Deception: The role of expectations. A Comment on Gneezy's “Deception: The Role of Consequences”*

Matthias Sutter[#]

Abstract. Gneezy (2005) has shown that the willingness of subjects to lie in a cheap-talk sender-receiver game is a matter of weighing costs and benefits. Subjects deceive others more often the higher the possible gains from deception and the lower the loss for the deceived. Analyzing senders’ expectations about receivers’ actions I show that even telling the truth should be classified as deception if the sender chooses the true message with the expectation that the receiver will not follow the sender’s (true) message. The data reveal a large degree of ‘sophisticated’ deception through telling the truth, both for individual and team senders.

JEL-classification: C72, C91, D82

Keywords: Deception, Expectations, Team decision making, Experiment

This version: 26th June 2006

* I would like to thank Martin Dufwenberg, Uri Gneezy, Stefan Haigner, Navin Kartik, Martin Kocher, Stephan Kroll, Topi Miettinen and Jörg Oechssler as well as two anonymous referees and Vincent Crawford for very helpful comments. Olga Mellizo was of great help in language editing. Financial support from the Austrian Science Foundation (FWF-Project No. P16617), from the Max Planck Society (for which the author was working when beginning this project) and from the Center of Experimental Economics at the University of Innsbruck (sponsored by *Raiffeisen Landesbank Tirol*) is gratefully acknowledged.

University of Cologne and University of Innsbruck. Address for correspondence: University of Cologne, Department of Economics (Staatswissenschaftliches Seminar), Albertus-Magnus-Platz, D-50923 Köln, Germany. e-mail: msutter@uni-koeln.de or matthias.sutter@uibk.ac.at

"Two Jewish merchants in Poland bump into each other at the train station at Warsaw one morning. They are competitors in the same trade, so they eye each other suspiciously, and one of them asks, 'So where are you traveling today?' 'To Lodz' comes the cautious answer. 'To Lodz, eh?' the first says skeptically. 'I know very well that you are only telling me that to make me think that you are actually going to Krakow. But – I happen to know that you really are *going to Lodz …' And after a little pause he adds: 'So tell me: Why are you trying to deceive me?' "*[1]

1 Introduction

In a recent paper, Uri Gneezy (2005) has shown that the willingness of subjects to lie to others is a matter of weighing costs and benefits. Subjects deceive others more often the higher the possible gains from deception and the lower the associated losses for the deceived subjects. Deception has been defined and measured by Gneezy as the relative frequency with which the sender in a cheap-talk sender-receiver game sends the wrong message about the consequences of a given action for the receiver.[2]

In this comment, I argue that Gneezy's measure of the frequency of deception does not capture the full amount of deception. Rather, by analyzing the interplay between senders' actions and expectations, I show that even telling the truth should be classified as intended deception[3] if the sender expects that the receiver will not follow the sender's (true) message.

[1] I thank Navin Kartik for bringing this folklore joke to my attention. It captures this paper's central issue, i.e. deception through telling the truth.

[2] In fact, Gneezy (2005) discusses also several other definitions of deception and lies. Yet, when analyzing his data he uses the definition at the top of p. 386 that implies the definition to which I refer in the opening paragraph. More generally, deception is always an issue when one party has an incentive to misrepresent information to another party. In the case of a conflict of both parties' interests, game theory predicts that zero-cost messages are not informative in such situations (Vincent Crawford and Joel Sobel, 1982). Recently, Crawford (2003) has shown why such messages are nevertheless sent and are, in fact, influential for behavior. If there are two types of players (sophisticated ones and boundedly rational ones) then deception can be used by the former type to exploit the latter type.

[3] In the following I will focus on the intention to deceive, even if I speak of deception. Note, however, that one can make a distinction between the *intention to deceive* and *deception*. The latter refers to a successful attempt at misleading the receiver, whereas the former concerns intentions alone – irrespective of the actual outcome of an interaction.

My experimental data corroborate Gneezy's general findings with respect to the influence of costs and benefits on the frequency of deception. However, my results also indicate that Gneezy reports only a lower bound of the frequency of deception in strategic interaction. By taking into account what I will call "sophisticated" deception through telling the truth, the frequency of deception increases by about 20 to 30 percentage points. Given this substantial increase, it seems warranted not to confine deception to simply telling a lie.

The rest of the paper is organized as follows: Section 2 introduces the experimental design and procedure. Section 3 presents the experimental results. Section 4 reports on a robustness test of my classification of deception by observing the discussion in teams before teams send a message. Section 5 relates the paper to other recent comments on Gneezy's seminal work and concludes the paper.

2 Experimental design

The cheap-talk game. There are two players of which only the sender is informed about the monetary consequences of two different options, A and B. The sender's only choice in the game is to send one of the following two messages to the receiver:

Message A: "Option A will earn you more money than option B."

Message B: "Option B will earn you more money than option A."

Knowing the sent message, the receiver has to pick one of the two options which is then implemented for payment. Note that after his decision, the receiver is only informed about the monetary payoff from the chosen option, but not about the payoff in the other option, nor about the sender's payoff in any of the options. Hence, the receiver cannot judge whether the sender has told the truth or not.

Procedure. Table 1 summarizes the 3 different treatments (T1, T2, T3) used in this experiment (and also in Gneezy's original paper). Treatments differ with respect to the possible gains for the sender, and the possible losses for the receiver. Note that Option B is always more favorable for the sender than Option A.

TABLE 1 – PAYOFFS (IN €) IN THE DIFFERENT TREATMENTS

Treatment	Option	Payoff to Sender	Payoff to Receiver
T1	A	5	6
	B	6	5
T2	A	5	15
	B	6	5
T3	A	5	15
	B	15	5

At the beginning of the experimental sessions subjects received written instructions (see the Appendix) and clarifying questions were answered privately by the experimenter. After senders had made their decisions and while receivers had time to consider their decision, senders were asked to answer the following two questions in turn.[4]

Question Q1: "Which option do you expect the receiver to choose?"

Question Q2: "Out of 100 receivers, how many do you think follow the sender's message on average?"

Experimental sessions were run from June to August 2005 at the Max Planck Institute of Economics in Jena. A total of 570 participants yielded 96 sender-receiver-pairs each in treatments T1 and T2, and 93 pairs in T3. Recruitment was done via ORSEE (Ben Greiner, 2004), and sessions were computerized using zTree (Urs Fischbacher, 1999). Average session length was 25 minutes.

[4] Answers to these questions were not monetarily incentivized. There is evidence that eliciting expectations with or without monetary rewards for accuracy does not yield significantly different results (see David Grether, 1992, or Colin Camerer and Robin Hogarth, 1999).

3 Experimental results

Table 2 reports the most important results of the experiment. The figures in row [1] indicate the relative frequency of lying, i.e. sending the incorrect message B. It is lowest in treatment T2 – where it costs the receiver 10 € but gains the sender only 1 € more – intermediate in T1, and highest in T3 – where the sender can gain 10 € more at the same cost for the receiver. Considering all three treatments, the frequencies of sending message B are significantly different across treatments ($\chi^2 > 10$, *d.f.* = 2, $p < 0.01$), as has also been found in Gneezy's experiment. However, there is no significant difference between treatments T1 and T2 in my experiment, even though the overall averages point in the same direction as in Gneezy's experiment. Hence, it seems that in my subject pool the treatment differences with respect to own payoffs in option B (compare T3 vs. T1 and T2) are relatively more important than the treatment differences with respect to the receiver's payoffs in option B (compare T1 vs. T2 and T3).

TABLE 2 – EXPERIMENTAL RESULTS (RELATIVE FREQUENCIES)

Treatment	T1	T2	T3
	$N = 96$	$N = 96$	$N = 93$
[1] Sender chooses message B	0.44	0.35	0.59
[2] Q1: Sender expects the receiver to implement chosen message	0.73	0.70	0.66
[3] Receiver implements message	0.67	0.75	0.74
[4] Sender expects receiver to choose option B	0.67	0.61	0.89

Row [2] of Table 2 reports the answers of senders to question Q1. It shows that across treatments senders expect the receiver to implement the option that is indicated in the chosen message as yielding the higher payoffs in about 70% of cases. There is no significant difference in these expectations across treatments. Row [3] of Table 2 then shows the relative frequency of receivers actually following the sender's message, and there is again no significant treatment effect.[5] Across treatments receivers follow the sender's message in 72% of cases. Thus, senders' expectations match receivers' actions remarkably well in the aggregate.[6]

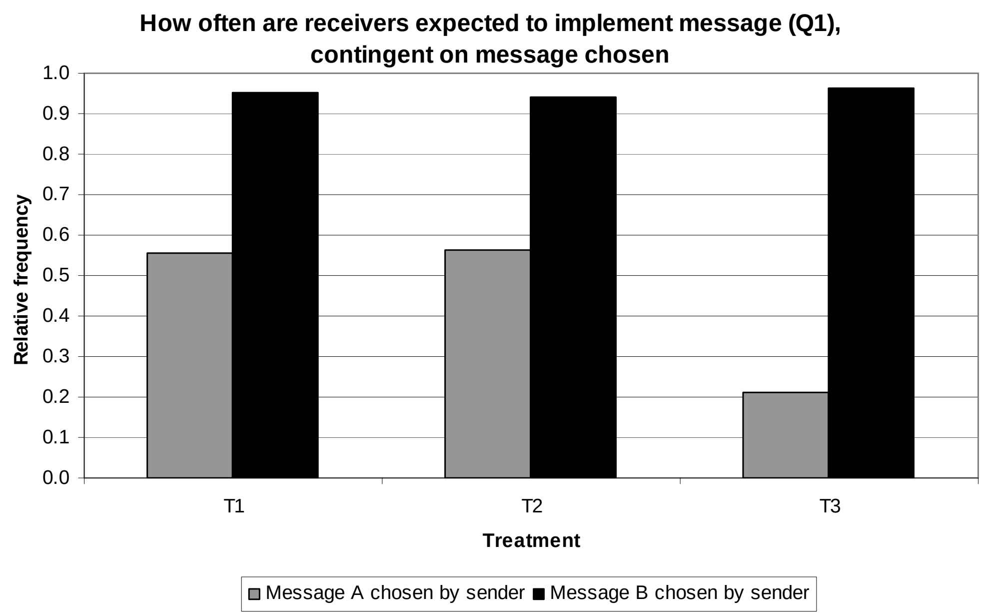

FIGURE 1 – EXPECTATIONS OF SENDERS (Q1) DEPENDING UPON CHOSEN MESSAGE

The details with respect to the sender's chosen message, however, matter significantly for the corresponding expectations, as is shown in Figure 1. The grey bars indicate the relative frequency with which senders expect the receivers to implement their message (question Q1) *if* the sender chooses the *true* message A. On the contrary, the black bars are based on senders

[5] Note that receivers had no information about the possible monetary payoffs in any of the treatments. Rather, receivers had exactly the same information conditions in all treatments. Hence, it would have been discomforting to find significant differences between treatments concerning the receivers' frequency to implement the received message.

[6] It is remarkable that senders and receivers seem to be very well locally calibrated both in Gneezy's experiment in Israel, mine in Germany and Sjaak Hurkens and Navin Kartik's (2006) one in Spain. Although the level of senders expecting the receiver to follow the message differs significantly across these three experiments (about 80% in Gneezy, 70% here, and 60% in Hurkens and Kartik), all populations are very well calibrated as the fraction of receivers following the message is very close to (and statistically not distinguishable from) the senders' expectations. In a more complex game with three options, Julian Rode (2006) finds considerably less perfect calibration, but adds the interesting finding that the relative frequency of receivers following the sender's message depends on the context in which subjects interact. In a competitive framing, less receivers follow the message than in a cooperative framing.

who send the *wrong* message B. Obviously, there is a marked difference in expectations, depending on the actually chosen message.

When sending message B, senders expect in about 95% of cases the receiver to implement the chosen message (black bars in Figure 1). If senders choose the true message A, then they expect receivers to implement option A in only 56% of cases both in treatments T1 and T2, and in only 21% (!) of cases in treatment T3 (grey bars). The latter value is particularly noteworthy as it is significantly smaller than the values in treatments T1 and T2 ($\chi^2 > 10$ in both comparisons, $d.f. = 1$, $p < 0.01$). Recall that senders in all conditions are aware of the fact that receivers do not have any information on the possible payoffs. Based on this fact it would be intuitive to conclude that one should not find significantly different expectations of senders about the actions of receivers across treatments. Nevertheless, in the particular case where senders choose message A, this is the case. One possible explanation would be that senders' expectations are influenced by the possible gains from the different options. Such a behavioral bias is known in the psychology literature as desirability bias or wishful thinking (Hogarth, 1987). Douglas McGregor (1938) was the first to document the existence of a wishful thinking bias by showing that the overestimation of the probability of a desirable future event was positively correlated to the importance of the desirable event to the forecaster. Applying the concept of wishful thinking to my experimental data, one should recall that in treatment T3 senders can gain 10 € more if the receiver implements option B rather than option A, whereas they can only gain 1 € more from option B in treatments T1 and T2. The tenfold gains from option B in treatment T3 might induce a wishful thinking bias that makes truthful senders in T3 much more "pessimistic" about the frequency with which receivers implement option A.

In the following, those senders who send message A and expect the receiver to implement option A are called "*benevolent*." Senders who send message A, but expect their receiver to implement option B are called "*sophisticated*." There is a strong and significant difference between benevolent and sophisticated senders in each treatment concerning their answers to question Q2 (Mann Whitney U-test, $p < 0.05$). The latter question asked for the relative frequency with which receivers follow the sender's message, i.e. implement the option that is said to be better in the chosen message.

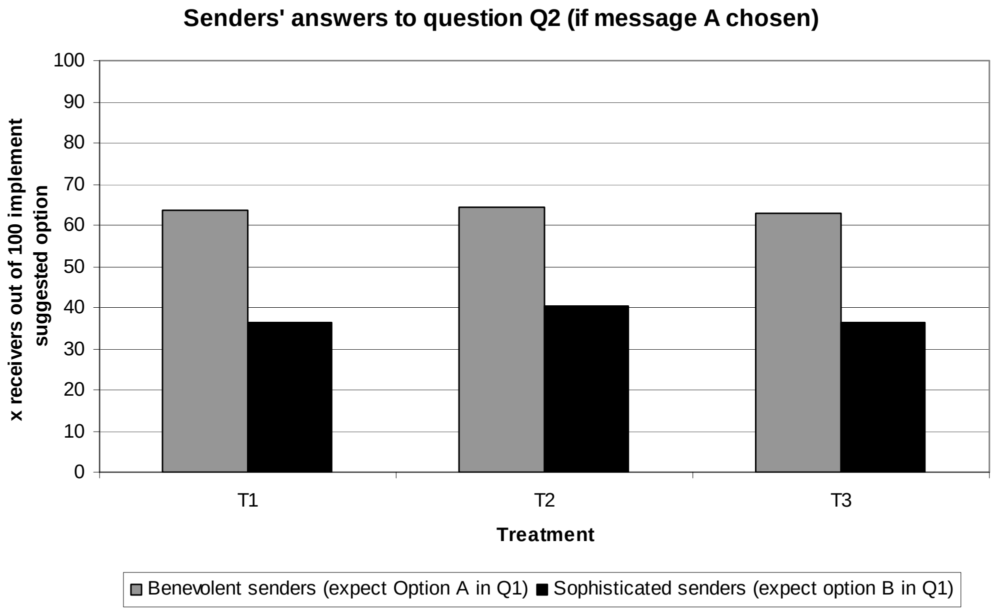

FIGURE 2 – TYPES OF TRUTHFUL SENDERS AND EXPECTATIONS (Q2)

One can see from the grey bars in Figure 2 that *benevolent* senders expect about two thirds of receivers to implement the suggested option (which is significantly larger than 50%; binomial test, $p < 0.05$). The black bars reveal that *sophisticated* senders expect only about one third of receivers to implement the suggested option (which is significantly smaller than 50%; binomial test, $p < 0.05$).

If I regard the *sophisticated* senders as subjects who actually intend to deceive the receiver (by sending message A!), then it seems that the frequency of senders picking the wrong message B is not an adequate measure of the frequency of deception in this game. Therefore, I propose the following

Broader definition of deception: *Deception includes all cases where a sender sends either of the two messages, but expects option B to be implemented by the receiver.*

This definition then encompasses practically all senders with message B (of which about 95% expect the receiver to implement option B), but additionally those senders with message A who expect the receiver to implement option B. Applying this broader definition I arrive at the figures in row [4] of Table 2. It shows that 67% of senders in treatment T1 expect option B to be chosen by the receiver. The corresponding figures are 61% in T2, and 89% in T3. The share of senders expecting to get option B is significantly larger in treatment T3 (with gains of 10 €) than in the other treatments T1 and T2 (with gains of 1 € only) ($\chi^2 > 10$ in both comparisons, *d.f.* = 1, $p < 0.01$). There is again no significant difference between treatments T1 and T2, confirming the earlier result (from row [1]) that the differences in own payoffs across treatments seem to have a stronger (and the only statistically significant) impact than the differences in receivers' payoffs across treatments.

Subtracting the figures in row [1] of Table 2 (narrow definition of deception) from those in row [4] (broader definition of deception) shows that the broader definition of deception increases the relative frequency of deception by about 20 to 30 percentage points. Hence, relying solely on row [1] as an indicator of deception would have grossly underestimated its extent.

4 A robustness test of the broader definition of deception – An experiment with teams as decision makers

My suggested measure for the overall frequency of deception hinges critically on the assumption that *sophisticated* senders (who send message A, but expect option B to be implemented) really intend to deceive and exploit the receiver. There are three possible alternative explanations, though, that might lead to the same behaviour, but would not be an indication of deception. First, there might be senders who tell the truth (message A) regardless of what they expect, because telling the truth is the "moral choice" that is separated from expectations. Second, there might be senders who, in fact, intended to benefit the receiver (wishing that he implemented option A), but expected the receiver to misread their true intention. If such senders were averse to lying (i.e. sending message B) they would stick to message A even though they expected the receiver to invert the message.[7] Third, there might be some senders that expected the receiver to choose a particular option independent of the sender's message.[8] Hence, if a sender expected the receiver to choose option B irrespective of the sent message, it would be unreasonable to interpret the combination of "Send message A and expect option B" as deception.

In order to discriminate between these alternative explanations and my broader definition of deception I ran additional sessions with 342 subjects in late August 2005. These participants were randomly assigned to *teams* of two or three subjects each, with the same payoffs *per capita* in a given treatment as in the sessions with individual decision makers. I obtained 22 pairs of sender and receiver teams for each of the treatments T1 and T2, and 24 such pairs for treatment T3. The sessions were conducted in the VideoLab of the Max Planck Institute of Economics in Jena. The discussions of sender teams were videotaped in order to detect the true intention behind sending message A.[9]

Table 3 summarizes the main results of the experiment with teams as decision makers. Row [1] indicates that teams send message B significantly less often than individual senders, namely in 23% of cases in T1 and T2, and in 25% in T3 ($p < 0.05$ in any comparison between

---

[7] If such senders did not care about lying, yet expected their message to be inverted and wanted to benefit the receiver, then they should send message B. However, it should be clear from Figure 1 that less than 5% of senders who send message B expect their message to be inverted.

[8] Using the strategy method, Hurkens and Kartik (2006) show that about one third of their receivers ignored the received message, but implemented a particular option that was independent of the message. Yet, it is important to stress that only one out of 30 senders expected his message to be ignored. Hence, almost all senders believe that their message has a systematic influence on the option finally implemented by the receiver.

[9] Sessions with individuals had also been run in the VideoLab of the Max Planck Institute of Economics in order to guarantee the same local environment for individuals and teams. Upon entering the VideoLab, participants were routinely informed that they might be videotaped, but that the tapes would only be used for scientific purposes. In fact, only sender cabins in the team condition were videotaped.

teams and individuals for a given treatment). Thus, at first sight it seems as if teams were the more "honest" decision makers. Yet, the fact that teams send the wrong message B less often than individuals is driven by different expectations about the receiver team's decision. Averaging over the figures in row [2] of Table 3 shows that sender teams expect the receiver team to follow their message in only 44% of cases. This relative frequency is significantly smaller than the corresponding 70% of individual senders ($\chi^2 > 10$, $p < 0.01$). Hence, teams are far more pessimistic concerning whether receivers follow the senders' message. Across the three treatments there is also an outstanding low figure in treatment T3 (29%) which might again be explained by wishful thinking.

TABLE 3 – DECISIONS OF TEAMS (RELATIVE FREQUENCIES)

Treatment	T1	T2	T3
	$N = 22$	$N = 22$	$N = 24$
[1] Sender team chooses message B	0.23	0.23	0.25
[2] Q1: Sender team expects receiver team to implement chosen message	0.55	0.50	0.29
[3] Receiver team implements message	0.64	0.45	0.67
[4] Sender team expects receiver team to choose option B	0.68	0.73	0.88

Concerning receivers' actions we see from row [3] of Table 3 that receiver teams follow the message on average in 59% of cases. This value is significantly smaller than the 72% of individual receivers ($\chi^2 = 4.4$, $p < 0.05$), which indicates that receiver teams are more skeptical about the truthfulness of the sender's message than individual receivers.

Matching messages and expectations we find that a total of 94% of teams that send the wrong message B expect the receiver team to implement option B. This means that those teams sending a lie are pretty confident that the lie will work out. When sending message A, however, sender teams expect the receiver team to implement option A only in 15 out of 52 cases (i.e. 29%). That means that in 37 out of 52 cases (i.e. 71%) with message A, the sender team expects the receiver team to choose option B. The crucial question is whether these teams have an intention to deceive the receiver in order to get more money (as implied by my broader definition of deception) or whether the behavior of these teams has to be subsumed under any of the above discussed three alternative explanations.

The videotapes of the 37 sender teams that sent message A, but expected option B, were scrutinized with respect to the driving motivation for the decision to send message A. The communication of team members reveals that 34 out of these 37 sender teams (i.e. 92%) explicitly mention the following argument: by sending the true message A the sender team will earn more money because the receiver team will not believe the received message, but rather invert it and implement option B. Consequently, sending message A and expecting message B can almost always be classified as an attempt to deceive the receiver team, meaning that my broader definition of deception is appropriate. In fact, only three out of 37 sender teams sent message A because they wanted to avoid lying, without any recourse to the desirability of the receiver team inverting the message.

Interestingly, some sender teams discussed that the argument described above might lead to an infinite number of reasoning steps. If the receiver team expected the sender team to send the true message because the sender team expects the receiver team not to believe the message, then the receiver team should in fact implement the received message. That, in turn, would make it more profitable for the sender team to send the wrong message B instead of the true message A. Yet, if sender teams expected receiver teams to anticipate this move and therefore invert the message, then sender teams would again prefer the true message A, and so on. The literature on level-k thinking or step-level of reasoning models explores such problems where decision makers have to guess the correct depth of reasoning of another decision maker. In general, such models examine a decision maker's best response given an (expected) distribution of other decision makers' level of reasoning. One of the best known applications of such reasoning models is the guessing- or beauty-contest game (Rosemarie

Nagel, 1995) where decision makers have to choose a number from an interval $I \equiv [0,100]$ and the winner is the decision maker whose number is closest to p times the average of all chosen numbers (with $p < 1$). Dale Stahl (1996) or Miguel Costa-Gomes and Crawford (2006) discuss level-k thinking in guessing games. Martin Kocher and Matthias Sutter (2005) have shown that teams apply higher levels of reasoning in guessing-games than individuals do. This might explain why teams send message A significantly more often than individuals in my experiment. Note, however, that in my experiment any levels of thinking that are two steps apart would result in the same behaviour.

Although individual and team senders differ quite a lot with respect to the frequency of sending the wrong message B or concerning the expectations about the receiver's actions, individual and team senders do not differ when my broader definition of deception is applied (see the figures in row [4] of Table 3). Checking the frequency of sender teams expecting option B to be chosen reveals that there are no longer any significant differences between individuals and teams (compare the entries in row [4] of Tables 2 and 3). Hence, the overall amount of deception in a given treatment does not depend on the type of decision maker being an individual or a team. Moreover, it is again treatment T3 – where senders can gain the most from receiving option B instead of option A – in which sender teams expect the receiver team to choose option B most often (i.e. in 88% of cases; with $p < 0.05$ compared to T1 or T2). Hence, the size of the possible gains from deception has a similar influence on team decision making as it has been found to have on individual decision making.

5 Conclusion

Gneezy (2005) has found that the willingness of subjects to lie in a cheap-talk sender-receiver game is a matter of weighing costs and benefits. Without doubt, sending a wrong message can be interpreted as a deliberate attempt to deceive the receiver. Corroborating

Gneezy's interpretation, my results show that 95% of senders choosing the wrong message B do so while expecting the receiver to follow their message. Like Gneezy, I have also found that individual senders lie significantly more often when they can gain relatively more by doing so and less often (though not significantly so) when the receiver loses relatively more. This obviously rather stable pattern of behavior has been explained in a recent comment by Gary Charness and Martin Dufwenberg (2005) by their theory of guilt aversion (Charness and Dufwenberg, 2006). In a nutshell, this theory states that a player i suffers from guilt to the extent that he believes that player $j \neq i$ receives a lower payoff than i believes j believes she will receive. If subjects are guilt averse, they might behave as if they have a cost of lying in contexts where they believe their lies will mislead. Since the payoffs for the receiver differ across treatments, the theory of guilt aversion can also explain the differences in the frequency of deception across the three different treatments.

In another recent comment, Hurkens and Kartik (2006) can confirm from their experiments in Spain that there is a statistically significant level of lying aversion. However, they argue that the level of lying aversion need not be dependent on the consequences from lying. Rather, they argue that Gneezy's data as well as their own are consistent with a hypothesis which claims that people are one of two kinds: either a person will never lie, or a person will lie whenever she prefers the outcome obtained by lying over the outcome obtained by telling the truth. Therefore, a person's decision to lie may be completely insensitive to changes in the induced outcomes. Hurkens and Kartik's (2006) paper offers a complementary – rather than contradictory – alternative to Gneezy's interpretation that the frequency of deception is a matter of costs and benefits – with the latter interpretation also accepted here and rationalized by Charness and Dufwenberg (2005).

The main point of this paper concerns the classification of what a lie or deception is, and therefore complements the papers by Gneezy (2005), Charness and Dufwenberg (2005) or Hurkens and Kartik (2006). I have argued that the concept of deception as measured by

Gneezy (2005) is too narrow to capture the full amount of deception prevalent in the two-person game under consideration, and probably in other games where misrepresentation of information is possible. I have proposed that telling the truth should be counted as an act of deception when the sender expects the receiver *not* to follow the sender's message *and* when the true message is sent for precisely this reason. Analyzing the discussions preceding the decision of teams I have found evidence that deception through telling the truth is, in fact, a widespread phenomenon which is obviously driven by payoff considerations and by the expectations about the receiver's actions.

References

Camerer, Colin and Hogarth, Robin M. (1999), The effects of financial incentives in experiments: A review and capital-labor-production framework. *Journal of Risk and Uncertainty* 19: 7-42.

Charness, Gary and Dufwenberg, Martin (2005), Deception: The role of guilt. Working Paper. University of Arizona, Tucson.

Charness, Gary and Dufwenberg, Martin (2006), Promises and partnerships. *Econometrica*, forthcoming.

Costa-Gomes, Miguel and Crawford, Vincent P. (2006), Cognition and behavior in two-person guessing games: An experimental study. *American Economic Review*, forthcoming.

Crawford, Vincent P. (2003), Lying for strategic advantage: Rational and boundedly rational misrepresentation of intentions. *American Economic Review* 93: 133-149.

Crawford, Vincent P. and Sobel, Joel (1982), Strategic information transmission. *Econometrica* 50: 1431-1451.

Fischbacher, Urs (1999), Z-tree: Zurich toolbox for readymade economic experiments. University of Zurich, Institute for Empirical Research in Economics. Working Paper No. 21.

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

Greiner, Ben (2004), An Online Recruitment System for Economic Experiments. In: Kurt Kremer and Volker Macho (eds.): *Forschung und wissenschaftliches Rechnen 2003*. GWDG Bericht 63. Gesellschaft für Wissenschaftliche Datenverarbeitung, Göttingen, 79-93.

Grether, David M. (1992), Testing bayes rule and the representative heuristic: Some experimental results. *Journal of Economic Behavior and Organization* 17: 623-638.

Hogarth, Robin M. (1987), *Judgment and Choice*. New York: Wiley.

Hurkens, Sjaak and Kartik, Navin (2006), (When) Would I lie to you? Comment on "Deception: The role of consequences". University of California San Diego. Working Paper.

Kocher, Martin and Sutter, Matthias (2005), The decision maker matters: Individual versus group behaviour in experimental beauty-contest games. *Economic Journal* 115: 200-223.

McGregor, Douglas (1938), The major determinants of the prediction of social events. *Journal of Abnormal Psychology* 33: 179-204.

Nagel, Rosemarie (1995), Unravelling in guessing games: An experimental study. *American Economic Review* 85: 1313-1326.

Rode, Julian (2006), Truth and trust in communication. An experimental study of behaviour under asymmetric information. University Pompeau Fabra Barcelona. Working Paper.

Stahl, Dale (1996), Boundedly rational rule learning in a guessing game. *Games and Economic Behaviour* 16: 303-330.

Appendix A1 – Instructions for senders
Welcome to this experiment on decision making. Please read the following instructions carefully. If you have any questions, an experimenter will answer them privately.

2 people interact with each other once
In this experiment, you and another person will interact with each other for exactly one round. Each of you will have to make one single decision after which the experiment will end and you both will be paid in cash.
Both of you are randomly assigned to one of two possible roles: either the role of Person 1 or the role of Person 2. Your role in this experiment is **Person 1**.

Your decision
There are two options that can be chosen in this experiment. Both options yield different payoffs for you and the Person 2 that is paired with you. The options are as follows:

In treatment T1 it read

Option A: 5 € for you (= Person 1) and 6 € for Person 2
Option B: 6 € for you (= Person 1) and 5 € for Person 2

In treatment T2 it read

Option A: 5 € for you (= Person 1) and 15 € for Person 2
Option B: 6 € for you (= Person 1) and 5 € for Person 2

In treatment T3 it read

Option A: 5 € for you (= Person 1) and 15 € for Person 2
Option B: 15 € for you (= Person 1) and 5 € for Person 2

Person 2 will make the choice which option to implement. Yet, note that Person 2 **will not** be informed about the monetary consequences of both options. Instead, Person 2 will only receive one of the two following messages:

Message 1: „Option A will earn you more money than Option B.“
Message 2: „Option B will earn you more money than Option A.“

You have to **choose which message** you want to send to Person 2.

Information conditions and payoffs
After Person 2 has implemented one of the two options, you and Person 2 will be paid the corresponding amount of money. Note that Person 2 will neither be informed about your payoff in the chosen option nor about the possible payoffs (for both persons) in the other option. Person 2 will only be informed about his monetary payoff earned in the implemented option.

After Person 1 had sent her message, she was asked the following two questions:
While Person 2 is deciding about which option to implement, would you please answer the following two questions:
Q1: “Which option do you expect the receiver to choose?”
Q2: “Out of 100 receivers, how many do you think follow the sender’s message on average?”

Appendix A2. Instructions for receivers

Welcome to this experiment on decision making. Please read the following instructions carefully. If you have any questions, an experimenter will answer them privately.

2 people interact with each other once

In this experiment, you and another person will interact with each other for exactly one round. Each of you will have to make one single decision after which the experiment will end and you both will be paid in cash.

Both of you are randomly assigned to one of two possible roles: either the role of Person 1 or the role of Person 2. Your role in this experiment is **Person 2**.

Two possible options and your decision

Two possible monetary payoffs are available to you and Person 1 in this experiment. The payoffs depend on the option chosen by **you**. We have already showed the two options to Person 1. The only information that you will have before making your choice is the message that Person 1 has sent to you.

Person 1 will choose one of the two following messages:

Message 1: „Option A will earn you more money than Option B.“
Message 2: „Option B will earn you more money than Option A.“

You will be informed about the message chosen by Person 1 on the screen.

We will then ask you to choose either option A or option B. Your choice will determine the payoffs in the experiment. You will never know what sums were actually offered in the option not chosen (i.e. whether the message sent by Person 1 is true or not). Furthermore, you will never know the amount that Person 1 has earned in the chosen option or would have earned in the option not chosen by you.

We will pay both persons the amount of money that corresponds to the option you chose.

Thank you for your participation in the experiment.