exo**	**T_endo**
First-order beliefs about receiver actions		Means	
Percentage of receivers accepting the	49.36	46.19	45.24
Second-order beliefs about relative		Percentages	
Higher or much higher than sender's payoffs	43.33	34.72	47.78
Peer group beliefs		Means	
Percentage of senders deceiving in the			
Scenario 1 (low+;low-)	21.03	18.85	20.55
Scenario 2 (low+;high-)	20.67	18.47	19.74
Scenario 3 (high+;high-)	26.23	23.66	25.20

Table 4
Sender beliefs across treatments (rounds 1-9)

A Wilcoxon matched-pairs signed-ranks test shows that the senders' expected probability of their counterpart accepting the message is moderately lower with reporting systems ($p=0.024$ for T_exo and $p=0.015$ for T_endo). There is no significant difference between the two report types. A possible explanation for why senders expect a lower acceptance rate with reporting systems is based on the fact that information about their dishonesty will potentially be stored in their report list and shown to their future counterparts. Receivers might therefore be more aware of the senders' overall dishonesty than in the baseline with only bilateral reputation, and might show less trust in the received messages. However, this decrease in expected acceptance amounts to less than five percentage points in both treatments with reporting systems.

The significant difference in second-order beliefs between T_base and T_exo ($p=0.020$) is in line with this explanation since significantly fewer senders think that their counterparts expect to gain relatively more from the message when computer reports are available. The percentage of the second-order beliefs is not significantly different between T_base and T_endo. However, it is puzzling that the fraction of senders with high second-order beliefs is significantly higher in T_endo

compared to T_exo ($p<0.01$) since there are no differences in first-order beliefs between these two treatments. In terms of relative payoff expectations it seems as if senders expect receivers to rely relatively more on the effect of reports when they are written by them or their peers. This expectation is actually supported by the general pattern in receiver behavior presented in the following subsection.

Result 5: *The expected probability of message acceptance is lower with reporting systems. More senders believe that their counterpart expects higher relative earnings with endogenous reports than with exogenous reports.*

Regarding peer beliefs we find that slightly fewer senders expect their peers to deceive in a specific scenario when reports are used, but these differences are not significant. The only exception is the difference between the baseline and computer reports in scenario 3, which is significant at the 0.01 level.

Furthermore, we used Cuzick trend tests to analyze the development of the elicited beliefs from round 1 to 9 and find that the senders' first-order and peer expectations do not change significantly in any treatment and scenario combination. On the other hand, we find a significant upward trend of the senders' second-order beliefs ($p<0.1$ in all treatments) that we present in Figure 3. This trend implies that senders expect their counterparts to trust them more over time.

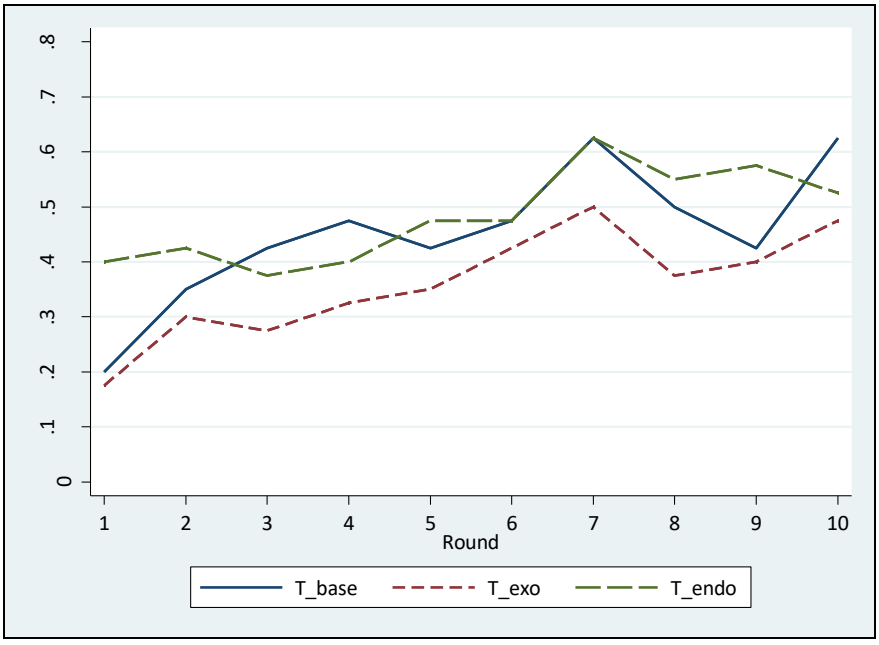

Figure 3
Senders' average second order beliefs per treatment

We turn now to an econometric model with which we analyze the effects of treatment modifications, individual beliefs and socio-economic factors on sender behavior. We use multi-level mixed-effects logistic regressions in order to account for the panel structure of our data and, at the same time, to control for correlation within the groups in which senders and receivers are matched. Since honesty was not the only alternative to deception in our design, we ran separate regressions for these two types of sender behavior. The dependent variable for deception (honesty) takes the value 1 in case the sender sent a deceptive (an honest) message and zero

otherwise. Since reports were only provided in rounds in which the whole payoff information was revealed to the receiver, a standard time variable would not cope with our aim to investigate the long-term effect of reporting sender behavior. Therefore, we use a more elaborated time variable named "revealed dishonesty", which captures a sender's public reputation by showing the number of times a dishonest action of the sender was reported, up to the respective round.[8]

All elicited belief types are included in the models. The belief variable First-order_follow captures the sender's subjective probability with which the receivers follow a message, while second-order_more is a binary variable that takes the value 1 in case a sender believes that his counterpart expects a relatively higher payoff from the message. Peer_group_lying captures the sender's subjective probability of other senders transmitting a deceptive message in a specific scenario.

Treatment and belief effects		Scenario 1 (low+;low-)		Scenario 2 (low+;high-)		Scenario 3 (high+;high-)	
Deception	T_exo	-0.664**	(0.260)	-0.791**	(0.348)	-0.561*	(0.289)
	T_endo	-0.344	(0.260)	-0.563	(0.346)	-0.290	(0.294)
	Revealed_dishonesty	0.062	(0.059)	0.013	(0.063)	-0.009	(0.060)
	First-order_follow	0.028***	(0.004)	0.025***	(0.004)	0.033***	(0.004)
	Second-order_more	0.002	(0.140)	0.129	(0.149)	-0.032	(0.144)
	Peer_group_lying	0.007***	(0.002)	0.006***	(0.002)	0.007***	(0.002)
	Female	0.814***	(0.152)	1.231***	(0.167)	0.284*	(0.152)
	Siblings_max one	0.081	(0.178)	0.312	(0.199)	-0.145	(0.184)
	Economics_Business	0.125	(0.151)	0.046	(0.163)	0.056	(0.156)
	Grant	0.129	(0.159)	-0.166	(0.172)	-0.554***	(0.163)
	Constant	-2.119***	(0.330)	-2.352***	(0.380)	-0.838**	(0.333)
Wald		91.43***		99.06***		100.71***	
Honesty	T_exo	0.608**	(0.248)	0.606**	(0.268)	0.439	(0.313)
	T_endo	0.401	(0.251)	0.599**	(0.270)	0.245	(0.317)
	Revealed_dishonesty	-0.209***	(0.061)	-0.193***	(0.059)	-0.117*	(0.070)
	First-order_follow	0.001	(0.004)	-0.003	(0.004)	-0.015***	(0.004)
	Second-order_more	0.314**	(0.138)	0.130	(0.137)	0.360**	(0.159)
	Peer_group_lying	-0.004*	(0.002)	-0.005**	(0.002)	-0.005**	(0.002)
	Female	-0.590***	(0.147)	-0.922***	(0.149)	-0.029	(0.170)
	Siblings_max one	-0.345*	(0.178)	-0.537***	(0.179)	-0.132	(0.202)
	Economics_Business	0.082	(0.148)	0.223	(0.149)	0.201	(0.175)
	Grant	-0.301*	(0.157)	-0.165	(0.156)	0.265	(0.179)
	Constant	-0.197	(0.308)	0.532*	(0.314)	-0.849**	(0.362)
Wald		47.11***		67.36***		29.91***	
N		1080		1080		1080	
Groups		30		30		30	

Note: *** p-value < 0.01; ** p-value < 0.05; * p-value < 0.1. Standard errors in parentheses.

Table 5
Multi-level mixed-effects logistic regression models for sending honest and deceptive messages (rounds 1-9)

[8] In T_endo, we included the actual receiver report based on our qualitative analysis presented in subsection 4.5, i.e. controlling for missing or false reports.

Furthermore, we use the binary variable siblings_max which takes the value 1 in case the subject has not more than one sibling. Since we invited subjects from different faculties, we also included a dummy variable for being an economics or business student. The variable grant captures if the subjects receive any kind of financial aid for their studies.

The results of the regressions are reported in Table 5. By controlling for other potential determinants of sender behavior, such as individual beliefs, we get a clear picture of the report types' efficacy in reducing deception. While computer reports lead to a significant decrease in the fraction of deceptive messages in scenarios 1 and 2 in addition to a marginally significant effect in the large stakes scenario 3, receiver reports do not show such an impact on deception in any of the scenarios. Consistently, the provision of computer reports significantly increased the rate of honest messages in the first two scenarios but not in scenario 3, while receiver reports only show a significant effect in the mean scenario 2. When we compare sender behavior between the different payoff scenarios in T_exo, we find that the reduction in deception is stronger in scenario 2 than in the other payoff scenarios. Regarding honesty, there is almost no difference between scenario 1 and 2 in T_exo.

Our time variable regarding revealed dishonesty does not affect the sending of deceptive messages in any scenario. On the other hand, the relative probability of senders being honest decreases over time with the number of reported dishonesty in scenarios 1 and 2. This effect is marginally significant in scenario 3. These results are in line with the time trends we observed in the previous subsection, i.e. that the fraction of deceptive messages exhibits an upward trend only in some treatment and scenario combinations whereas honesty decreases in almost all cases.

Result 6: *Exogenous reports are a more effective measure to reduce deception in favor of increasing honesty than endogenous reports also when we control for individual beliefs.*

We now have a look at the coefficients of the two counterpart-related belief types, which show an interesting discrepancy between deception and honesty. Observe in Table 5 that deceptive senders are affected by their first-order beliefs in all cases and that second-order beliefs do not correlate with their decision to deceive in any of the scenarios. By contrast, honest senders are indeed influenced by their beliefs about their counterparts' payoff expectations in scenario 1 and 3, while first-order beliefs only affect them, to a comparatively lower extent, in scenario 3. This discrepancy implies that selfish senders are mostly driven by strategic considerations in terms of the actual message acceptance since receivers have a say in the outcome of the game, whereas honest senders tend to incorporate their counterparts' expectations when making decisions, which is closely related to guilt aversion. The senders' peer beliefs correlate significantly with their behavior in the sense that the relative probability of sending a deceptive message is higher when a sender expects others to deceive the receiver in the respective scenario, and vice-versa for honesty. This effect is only marginally significant regarding honesty in the first scenario. However, the coefficients show that these beliefs have a relatively small effect on sender behavior in our experiment.

Result 7: *Deceptive senders are driven by first-order belief, which play an important role for strategic considerations, while honest senders are mainly influenced by second-order beliefs, which are related to guilt aversion.*

Finally, we find that socio-economic factors have only a limited impact on sender behavior. The only effects that are significant at conventional levels consist of subjects being honest with a lower relative probability in the mean scenario 2 when they have more than one sibling and senders being less inclined to send deceptive messages in the high stakes scenario 3 when they receive a grant. Surprisingly, we find that the relative probability of sending deceptive (honest) messages increases (decreases) significantly for female subjects. This finding is in contrast to the results of the one-shot version of our game in Behnk et al. (2014), where no significant gender differences were found at conventional levels, similar to Childs (2012). Furthermore, the finding is in contract with a series of studies that find a higher rate of deceptive messages for men in sender-receiver games, such as Dreber and Johannesson (2008) or Houser et al. (2012). When we compare the average rates of messages between men and women, we find more women lying in each scenario and treatment combination except for scenario 3 in T_exo and T_endo, where fractions are almost identical. Since the gender effect is also present in the baseline, it cannot be attributed to the different report types. We propose that the gender effect is due to the repeated game structure but this suggestion needs further exploration in future studies.

4.3 Receiver behavior

We turn now to the receiver behavior and illustrate the development of the acceptance rates in Figure 4. On average, 44.4% of the receivers accepted the transmitted message in T_base.[9] The average acceptance rate is only slightly higher in T_exo, 49.7%, but the difference is not significant according to a McNemar test. On the other hand, receivers accept on average significantly more messages in T_endo compared to the baseline ($p<0.01$). Therefore, we can confirm our Hypothesis H4 only partly. We do not observe a significant difference between the two report types. These results imply that receivers do not anticipate the relatively stronger effect of computer reports on deceptive behavior and rather seem to trust those reports that were written by them and their peers. Hence, we cannot confirm our hypothesis H5.

Result 8: *The acceptance rate is significantly higher than in the baseline only when endogenous reports are provided.*

[9] In order to make the analyses comparable to the ones used for sender behavior and to control for a possible endgame effect, we excluded the last round from our tests and regressions models.

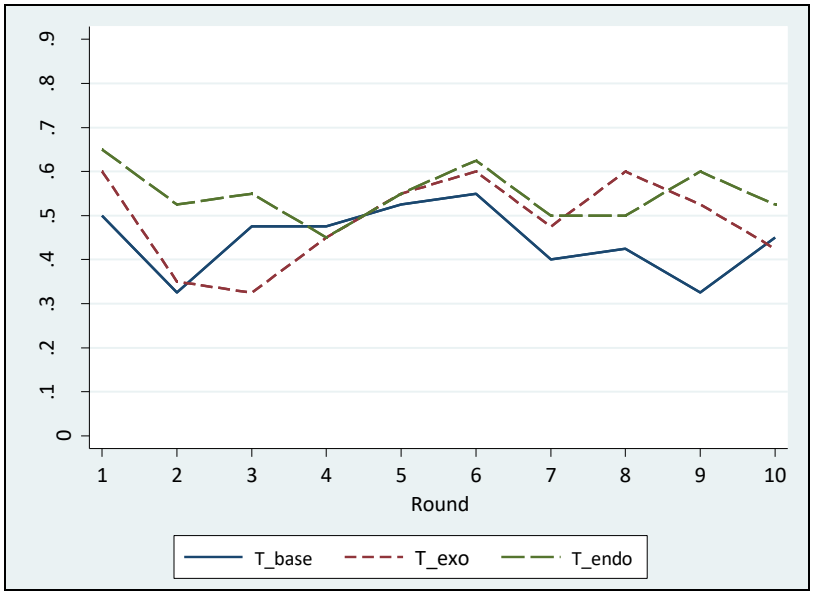

Figure 4
Acceptance rates per treatment

With regard to the development of receiver behavior over time, we calculated the fractions of receivers accepting a message at the group level in each round and used the Cuzick trend test. We find that acceptance rates follow a significant upward trend from rounds 1 to 9 in T_exo ($p<0.01$) while the fractions are stable in T_base and T_endo. Although the overall message acceptance is only higher with receiver reports, the computer reports seem to gradually enhance the level of trust over time. Therefore, it is possible that computer reports reach a substantially higher trust level with an increased number of interactions among the group members.

4.4 Receiver beliefs

In line with our analysis of sender behavior, we also explore the impact of individual beliefs on receiver decisions in our experiment. Table 6 shows the means of the receivers' first-order beliefs regarding percentages of honest senders, which vary between 38% and 45% among the treatments. Both mean values in T_exo and T_endo are significantly higher compared to the baseline ($p<0.01$ in both cases) but there is no significant difference between the two report types according to Wilcoxon matched-pairs signed-ranks tests. These results show that receivers assign both reporting systems an equally positive effect on sender actions.

We also elicited the receivers' second-order beliefs in terms of their counterparts' relative payoff expectations. On average between two thirds and three quarters of the receivers believe that the senders expect to gain relatively more from the message they sent. McNemar tests show that only the difference between T_base and T_endo is significant at conventional levels ($p=0.033$). This finding shows that receivers assign endogenous reports a positive impact on the senders' intentions and, hence, reflects the relatively higher acceptance rate in T_endo. Furthermore, receivers expect on average that only a small fraction of the other players in their own role, between 14% and 15%, accept the message they received. These fractions are similar across the three treatments according to Wilcoxon matched-pairs signed-ranks tests but substantially lower than the actual acceptance rates we presented in Table 5.

Receiver beliefs	Treatments		
	T_base	T_exo	T_endo
First-order beliefs about sender actions		Means	
Percentage of senders sending honest	37.91	43.13	44.53
Second-order beliefs about relative		Percentages	
Higher or much higher than receiver's payoffs	73.89	71.39	66.39
Peer group beliefs		Means	
Percentage of receivers accepting the	13.95	14.85	14.87

Table 6
Receiver beliefs across treatments (rounds 1 to 9)

Cuzick trend tests are used to analyze possible trends in the belief development over time. We find that all receiver beliefs are stable from rounds 1 to 9 except for a significant decrease of first-order beliefs in T_endo (p=0.014), i.e. with an increasing number of interactions, receivers think that a lower percentage of senders will transmit an honest message when the reports are endogenous. This result is in line with the lower impact of receiver reports on sender behavior.

In the next step, we include treatment modifications, individual beliefs and socio-economic variables in a multi-level mixed-effects logistic regression model as we did in the analysis of sender behavior. The results are presented in Table 7. It turns out that receivers are mainly driven by their beliefs when they decide whether to accept a message, while the provision of reports does not correlate with the message acceptance in none of the two systems. Accordingly, receivers do not seem to care about the sender's long-term reputation in terms of the number of dishonesty revelations.

Treatment and belief effects on message acceptance		
T_exo	0.081	(0.201)
T_endo	0.220	(0.204)
Revealed_dishonesty	-0.090	(0.062)
First-order_honest	0.028***	(0.004)
Second-order_more	-1.903***	(0.176)
Peer_group_accept	0.007**	(0.003)
Female	-0.035	(0.150)
Siblings_max one	-0.091	(0.164)
Economics_Business	-0.006	(0.149)
Grant	-0.094	(0.157)
Constant	0.249	(0.329)
Wald	198.67***	
N	1080	
Groups	30	

Note: *** p-value < 0.01; ** p-value < 0.05; * p-value < 0.1.
Standard errors in parentheses.

Table 7
Multi-level mixed-effects logistic regression models for the receiver's behavior and beliefs (rounds 1-9)

First-order beliefs about the sender's honesty have a significantly positive impact on receiver behavior in the sense that receivers are more likely to accept a message when they think that senders send an honest message. Consistently, the relative probability of message acceptance

decreases significantly when receivers think that their counterpart expects to obtain relatively higher earnings from the message. These second-order beliefs are the strongest determinant of acceptance rates. Furthermore, receivers are more likely to trust a sender when they expect their peers to accept the messages. We do not find significant gender differences among receivers. The same is true for the remaining socio-economic variables.

Result 9: *The difference in the receiver behavior between the two reporting systems is mainly driven by their second-order beliefs.*

4.5 Qualitative analysis of receiver reports

In Table 8 we present the results of our qualitative analysis of the receiver reports. Altogether, the sender behavior was revealed to the affected receivers 212 times in T_endo. Strikingly, the reports were not adequately provided by the receivers in almost one third of the reporting possibilities. We found that nearly every fifth affected receiver did not provide a report at all, leading to a blank line for the respective round in the sender's personal report list. One possible explanation for this phenomenon is that some receivers exhibit an aversion to damaging the reputation of others in the long run, since 85% of the missing reports appeared after the revelation of a dishonest message, although only in 67% of the revelations the senders were actually dishonest.

Reporting possibilities	**212**	**100%**
Missing, incorrect or unrelated reports	**68**	**32,1%**
No report provided	40	18.9 %
(missing report after revealed dishonesty)	34/40	85%
Incorrect reports	25	11.8%
Unrelated information	3	1.4%
Other information		
Emotional wording and/or emoticons	25	11.8%
Ethical comment	25	11.8%
Justification of own decision	6	2.8%
History-related information	16	7.5%
Reference to previous reports	3	1.4%
Sender number mentioned	13	6.1%
Scenario-specific information	17	8.0%
Payoff-equalization reported	17	8.0%

Table 8
Qualitative analysis of receiver reports

Another 12% of the receivers provided an incorrect report, i.e. they wrote that the sender transmitted a dishonest message when in fact he sent an honest message and vice-versa. We cannot rule out that these reports were given because a receiver confused the payoffs shown to him in the end of the round. However, the majority of these reports were given after the receiver already provided other reports in a correct way, which demonstrates that she understood the general procedures of the game. Finally a small fraction of 1.4% of the receivers provided completely unrelated information in their reports, for instance by referring to the sender's risk attitudes.

Result 10: *A substantial fraction of all possible receiver reports was either not provided or included unrelated or incorrect information.*

Apart from the inadequate reporting, we also analyzed the way how reports were written and how the information was framed. In general, the language used in the reports was overall quite factual. Receivers used an emotional wording only in 12% of their reports. Around 15% of the senders made an ethical comment or justified their own decision, for instance, by commenting on the spitefulness of deception in scenario 2. Since the computer reports were standardized and did not include any information about the scenarios' degree of spitefulness, the receiver reports include additional information that might have intimidated selfish senders especially regarding scenario 2 and, hence, might have played a role in the stronger effect of endogenous reports in this scenario compared to the other scenarios. Other receivers mentioned their indifference regarding selfish sender behavior in the first scenario, where the payoff misalignment was relatively low, or even approved such a behavior. However, none of the subjects used swearwords or insulted other participants, which they knew was prohibited and would result in final earnings of zero. Furthermore, receivers were quite focused on the sender behavior in the present round. Not even 10% of them related their information to previous interactions with the same sender or to other reports in his personal list. Finally, 16% of the receivers provided payoff-specific information in their report and showed thereby that they learned at least some of the different payoff distributions over time.

5. CONCLUSION

In this study, we conducted an experiment to investigate the effect of reporting systems on deceptive behavior. We used a repeated sender-receiver game with various payoff scenarios to compare two different report types, exogenous ones that are generated by the computer in a standardized way and endogenous reports in form of an individual text written by the receivers. We were mainly interested in answering the question if the reliability or the personal character of reports more effective in reducing deception.

Our analysis of sender behavior shows that the use of both report types leads to a significant lower level of deception compared to the baseline without reports. Furthermore, we find that computer reports lead to a significantly lower deception level than receiver reports in each payoff scenario. Regarding the development over time, our analysis shows that deception even increases with receiver reports, while the rate of deceptive messages does not change when computer reports are provided. It turns out that deceptive behavior is rather driven by the strategic first-order beliefs whereas honesty is mostly determined by the guilt aversion-related second-order beliefs.

Receivers show relatively more trust compared to the baseline when endogenous reports are provided, whereas exogenous reporting does not have a significant effect on trust. This finding reveals an interesting discrepancy between the agents' and theirs principals' perception of the two report types. On the other hand, we find that computer reports gradually enhance the level of trust over time and conclude that receivers might realize in the long run that the reliability of the

exogenous reports lead to comparatively less deception. Receiver behavior is mainly driven by changes in their second-order beliefs.

A qualitative analysis of the receiver reports provides a potential explanation for the lower deception level with computer reports. It turns out that a substantial fraction of the receivers did not use the opportunity to report dishonest senders in a proper way. We suppose that the comparably lower impact of the endogenous reports is due to an anticipation of the high number of missing, incorrect or unrelated reports by the senders. Interestingly, the high number of missing reports after successful deception implies that some receivers experience an aversion to damage the reputation of others.

Altogether, we can conclude that the higher reliability of the exogenous reports has a higher impact on pro-social behavior in our setting than the personal character of endogenous reports. Strikingly, receivers seem to trust the wrong system. Especially in the light of the popular consumer ratings in E-commerce (Dellarocas, 2003), we propose that the importance of an externally regulated reporting system should not be underestimated.

Since we let the subjects play the repeated sender-receiver game over 10 rounds and implemented a mechanism that allowed for reporting only in 50% of the cases, we plan to investigate the development of reputation and its long-term effect of deceptive behavior with a larger time frame and to compare the cost-effectiveness of both systems in a future study. Once an infrastructure is provided, endogenous reporting systems might not be as costly as the establishment of an external authority that is in charge of preparing the reports. However, this advantage comes at the cost of less reliable reporting and, hence, can reduce deception to a substantially lower extent than the exogenous reporting system. Another question which needs further exploration is why the probability of subjects to deceive others is relatively higher for females compared to men in this repeated version of the game.

REFERENCES

Abraham M., Grimm, V., Neeß, C., Seebauer, M., 2016. Reputation formation in economic transactions. Journal of Economic Behavior & Organization 121, 1-14.

Akerlof, G.A., 1970. The Market for 'Lemons': Quality uncertainty and the market mechanism. The Quarterly Journal of Economics, 84(3), 488-500.

Angelova, V., Regner, T., 2013. Do voluntary payments to advisors improve the quality of financial advice? An experimental sender-receiver game. Journal of Economic Behavior and Organization 93, 205-218.

Behnk, S., Barreda-Tarrazona, I., García-Gallego, A., 2014. The role of ex post transparency in information transmission - An experiment. Journal of Economic Behavior and Organization 101, 45-64.

Bicchieri, C., Sontuoso, A., 2015. I cannot cheat on you after we talk. In Peterson, M. (Ed.), The Prisoner's Dilemma. Cambridge: Cambridge University Press.

Blume, A., DeJong, D.V., Kim, Y.G., Sprinkle, G.B., 1998. Experimental evidence on the evolution of meaning of messages in sender-receiver games. American Economic Review 88, 1323-1340.

Blume, A., DeJong, D.V., Neumann, G. R., Savin, N.E., 2002. Learning and communication in sender-receiver games: an econometric investigation. Journal of Applied Econometrics 17(3), 225-247.

Brier, G., 1950. Verification of forecasts expressed in terms of probability. Monthly Weather Review 78(1), 1-3.

Cain, D.M., Loewenstein, G., Moore, D.A., 2011. When sunlight fails to disinfect: Understanding the perverse effects of disclosing conflicts of interest. Journal of Consumer Research 37(5), 836-857.

Charness, G., Dufwenberg, M., 2006. Promises and partnership. Econometrica 74(6), 1579-1601.

Charness, G., Dufwenberg, M., 2010. Bare promises: An experiment. Economics Letters 107(2), 281-283.

Childs, J., 2012. Gender differences in lying. Economics Letters 114(2), 147-149.

Church, B.K., Kuang, X., 2009. Conflicts of interest, disclosure, and (costly) sanctions: Experimental evidence. The Journal of Legal Studies 38(2), 505-532.

Cuzick, J., 1985. A wilcoxon-type test for trend. Statistics in medicine 4(4), 543-547.

Danilov, A., Sliwka, D., forthcoming. Can Contracts Signal Social Norms - Experimental Evidence. Management Science.

Dellarocas, C., 2003. The digitization of word of mouth: Promise and challenges of online feedback mechanisms. Management Science 49(10), 1407-1424.

Dellarocas, C., Wood, C.A., 2008. The sound of silence in online feedback: Estimating trading risks in the presence of reporting bias. Management Science 54(3), 460-476.

Dreber, A., Johannesson, M., 2008. Gender differences in deception. Economics Letters 99(1), 197-199.

Fischbacher, U., 2007. z-Tree: Zurich toolbox for ready-made economic experiments. Experimental Economics 10(2), 171-178.

Fischbacher, U., Föllmi-Heusi, F., 2013. Lies in disguise - an experimental study on cheating. Journal of the European Economic Association 11(3), 525-547.

Gibson, R., Tanner, C., Wagner, A.F., 2013. Preferences for truthfulness: Heterogeneity among and within individuals. American Economic Review, 103, 532-548.

Gino, F., Krupka, E.L., Weber, R.A., 2013. License to cheat: Voluntary regulation and ethical behavior. Management Science 59(10), 2187-2203.

Gneezy, U., 2005. Deception: The role of consequences. American Economic Review 95(1), 384-394.

Gneezy, U., Rockenbach, B., Serra-Garcia, M., 2013. Measuring lying aversion. Journal of Economic Behavior & Organization 93, 293-300.

Greiner, B., 2004. An online recruitment system for economic experiments. In: Kremer, K. and Macho, V. (Eds.). Forschung und wissenschaftliches Rechnen, GWDG Bericht 63, Gesellschaft für wissenschaftliche Datenverarbeitung, Göttingen, 79-93.

Houser, D., Vetter, S., Winter, J., 2012. Fairness and cheating. European Economic Review 56(8), 1645-1655.

Kimbrough, E.O., Rubin, J., 2013. Sustaining group reputation, Discussion Papers dp13-02, Department of Economics, Simon Fraser University.

Kimbrough, E.O., Sheremeta, R.M., Shields, T.W., 2014. When parity promotes peace: Resolving conflict between asymmetric agents. Journal of Economic Behavior & Organization 99, 96-108.

Koch, C., Schmidt, C., 2010. Disclosing conflicts of interest - Do experience and reputation matter? Accounting, Organizations and Society 35(1), 95-107.

López-Pérez, R., Spiegelman, E., 2013. Why do people tell the truth? Experimental evidence for pure lie aversion. Experimental Economics 16(3), 233-247.

Muehlheusser, G., Roider, A., 2008. Black sheep and walls of silence. Journal of Economic Behavior & Organization 65, 387-408.

Nyarko, Y., Schotter, A., 2002. An experimental study of belief learning using elicited beliefs. Econometrica 70, 971-1005.

Palfrey, T.R., Wang, S., 2009. On eliciting beliefs in strategic games. Journal of Economic Behavior & Organization 71(2), 98-109.

Peeters, R., Vorsatz, M., Walzl, M., 2012. Beliefs and truth-telling: A laboratory experiment. University of Innsbruck, Working Papers in Economics and Statistics 2012-17.

Peeters, R., Vorsatz, M., Walzl, M., 2013. Truth, trust, and sanctions: On institutional selection in sender-receiver games. Scandinavian Journal of Economics 115(2), 508-548.

Rockenbach, B., Milinski, M., 2006. The efficient interaction of indirect reciprocity and costly punishment. Nature 444(7120), 718-723.

Ruffle, B.J., Tobol, Y., 2014. Honest on Mondays: Honesty and the temporal separation between decisions and payoffs. European Economic Review 65(C), 126-135.

Sánchez-Pagés, S., Vorsatz, M., 2007. An experimental study of truth-telling in a sender-receiver game. Games and Economic Behavior 61(1), 86-112.

Sánchez-Pagés, S., Vorsatz, M., 2009. Enjoy the silence: an experiment on truth telling. Experimental Economics 12(2), 220-241.

Sobel, J., 1985. A theory of credibility. The Review of Economic Studies 52(4), 557-573.

Sutter, M., 2009. Deception through telling the truth?! Experimental evidence from individuals and teams. Economic Journal 119(534), 47-60.

White, L.J., 2010. Markets: The credit rating agencies. The Journal of Economic Perspectives 24(2), 211-226.

Zylbersztejn, A., 2014. Strategic signaling or emotional sanctioning? An experimental study of ex post communication in a repeated public goods game. Department of Economics Working Papers wuwp161, Vienna University of Economics and Business, Department of Economics.

APPENDIX A

Instructions for the experimental subjects (translated from Spanish)

Welcome to this experiment, we greatly appreciate your participation. From this moment on, please switch off your cell phone and do not talk or communicate in any way with the other participants. Read these instructions carefully and raise your hand if you have any questions during the session. One of the officials of the experiment will answer your questions individually.

Your decisions in this experiment will allow you to earn a certain amount of money that we will pay you in cash at the end of the session.

You will be a player in a two-player game which will be played for ten rounds. Therefore, the computer will assign the players to groups of eight and you will part of one of the groups. Within the group, four of you (labeled I to IV) will be randomly assigned the role of "Player 1" and the other four (also labeled I to IV) the role of "Player 2". Your role and your number will be the same until the session ends. In each round, the computer will randomly match you with one of the other participants with the opposite role in your group. None of the players will ever know the real identity of the other players, only the number of the player he/she is matched with in a round.

You will make your decisions during the game through the computer in front of you. After the ten rounds, the experiment will end and you will be asked to fill in a short questionnaire.

Decision Making Player 1

In each round, we will present three scenarios to player 1, each one them contains three options. Each option consists of a payoff for both players. This is the general structure of the options in each scenario that will be presented to Player 1:

Option A: Player 1 receives ... euros and player 2 receives ... euros.

Option B: Player 1 receives ... euros and player 2 receives ... euros.

Option C: Player 1 receives ... euros and player 2 receives ... euros.

We will present to Player 1 (and only to him/her) the payoffs for both players of each option and in each scenario (the order of the options is at random). By contrast, Player 2 will not get this information. Player 1's task is to choose one of the following three messages that will be sent to Player 2 afterwards:

Message 1: Option A will earn you more money than the other two options.

Message 2: Option B will earn you more money than the other two options.

Message 3: Option C will earn you more money than the other two options.

Remember that there are three scenarios. In each scenario, Player 1 has to decide which message he/she wants to be sent to Player 2 in case this scenario will be selected.

After Player 1 has chosen a message for each scenario, the computer will randomly select one of the scenarios. This scenario will then be implemented and the specific message that Player 1 chose for this scenario will be sent to Player 2. From this moment on, it will depend on the decision of Player 2 which of the three corresponding options will be implemented and, according to this, which amount of money both players will earn.

Decision Making Player 2

Player 2 knows about the three options in the selected scenario but he/she knows nothing about the payoffs associated with each option. He/she receives the message that Player 1 chose for the implemented scenario. After receiving the message, Player 2 has to decide whether to "accept" or "reject" the message:

- To "accept" the message means that Player 2 accepts the information of the message and that the option mentioned in the message determines the earnings of the two players.
- To "reject" the message means that Player 2 does not want the option mentioned in the message but one of the other option to determine the earnings of both players.

Therefore, if Player 2 accepts the message, the option in the message will be implemented and determines the payoffs of the players in this round. In the case that Player 2 rejects the message, one of the remaining options of the selected scenario will be randomly implemented by the computer in order to determine the earnings of both players in this round.

[***Exogenous treatment**:*

From the second round on, and before making his/her decision, we will show a display to Player 2. In this display we will present information about Player 1's decisions to send honest or dishonest messages to his/her respective counterparts in the previous rounds. In those rounds in which his/her behavior has not been revealed, which happens with 50% probability, no information will be shown.

There are three possible reports with which the computer classifies Player 1's actual behavior in each of the previous rounds:

"Player 1 has sent an honest message"

"Player 1 has sent a dishonest message"

"No information" (for rounds in which the behavior of Player 1 has not been revealed)

The following display is an example of how the information is presented to Player 2 (referring to the first six rounds including the three possible reports mentioned before; the displayed order is just an example):

Round	Player 1 ...
1	- No information -
2	- ... has sent a dishonest message -
3	- ... has sent an honest message -
4	- No information -
5	- ... has sent a dishonest message -
6	- ... has sent an honest message -

Player 1 will also get this information at the beginning of each round.]

[***Endogenous treatment**:*

From the second round on, and before making his/her decision, we will show a display to Player 2. In this display we will present information about Player 1's decisions to send honest or dishonest messages to his/her respective counterparts in the previous rounds. The presented reports have been written voluntarily by the Player 2s with whom this specific Player 1 was matched in the previous rounds.

In each round in which his/her behavior has not been revealed, which happens with 50% probability, there is no information available and as a consequence, the respective Player 2 did not have the possibility to write a report. Additionally, Player 2 has the possibility to not write a report. In this case there is also no information available.

The following display is an example of how the information is presented to Player 2 (referring to the first six rounds including the available reports; the displayed order is just an example):

Round	Player 1 ...
1	- No information -
2	*Player 2's report in round 2*
3	*Player 2's report in round 3*
4	- No information -
5	*Player 2's report in round 5*
6	*Player 2's report in round 6*

Player 1 will also get this information at the beginning of each round.

Accordingly, after making his/her decision, Player 2 has the possibility to write a text with a maximum of 140 characters in each round that informs the following Player 2s about whether Player 1 has sent an honest or a dishonest message in the current round. The text can be written freely and individually, taking into account that any swearwords or insults are totally prohibited. In case a participant does not stick to this rule, his/her earnings will be zero.]

Earnings

In each round you will answer some short questions once you have made your decisions but before your earnings are shown on the screen. We will pay you an additional amount depending on the precision of your answers.

After that, at the end of each round,

- Player 1 will receive information about the acceptance or rejection of her message, the implemented option corresponding to the scenario that was selected by the computer and the earnings of both players.
- In principle, Player 2 will only receive information about her own payoff. Furthermore, a possibility exists that the computer decides to provide additional information to player 2 about all potential payoffs for both players of each option in the implemented scenario. The probability of this happening is 50%.

After the end of the ten rounds, we will pay you your earnings from two rounds that will be randomly chosen by the computer. On the last screen we will present your final earnings that we will pay you anonymously in cash at the end of the session.

Do you have any questions about these instructions? If so, please raise your hand. If you do not have any questions, remain silent until you get instructions from the experimenter.

Summary

- The game will be played for ten rounds. In each round each Player 1 will be matched randomly with one Player 2 from the same group (which consists of eight players from this session).

- There are three scenarios, each one them contains three options. Player 1 knows the payoffs for both players of each option. Player 2 does not receive this information.

- After Player 1 has chosen a message for each scenario, the computer will randomly select one of the scenarios and the corresponding message will be sent to Player 2.

[***Exogenous treatment****:*

- Player 2 will receive the message and, from the second round on, a display with reports about Player 1's behavior in the previous rounds in which his/her behavior has been revealed.]

[***Endogenous treatment****:*

- Player 2 will receive the message and, from the second round on, a display with reports about Player 1's behavior in the previous rounds in which his/her behavior has been revealed. The reports have been written by the respective Player2s.]

- Player 2 decides whether to accept or reject the received message. In the case of acceptance the option mentioned in the message will be implemented. In case of rejection, the computer will randomly select one of the remaining options in order to determine the earnings of both players.

- With 50% probability, Player 2 will receive information about all options and payoffs in the implemented scenario at the end of each round.

[***Endogenous treatment****:*

In this case, after making his/her decision, Player 2 has the possibility to write a text that informs the following Player 2s about if Player 1 has sent an honest or a dishonest message in the current round.]