Volume 33, Issue 1

Does the direct-response method induce guilt aversion in a trust game?

David Amdur
Muhlenberg College

Ethan Schmick
Muhlenberg College

Abstract

We compare the strategy and direct-response methods in a one-shot trust game with hidden action. In our experiment, the decision elicitation method is not statistically associated with participants' behavior or beliefs. We find no evidence that the direct-response method induces guilt aversion.

1. Introduction

Individuals often behave in trustworthy ways, despite opportunities to earn higher material payoffs by betraying another's trust.[1] Why do people repay the trust of others? One theory is that they would feel guilty if they did not (Charness and Dufwenberg 2006, Battigalli and Dufwenberg 2007). According to the theory of *guilt aversion*, people experience disutility when they fail to live up to the expectations of others. What factors induce guilt? Must a trustee make – and break – an explicit promise, or can guilt be induced by the revealed expectation of a truster? The answer has important implications for efforts aimed at promoting trust and cooperation.

To examine the link between revealed expectations and trustworthy behavior, we conduct a one-shot trust game with hidden action, similar to Charness and Dufwenberg (2006). Player A (the "truster") chooses whether or not to trust Player B (the "trustee"); and Player B, if trusted, chooses whether or not to reciprocate. We use two different decision elicitation treatments. Under the strategy method, Player B decides whether to reciprocate *contingent on being trusted by Player A*, but without observing A's actual decision. Under the direct-response method, B makes her choice *after observing A's actual decision*.[2] We hypothesize that B's who observe a trusting decision by A's in the direct-response treatment will manifest more "guilt aversion" than B's in the strategy method treatment. In fact, we find no evidence of this. In our experiment, observability of the truster's decision is not statistically associated with the frequency of trustworthy behavior or participants' beliefs about this behavior.

2. Guilt Aversion in a Trust Game

Our experiment is similar to the one conducted by Charness and Dufwenberg (2006), henceforth "CD". The game tree is shown in Figure 1. Payoffs are in dollars, not necessarily utilities that may depend on psychological factors such as guilt.[3] Player A first chooses *In* or *Out*. If Player A chooses *In*, then Player B chooses whether to *Roll* or *Don't Roll* a six-sided die. If A and B choose the strategy profile (*In*, *Roll*), then the die roll determines the final monetary payoffs. Assuming risk-neutral and narrowly self-interested preferences, the only subgame perfect Nash equilibrium is (*Out*, *Don't Roll*). Note, however, that the equilibrium is inefficient. Both players would get higher expected payoffs if they could commit, *ex-ante*, to the strategy profile (*In*, *Roll*). Commitment is ruled out by the assumption that A cannot observe B's decision.

CD show that when Player B can send a non-binding, free-form message to Player A prior to the game, B players are more likely to choose *Roll*. CD argue that guilt aversion can explain this result. Their hypothesis is that B experiences guilt whenever she believes that she has failed to live up to A's expectation. CD show that pre-game messages from B to A are associated with higher B beliefs about A's expectation that B will choose *Roll*, consistent with guilt aversion.[4]

---

[1]See Fehr and Gachter (2000) for a survey of early work on trust and reciprocity.

[2]The strategy method was first proposed by Selten (1967). See Brandts and Charness (2011) for a review of experimental studies comparing the strategy and direct-response methods.

[3]Our monetary payoffs are 30% of those in CD.

[4]Charness and Dufwenberg (2010) document more modest effects in the same direction using constrained messages.

Figure 1: The game tree.
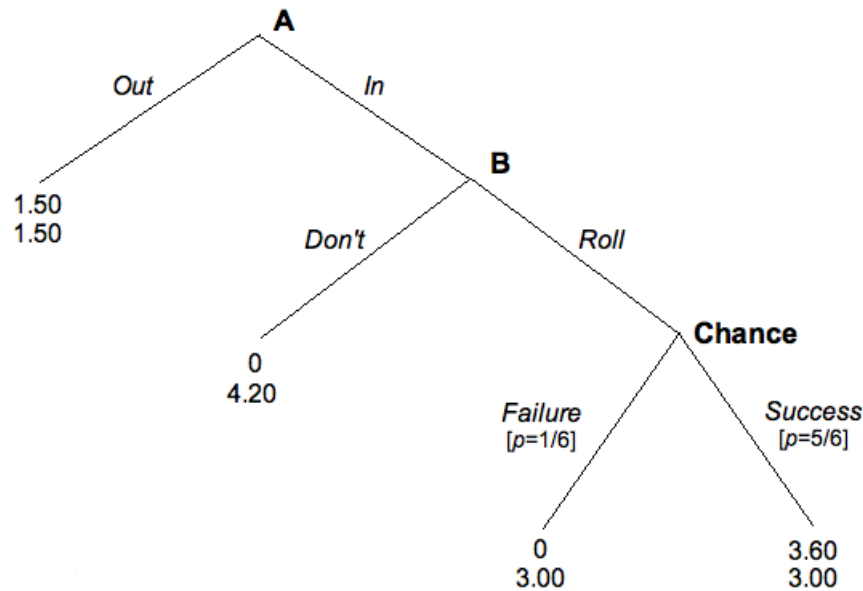
CD conducted their experiment using the strategy method. In particular, Player B made a contingent decision without observing Player A's choice, with the understanding that B's decision would only matter if A chose *In*. We play the game without any pre-game messaging, using the strategy method in one treatment and the direct-response method in another. In the direct-response treatment, A plays first, and B makes her choice after observing A's decision. We hypothesize that Player B's observation of an *In* decision from Player A will induce guilt aversion. The idea is that the revelation of an *In* decision to B will raise B's prior about A's belief that B will *Roll*, thereby increasing B's guilt from choosing *Don't Roll*. In effect, we ask whether direct-response can "substitute" for B-to-A communication as a vehicle for inducing guilt aversion.

We recognize that a fully rational (but guilt-averse) Player B should see no distinction between the strategy and direct-response methods. Under the strategy method, B makes her decision *contingent on A choosing In* – even though B never sees A's decision. Similarly, under the direct response method, B knows that her decision only matters if she *sees* that her partner chose *In*. Our premise is that in practice, players may find it difficult to estimate separate priors for different types of agents (here, trusting and non-trusting types) before a concrete Player A has revealed which group he belongs to.[5] In the direct-response treatment, the arrival of a sheet of paper with the word *In* on it serves as a visceral reminder that B's partner is counting on her. Our hypothesis, then, reflects the assumption that concrete evidence of trust should induce guilt to a greater degree than abstract contemplation of trust.

---

[5]We are grateful to an anonymous referee for suggesting this line of reasoning.

The key tests will be (1) is B more likely to *Roll* under direct-response (conditional on seeing *In*) than under the strategy method, and (2) does B have higher second-order beliefs under direct-response than under the strategy method. The answers to both questions turn out to be negative.

3. Related Literature

Ellingsen *et. al.* (2010) ("EJTT") repeat CD's experiment with a twist: They reveal each A player's first-order belief to the B partner before B makes her decision. Guilt aversion predicts a positive relationship between A's revealed belief and B's likelihood of rolling. EJTT find no evidence of this. Our approach is different: We reveal A's *decision* to B, while keeping A's belief private. Nonetheless, our result is similar: We find no relationship between the observability of A's decision and B's likelihood of rolling. CD report a related result. They find that non-binding, free-form messages from A to B are not associated with a higher likelihood of rolling, in contrast with messages from B to A.

Kawagoe and Narita (2011) ("KN") note that in the existing literature, the evidence for guilt aversion is strongest when trustees can make – and possibly break – a promise.[6] This motivates KN to propose the following refinement of the guilt aversion hypothesis: "people feel guilty when they betray another person's expectation, *with that expectation having been raised by their very own actions*, typically by their promises" (italics theirs). Our result is consistent with KN's formulation in the sense that we have no pre-play communication and hence no opportunity for B to proactively raise A's expectation. Viewed this way, the direct-response method resembles other forms of "A-to-B information transfer" in the literature.

4. Design

The experiment was conducted at Muhlenberg College. There were four sessions, two each of the strategy method and direct-response treatments. Each session had between 25 and 37 participants; no one participated in more than one session. Average earnings were about $4 (including a $1 show-up fee).

The experimental instructions are included in an online appendix (Amdur and Schmick 2013). We randomly assigned participants to A and B roles based on identification numbers that we handed out (A's were even, B's were odd). A and B participants sat in separate classrooms, and the A-B pairings were anonymous to all participants throughout. In the strategy method sessions, A and B participants made their decisions simultaneously. In the direct-response sessions, we informed each B participant of her A partner's decision while

---

[6]In Vanberg (2008), players in a dictator game behave more generously when they have made a promise to their paired recipient. In contrast, when dictators are merely aware that their recipient received a promise from another party, they are less generous. EJTT find little evidence of guilt aversion in settings in which the truster cannot make a promise. In Charness and Dufwenberg (2006), the correlation between trustworthy behavior and second-order beliefs *within* treatments could reflect guilt aversion without promises. However, as emphasized by Vanberg (2008), the "false consensus effect" could also explain this correlation. Reuben *et. al.* (2009) is an exception to KN's rule. They find evidence of guilt aversion without promises in a trust game that is arguably free of the false consensus effect.

Table I: Tests for the effect of the decision elicitation method on behavior.

A's *In* Rate			B's *Roll* Rate			(*In, Roll*)		
SM	DR	Z Stat	SM	DR	Z Stat	SM	DR	Z Stat
15/24	18/33	0.60	10/24	10/18	0.89	7/24	10/33	0.09
(63%)	(55%)		(42%)	(56%)		(29%)	(30%)	

preserving anonymity.[7] Note that under direct response, B's decision is only meaningful if B saw that A chose *In*. If B saw that A chose *Out*, then B knew with certainty that her own decision would not matter, making it difficult to interpret.

After the B's had made their decisions, each B rolled a six-sided die. As in CD, we made this clear to the participants in advance, to preserve public anonymity for B's who chose *Don't Roll*. We then collected the decision sheets (with die roll results noted) and calculated the payouts while the participants filled out a questionnaire. Following CD, we elicited beliefs by asking A's to guess the percentage of B's who chose *Roll*, and B's to guess the average guess made by A's who chose *In*.[8] We handed payoffs to students in sealed envelopes.

5. Results

Table I summarizes the decisions across the two treatments. First consider B's behavior. In the strategy method (SM) treatment, 10 of 24 (42%) B's chose *Roll*; while in the direct-response (DR) treatment, 10 of 18 (56%) B's who observed *In* chose *Roll*. This difference is not statistically significant at any of the conventional significance levels ($p=0.37$ on a test of proportions for two populations, two-tailed). Next consider A's behavior. 15 of 24 (63%) A's chose *In* in the SM treatment, compared to 18 of 33 (55%) A's in the DR treatment, a difference that is not significant either ($p=0.55$). The fraction of pairs choosing (*In, Roll*) was almost identical across treatments: 29% (7 of 24 pairs) under SM and 30% (10 of 33 pairs) under DR ($p=0.93$). The direct-response method did not appear to have a significant effect on behavior.[9]

Table II reports elicited beliefs by treatment. A's average guess about the percentage of B's who chose *Roll* was 32.3 in the SM treatment and 29.8 in the DR treatment. This difference is not statistically significant at any of the conventional significance levels ($p=0.54$

---

[7]Specifically, each A indicated his decision of *In* or *Out* on a decision sheet with his id number. We then transcribed this decision onto another sheet with only B's id number and delivered the transcribed sheet to B.

[8]Due to budget limitations, we were unable to offer monetary rewards for accurate guesses. We also asked some demographic questions on the questionnaire, as well as questions about economics courses taken.

[9]The behavioral results from both of our treatments are very similar to those reported in the baseline treatment of CD (no messages). In that treatment, CD reported A's *In* rate as 56% (25 of 45), B's *Roll* rate as 44% (20 of 45), and the (*In, Roll*) rate as 20% (9 of 45 pairs). CD also conducted a treatment in which the A participant could send a non-binding, free-form message to B prior to the game. This somewhat resembles our DR treatment in the sense that guilt aversion could have been introduced by Player A. In this treatment, CD reported A's *In* rate as 67% (31 of 46), B's *Roll* rate as 39% (18 of 46), and the (*In, Roll*) rate as 26% (12 of 46 pairs). None of these rates differ significantly from those in our treatments.

Table II: Tests for the effect of the decision elicitation method on beliefs.

A's Average Guess			B's Average Guess		
SM	DR	Z Stat	SM	DR	Z Stat
32.3	29.8	0.61	43.2	46.4	0.18

on a Wilcoxon-Mann-Whitney rank sum test, two-tailed). B's average guess about A's average guess was 43.2 in the SM treatment and 46.4 in the DR treatment, also not significant (p=0.86).[10] Neither participants' choices nor their beliefs varied significantly across treatments. The direct-response method did not appear to induce guilt aversion in our experiment.

Do we find evidence of guilt aversion when we abstract from the treatment? Not as such. Pooling treatments, the average belief of B's who chose *Don't Roll* was 48.6, and the average belief of B's who chose *Roll* was 40.2. So B's who reciprocated actually had lower average beliefs about A's expectation, although the difference is not significant (Z=0.81, p=0.42). Interestingly, A's average belief was significantly higher for A's choosing *In* (38.2) as opposed to *Out* (20.8) (Z=3.12, p<0.01). While not a test of guilt aversion per se, this suggests that A's who chose *In* had more favorable beliefs about their B partners.[11]

6. Conclusion

We conduct a one-shot trust game with hidden action using two decision elicitation treatments: strategy and direct-response. We find no significant differences in behavior or beliefs across elicitation methods. By itself, the direct-response method did not appear to induce guilt aversion.

---

[10]In the DR treatment, "B's Average Guess" includes only B's who saw A choose *In*. If we also include the guesses of B's who saw A choose *Out*, then B's average guess is 42.9, also not significantly different from the SM value (Z=0.19, p=0.85).

[11]Beliefs across decisions *within* the DR treatment were similar. Beliefs *within* the SM treatment were fairly flat across decisions. Results are available on request.

References

Amdur, D. and E. Schmick (2013) "Appendix for: Does the Direct-response Method Induce Guilt Aversion in a Trust Game?" Mimeo, Muhlenberg College.

Battigalli, P. and M. Dufwenberg (2007) "Guilt in Games" *American Economic Review* **97**, 170-176.

Brandts, J. and G. Charness (2011) "The Strategy Versus the Direct-response Method: A First Survey of Experimental Comparisons" *Experimental Economics* **14**, 375-398.

Charness, G. and M. Dufwenberg (2010) "Bare Promises: An Experiment" *Economics Letters* **107**, 281-283.

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

Ellingsen, T., M. Johannesson, S. Tjotta, and G. Torsvik (2010) "Testing Guilt Aversion" *Games and Economic Behavior* **68**, 95-107.

Fehr, E. and S. Gachter (2000) "Fairness and Retaliation: The Economics of Reciprocity" *Journal of Economic Perspectives* **14**, 159-181.

Kawagoe, T. and Y. Narita (2011) "Guilt Aversion Revisited: An Experimental Test of a New Model" Mimeo, Future University – Hakodate.

Reuben, E., P. Sapienza, and L. Zingales (2009) "Is Mistrust Self-fulfilling?" *Economics Letters* **104**, 89-91.

Selten, R. (1967) "Die Strategiemethode zur Erforschung des Eingeschränkt Rationalen Verhaltens im Rahmen eines Oligopolexperimentes" in *Beiträge zur Experimentellen Wirtschaftsforschung* by H. Sauermann, Ed., Mohr Siebeck: Tübingen 136-168.

Vanberg, C. (2008) "Why Do People Keep Their Promises? An Experimental Test of Two Explanations" *Econometrica* **76**, 1467-1480.