Netspar
Network for Studies on Pensions, Aging and Retirement

Netspar DISCUSSION PAPERS

Huseyn Ismayilov and Jan Potters

Disclosing Advisor's Interests Neither Hurts Nor Helps

DP 09/2012-058

Disclosing Advisor's Interests Neither Hurts Nor Helps*

Huseyn Ismayilov † Jan Potters‡

September 15, 2012

Abstract

An informed advisor may have incentives to send a biased advice to an uninformed decisionmaker. We set up an experiment to study whether disclosure of the advisor's interests can foster truthfulness and trust. A key feature of our experiment is that we measure how advisors expect decisionmakers to react to their advice. This allows us to distinguish between strategic and nonstrategic (i.e. moral) reactions to disclosure by advisors. We find that advisors do not expect decision makers to react drastically to disclosure. Also, advisors do not find deceiving morally more acceptable with disclosure than with no disclosure. Overall, disclosure neither hurts nor helps; deceptive advice and mistrust are equally frequent with as without disclosure. Relative to the aims of disclosure policies this is 'bad news'; relative to studies suggesting that disclosure may actually increase the rate of deception this may be regarded as 'good news'.

*We thank Dirk Engelmann, George Loewenstein, participants at the 2012 ESA Annual meeting, the M-BEES 2012 at the Maastricht University, the "Deception, Incentives, and Behavior" conference at UC San Diego, and the Netspar "Economics and Psychology of life cycle decision-making" meeting in Amsterdam for helpful comments and discussions. Financial support from Netspar is gratefully acknowledged.

†CentER and Netspar, Tilburg University. Address: PO Box 90153, 5000LE Tilburg, The Netherlands. E-mail: H.I.O.Ismayilov@uvt.nl

‡CentER and Netspar, Tilburg University. Address: PO Box 90153, 5000LE Tilburg, The Netherlands. E-mail: J.J.M.Potters@uvt.nl

1 Introduction

Conflicting interests may provide advisors with incentives to give biased advice to their clients. Insurance agents, for example, may be led by the commissions they receive on different products and not just by the interests of their customers. Besides the interests of their patients, physicians may be affected by their relationship with pharmaceutical companies. One of the solutions suggested to mitigate such problems is that advise recipients be informed about matters that present a potential conflict of interest. Mandatory disclosure rules exist in many domains, including accounting, retail finance, medicine, and academia.[1] In all cases, it is expected that disclosure of the advisor's interests reduces the information gap between the advisor and the advisee and will allow the latter to make better decisions.

In this paper, we test how disclosure affects advisors and advice recipients in a simple sender-receiver game based on Gneezy's (2005) deception experiment. The receiver has to choose between two options without knowing the associated payoffs. The sender knows the payoffs of each option, and sends a message stating which option is better for the receiver. In our baseline treatment, the receiver has no information on the sender's payoffs (as in Gneezy 2005). In our disclosure treatment, the receiver is informed about the sender's payoffs for each of the two options. Comparing the two treatments allows us to see how disclosure affects the sender's advice and how the receiver uses the advice.

Interestingly, previous experimental studies have suggested that disclosing conflict of interests may actually hurt advice recipients (Cain et al., 2005, Cain et al., 2011, Inderst et al., 2010, Koch and Schmidt, 2009, Rode, 2010). With

[1]For example, the Insurance Conduct of Business sourcebook in the UK requires "a firm to provide its customers with details about the amount of any fees other than premium monies for an insurance mediation activity" (FSA, Section 4.3.1), and the EU Market in Financial Instruments Directive (MiFID) has similar provisions; several states in the US mandate disclosure of payments made to physicians by pharmaceutical companies (Ross et al., 2007), and the Physician Payment Sunshine Act aims to lift this to federal level; many academic journals and granting bodies such as the NSF require their reviewers to issue a statement of no conflict of interest.

disclosure, advisors bias their advice more than they do without disclosure, and advice recipients fail to account for this sufficiently. As a result, disclosure makes advice recipients worse off compared to no disclosure. Cain et al. (2005, 2011) provide two possible explanations for the increased exaggeration by advisors. One is, what they call, moral licensing, according to which advisors find it less unethical to send deceptive messages once their own interests are revealed to advice recipients. An alternative explanation discussed by Cain et al. (2005, 2011) is that the increased bias is done strategically to compensate for the anticipated reaction to disclosure by the advisees. An important feature of our experiment is that we measure the beliefs of the sender about the receiver's reaction to her messages. This allows us to distinguish between the two reasons for why senders might change their advice in response to disclosure, since the sender's beliefs provide us with a direct measure of the strategic motive. Note that another feature of our design is that with disclosure the receiver knows the sender's interests but not that there is a conflict of interest.[2]

Our results indicate that disclosure does not significantly increase the frequency of deception by senders. The beliefs data show that senders do not expect the receivers to change their reaction to messages substantially in response to disclosure. We also do not find any support for the moral licensing argument. Controlling for the expected benefit of deception, receivers deceive more nor less with disclosure than without. On the receiver side, we find no significant effect of disclosure. Receivers do not discount advice much differently when they know that it is, or is not, in line with the self-interest of the sender. They largely ignore this information.

Furthermore, we run an additional treatment in which disclosure is not automatic (mandated) but must be requested by the receiver. This treatment is inspired by circumstances in which clients have to explicitly ask for the disclo-

[2]In this respect our experiment is similar to de Meza et al. (2011) who study how disclosing seller interests to a buyer affects trust and persuasion in insurance markets. Unlike our experiment, there is no asymmetric information about the mis(alignment) of the interests of the seller and the buyer in de Meza et al.(2011).

sure of the advisor's interests.[3] In line with the 'hidden costs of control' (Falk and Kosfeld, 2006), we hypothesize that solicited disclosure is particularly prone to increase the moral license to deceive felt by the sender. Again, however, we find that the rate of deception does not depend on whether the receiver asks the sender's interests to be disclosed or not. At the same time, receivers who do not demand disclosure are much more likely to follow the senders' advice than those who do.

Overall, although we do not find that disclosure hurts advice recipients, it also does not help. This suggests that disclosing advisor's interests will not be enough to solve the problem of biased advice.

2 Experimental Design and Procedure

Our experimental design is based on the two player sender-receiver game from Gneezy (2005). In this game one of the parties, the sender, observes payoffs to both players associated with two options, Option A and Option B. The sender sends one of the two possible messages to the other player, the receiver:

Message 1: "Option A will earn you more money than option B."

Message 2: "Option B will earn you more money than option A."

After receiving the message from the sender, the receiver chooses one of the two options and both players are paid according to the chosen option. In our *No disclosure* treatment, as in Gneezy (2005), the only information available to the receiver is the message sent by the sender. The receiver observes neither the payoffs to the sender nor the payoffs to himself. In the *Disclosure* treatment in addition to the message sent by the sender the receiver observes the payoffs to the sender for each option but not the payments to himself. Thus, the only difference between the two treatments is that the receiver observes the sender's interests in the *Disclosure* treatment but not in the *No disclosure* treatment. Our goal is

[3]For example, the Insurance Conduct of Business Sourcebook in the UK requires "that an insurance intermediary must, on a commercial customer's request, promptly disclose the commission that it and any associate receives in connection with the policy (FSA, 2012, Section 4.4.1).

to study how the disclosure of the sender's interests to the receiver affects the behavior of each party.

We also implement another treatment where the receiver decides himself whether he wants to know the interests of the sender before the sender sends a message. Once the receiver makes a decision, the sender is informed about it. With this treatment we want to test if leaving the decision to disclose the potential conflicts of interest to the receiver leads to different outcomes. We call this the *Endogenous* treatment. Depending on the receiver's decision whether or not to have the sender's interests disclosed we will have two conditions within the *Endogenous* treatment: *Endogenous No Disclosure* and *Endogenous Disclosure*. For convenience, we call the latter two 'treatments' instead of 'conditions' in what follows. Thus, overall we have four treatments: *No disclosure*, *Disclosure*, *Endogenous No Disclosure*, and *Endogenous Disclosure*.

To test the robustness of our results we implement two different payoff structures: Low Incentive and High Incentive. Table I provides details of both payoff structures. Note that under both payoff structures there is a conflict of the interests. We are interested in how the receivers will react to the disclosure depending on the magnitude of potential conflict of interests.

Table I

Low and High Incentive payoff structures

		Payoff to	
Payoff structure	Option[a]	Sender	Receiver
Low incentive	A	8	3
	B	6	6
High incentive	A	15	5
	B	5	15

[a] In this table Option A gives higher payoff to the sender. In the experiment the option with higher payoff for the sender could be either A or B.

Another important feature of our experiment is that we measure beliefs of the sender about the receiver reaction to each of the possible messages. After

choosing a message, the sender guesses how likely it is that the receiver will follow Message 1 and Message 2 (i.e. also for the message that is not sent) by choosing one of the five columns as shown in Table II . To be able to incentivize sender guessing for both messages we ask the receivers to make a choice conditional on each message (i.e. the strategy method). Measuring beliefs allows us to explore the motivations behind the potential effects of disclosure on the sender. For example, we can identify whether the sender's reaction to disclosure is strategic, non-strategic (as suggested by Cain et al. 2005, 2011) or maybe a combination of both.

Table II

Belief Elicitation[a, b]

	(1)	(2)	(3)	(4)	(5)
Your guess	Almost certainly will follow	Probably will follow	Not sure	Probably will not follow	Almost certainly will not follow
Your bonus if the receiver (would) follow your message	€1.30	€1.20	€1.00	€0.70	€0.40
Your bonus if the receiver (would) not follow your message	€0.40	€0.70	€1.00	€1.20	€1.30

[a] Subjects chose one of the five columns first for Message 1 and then for Message 2.

[b] Under risk neutrality columns correspond to intervals with midpoints at the probabilties of 87.5%, 67.5%, 50%, 37.5%, and 12.5% that the message will be followed by the receiver.

The experiment was ran in September 2011 at Centerlab, Tilburg University. Subjects were students recruited via email. Upon arrival subjects were seated behind partitioned workstations and randomly assigned one of the two roles, player 1 (the sender) or player 2 (the receiver), and formed a pair with one of the participants in the other role. The experiment was computerized using the

Z-tree software (Fischbacher 2007). To increase the number of observations each subject played the game twice with different partners. No feedback was provided after the first period was played. Each subject played both the low incentive and the high incentive payoff structures. Those who played the low incentive payoff structure in the first period played the high incentive payoff structure in the second period and vice versa. The order was randomized. As mentioned above we also randomized which of the two options gave a higher payoff to the sender. At the end of the second period subjects were provided with feedback for both periods. One of the periods was randomly selected and subjects were paid their earnings in that period.

The experiment lasted for approximately 40 minutes and subjects earned 8.9 euros on average. In total 170 students participated in 9 sessions. We ran 2 sessions (18 pairs) in the *No Disclosure* treatment, 3 sessions (31 pairs) in the *Disclosure* treatment, and 4 sessions (36 pairs) in the *Endogenous* treatment. More sessions were run in the *Endogenous* treatment because this treatment would be split into two treatments depending on the decisions of the receivers.

3 Hypotheses

In this section we analyse how the disclosure of the sender's interests to the receiver might affect each party. We discuss moral licensing (Cain et al. 2005, 2011) and strategic effects of disclosure. Without loss of generality, when analyzing the sender behavior we assume that Option A gives a higher payoff to the sender relative to Option B.

We start by analysing sender behavior in the *No Disclosure* treatment. The sender can send either the deceptive message (Message A: "Option A will earn you more money than Option B") or the truth-telling message (Message B: "Option B will earn you more money than Option A"). We assume that there is a cost, c, to the sender of sending the deceptive message (Gneezy 2005). The expected payoff from sending each message for the sender is:

$$E(\pi)_{deceptivemessage} = p_A * \pi_A + (1 - p_A) * \pi_B - c, \tag{1}$$

$$E(\pi)_{truthtellingmessage} = p_B * \pi_A + (1 - p_B) * \pi_B. \quad (2)$$

In the above equations, $p_A(p_B)$ denotes the probability that the receiver will choose Option A (i.e., the one that is best for the sender) conditional on receiving Message A (Message B) and $\pi_A(\pi_B)$ stands for the sender's payoff of Option A (Option B). Since in the *No Disclosure* treatment the receiver does not know which option is in the sender's self-interest and the sender knows this, it is reasonable to assume that $p_A = 1 - p_B$ (i.e. both messages are equally likely to be followed by the receiver). From equations (1) and (2) it follow that the sender will lie whenever

$$(p_A - p_B)(\pi_A - \pi_B) \geq c \quad (3)$$

In what follows, we call the expression on the left hand side of equation (3) the *expected benefit of lying*. By equation (3), the sender will lie whenever the *expected benefit of lying* is larger than the cost of lying.

We now discuss how the disclosure of the sender's interests might affect c, the sender's cost of lying, and, p_A and p_B, the probabilities that the receiver will follow the sender messages. Cain et al.(2005, 2011) argue that once the interests of advisors are revealed, advisors find lying less immoral and will feel morally more licensed to deceive the advice recipient. In our setup the moral licensing argument implies that the cost of lying, c, decreases with disclosure. From equation (3), for given *expected benefit of lying*, $(p_A - p_B)(\pi_A - \pi_B)$, a decrease in c should make deception more likely. Thus, we can formulate the following hypothesis:

Moral Licensing Hypothesis: *Controlling for the expected benefit of lying, the deception rate increases with disclosure.*

In Appendix A we present a theoretical analysis to study the impact of disclosure on p_A, p_B, and $(p_A - p_B)(\pi_A - \pi_B)$. Note that with disclosure the receiver observes the option that is in the sender's self interest (Option A) and the option that is not in the sender's self interest (Option B) and the sender knows this. Let p_A^D and p_B^D stand for p_A and p_B in the *Disclosure* treatment. Our theoretical

analysis shows that in equilibrium we have $p_A^D < p_A$ and $p_B^D = 0 < p_B$. Once disclosed, the sender's self-interest message A is less likely to be followed by the receiver. On the other hand, if the sender advises the option that is not in her self interest, the receiver follows this advice. The model also shows that the effect of disclosure on the *expected benefit of lying* is ambiguous and can go in either direction depending on the distribution of the lying costs of the senders. Thus, we do not make any predictions regarding the effect of disclosure on the expected benefit of lying. For the empirical analysis we can rely on the sender's subjective beliefs about p_A and p_B.

For the endogenous treatment, with disclosure one would expect the moral licensing effect to become more pronounced. The experimental literature has shown that signalling mistrust can backfire for the mistrusting party (see, for example, Falk and Kosfeld 2006). A request by the receiver to have the sender's interests revealed, may be perceived by the sender as a signal of mistrust. We expect that this will increase the importance of the moral licensing argument relative to the exogenous disclosure case.

4 Results

For pairs of senders and receivers we had 36 observations in the *No Disclosure* treatment, 62 observations in the *Disclosure* treatment, and 72 observations in the *Endogenous* treatment. For the analysis, however, we use only 34 observations for senders and 36 observations for receivers in the *No Disclosure* treatment, 61 observations for senders and 62 observations for receivers in the *Disclosure* treatment, and 72 observations for senders and 70 observations for receivers in the *Endogenous* treatment.[4]

[4]We exclude three second-round observations for senders due to an input error in the parameter table of the Z-tree. Due to this error, these three senders played High Incentive payoff structure in both periods. We exclude their second period choices from the analysis below, because all the other senders who played High Incentive payoff structure in the second period played Low Incentive payoff structure in the first period. We do not exclude the first period choices by these 3 senders as these are comparable to the first period choices by the other

4.1 Sender behavior

We start by comparing sender deception rates in the *No Disclosure* treatment to deception rates in the *Disclosure* treatment. Figure 1 reports deception rates in the two treatments for each payoff structure separately. Disclosure increases the deception rate by 9% with the Low Incentive payoffs and by 2% with the High Incentive payoffs. None of the differences is significant, though (for Low Incentive p=0.56, two-tailed Chi-square test, and for High Incentive p=0.86, two-tailed Chi-square test). Thus, we do not observe a significant increase in sender deception rates with disclosure.

The figure also shows that senders lie more with the High Incentive payoffs than with the Low Incentive payoffs in both the *No disclosure* and the *Disclosure* treatments. The differences are marginally significant for each treatment separately and highly significant for the combined data (p=0.08, one-tailed, Chi square test, for *No disclosure* treatment, p=0.08, one-tailed, Chi-square test, for *Disclosure* treatment and p=0.02, one tailed, for both treatments combined)[5]. Gneezy (2005) and Sutter (2009) also show that senders lie more the higher are the incentives to do so.

In Figure 2 we report average beliefs of the senders about the receiver's reaction to each of the messages. In the *No Disclosure* treatment, one would not expect any difference in the receiver reaction to the self-interest message and the non-self interest message (because the receiver does not know which message is in the sender's self-interest) and this is what we observe. Interestingly, with disclosure senders do not expect drastic changes in the receiver's reaction to the messages. Senders expect that receivers are slightly more likely to follow the non-self interest message than the self-interest message. This difference, however,

senders who played High Incentive in the first period. Note there was no error in the input for receivers and thus no observations are excluded for receivers. In addition, one of the subjects in the receiver role participated previously in our pilot session. We exclude the decisions made by this receiver (2 observations).

[5]The High Incentive payoff structure in the *No Disclosure* treatment is the same as Treatment 3 in Gneezy (2005). We observe a deception rate (0.56) similar to Gneezy (2005) in this case (0.52).

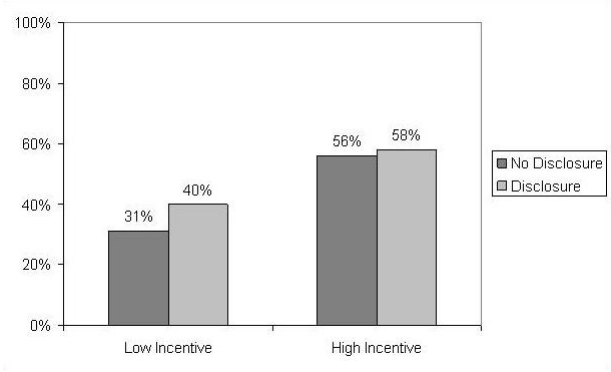
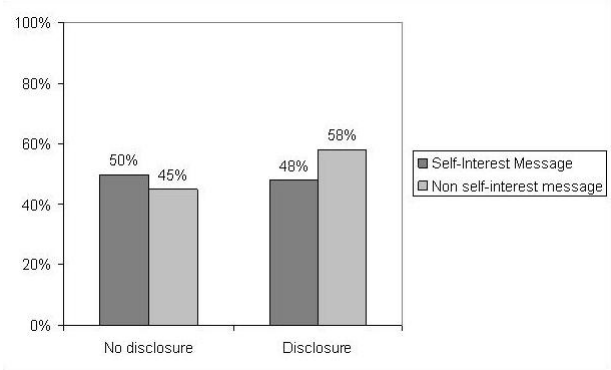


Figure 1: The impact of disclosure on the frequency of lies.

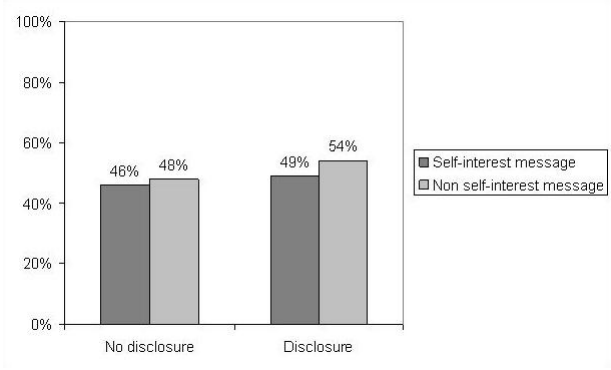
is significant only for the Low Incentive payoffs (p=0.04, Wilcoxon signed rank test). Another interesting observation is that senders think that receivers are as likely to follow the sender's self-interest message with disclosure as any of the two messages with no disclosure. In other words, senders do not expect that receivers will mistrust a message which is in the sender's self-interest, once these interests are revealed to the receiver.

Next, we discuss the results for the *Endogenous* treatment. In 55 out of 72 cases receivers asked to reveal the sender's interests. This results in 17 observations in the *Endogenous No Disclosure* treatment and 55 observations in the *Endogenous Disclosure* treatment. Figure 3 shows that senders do not lie more when the receivers request disclosure of the sender's interests (p=0.89, two-tailed Chi-square, for the Low Incentive payoffs, and p=0.93, two tailed Chi-square, for the High Incentive payoffs). Contrary to what we expected, the senders do not "punish" the receivers for asking to reveal their interests. Overall, the results with respect to the effect of disclosure are similar to the exogenous case.

In Table III we report results of a regression analysis of our combined experimental data for senders. The regression reported in column (1) reiterates that disclosure, whether exogenous or endogenous, does not significantly affect the likelihood of deception. The regression in column (2) includes the *expected benefit of lying* to test for the moral licensing argument suggested by Cain et al. (2005, 2011). The *expected benefit of lying* for each sender is calculated as $(p_A - p_B)(\pi_A - \pi_B)$, using the sender's stated beliefs that the receiver will follow



(a) Low Incentive



(b) High Incentive

Figure 2: Average sender beliefs about the receiver following each of the messages.

each of the two messages. If the moral licensing argument is correct, then controlling for the *expected benefit of lying* senders should lie more in the *Disclosure* treatment than in the *No disclosure* treatment. However, we observe no effect of disclosure even when we control for the *expected benefit of lying*. This means that the likelihood of lying does not increase once we control for strategic reasons and is not in support of the moral licensing argument. Note that the coefficient of the *expected benefit of lying*, although positive, does not achieve statistical significance (z=1.49, p=0.14). In column (3) we interact the *expected benefit of lying* with the disclosure dummy. The coefficient on the *expected benefit of lying* becomes significant (z=1.95, p=0.051) and the interaction variable is negative

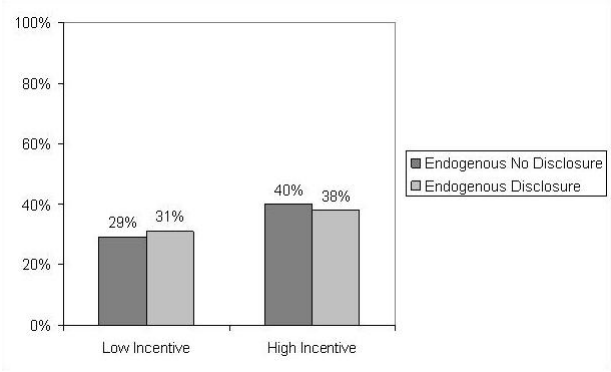


Figure 3: The impact of endogenous disclosure on the frequency of sender lies.

but insignificant (z=-1.52, p=0.13). This suggests that, with disclosure, senders are less likely to base their decision on the perceived private benefits of deception than without disclosure.

4.2 Receiver behavior

As mentioned above we asked receivers to make a choice conditional on each message they might receive from the sender. In Figure 4 we report the proportion of receivers who follow the sender's message in the *No Disclosure* and *Disclosure* treatments for each payoff structure separately. Since the receivers do not observe the sender's interests in the *No Disclosure* treatment, we report the average of the following rates of the sender's self-interest message and the sender's non self-interest message for this treatment. For the *Disclosure* treatment the following rates are shown separately for the sender's self-interest and non self-interest messages.

From the figure we observe that in the *Disclosure* treatment with the Low Incentive payoffs the sender's self-interest message is followed slightly less than the messages in the *No Disclosure* treatment. The difference is not significant, though (78% vs 68%, p=0.45, two-tailed Chi-squared). With the High Incentive payoffs the sender's self-interest message is actually followed a bit more than the messages in the *No Disclosure* treatment (74% vs 72%). Moreover, with disclosure some receivers do not follow the sender's advice even when it is not

Table III

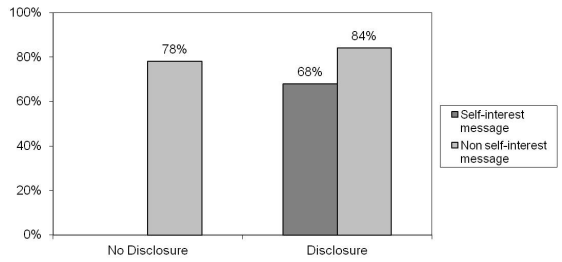
Regression Analysis - Sender Behavior[a]

Variables	(1) Lie	(2) Lie	(3) Lie
Disclosure	0.22	0.17	0.16
	(0.44)	(0.44)	(0.45)
High Incentive	0.64**	0.69**	0.70**
	(0.32)	(0.33)	(0.33)
2nd period	-0.05	-0.07	-0.03
	(0.32)	0.32	(0.33)
Endogenous	-0.41	-0.30	-0.20
	(0.62)	(0.63)	(0.65)
Endogenous*Disclosure	-0.18	-0.27	-0.40
	(0.73)	(0.74)	(0.76)
Expected benefit of lying		0.12	0.25*
		(0.08)	(0.13)
Expected benefit of lying*Disclosure			-0.25
			(0.17)
Number of observations	167	167	167

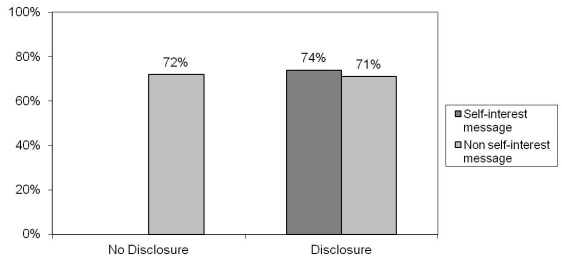
[a] Standard errors are in parentheses. *, **, and *** denote significance at p<0.10, p<0.05, and p<0.01 respectively. Constants are omitted.

in the self-interest of the sender (16% of the receivers with the Low Incentive payoffs and 21% of the receivers with the High Incentive payoffs). Interestingly, the sender's self-interest message is not followed less with the High Incentive payoffs than with the Low Incentive payoffs (74% with High Incentive payoffs vs 68% with Low Incetive payoffs). This suggests that the magnitude of the potential conflict of interest does not make a difference for receiver trust.

Finally, we discuss the receiver behavior in the *Endogenous* treatment. We have 17 observations in the *Endogenous No Disclosure* treatment and 53 observations in the *Endogenous Disclosure* treatment. Figure 5 reports the receiver



(a) Low Incentive

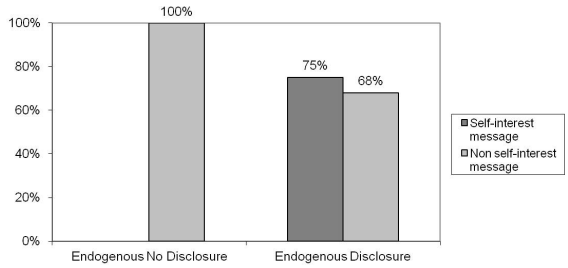


(b) High Incentive

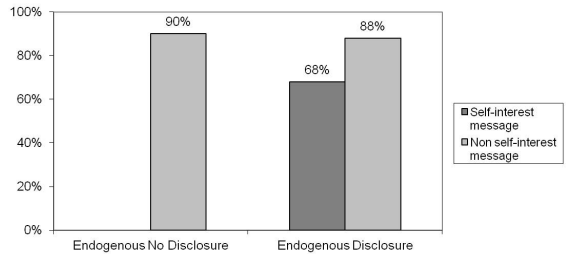
Figure 4: The proportion of receivers who follow the sender's message with and without disclosure.

behavior in both treatments. When receivers do not ask to disclose the sender's interests, they almost always follow the advice the sender sends. With endogenous disclosure, on the other hand, the receiver following rates are lower. The sender's self-interest message is followed less with endogenous disclosure than with endogenous non disclosure. However, the difference is marginally significant for the combined data only (100% vs 75% with Low Incentive payoffs, $p<0.14$, two-tailed Chi-square test, 90% vs 68% with High Incentive payoffs, $p=0.18$, two-tailed Chi-square test and 94% vs 72% for combined data, $p=0.055$). Overall, the results of the *Endogenous* treatment suggest that the receivers who do not request disclosure are very credulous and do not doubt the sender's good intentions. Alternatively, they may want to signal trust and expect senders to reward them by sending the truthful message. However, in the previous subsection we

saw that sender deception rates are similar in the *Endogenous No Disclosure* and *Endogenous Disclosure* treatments.



(a) Low Incentive



(b) High Incentive

Figure 5: The proportion of receivers who follow the sender's message with endogenous non disclosure and endogenous disclosure.

5 Conclusion

In this paper, we test for possible perverse effects of disclosing potential conflicts of interests in a simple setup with a binary choice. We find that the senders are no more likely to deceive with disclosure than with no disclosure. In addition, we find that disclosure of the sender's interests does not lead to moral licensing. Controlling for the senders' beliefs about the private material benefits of lying, deception rates do not increase with disclosure. If anything, disclosure renders senders less responsive to their own gains from lying. Moreover, the rate at which

the receivers follow the sender's advice is also not affected by the disclosure of sender interests.

We also test what happens when the decision to disclose or not to disclose the sender's interests is left to the receivers. Senders do not punish receivers for disclosing sender's interests and the receivers who do not reveal sender's interests are more likely to follow sender's advice than the receivers who do look at sender's interests. This suggests that there is a substantial fraction of gullible advisees, who are particularly vulnerable to deceptive advisors.

To summarize, we do not find any perverse effects of disclosure in our setup as reported in the literature. However, our results also show that disclosure of potential conflicts of interests does not help advice recipients. This suggests that other measures are necessary to protect advice recipients from biased advice.

A Model

In this section we present a theoretical analysis of the sender-receiver game with and without disclosure. Our main goal is to analyse the strategic effect of disclosing the sender's interests to the receiver on sender deception rate. The results show that, unlike the moral licensing effect, the strategic effect of disclosure on sender deception rate is ambiguous (i.e. can go in either direction).

There are two players: the sender (she) and the receiver (he). The receiver has to choose one of two options, Option A or Option B, but does not observe the payoffs. The sender observes the payoffs to both players for each option and sends one of the two possible messages: m = A ("Option A will earn you more than Option B"), and m = B ("Option B will earn you more than Option A").

A.1 States

There are four possible states: AA, AB, BA, and BB, where the first letter shows the option that gives the highest payoff to the sender and the second letter denotes the option that gives the highest payoff to the receiver. For example, at state AA Option A gives a higher payoff than Option B for both the sender and the receiver. To simplify analysis, for both the sender and the receiver we normalize payoffs such that the higher payoff is 1 and the lower payoff is 0.[6]

A.2 Senders

As mentioned above the sender observes the state and sends one of the two possible messages to the receiver. We assume that the sender incurs a cost,

[6]We assume that even in the no disclosure case the receiver knows that the sender higher payoff option gives 1 to the sender and the sender lower payoff option gives 0 to the sender. An alternative way is to assume that the sender payoffs for the higher and the lower payoff options are drawn from some distribution and the receiver forms an expectation based on this distribution. In this case there will be an additional effect of the disclosure, the receiver will know the exact size of the sender payoffs. We do not consider this effect because it complicates our model and does not change our main conlcusion that the strategic effect of disclosure can go either way.

c, from lying (sending the untruthful message). The cost of lying differs among senders and has a cumulative distribution function $F(c)$. Taking into account the cost of lying, the sender sends the message that gives her the highest expected payoff.

By σ_t we denote the proportion of senders who send Message A when the state is t. We assume that $\sigma_{AA} = 1$ and $\sigma_{BB} = 0$, i.e., that the senders send the truthful message when the interests are aligned. As will be seen later, given the equilibrium strategies of the receiver, the sender has no incentive to deviate from these strategies. By symmetry, $\sigma_{AB} = 1 - \sigma_{BA}$. For simplicity we will denote σ_{AB}, the proportion of senders who lie, by σ in what follows.

In the analysis of the sender behavior below, without loss of generality, we will assume that Option A gives a higher payoff to the sender than Option B.

A.3 Receivers

Using the principle of insufficient reason we assume that the receiver's prior belief that Option A gives him a higher payoff than Option B is $\frac{1}{2}$. In addition, let α denote the prior belief that the interests are aligned (that the state is either AA or BB). This gives the receiver's prior belief of being at each state: $\frac{1}{2}\alpha$ for state AA, $\frac{1}{2}(1-\alpha)$ for state AB, $\frac{1}{2}(1-\alpha)$ for state BA, and $\frac{1}{2}\alpha$ for state BB. We assume that α may differ across the receivers and is drawn from a distribution $G(\alpha)$.

As the game proceeds the receiver updates his beliefs using Bayes rule. By β_A we denote the receiver's belief that Option A is better for him than Option B conditional on receiving Message A and by β_B we denote the receiver's belief that Option A is better for him than Option B conditional on receiving Message B. In the analysis below, we assume that conditional on the message sent by the sender the receiver chooses the option that gives him the highest expected payoff.

A.4 Equilibrium with No Disclosure

We start by calculating the receiver's belief that A is the higher payoff option conditional on receiving Message A from the sender. By Bayes rule:

$$\beta_A = \frac{Pr((t = AA \text{ or } t = BA) \cap m = A)}{Pr(m = A)}$$
$$= \frac{1 * Pr(t = AA) + (1 - \sigma) * Pr(t = BA)}{1 * Pr(t = AA) + (1 - \sigma) * Pr(t = BA) + \sigma * Pr(t = AB) + 0 * Pr(t = BB)} \tag{1}$$

This gives $\beta_A = \alpha + (1 - \alpha)(1 - \sigma)$. The receiver's expected payoff from choosing Option A conditional on receiving Message A is $\beta_A * 1 + (1 - \beta_A) * 0 = \beta_A$. Likewise, the expected payoff from choosing Option B conditional on receiving Message A is $\beta_A * 0 + (1 - \beta_A) * 1 = 1 - \beta_A$. This means that the receiver will follow Message A when $\beta_A \geq 1 - \beta_A$ and will not follow otherwise. Substituting for β_A and rearranging, we have that the receivers with $\alpha \geq 1 - \frac{1}{2\sigma}$ will follow the sender message. This gives $1 - G(1 - \frac{1}{2\sigma})$ as the proportion of receivers who follow message A. Note that as the proportion of senders who lie, σ, increases, the proportion of receivers who follow the message decreases and vice versa. By symmetry, the proportion of receivers who follow message B is equal to the proportion of receivers who follow message A.

Next, we analyze senders. Without loss of generality, we consider the case where Option A gives a higher payoff to the sender than Option B and will derive $\sigma (= \sigma_{AB} = 1 - \sigma_{BA})$, the probability that the sender sends the deceitful message $m = A$. Let p_A denote the probability that the receiver will choose Option A conditional on receiving message A and p_B the probability that the receiver will choose Option A conditional on message B. From above we have that $p_A = 1 - G(1 - \frac{1}{2\sigma})$ and $p_B = G(1 - \frac{1}{2\sigma})$ because it is equal to the complementary probability of p_A by symmetry.

The sender lies whenever the expected payoff of lying minus the cost of lying is higher than the expected payoff of sending the truthful message. The sender receives $p_A - c$ from lying and the expected payoff of sending the truthful message is p_B. This means the sender lies when $(p_A - p_B) - c > 0$. By rearranging we obtain that all senders with $c < p_A - p_B$ lie. Thus, we have

$$\sigma = F(p_A - p_B). \tag{2}$$

Note that from above we also have that

$$p_A - p_B = 1 - 2 * G(1 - \frac{1}{2\sigma}) \tag{3}$$

By solving equations (2) and (3) simultaneously we can find the equilibrium values of σ and $p_A - p_B$. Note that the equilibrium values depend on the functional forms of the cumulative distribution functions G and F. In Figure A.1 we illustrate the equilibrium in (σ, $p_A - p_B$) plane for the case when α is uniformly distributed between 0 and 1 and c is uniformly distributed between 0 and $\frac{3}{4}$.

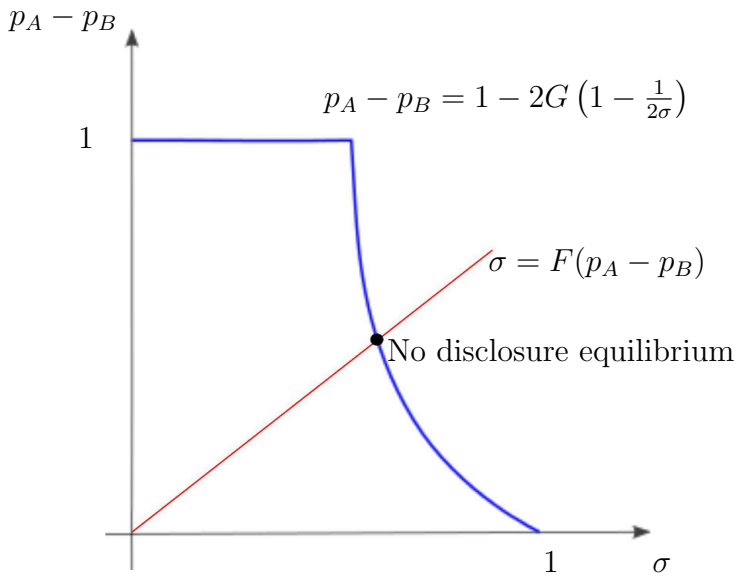


Figure A.1: No Disclosure Equilibrium

A.5 Equilibrium with Disclosure

Since the sender gets a higher payoff from Option A than from Option B, with disclosure the receiver knows that the state is either AA or AB. We start by calculating β_A^D and β_B^D using the Bayes rule.

$$\begin{aligned} \beta_A^D &= \frac{Pr(t = AA \cap m = A)}{Pr(m = A)} \\ &= \frac{1 * Pr(t = AA)}{1 * Pr(t = AA) + \sigma^D * Pr(t = AB)} = \frac{\alpha}{\alpha + \sigma^D(1 - \alpha)}, \end{aligned} \tag{4}$$

and

$$\begin{aligned}\beta_B^D &= \frac{Pr(t = AA \cap m = B)}{Pr(m = B)} \\ &= \frac{0 * Pr(t = AA)}{0 * Pr(t = AA) + (1 - \sigma^D) * Pr(t = AB)} = 0.\end{aligned} \tag{5}$$

Thus, with disclosure the non self-interest message is revealing. Similar to the *No Disclosure* case, the receiver will follow Message A when $\beta_A^D \geq (1 - \beta_A^D)$ and will not follow otherwise. This means that with disclosure the receiver follows the sender's self interest message when $\alpha \geq 1 - \frac{1}{1+\sigma^D}$. On the other hand, all receivers will follow message B because $\beta_B^D \leq (1 - \beta_B^D)$.

The sender will send the message that gives her the higher expected payoff (taking into account the cost of lying). Similar to the *No Disclosure* case, let p_A^D denote the probability that the receiver will choose Option A conditional on receiving message A and p_B^D the probability that the receiver will choose Option A conditional on receiving message B. We have that $p_A^D = 1 - G(1 - \frac{1}{1+\sigma^D})$ and $p_B^D = 0$. This gives us

$$p_A^D - p_B^D = 1 - G(1 - \frac{1}{1 + \sigma^D}). \tag{6}$$

The sender will lie to the receiver when $c \leq p_A^D - p_B^D$. Thus, the proportion of senders who lie with disclosure is given by equation

$$\sigma^D = F(p_A^D - p_B^D). \tag{7}$$

Solving equations (6) and (7) one can find the equilibrium values of σ^D and $p_A^D - p_B^D$. As an example, in Figure A.2 we show graphically the equilibria with and without disclosure for the specific functional forms of $G(\alpha)$ and $F(c)$ we assumed above. Note that for the given G the impact of the disclosure on the proportion of senders who lie depends on the shape of the cumulative distribution function, F, and can go in either direction. In the example we draw the proportion of senders who lie increases with disclosure.

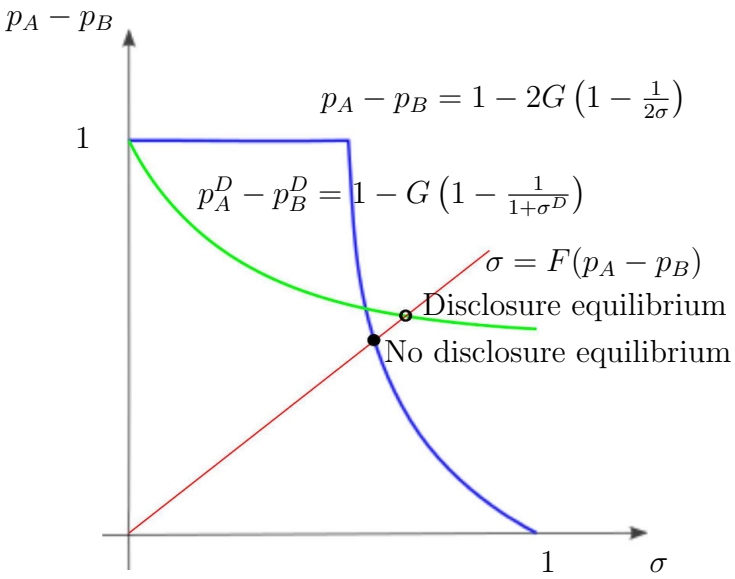


Figure A.2: No Disclosure and Disclosure equilibria

We can also illustrate the moral licencing effect in our model. Let F^D be the cumulative distribution function of lying costs with disclosure. Assume that with disclosure senders find lying morally more acceptable than without disclosure. This can be captured by assuming that for given c we have $F(c) \leq F^D(c)$. In other words, the cumulative distribution function F^D first order stochastically dominates F. This means that the graph of equation (7) will move to the right and this will increase the proportion of senders who lie. In Figure A.3 we illustrate the moral licensing effect assuming that c is distributed uniformly between 0 and $\frac{3}{4}$ without disclosure and uniformly between 0 and $\frac{5}{8}$ with disclosure.

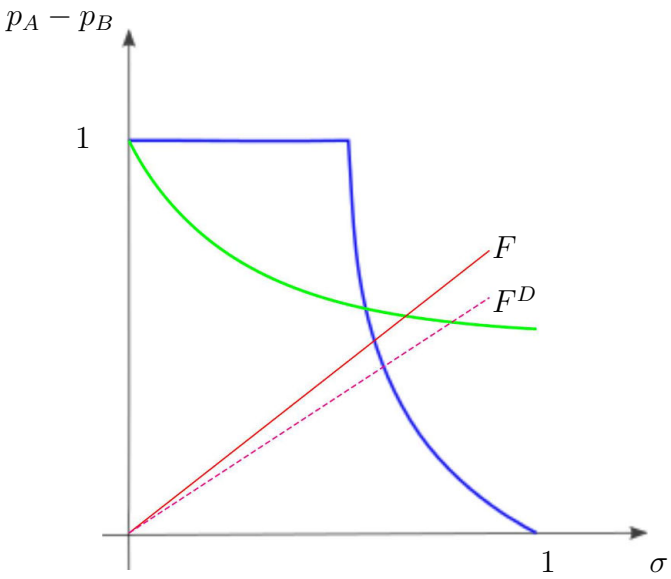


Figure A.3: The Moral Licensing effect

B Instructions

B.1 Treatment No Disclosure

General Instructions

Thank you for participating in this experiment. The experiment consists of two rounds. In each round, you will be paired with one other participant. In each pair, one person will have the role of player 1, and the other will have the role of player 2. Your role will be the same in each of the two rounds. The participant in the other role will be different in round 1 and round 2. No participant will ever know the identity of his or her counterpart in any round.

At the end of the experiment, one of the two rounds will be chosen at random. The amount of money you earn in this experiment will be equal to your payments in the chosen round. These payments depend on the decisions made in your pair in that round. The money you earn will be paid to you privately and in cash at the end of the experiment.

You are not allowed to talk or communicate to other participants. If you have a question, please raise your hand and I will come to your table.

(Player 1 instructions)

You are player 1

In each round, two possible monetary payments will be available to you and your counterpart in that round. These payment options are labeled Option A and Option B. Note that the payments corresponding to Option A and Option B are not necessarily the same in round 1 and round 2. At the beginning of the round you will see the payments to you and your counterpart for Option A and Option B on your computer screen.

The choice rests with the other participant who will have to choose either Option A or Option B. The only information your counterpart will have is information sent by you in a message. That is, he or she does not know the monetary payments associated with each option.

After you are informed about the payments corresponding to Options A and Options B, you can choose one of the following two messages to send to your counterpart:

Message 1: "Option A will earn you more money than Option B."

Message 2: "Option B will earn you more money than Option A."

Your message will be sent to your counterpart, and he or she will choose either Option A or Option B. This is done as follows. Before your counterpart receives your message, he or she has to decide which option (A or B) he or she wants to choose in case you send Message 1 and which option (A or B) he or she wants to choose in case you send Message 2. After your message is sent, the option chosen by your counterpart (Option A or Option B) is implemented.

Your message will be sent to your counterpart as soon as all participants in the experiment have entered their decisions.

To repeat, in each round your counterpart's choice will determine the payments of that round. Note however that your counterpart will never know what his or her payment was in the option not chosen (that is, he or she will never know whether your message was true or not). Moreover, he or she will never know your payments of the different options.

You and your counterpart will not get any information on the outcomes of the first round until after the second round is finished. Once the two rounds are over, one of the rounds will be chosen randomly and the outcome of that round will determine your payments.

At certain points during the experiment you will have an opportunity to earn a small bonus by making guesses about what your counterpart will choose. You will receive more information on your screen.

(Player 2 instructions)

You are player 2

In each round, two possible monetary payments are available to you and your counterpart in the round. These payment options are labeled Option A and Option B. The actual payments depend on the option you choose. We show the two payment options on the computer screen of your counterpart for that round, that is, he or she knows his or her own payments and also your payments for Option A and Option B. Note that the payments corresponding to Option A and Option B are not necessarily the same in round 1 and round 2. The only information you will have is the message your counterpart for that round sends to you. Two possible messages can be sent:

Message 1: "Option A will earn you more money than Option B."

Message 2: "Option B will earn you more money than Option A."

Before you receive the message, you will be asked which option (A or B) you want to choose in case you receive Message 1, and which option (A or B) you want to choose in case you receive Message 2.

You will receive the message of your counterpart as soon as all participants in the experiment have entered their decisions.

To repeat, in each round your counterpart for the round will send one of two possible messages to you. You decide which choice you want to make in that case: Option A or Option B. Your choice will determine the payments for the round. You will never know what payments were actually offered in the option not chosen (that is, whether the message sent by your counterpart was true or

not). Moreover, you will never know the payments to your counterpart in the two options.

You and your counterpart will not get any information on the outcomes of the first round until after the second round is finished. Once the two rounds are over, one of the rounds will be chosen randomly and the outcome of that round will determine your payments.

B.2 Treatment Disclosure

General Instructions

Thank you for participating in this experiment. The experiment consists of two rounds. In each round, you will be paired with one other participant. In each pair, one person will have the role of player 1, and the other will have the role of player 2. Your role will be the same in each of the two rounds. The participant in the other role will be different in round 1 and round 2. No participant will ever know the identity of his or her counterpart in any round.

At the end of the experiment, one of the two rounds will be chosen at random. The amount of money you earn in this experiment will be equal to your payments in the chosen round. These payments depend on the decisions made in your pair in that round. The money you earn will be paid to you privately and in cash at the end of the experiment.

You are not allowed to talk or communicate to other participants. If you have a question, please raise your hand and I will come to your table.

(Player 1 instructions)

You are player 1

In each round, two possible monetary payments will be available to you and your counterpart in that round. These payment options are labeled Option A and Option B. Note that the payments corresponding to Option A and Option B are not necessarily the same in round 1 and round 2. At the beginning of the round

you will see the payments to you and your counterpart for Option A and Option B on your computer screen.

The choice rests with the other participant who will have to choose either Option A or Option B. Your counterpart knows *your* payments for Option A and Option B, but does not know her or his *own* payments for Option A and Option B. The only other information your counterpart will have is a message sent by you.

After you are informed about the payments corresponding to Options A and Options B, you can choose one of the following two messages to send to your counterpart:

Message 1: "Option A will earn you more money than Option B."

Message 2: "Option B will earn you more money than Option A."

Your message will be sent to your counterpart, and he or she will choose either Option A or Option B. This is done as follows. Before your counterpart receives your message, he or she has to decide which option (A or B) he or she wants to choose in case you send Message 1 and which option (A or B) he or she wants to choose in case you send Message 2. After your message is sent, the option chosen by your counterpart (Option A or Option B) is implemented.

Your message will be sent to your counterpart as soon as all participants in the experiment have entered their decisions.

To repeat, in each round your counterpart's choice will determine the payments of that round. Note however that your counterpart will never know what his or her payment was in the option not chosen (that is, he or she will never know whether your message was true or not).

You and your counterpart will not get any information on the outcomes of the first round until after the second round is finished. Once the two rounds are over, one of the rounds will be chosen randomly and the outcome of that round will determine your payments.

At certain points during the experiment you will have an opportunity to earn a small bonus by making guesses about what your counterpart will choose. You will receive more information on your screen.

(Player 2 instructions)

You are player 2

In each round, two possible monetary payments are available to you and your counterpart in the round. These payment options are labeled Option A and Option B. The actual payments depend on the option you choose. We show the two payment options on the computer screen of your counterpart for that round, that is, he or she knows his or her own payments and also your payments for Option A and Option B. Note that the payments corresponding to Option A and Option B are not necessarily the same in round 1 and round 2.

In each round you will know the payments of your counterpart for Option A and Option B, but you will not know what your own payments for Option A and Option B. The only information you will have about your payments is the message your counterpart for that round sends to you. Two possible messages can be sent:

Message 1: "Option A will earn you more money than Option B."

Message 2: "Option B will earn you more money than Option A."

Before you receive the message, you will be asked which option (A or B) you want to choose in case you receive Message 1, and which option (A or B) you want to choose in case you receive Message 2.

You will receive the message of your counterpart as soon as all participants in the experiment have entered their decisions.

To repeat, in each round your counterpart for the round will send one of two possible messages to you. You decide which choice you want to make in that case: Option A or Option B. Your choice will determine the payments for the round. You will never know what payments were actually offered in the option not chosen (that is, whether the message sent by your counterpart was true or not).

You and your counterpart will not get any information on the outcomes of the first round until after the second round is finished. Once the two rounds are over, one of the rounds will be chosen randomly and the outcome of that round will determine your payments.

B.3 Treatment Endogenous Disclosure

General Instructions

Thank you for participating in this experiment. The experiment consists of two rounds. In each round, you will be paired with one other participant. In each pair, one person will have the role of player 1, and the other will have the role of player 2. Your role will be the same in each of the two rounds. The participant in the other role will be different in round 1 and round 2. No participant will ever know the identity of his or her counterpart in any round.

At the end of the experiment, one of the two rounds will be chosen at random. The amount of money you earn in this experiment will be equal to your payments in the chosen round. These payments depend on the decisions made in your pair in that round. The money you earn will be paid to you privately and in cash at the end of the experiment.

You are not allowed to talk or communicate to other participants. If you have a question, please raise your hand and I will come to your table.

(Player 1 instructions)

You are player 1

In each round, two possible monetary payments will be available to you and your counterpart in that round. These payment options are labeled Option A and Option B. Note that the payments corresponding to Option A and Option B are not necessarily the same in round 1 and round 2. At the beginning of the round you will see the payments to you and your counterpart for Option A and Option B on your computer screen.

The choice rests with the other participant who will have to choose either Option A or Option B. Your counterpart does not know her or his own payments for Option A and Option B. The only information your counterpart will have is a message sent by you.

After you are informed about the payments corresponding to Options A and Options B, you can choose one of the following two messages to send to your counterpart:

Message 1: "Option A will earn you more money than Option B."

Message 2: "Option B will earn you more money than Option A."

Your counterpart can request that *your* payments for Option A and Option B are revealed to him or her. You will be informed whether or not your counterpart made this request before you decide which message to send. Note that your counterpart will still not know his or her *own* payments for Option A and Option B if he or she enters the request.

Your message will be sent to your counterpart, and he or she will choose either Option A or Option B. This is done as follows. Before your counterpart receives your message, he or she has to decide which option (A or B) he or she wants to choose in case you send Message 1 and which option (A or B) he or she wants to choose in case you send Message 2. After your message is sent, the option chosen by your counterpart (Option A or Option B) is implemented.

Your message will be sent to you counterpart as soon as all participants in the experiment have entered their decisions.To repeat, in each round your counterpart's choice will determine the payments of that round. Note however that your counterpart will never know what his or her payment was in the option not chosen (that is, he or she will never know whether your message was true or not).

You and your counterpart will not get any information on the outcomes of the first round until after the second round is finished. Once the two rounds are over, one of the rounds will be chosen randomly and the outcome of that round will determine your payments.

At certain points during the experiment you will have an opportunity to earn a small bonus by making guesses about what your counterpart will choose. You will receive more information on your screen.

(Player 2 instructions)

You are player 2

In each round, two possible monetary payments are available to you and your counterpart in the round. These payment options are labeled Option A and Option B. The actual payments depend on the option you choose. We show the

two payment options on the computer screen of your counterpart for that round, that is, he or she knows his or her own payments and also your payments for Option A and Option B. Note that the payments corresponding to Option A and Option B are not necessarily the same in round 1 and round 2. The only information you will have about your payments is the message your counterpart for that round sends to you. Two possible messages can be sent:

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 can request that the payments of your counterpart for Option A and Option B are revealed to you. Your counterpart will be informed whether or not you made this request before he or she decides about the message to you. Note that you will still not know your own payments for option A and option B if you enter the request.

Before you receive the message, you will be asked which option (A or B) you want to choose in case you receive Message 1, and which option (A or B) you want to choose in case you receive Message 2.

You will receive the message of your counterpart as soon as all participants in the experiment have entered their decisions.

To repeat, in each round your counterpart for the round will send one of two possible messages to you. You decide which choice you want to make in that case: Option A or Option B. Your choice will determine the payments for the round. You will never know what payments were actually offered in the option not chosen (that is, whether the message sent by your counterpart was true or not).

You and your counterpart will not get any information on the outcomes of the first round until after the second round is finished. Once the two rounds are over, one of the rounds will be chosen randomly and the outcome of that round will determine your payments.

References

[1] Cain, Daylian M., George Loewenstein, and Don A. Moore. (2005): "The Dirt on Coming Clean: Perverse Effects of Disclosing Conflicts of Interest." *Journal of Legal studies*, 34(1), 1-25.

[2] Cain, Daylian M., George Loewenstein, and Don A. Moore. (2011): "When Sunlight Fails to Disinfect: Understanding the Perverse Effects of Disclosing Conflicts of Interest." *Journal of Consumer Research*, 37(5).

[3] Crawford, Vincent P., and Joel Sobel. (1982): "Strategic Information Transmission." *Econometrica*, 50(6): 1431-51.

[4] Financial Services Authority (2012), Insurance: Conduct of Business sourcebook.

[5] Fischbacher, Urs. (2007): "z-Tree: Zurich Toolbox for Ready-Made Economic Experiments," *Experimental Economics*, 10, 171-178.

[6] Gneezy, Uri. (2005): "Deception: the role of consequences," *American Economic Review*, 95, 384-394.

[7] Inderst, Roman, Rajko Alexander, and Ockenfels Axel. (2010): "Transparency and Disclosing Conflicts of Interest: An Experimental Investigation." *Discussion Paper No. 10-20.*

[8] Koch, Christopher, and Carsten Schmidt. (2009): "Disclosing Conflicts of Interest - Does Experience and Reputation matter?." *Accounting, Organisations, and Society*, 35(1): 95-107.

[9] Loewenstein, George, Daylian M. Cain, and Sunita Sah. (2011): "The Limits Of Transparency: Pitfalls and Potential of Disclosing Conflicts of Interest." *American Economic Review: Papers and Proceedings*, 101(3), 423-428.

[10] de Meza, David, Bernd Irlenbusch, and Diane Reyniers. (2011): "Disclosure, Trust and Persuasion in Insurance Markets." *IZA Discussion Paper No. 5060.*

[11] Rode, Julian. (2010): "Truth and trust in communication - Experiments on the effect of a competitive context." *Games and Economic Behavior*, 68, 325-338.

[12] Ross Joseph S., Josh E. Lackner, Peter Lurie, Cary P. Gross, Sidney Wolfe, and Harlan M. Krumholz (2007): "Pharmaceutical company payments to physicians. Early Experiences With Disclosure Laws in Vermont and Minnesota" *JAMA*, 297(11), 1216-1223.

[13] Sutter, Matthias. (2009): "Deception through telling the truth? Experimental evidence from individuals and teams." *Economic Journal*, 119, 47-60.