Do I Care if You Know I Betrayed You?

James C. Cox*
Experimental Economics Center
and Department of Economics
Georgia State University

Danyang Li†
Experimental Economics Center
and Department of Economics
Georgia State University

Abstract

It has been reported that betrayal aversion influences the trust decision (Bohnet and Zeckhauser 2004; Bohnet et al. 2008). This paper adds to the literature by examining how concern for others' disutility from betrayal can affect the decision to repay trust. We compare trustees' behavior when betrayal is obfuscated to an identical monetary payoffs situation where betrayal is revealed. We find that more trustees choose to defect in our experiment when betrayal is obfuscated than when it is revealed. Our result suggests that concern for betrayal costs influences not only the decision to trust but also the decision to repay trust.

JEL Codes: C72, C91

Keywords: Experiments; Betrayal Cost; Trust; Cooperation

*Address: 14 Marietta Street, Atlanta, GA 30303. Email: jccox@gsu.edu. Phone: +1-404-413-0200. Fax: +1-404-413-0195.

†Corresponding Author. Address: 14 Marietta Street, Atlanta, GA 30303. Email: dli@gsu.edu. Phone: +1-404-413-0198. Fax: +1-404-413-0145.

1. Introduction

In a social context, people may display an aversion to betrayal. First mover betrayal aversion in trust games has been reported by Bohnet and Zeckhauser (2004). They experiment with subjects' decision making in paired trust games played either with another person or a computer. They find that individuals are less willing to "trust" when the outcome is determined by another person than when it is determined by random draw by a computer. This result indicates an aversion to being betrayed by another human being, which is referred to as "betrayal aversion" by Bohnet and Zeckhauser. Bohnet et al. (2008) further examine this phenomenon across six countries. They report that betrayal aversion is a broad-based phenomenon.

In this paper, we seek to extend these findings by looking inversely at how the concern for betrayal aversion affects the behavior to repay trust. Assuming that the "trustee" takes others' betrayal costs into account, he is less likely to repay trust if betrayal is obfuscated from others. We examine whether the trustee's decision to repay trust is partially motivated by concern for others' disutility from betrayal.

We conduct a modified trust game experiment to test our hypothesis. In the trust game, the first mover's (or "trustor's") decision to trust the second mover (or "trustee") is "productive" (Deck 2009), which refers to the increase in total money payoff compared with the alternative choice.[1] We modify the trust game by introducing a move of nature in between the trustor's decision and the trustee's decision, which randomly determines the productivity level of trust. The trustor does not observe nature's move or the trustee's decision unless they are revealed by his own payoff. The key feature of the modified trust game is that when the trustee repays low-productivity trust, the trustor receives the same payoff as when the trustee betrays high-productivity trust. That is, the trustor cannot infer betrayal of high-productivity trust.

We recruited subjects to participate in two experimental treatments. Subjects play the modified trust game in treatment 1. In treatment 2, different subjects play a game identical to the modified trust game except that the trustor observes nature's move at the end of the game, hence the trustor eventually has perfect information about the trustee's decision. If the trustee cares whether the trustor suffers from betrayal costs, they are less likely to betray high-productivity trust in treatment 2. The results of the experiment support our hypothesis. We find that subjects repay high-productivity trust more frequently in treatment 2 than in treatment 1. Our findings suggest that the concern for others' betrayal costs partly motivates the decision to repay trust.

The next section of the paper describes some related literature. Section 3 describes the experimental design and protocol. Section 4 reports the results from the experiment. The final section of the paper concludes.

[1] Henceforth, we use the common, although questionable, designation of the first mover as "trustor" and the second mover as "trustee" even though the first mover's motivation may be altruism rather than trust (Cox 2004).

2. Some related literature

Traditional economic models assume that individuals' actions are exclusively motivated by material self-interest. The narrow material self-interest assumption is quite good at predicting behavior in many contexts. However, in some context, this assumption does not work well. Examples include ultimatum games (Güth et al. 1982; Slonim and Roth 1998), dictator games (Forsythe et al. 1994; Andreoni and Miller 2002), and investment games (Berg et al. 1995; Cox 2004; Cox and Deck 2005). Experimental studies of such games show that individuals often behave in a way that is inconsistent with narrow material self-interest.

As a result of these findings, researchers have proposed to develop models of other-regarding preferences. These models assume individuals have genuine concern for others' material payoffs. This literature broadly falls into two classes: outcome-based models and models of reciprocity. The outcome-based (distributional) models assume that individuals care about their own and others' material payoffs. Examples include inequality aversion models (Fehr and Schmidt 1999; Bolton and Ockenfels 2000) and quasi-maximin models (Charness and Rabin 2002; Engelmann and Strobel 2004), convex other-regarding preferences models (Andreoni and Miller 2002), and the egocentric altruism model (Cox and Sadiraj 2007; 2012). Alternatively, models of reciprocity assume that individuals prefer to repay kind actions by others with similar actions themselves and, also, to repay unkind actions with similarly unkind ones. For example, revealed altruism theory (Cox et al. 2008) assumes that one's generous action may change others' preferences and trigger a reciprocal response.

In our modified trust game experiment, we find that fewer subjects cooperate when betrayal is obfuscated from their partners. This finding cannot be explained by any of the above-cited other-regarding preferences models. Trustees with preferences consistent with those models should behave consistently across our treatments because the feasible sets of material payoffs alternatives are identical across treatments. Revealed altruism theory cannot be applied here without modification because opportunity sets in our modified trust game are not exclusively chosen by other (human) players. We attribute our findings to concern for others' betrayal costs.

The idea of betrayal aversion has been nicely addressed by Bohnet and Zeckhauser (2004). They conduct a trust game experiment in which they elicit subjects' minimum acceptance probability (MAP) to measure their willingness to bear the risk of transferring money to another person. In their trust game, the trustor has to choose between a sure option and a risky transfer to another subject in the experiment. Whether the risky transfer yields a good or bad outcome depends on the trustee's decision. Bohnet and Zeckhauser compare the trust game with a structurally identical risky dictator game. The only difference between these two games is that in the risky dictator game the trustee's decision is (pseudo-) randomly selected by a computer. They find that the trustors state significantly higher MAPs in the trust game that in the risky dictator game, indicating that subjects are averse to human-generated betrayal as distinct from computer-generated risk. Bohnet et al. (2008) follow the same design to examine whether betrayal aversion is a robust feature beyond the United States. Their results support betrayal aversion as a broad-based phenomenon across countries.

Bohnet and Zeckhauser demonstrate that social preferences toward betrayal influence the behavior of trustors in a trust game. In this paper, we take a step further by asking whether concern for betrayal costs of trustors affects the choices made by *trustees* in a trust game. We ask whether the behavior of trustees shows that betrayal aversion influences not only the trust decision but also the decision to repay trust. We find that the trustees are more likely to behave self-interestedly when betrayal costs of trustors are avoidable. Our results suggest that the decision to repay trust is partially motivated by concern for other's betrayal costs. This result also supports previous findings that people are more likely to be selfish in the situation of moral hazard (Castillo and Leo 2010), or where they can avoid being perceived negatively (Dana et al. 2006).

Our work provides a complementary explanation for cooperation. We are not denying that trustees may be motivated by distributional concerns or reciprocity. However, this possibility may be overstated by results of prior research. Our results suggest that sometimes trustees who repay trust can be motivated by concern for others' betrayal costs.

3. Experimental design and protocol

The extensive form of the modified trust game is represented in Figure 1. A first mover (the trustor) can choose a sure option ("exit" the game) that gives both movers a payoff of 10, or he can choose to trust (or "engage" into the game). The productivity level of trust is determined by nature's move. Fifty percent of the time, nature moves left and yields low-productivity trust. In this case, trust increases the total money payoff from 20 to 30. Another fifty percent of the time, nature moves right which selects high-productivity trust that increases the total money payoff from 20 to 50. After observing nature's move, a second mover (the trustee) has to choose between cooperate with the trust or defect. Cooperation with low-productivity trust gives both movers a payoff of 15, while defecting results in a payoff of 0 for the trustor and 30 for the trustee. Cooperating with high-productivity trust ends with a payoff of 25 for both movers, while defecting yields a payoff of 15 for the trustor and 35 for the trustee.

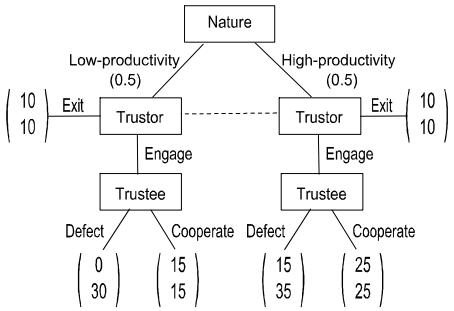

Figure 1: The modified trust game

Interestingly, when the trustee chooses to cooperate with low-productivity trust, the trustor receives the same payoff – a payoff of 15 – as when trust is highly-productive but the trustee defects. The trustor knows neither the choice nor the payoff of the trustee, unless they are revealed by his own payoff. The trustor also cannot observe nature's move. Imagine you are the trustor who receives a payoff of 15. You may want to believe that the trustee chose to cooperate and encountered an unlucky move of nature. However, it is also possible that you faced a greedy partner who relied on the presence of nature's move for obfuscation. Consequently, the trustee's action is not revealed.

We experiment with two treatments: (1) the modified trust game; (2) a game identical to the modified trust game except that the trustor observes nature's move at the end of the game. We compare data from these two treatments to capture the effect of the concern for betrayal costs. Compared with treatment 2, the trustee in treatment 1 is able to hide betrayal when trust is highly-productive. Since betrayal is not revealed, the trustee may believe that the trustor does not experience betrayal costs. This belief may allow the trustee to justify the choice to defect, and thereby lead to more self-interested actions.

Before the experiment started, the experimenter read the instructions out loud to the subjects. Whether betrayal would be revealed was made clear to all subjects. After the experiment began, subjects were reminded of whether betrayal would be revealed again on the decision screen. The actual decision screen for trustees in treatment 2 is shown below in Figure 2.

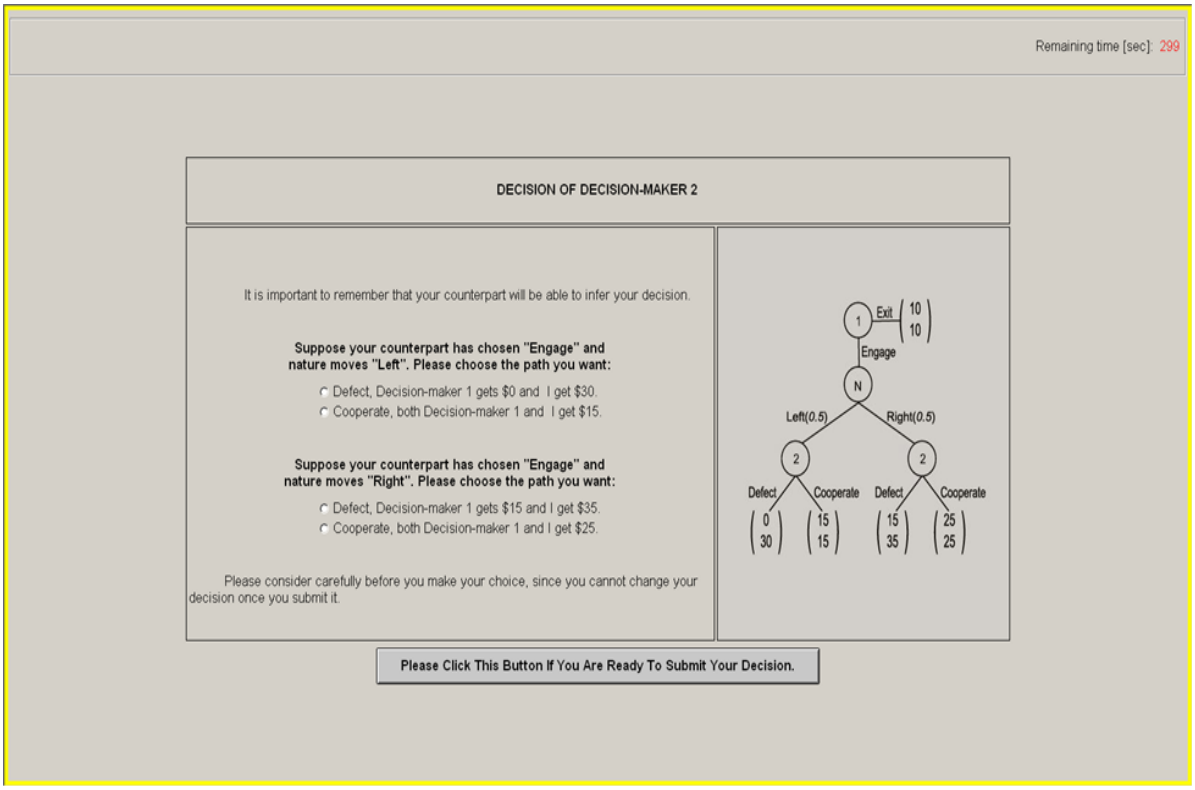

Figure 2: Decsion screen for trustees in treatment 2

In both treatments, we implemented the strategy method. This method allows us to observe trustees' choices for both possible move of nature even when the trustor chooses the sure option. In each treatment, trustors had to decide whether to trust. At the same time, trustees chose to cooperate or defect for each of the possible moves of nature. The earnings of each pair of subjects were determined by choices of both parties and nature's move. Trustees were informed that their choices were only determinative for the final payoff if their paired trustors chose to trust.

After subjects finished making decisions in each treatment, the experimenter flipped a coin to determine nature's move in the presence of all of the subjects. The difference between the two treatments in the information provided to the subjects was as follows. In treatment 2, *all subjects* were informed whether "Heads" on the coin flip meant that nature mover left or right. In treatment 1, only trustees were informed whether Heads corresponded to left or right for nature's move.

We used a double-blind subject payment protocol in which subjects' choices are anonymous to both other subjects and the experimenter. This protocol is implemented by first asking each subject to select one from a box full of identical, sealed envelopes. Each envelope contains a key with a unique number. Subjects are asked to use these numbers as their (only) identifiers in the experiment. At the end of the experiment, subjects exit the lab individually and collect their earnings in private from a mailbox with a number that corresponds to their key number. Payoffs are contained in sealed envelopes. Subjects are asked to exit the building before opening their envelopes. While waiting for pay envelopes to be filled with money and put in the mailboxes, subjects are asked to complete a questionnaire on demographic characteristics. Subjects' questionnaire responses are linked with their decisions by their mailbox key numbers.

4. Results

Subjects who participated in the experiment were recruited from undergraduate students at Georgia State University. A total of 142 subjects participated in the experiment, 72 in treatment 1 and 70 in treatment 2. Subjects earned on average \$22.61 (including a \$5 show-up fee). There were 2 sessions in each treatment. The treatments were implemented with a between-subjects design. Subjects were randomly assigned to one role, either the trustor or the trustee. Each trustor was randomly paired with a trustee. In each session, subjects played the game only once. Results from the experiment are as follows.

Table 1 reports the frequency of the choice to defect for each treatment. The results are consistent with our hypothesis that subjects choose to defect more frequently when betrayal is not revealed. When nature led to high-productivity trust, 29 out of 36 trustees in treatment 1 chose to defect, while 23 out of 35 trustees chose to defect in treatment 2. This difference across treatments is weakly significant ($z = 1.4122$, p-value $= 0.0789$). The observed difference across treatments provides some support for the conjecture that concern for others' betrayal costs is a motivation for choosing cooperation.

Table 1 Frequencies of the choice to defect by treatment

	Sample Size	Low-productivity trust			High-productivity trust		
		Obs	Percent	SD	Obs	Percent	SD
Treatment 1	36	20	55.56%	50.40	29	80.56%	40.14
Treatment 2	35	22	62.86%	49.02	23	65.71%	48.16
Proportion test across treatments: z-test (p-value)							
Total		-0.6258(0.7343)			1.4122(0.0789)		

Notes: The null hypothesis: difference in behavior across two treatments=0; the alternative hyphothesis: difference in behavior across two treatments¿0.

When trust was lowly productive, the choice to defect was revealed in both treatments. We did not observe a statistically significant difference in choice of defect across treatments. In treatment 1, 20 out of 36 trustees chose to defect. In treatment 2, 22 out of 35 trustees chose to defect. The difference across treatments is not statistically significant ($z = -0.6258$, p-value $= 0.7343$). This finding of *insignificant* difference for the (trustor-revealing) choices with low-productivity trust, together with the *significant* difference for the (trustor-obscuring) choices with high-productivity trust, provides further support for our conjecture that it is trustees' concern for betrayal costs that explains differences between behaviors across treatments.

Table 2 Probit regressions of the choice to defect

	Dependent variable: choice to defect			
	Low-productivity trust		High-productivity trust	
	(1)	(2)	(3)	(4)
Treatment	0.131	0.098	-0.176*	-0.224**
	(0.121)	(0.128)	(0.105)	(0.109)
Defect with high-productivity trust	0.333**	0.299**		
	(0.130)	(0.138)		
Defect with low-productivity trust			0.267**	0.240**
			(0.109)	(0.114)
Gender		-0.052		-0.009
		(0.121)		(0.106)
Race		-0.211*		-0.119
		(0.120)		(0.109)
Experience		-0.033		-0.172
		(0.148)		(0.140)
Log likelihood	-44.788	-43.230	-37.206	-35.984
Pseudo R2	0.0672	0.0997	0.0979	0.1275
N	71	71	71	71

Notes: The base treatment is treatment 1. Reporting marginal effects. Standard error in parentheses. *** significant at 1% level, ** significant at the 5% level, * significant at the 10% level.

Table 2 presents a series of probit regressions on trustees' decisions to defect. In regressions (1) and (2), we analyze trustees' decision when trust is lowly productive, while regressions (3) and (4) produce the anaylsis of trustees' decision with high-productivity trust. The dependent varibale of all regressions is a dummy variable indicating trustees' decision, which is equal to one if trustees chose to defect. The right-hand side of the regressions first includes a dummy variable "Treatment", which is equal to one in treatment 2. Since trustees made simultaneous decisions for each possible outcome of nature with the strategy method, regressions under low-productivity trust have controls on trustees' decisions for high-productivity trust and vice versa.

Regression (1) and (3) indicate that trustees only behave differently across treatments when trust is highly productive (variable "Treatment"). Trustees are 18% more likely to defect in treatment 1 where betrayal is not revealed and the result is statistically significant at 10% level. However, the treatment does not affect the likelihood to defect where betrayal is revealed (p-value = 0.284).

Regressions (2) and (4) include additional subject characteristic variables such as "Gender" (female=1), "Race" (black=1), and "Experience" (having previous experience in other behavior experiments=1). Again, we only see behavioral difference across treatments when trust is highly-productive. The coeffient on variable "Treatment" in regression (4) is negative and significant at 5% level, suggesting that trustees are 22% more likely to defect in treatment 1. However, regression (2) indicates that when trust is lowly productive, there is no treatment effect.

5. Conclusion

In this paper, we report results derived from a laboratory experiment based on a modified trust game design. This design allows us to examine whether the behavior to repay trust is influenced by concern for others' betrayal aversion. In the modified trust game, we introduce a random move of nature between decisions of trustors and trustees. By varying trustors' ability to observe nature's move, we allow trustees to be able to hide betrayal in one treatment. We find a significant increase in the choice to defect when trustees can hide betrayal, but no significant behavioral change when betrayal is revealed in both treatments. This result supports our hypothesis that concern for betrayal costs partially motivates the trustee to cooperate. Our finding suggests that concern for betrayal costs influences not only the decision to trust, but also the decision to repay trust.

References

Andreoni, J., & Miller, B. D. (2002). Giving According to GARP: An Experimental Test of the Consistency of Preferences for Altruism. Econometrica, 70(2), 737-753.

Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, Reciprocity, and Social History. Games and Economic Behavior, 10(1), 122-142.

Bohnet, I., & Zeckhauser, R. (2004). Trust, Risk and Betrayal. Journal of Economic Behavior & Organization, 55(4), 467-484.

Bohnet, I., Greig, F., Herrmann, B., & Zeckhauser, R. (2008). Betrayal Aversion: Evidence from Brazil, China, Oman, Switzerland, Turkey, and the United States. American Economic Review, 98(1), 294-310.

Bolton, G., & Ockenfels, A. (2000). ERC: A Theory of Equity, Reciprocity, and Competition. American Economic Review, 90(1), 166-193.

Castillo, M., & Leo, G. (2006). Moral Hazard and Reciprocity. Southern Economic Journal, 77(2), 271-281.

Charness, G., & Rabin, M. (2002). Understanding Social Preferences with Simple Tests. Quarterly Journal of Economics, 117(3), 817-869.

Cox, J. C. (2004). How to Identify Trust and Reciprocity. Games and Economic Behavior, 46(2), 260-281.

Cox, J. C., & Deck, C. A. (2005). On the Nature of Reciprocal Motives. Economic Inquiry, 43(3), 623-635.

Cox, J. C., Friedman, D., & Sadiraj, V. (2008). Revealed Altruism. Econometrica, 76(1), 31-69.

Cox, J. C., & Sadiraj, V. (2007). On Modeling Voluntary Contributions to Public Goods. Public Finance Review, 35(2), 31-69.

Cox, J. C., & Sadiraj, V. (2012). Direct Tests of Individual Preferences for Efficiency and Equity. Economic Inquiry, in press.

Dana, J., Cain, D. M., & Dawes, R. M. (2006). What You Don't Know Won't Hurt Me: Costly (but quiet) Exit in Dictator Games. Organizational Behavior and Human Decision Processes, 100(2), 193-201.

Deck, C. A. (2009). An Experimental Analysis of Cooperation and Productivity in the Trust Game. Experimental Economics, 12(1), 1-11.

Engelmann, D., & Strobel, M. (2004). Inequality Aversion, Efficiency, and Maximin Preferences in Simple Distribution Experiments. American Economic Review, 94(4), 857-869.

Fehr, E., & Schmidt, K. (1999). A Theory of Fairness, Competition, and Cooperation. Quarterly Journal of Economics, 114(3), 817-868.

Fischbacher, U. (2007). z-Tree: Zurich Toolbox for Ready-made Economic Experiments. Experimental Economics, 10(2), 171-178.

Forsythe, R., Horowitz, J. L., Savin, N. E., & Sefton, M. (1994). Fairness in Simple Bargaining Experiments. Games and Economic Behavior, 6(3), 347-369.

Güth, W., Schmittberger, R., & Schwarze, B. (1982). An Experimental Analysis of Ultimatum Bargaining. Journal of Economic Behavior & Organization, 3(4), 367-388.

Slonim, R., & Roth, A. E. (1998). Learning in High Stakes Ultimatum Games: An Experiment in Slovak Republic. Econometrica, 66(3), 569-596.