



Third party assessments in trust problems with conflict of interest: An experiment on the effects of promises



Mitesh Kataria*, Fabian Winter

Max Planck Institute of Economics, Strategic Interaction Group, Kahlaische Strasse 10, D-07745 Jena, Germany

HIGHLIGHTS

- We study the effect of promises and incentives on truthful assessments in a 3-player trust game.
- An assessor assesses the trustworthiness of his friend (i.e. the trustee).
- Trustors can condition their choice on this assessment.
- Assessor's promises to give a truthful assessment sign reduce favorable assessments.
- Promises reduce misreporting to the same extent as incentivized assessments.

ARTICLE INFO

Article history:

Received 11 December 2012

Received in revised form

28 February 2013

Accepted 15 March 2013

Available online 25 March 2013

JEL classification:

A13

C91

D03

Keywords:

Trust game

Communication

Cheap-talk

Promises

Oath

ABSTRACT

Is it possible to elicit reliable assessment from an assessor having a conflict of interest (e.g. a professor that writes a recommendation letter for a formal PhD student)? We propose an experimental test and show that compared to a not-incentivized assessment, a promise to give a truthful assessment reduces misreporting to the same extent as an incentivized assessment (i.e. when the assessor gains higher payoff if the assessment is correct).

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

Recruiting new employees is an unavoidable task for firms, governments as well as universities. One difficulty is that information about the applicants' abilities is usually limited. In many universities, taking an example, it is common to ask for a recommendation letter by a professor that has worked together with the applicant. For such information to be useful, honest reporting is required. Some professors, however, develop a friendship with their Ph.D. students, which creates a conflict of interest and the risk that the information is biased in favor of the applicant (Leising et al., 2010). Assessments play a crucial role in many job markets to inform the

employers about the applicants' abilities, and biased recommendations may have serious implications: Exaggerating the applicant's abilities increases the risk that the employer's expectations remain unfulfilled, often to the disadvantage of other, better candidates.

In this study we test a simple mechanism to overcome such biased assessments. We observe the behavior of Assessors in three treatments: when they are *not (monetary) incentivized* to tell the truth, when they have monetary *incentives* to tell the truth, and when they are not incentivized but sign a statement of honesty (*oath* henceforth).

There is evidence that promises work. Charness and Dufwenberg (2006) experimentally examine the impact of communication on trust and cooperation. They suggest that a promise works because of guilt aversion: A guilt-averse person does not want to let down others' expectations and is therefore motivated by beliefs about others' beliefs. An alternative explanation is that people

* Corresponding author. Tel.: +49 3641 686 632.

E-mail address: kataria@econ.mpg.de (M. Kataria).

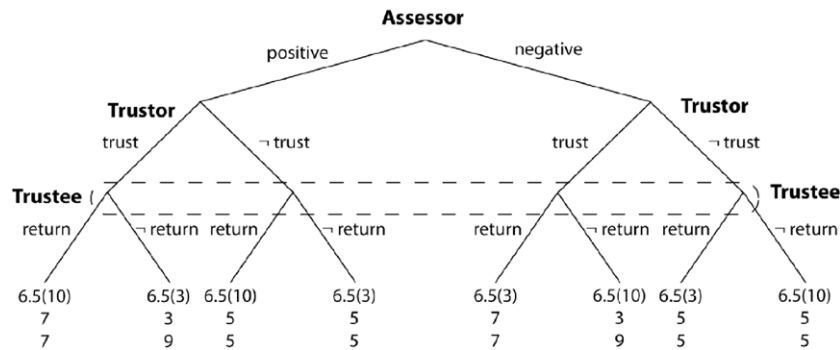


Fig. 1. Trust game with Assessor: numbers in parentheses indicate payoffs in the *Incent* treatment. Assessor and Trustee are friends with each other.

may have a taste for keeping their word (e.g. Ellingsen and Johannesson, 2004). Using a novel design, Vanberg (2008) found support for the latter explanation, i.e. people have preferences for promise-keeping per se.¹ Note that free form communication was used in all three cited studies. Interestingly, in Charness and Dufwenberg (2010) the free-form communication is replaced with an opportunity to send an exogenously given message of promise, and limited support for lying- and guilt aversion is found.

Jacquemet et al. (2013) explore the impact of oaths in an incentive-compatible second price auction. The oath treatments presented subjects with the opportunity to sign an oath prior to participating in the auction. By signing the oath, subjects “swear on their honor” to tell the truth and provide honest answers. Subjects who took the oath were on average less likely to either overstate or understate their bids. Carlsson et al. (forthcoming) tested an oath-like procedure in the field using non-market valuation surveys and found that the share of zero WTP responses and extremely high WTP responses decreases, which could be interpreted as reduced dishonesty. Shu et al. (2012) find that signing a statement of honesty at the beginning instead of at the end of a self-report serves as a commitment and leads to significant reductions in misreporting.²

We extend the existing literature on promises and oaths by asking for statements not about own intentions, but about information concerning the trustworthiness of friends.

2. The experiment

2.1. The trust game and treatments

In this study we extended a binary trust game to include an additional third player, called the Assessor. The game starts with the Assessor, who has private information about the trustee, because the Assessor and the Trustee are friends and know each other prior to the experiment, which is common knowledge. The Assessor has to assess whether the Trustee will later return the trust or not (i.e. give a *positive* or *negative* assessment of the Trustee’s trustworthiness).³ Next, the Trustor makes his choice whether to grant trust or not, and is free to condition his choice on this assessment.

The Trustor’s choice is elicited using the strategy method.⁴ Finally, without knowing both preceding players’ choices, the Trustee decides whether to return the trust or not. The game is played one-shot and shown in Fig. 1 (terminal nodes show the payoffs in Euros).

We implemented three treatments which differ in how the Assessor’s choice was framed and incentivized:

NotIncent: The Assessor is asked to assess whether the Trustee will return the trust (positive assessment) or not (negative assessment). Assessors receive 6.5 Euro for their assessment.

Incent: As *NotIncent*, except that the Assessor receives 10 (3) Euro for a correct (wrong) assessment. The assessment is thus incentivized.

Oath: As *NotIncent*, except that the Assessor is asked to sign a statement of honesty with the following wording before making the assessment:

I hereby give my word of honor that, to the best of my knowledge, I will provide an honest assessment of my friend’s decision.

In an additional sentence just after the statement of honesty we made clear to the subjects that their statement had no legal consequences, reading: “*I am aware that this assessment has no legal consequences*”. After every Assessor had signed the statement, the document was collected and it was announced that everyone had signed the oath.

2.2. Procedure

We ran twelve sessions with 18–30 subjects in each session. The subjects earned between 5.5 and 12.5 euros with an average of 8.6 euros. The experiment lasted for approximately 50 min. In total, 339 students participated in the experiment with 108 (*NotIncent*), 117 (*Incent*) and 114 (*oath*) subjects respectively.

¹ Sending a promise alters the sender’s beliefs about what the receiver will expect him to do and behavior in line with the promise made can be observed. The tendency to stand by one’s promise, however, depends largely on the fact that *oneself* made the promise, rather than on the receiver’s expectation: If a sender gave a promise to one receiver, but was later matched to another receiver who had received the same promise from a third player, behavior was not adjusted to match the promise, despite the belief that the promise raised the receivers’ expectations.

² Signing at the end is currently the predominant practice for e.g. tax returns or insurance policy forms.

³ In the experiment we used a neutral framing (e.g. choose left or right) and did not mention the word trust.

⁴ The strategy method has considerable advantages compared to the direct response method, e.g. it facilitates collection of richer datasets. A potential disadvantage, however, is that it may induce experimental demand effects (Zizzo, 2010). Although we acknowledge the problems in using the strategy method, we are less concerned about the demand effects in our experiment since we use the strategy method to elicit the Trustor’s choices, while our main hypothesis focuses on the Assessor’s behavior. For the latter, we rely on the direct response method to elicit choices. Also note that to use the direct response method to elicit the Trustor’s choice would be rather difficult without altering the Assessor’s monetary incentives. In the case that the Trustor does not grant trust and the game ends (and the Assessor’s performance cannot be evaluated), the Assessor’s payoff still has to be defined. Such incentives, however, would conflict with our aim to incentivize truthful recommendations in the incentivized treatment since depending on the Assessor preferences to earn safe money (which is earned if the Trustor chooses not to grant trust), he would have incentives to over- or understate his friend’s trustworthiness. This is avoided with the strategy method.

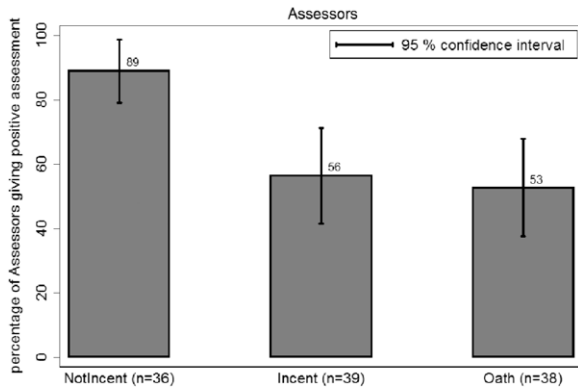


Fig. 2. Share of Assessors giving a positive assessment.

The subjects were undergraduate students from the Friedrich Schiller University Jena (Germany) and were recruited using ORSEE (Greiner, 2004). Invitations to subjects in the role of Assessors and Trustees asked these subjects to come to the lab together with a same gender friend. The experiment was programmed and conducted in z-Tree (Fischbacher, 2007).

3. Results

The results are presented starting with the choices of the Assessors, followed by the Trustees', and the Trustees' choices. Fig. 2 gives a visual impression of the assessments made by the Assessors.

The share of Assessors giving a positive assessment of their friend's trustworthiness is greatest in treatment *NotIncent* (89% positive assessments). Incentivizing honest reporting in the *Incent* treatment significantly reduces positive assessments to 56% ($z = 3.13, p < 0.006$).⁵ A very similar and significant reduction of positive assessments can be observed in *Oath*, with 53% positive assessments, ($z = 3.13, p < 0.002$), such that the differences between *Incent* and *Oath* are negligible and insignificant ($z = 0.33, p = 0.739$). Since misreporting for the friend's benefit obviously implies a positive assessment, this result suggests that both, incentives and oaths are successful means to increase honesty.

The share of Assessors that assessed these choices correctly was 56% (*NotIncent*), 64% (*Incent*) and 61% (*Oath*). The differences between the treatments are not significant. Since following the assessment is only beneficial to the Trustor if it is actually correct, these figures also represent an upper bound of Trustees that *could have maximized* their payoff by granting trust after a positive assessment and vice versa.⁶

Fig. 3 gives a visual impression of the Trustees' decisions. After a positive assessment, Trustees grant trust in 72% (*NotIncent*), 64% (*Incent*) and 66% (*Oath*) of the cases. These differences are, however, not statistically significant using the two-sided difference-in-proportions tests (*NotIncent* vs. *Incent*: $z = 0.75, p = 0.4514$; *NotIncent* vs. *Oath*: $z = 0.60, p = 0.5501$; *Incent* vs. *Oath*: $z = -0.016, p = 0.876$). On the other hand, the share of Trustees granting trust after a negative assessment is only 8%, 3% and 3%.

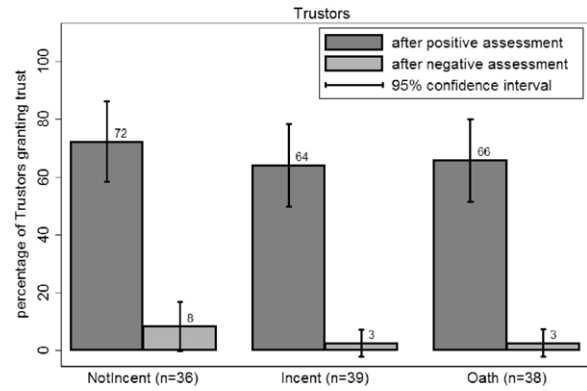


Fig. 3. Share of Trustees granting trust.

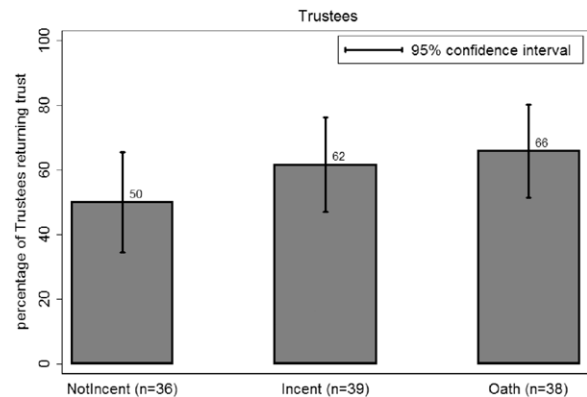


Fig. 4. Share of Trustees returning trust.

Again, the differences between treatments are insignificant (*NotIncent* vs. *Incent*: $z = 1.11, p = 0.2666$; *NotIncent* vs. *Oath*: $z = 1.08, p = 0.2783$; *Incent* vs. *Oath*: $z = 0.02, p = 0.9852$). Finally, it is evident that the Trustees are much more likely to trust the Trustees after a positive assessment. Since the confidence intervals do not overlap we conclude that the effect is significant at 5% level directly from Fig. 3.⁷

The share of Trustees proving trustworthy by returning trust is 50% (*NotIncent*), 62% (*Incent*), and 66% (*Oath*), respectively (see Fig. 4). The differences between treatments are never significant for any conventional significance level using two-sided proportional tests (*NotIncent* vs. *Incent*: $z = -1.01, p = 0.3145$; *NotIncent* vs. *Oath*: $z = -1.38, p = 0.1688$; *Incent* vs. *Oath*: $z = -0.39, p = 0.698$).

4. Concluding discussion

The truthfulness of an assessment by a professor considering the abilities of a former Ph.D. student, or an employer about a former employee, are two applications highlighting the importance of reliable communication in economic interactions. Relying on such assessments involves the risk that assessments may be biased in favor of the applicants, and hired applicants could turn out to be unreliable themselves.

⁵ The reported treatment effects are tested using two-sided difference in proportions tests. Normal distributed z-score based on differences in proportions is calculated by $z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}_p \hat{q}_p (1/n_1 + 1/n_2)}}$, where the numerator states the difference in proportion between the two treatments $\hat{q}_p = 1 - \hat{p}_p$, $\hat{p}_p = (x_1 + x_2)/(n_1 + n_2)$, and x_1 and x_2 are the total number of successes in the two treatments.

⁶ The joint probability that a Trustor follows an assessment and that the assessment is correct is 39%, 51%, and 42%, respectively (no significant differences).

⁷ It is well known that non-overlapping confidence intervals overstate significance level. The results are also confirmed using two-sided, within-sample test of proportions for any conventional significance level.

In this study, we augmented a trust game to test if elicitation under oath can mitigate biases in such assessments. Three key findings have emerged. First, the Assessors were more likely to give positive assessments by stating that the trustees will return trust in the *Not Incentivized* treatment compared to the *Incentivized* treatment and the *Oath* treatment. We interpret this finding as a reduced bias in assessments in the later treatments. Second, there is no difference in the assessments between the *Incentivized* and *Oath* treatment. Third, we find that Trustors significantly base their decision to trust or not on these assessments. Trustors are much more likely to grant trust after a positive assessment than they are after a negative one. This is, however, not affected by the Assessors' incentive structure. They choose to trust after a positive assessment at the same high level across all treatments, just as they choose not to trust after a negative assessment.

In conclusion, our results suggests that the bias in assessments can be significantly reduced using statements of honesty, a cheap and simple mechanism that appeals to psychological aspects of emotional commitments.

Acknowledgments

Helpful comments by an anonymous reviewer and the help of Adrian Liebtrau, Andreas Matzke, Lisa Schöttl and Ria Stangneth in conducting the experiments is gratefully acknowledged.

References

- Carlsson, F., Kataria, M., Krupnick, A., Lampi, E., Löfgren, A., Qin, P., Sterner, T., 2013. The truth, the whole truth, and nothing but the truth: a multiple country test of an oath script. *Journal of Economic Behavior & Organization* (forthcoming). Available online 27 February 2013, ISSN 0167-2681, <http://dx.doi.org/10.1016/j.jebo.2013.02.003>.
- Charness, G., Dufwenberg, M., 2006. Promises and partnership. *Econometrica* 74, 1579–1601.
- Charness, G., Dufwenberg, M., 2010. Bare promises: an experiment. *Economics Letters* 107, 281–283.
- Ellingsen, T., Johannesson, M., 2004. Promises, threats and fairness. *The Economic Journal* 114, 397–420.
- Fischbacher, U., 2007. z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics* 10 (2), 171–178.
- Greiner, B., 2004. An online recruitment system for economic experiments. In: Kremer, Kurt, Macho, Volker (Eds.), *Forschung und wissenschaftliches Rechnen 2003*. In: GWDG Bericht, vol. 63. Ges. für Wiss. Datenverarbeitung, Göttingen, pp. 79–93.
- Jacquemet, N., Joule, R.V., Luchini, S., Shogren, J., 2013. Preference elicitation under oath. *Journal of Environmental Economics and Management* 65, 110–132.
- Leising, D., Erbs, J., Fritz, U., 2010. The letter of recommendation effect in informant ratings of personality. *Journal of Personality and Social Psychology* 98, 668–682.
- Shu, L., Mazar, N., Gino, F., Ariely, D., Bazerman, M., 2012. Signing at the beginning makes ethics salient and decreases dishonest self-reports in comparison to signing at the end. *Proceedings of the National Academy of Sciences of the United States of America* 109 (38), 15197–15200.
- Vanberg, C., 2008. Why do people keep their promises? An experimental test of two explanations. *Econometrica* 76, 1467–1480.
- Zizzo, D.J., 2010. Experimenter demand effects in economic experiments. *Experimental Economics* 13 (1), 75–98.