Is mistrust self-fulfilling?

Reuben, Ernesto and Sapienza, Paola and Zingales, Luigi

Northwestern University, University of Chicago

May 2008

IS MISTRUST SELF-FULFILLING?*

ERNESTO REUBEN

Northwestern University

PAOLA SAPIENZA

Northwestern University, NBER, and CEPR

LUIGI ZINGALES

University of Chicago, NBER, and CEPR

ABSTRACT

We study experimentally the effect of expectations on trustworthiness. Most subjects respond with untrustworthy behavior if they find out that little is expected from them. This suggests that guilt aversion plays an important role in inducing trustworthiness.

Date of this version: May 2008

Keywords: trust; trustworthiness; reciprocity; guilt aversion; experiment

JEL classification: C72; C91; Z13

---

* We thank the Templeton Foundation for financial support.

1. Introduction

Countries differ enormously in their level of trust. According to the World Value Survey, 60% of Norwegians think that most people can be trusted, while only 5% of Brazilians think so (Inglehart, 1999). A compelling explanation for these differences is that they are culturally based. Citizens of trusting countries expect others to be trustworthy while citizens of less trusting countries expect others to be untrustworthy (Guiso et al., 2004). It is clear how low expectations affect someone's willingness to trust. An interesting question, however, is whether low expectations (if manifested) can affect trustworthiness. The answer to this question has important implications for the stability of trust (or lack thereof).

To observe whether expectations impact trustworthiness we ran an experiment based on the trust game (Berg et al., 1995). In the game, a *sender* is endowed with $50, which he can keep or send to a *receiver*. If the sender sends, the receiver receives $150 of which she can return any amount to the sender. Since unreturned money is kept by the receiver, the amount returned can be considered a measure of trustworthiness.

Subjects play the game twice as receivers. The first time, they do not know the amount the sender expects back. This gives us a measure of their trustworthiness. The second time, they are informed of the sender's expectation. Half of them are matched with senders with high expectations and the rest with senders with low expectations. We test whether observing a high/low expectation increases/decreases the amount returned.

In this context, it is useful to consider models where the senders' expectations have a direct impact on trustworthiness. In models of reciprocity (Dufwenberg and Kirchsteiger 2005; Falk and Fischbacher, 2006), trustworthiness is the result of receivers reciprocating kind actions by senders. Since senders with low expectations who nevertheless send are being particularly kind—given their expectation they are giving away money—facing such senders motivates receivers to be *more* trustworthy. In contrast, in models of guilt aversion (Charness and Dufwenberg, 2006), trustworthiness

is the result of receivers not wanting to disappoint the sender. Hence, since senders with low expectations are not disappointed by low returns, facing such senders motivates receivers to be *less* trustworthy.

2. Design

The experiment was conducted in October 2007 with MBA students from Northwestern University. Average earnings in the game were $47.39. The detailed experimental procedures are found in Reuben et al. (2008).

In order to convey the senders' true expectations in an incentive-compatible way, subjects made four decisions. Decision 1 consists of playing the trust game as a sender. As decision 2, subjects are asked to predict the amount returned by their receiver.[1] As decision 3, subjects play the trust game as receivers and choose how much to return conditional on receiving the $150. Decision 4 consists of playing the trust game once more. Senders are told that their receiver will be informed of their expected return (elicited in decision 2). Receivers are told their sender's expectation. This procedure prevents senders from manipulating their expectation but allows them to avoid any adverse effects of having it revealed. The role of sender is given at random to subjects with a low[2] or high expectation (less than $10 or more than $70).[3]

To avoid subjects from influencing each other in-between decisions, they are not informed of the behavior of others nor are they told what the future decisions are. Subjects know there are four decisions and that one will be paid at random.

---

[1] Subjects are paid according to their accuracy: they get max{$75 – |actual trustworthiness – predicted trustworthiness|, $0}. Receivers choose using the strategy method. Hence, they always submit an amount.

[2] One might worry that the receivers' choice is irrelevant for senders with low expectations. However, given the various motivations for sending (e.g. altruism), we consider there is always a positive probability that the sender sends. In the experiment, 25% of those expecting back $10 or less send money.

[3] Subjects are unaware that decision 2 affects future roles. They are simply informed of the role they were assigned. A similar procedure is used by, for example, Gächter and Thöni (2005).

3. Related work

There are a few papers that study the effect of expectations on trustworthiness. Dufwenberg and Gneezy (2000), Charness and Dufwenberg (2006), and Bacharach et al. (2007) study variations of trust games and measure the receivers' expectation of the senders' expectation concerning their trustworthiness. They find that receivers with high second-order beliefs are more trustworthy.

The paper that comes closest to our study is Ellingsen et al. (2007). The authors report results of a trust game in which the senders' expectations are communicated to the receivers (senders are unaware that their expectation is revealed). They do not find a significant correlation between the revealed expectations and trustworthiness. They suggest that the relation between second-order beliefs and returns is driven by the "false consensus effect". Our design provides three improvements. First, measuring each subject's initial level of trustworthiness and then observing the *change* induced by a low/high expectation, allows us to control for other motivations to return money (e.g. altruism). Second, concentrating on low/high expectations makes it easier to observe their impact. Third, eliciting expectations from all subjects ensures that receivers understand they are observing the senders' true expectation.

4. Results

For the statistical analysis, we concentrate on the 52 subjects who were receivers in decisions 3 and 4. In the latter, 24 (28) subjects faced a sender with a low (high) expectation.

Overall, 67.3% of the subjects send money as senders. Moreover, on average, their expectations are relatively accurate. The mean expected return is \$56.67, whereas the actual mean return in decision 3 is \$52.88.

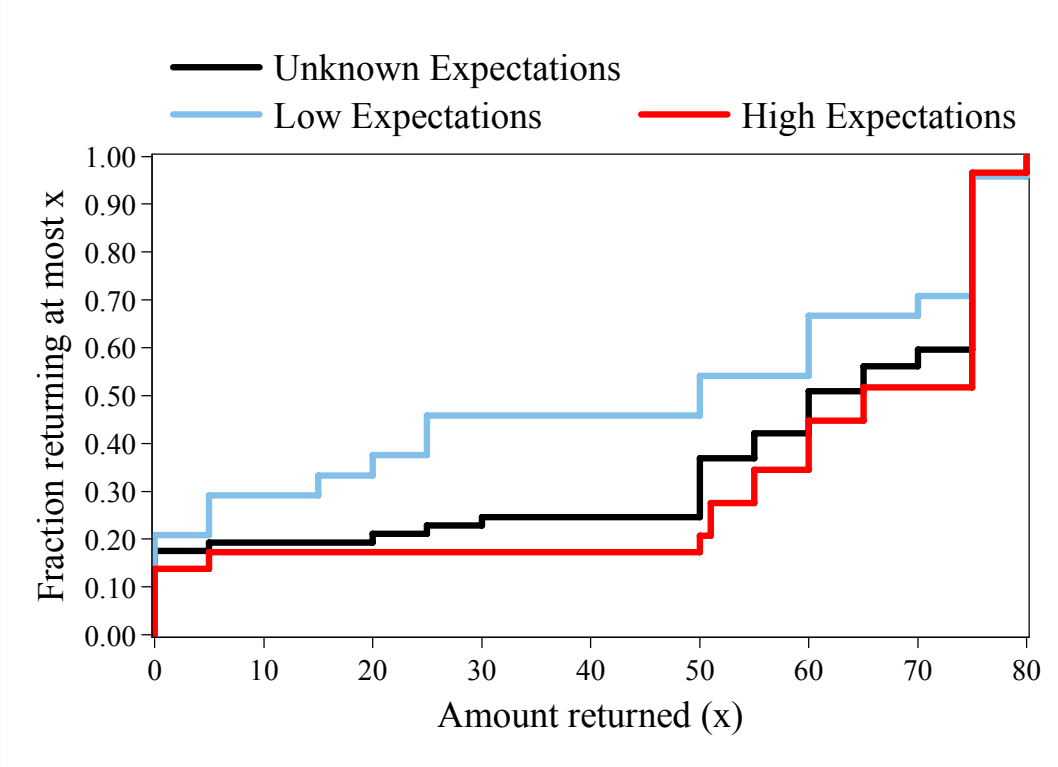

FIGURE 1 – CUMULATIVE DISTRIBUTIONS OF TRUSTWORTHINESS

The senders' expectations—observed in decision 4—produce a stark difference in trustworthiness. On average, receivers who observe a low expectation decrease their return from \$53.75 to \$40.63. The decrease is significant with a Wilcoxon-signed-ranks test (p = 0.030). A majority of them, 54%, decrease the amount returned, 33% keep it constant, and 13% increase it.[4]

In contrast, receivers who observe a high expectation increase their return from \$52.14 to \$58.29. The increase, however, is not significant with a two-sided Wilcoxon-signed-ranks test (p = 0.125). In this case, 61% of the receivers keep their return constant, 28% increase it, and 11% decrease it.

[4] This effect seems to be due to the sender's expectation and not to receivers typing the number on the screen. Of those who change their amount returned, only 30% return the observed expectation. Thus, a considerable majority are clearly thinking of how to adjust their decision and are not blindly complying with the sender's expectation or using it as an anchor. Ellingsen et al. (2007) directly test the anchoring hypothesis and find no effect.

The difference in trustworthiness is clearly seen in Figure 1, which reports the cumulative distribution functions of the returned amounts in the three situations (the "unknown expectations" line is the CDF of decision 3). It can be seen that the CDF under low expectations first-order stochastically dominates the CDF under unknown expectations, which in turn first-order stochastically dominates the CDF under high expectations.

In fact, given the structure of the game, the observed effect of expectations can be considered as a lower bound of the total effect. Receivers who return $0 when expectations are unknown cannot further decrease their return if they see a low expectation. Similarly, since it seems most receivers consider the equal split as an upper bound to the amount returned (36.5% return exactly $75, 59.6% return less, and only 3.9% return more), receivers who initially returned $75 cannot further increase their return if they observe a high expectation. To take this into account and to control for additional variables, we ran Tobit regressions with the amount returned in decision 4 as the dependent variable (censoring at $0 and $75, see Table 1).

In column I we use four independent variables: a dummy variable indicating whether the subject sees a low expectation, the subject's trustworthiness (under unknown expectations), a dummy variable indicating whether the subject sent money in the sender role, and the subject's expectation of the amount returned by others.

Compared to a sender with high expectations, playing with a mistrusting sender is associated with a $31.77 lower return. This is a striking 43% difference in trustworthiness. Also strong is the effect of the subjects' expectation of the amount returned by others: those who expect others to be trustworthy return higher amounts. Finally, the amount returned under unknown expectations has a positive effect.

In column II we run the same regression including the only significant interaction term: the interaction between observing a low expectation and the subject's trustworthiness under unknown expectations. In this regression, the dummy variable for low expectations loses significance, which is picked up by the interaction variable.

TABLE 1 – ESTIMATION OF TRUSTWORTHINESS

	I	II
Low expectations	-31.77**	26.09
	(13.66)	(27.91)
Trustworthiness (unknown expectations)	0.51*	0.97***
	(0.29)	(0.35)
Sent money	2.96	3.59
	(16.82)	(15.92)
Expected trustworthiness	1.13**	1.19**
	(0.49)	(0.48)
Trustworthiness (unknown expectations) ×		-1.11**
low expectations		(0.49)
Constant	-17.1	-44.74
	(23.96)	(27.65)

Note: Standard errors in parentheses, *, **, *** indicate statistical significance at 10%, 5%, and 1%.

The coefficients suggest that individuals who observe a high expectation keep constant their level of trustworthiness (the coefficient is not different from one). Observing a low expectation, however, eliminates completely the effect of their original trustworthiness (the sum of the two coefficients is not different from zero). Interestingly, the receivers' own expectations do not significantly interact with the low-expectations dummy. This suggests that the decrease in trustworthiness is driven by the knowledge of the sender's mistrust.

5. Conclusions

In this paper we study the effect on trustworthiness of the senders' expectations in a trust game. In line with the guilt-aversion hypothesis, we show that most receivers respond with lower returns when they face a sender with low expectations. These results suggest that reciprocity and motivations such as altruism are not sufficient to

explain trustworthiness. Receivers are also driven by a desire to comply with the sender's expectation.

Our results suggest that expecting others to be untrustworthy (if manifested) causes untrustworthy behavior, which further confirms the initially low expectations. This makes mistrust self-fulfilling and a mistrust equilibrium particularly difficult to change.

References

Bacharach, M., Guerra, G., and Zizzo, D.J. (2007). The Self-Fulfilling Property of Trust: An Experimental Study. *Theory and Decision* 63: 349-388.

Berg, J., Dickhaut, J., and McCabe, K. (1995). Trust, Reciprocity and Social History. *Games and Economic Behavior* 10: 122-142.

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

Dufwenberg, M. and Gneezy, U. (2000). Measuring Beliefs in an Experimental Lost Wallet Game. *Games and Economic Behavior* 30: 163-82.

Dufwenberg, M. and Kirchsteiger, G. (2005). A theory of sequential reciprocity. *Games and Economic Behavior* 47: 268-298.

Ellingsen, T., Johannesson, M., Tjøtta, S., and Torsvik, G. (2007). Testing Guilt Aversion. SSE/EFI Working Paper.

Falk, A. and Fischbacher, U. (2006). A theory of reciprocity. *Games and Economic Behavior* 54: 293-315.

Gächter, S. and Thöni, C. (2005). Social Learning and Voluntary Cooperation Among Like-Minded People. *Journal of the European Economic Association* 3: 303-314.

Guiso, L., Sapienza, P., and Zingales, L. (2004). Cultural Biases in Economic Exchange. NBER WP-11005.

Inglehart, R. (1999). Trust, well-being and democracy. In Warren, M.E. (ed) *Democracy and Trust*, Cambridge University Press.

Reuben, E., Sapienza, P., and Zingales, L. (2008). Is Mistrust Self-fulfilling? Supplementary Material. Mimeo.